Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_5-1

Commercial Antenna Design Tools

Qing Huo Liu*
Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA

Abstract

Antenna design traditionally relies on physical understanding of electromagnetic radiation, intuition, and experience, as well as trial-and-error experimentations. With the advent of computers and increasingly sophisticated numerical methods, however, computer-aided design tools play a central role in today's antenna design and optimization process. This chapter presents a summary of commonly used commercial antenna design simulation tools and their underlying computational electromagnetics methods.

Keywords

Antenna design; Computational electromagnetics; Finite element method; Finite difference time domain method; Method of moments; Computer-aided design

Introduction

Antenna design is an art that dates back to the early work of Hertz in 1886–1889 when he first experimentally verified the existence of electromagnetic waves, using his electric dipole antenna transmitter and ring receiver. Since then, traditional antenna design relies heavily on physical intuition and trial-and-error experimentation.

During the last few decades, computer technologies and computational science and engineering have revolutionized the design and optimization of antennas. Today, one can easily use commercially available antenna design tools to study how the antenna structure interacts with electromagnetic waves so that the structure and materials can be optimized to achieve the design goals.

Antenna design and optimization rely on solutions of Maxwell's equations. Unfortunately, only problems with canonical geometries (such as spheres, infinitely long circular cylinders, and planar layers) admit analytical solutions. Almost all antenna design problems require numerical solutions of Maxwell's equations, either exactly with full-wave solution methods or approximate methods such as high-frequency asymptotic techniques.

Computational electromagnetics is an important area that develops effective numerical methods for Maxwell's equations for applications in antenna design and optimization, in addition to many other application areas.

The goal of this chapter is to give an outline of the numerical methods commonly used in antenna design tools. First, several commercial antenna design tools are listed, followed by the basic electromagnetic theory. Then three numerical techniques widely used in these tools, i.e., the method of moments for surface integral equation solvers, finite element method, and finite difference time domain method, are discussed in more detail.

*Email: qing.liu@duke.edu

DOI 10.1007/978-981-4560-75-7_5-1

Table 1 Commonly used commercial antenna design tools

Abbreviation	Primary method	Additional methods	Company
COMSOL	FEM	FETD, DDM	COMSOL, Inc.
HFSS	FEM		ANSYS
EMCoS	MOM	FEM, FDTD, PO, GO, UTD	EMCoS
FEKO	MOM/MLFMM	High-frequency methods	Altair
newFASANT	MOM		FASANT
Microwave Studio	FIT	FEM, MOM, MLFMM	CST
SEMCAD	FDTD	Sub-gridding	SPEAG
XFDTD	FDTD	GTD, UTD	Remcom
Wavenology	DGTD/FDTD	SETD, FETD, SPICE	WCT

Commonly Used Commercial Antenna Tools

The commonly used commercial antenna design tools are listed in Table 1. Note that this may be a non-exhaustive list as there are certainly other software packages capable of doing antenna design simulations. Also note that open-source but noncommercial tools, such as the very well-known antenna design tool NEC (www.nec2.org) (NEC 2015) based on the method of moments (MOM), have not been included in the table.

Although most software packages use multiple numerical techniques, according to their primary methods involved, methods used by these commercial tools can be categorized into four classes: (a) frequency domain finite element method (FEM), (b) frequency domain method of moments, (c) finite difference time domain (FDTD) method, and (d) multiscale discontinuous Galerkin time domain (DGTD) method. Complementary methods including high-frequency asymptotic techniques have also been used to simulate electrically very large structures where the full-wave solutions are difficult to apply because of the sheer size of the problem.

General Flow of an Antenna Simulation Tool

A typical antenna simulation tool consists of following parts:

- **Graphic User Interface (GUI):** GUI is the central part of the software that the user interacts with. In general, the GUI provides the following functions:
  - **Problem definition:** The GUI allows the user to enter easily geometries, materials, sources, observers, boundary conditions, required results, and other simulation parameters.
  - **Problem modification:** The GUI also enables the user to modify the problem settings easily. This may include modifications to any parameters, including editing the geometries in the solid modeling.
  - **Preprocessing:** The GUI in general will enable the necessary preprocessing of the problem to validate the model and to prepare for the actual computation.
  - **Results and post-processing:** Finally, GUI allows the user to view results and to perform various post-processing.
  - **Design optimization and parameter sweeping:** Through GUI, the user can further perform design optimization of the antenna by optimizing certain parameters either through an optimization algorithm or by sweeping certain antenna parameters to search for the best configuration.
- **Preprocessing:** Typical preprocessing includes validation of the configuration, mesh generation, and other pre-computations including the source excitations.
- **Computational Engine:** This is the core of the numerical solver that gives the solutions to Maxwell's equations for the problem prescribed by the GUI. These methods will be further discussed in sections

DOI 10.1007/978-981-4560-75-7_5-1

"Method of Moments for Surface Integral Equations," "Finite Element Method," and "Finite Difference Time Domain Method."

- **Post-Processing:** The computational engine provides the basic near-field solutions, which can then be processed to arrive at the antenna performance characteristics including radiation pattern, gain, and efficiency. Near-to-far-field transformation is necessary to obtain the far fields from the computed near fields.

Numerical Methods for Antenna Simulation

Commonly used numerical methods can be divided into the frequency domain solvers and time domain solvers:

- **Frequency domain methods:** Antenna designers usually need to know the scattering parameters and other antenna performance characteristics over a frequency band. Thus, the antenna simulation is very often directly obtained in the frequency domain by assuming a time-harmonic excitation with a given frequency. By changing the input frequency, one can obtain the results over a frequency band. If the transient results are needed (especially for ultra-wideband antennas whose waveforms are often sought for), one can perform the inverse Fourier transform on the frequency domain results to arrive at time domain results. The frequency domain methods can be further divided into the following:
  - Partial differential equation solvers: They normally solve the 2nd-order vector Helmholtz equation that is derived from 1st-order Maxwell's equations (usually just Faraday's law and Ampere's law). Methods include finite element method (FEM), spectral element method, finite difference frequency domain method, and pseudospectral frequency domain method. As the FEM is the mostly commonly used frequency domain technique in commercial tools, it will be discussed in more detail in section "Finite Element Method."
  - Integral equation solvers: They solve the surface or volume integral equations derived from Maxwell's equations and Green's function for a background medium. Surface integral equation solvers based on the MOM are the mostly commonly used integral equation technique, so it will be discussed in more detail in section "Method of Moments for Surface Integral Equations."

- **Time domain methods:** If antenna characteristics are desired over a wide frequency range, it may be more advantageous to solve Maxwell's equations directly in time domain to arrive at the wideband waveforms. The frequency domain results such as scattering parameters often required by antenna designers can be obtained by Fourier transform of the transient results. Therefore, time domain methods often are preferred for wideband application as one can obtain all frequency domain results with a single simulation. On the other hand, if a narrow band result is needed, especially when the system has high resonances within this band, the frequency domain solvers are preferred. The time domain methods can be further divided into the following:
  - Partial differential equation solvers: Time domain PDE solvers work with either the 2nd-order vector wave equation or 1st-order Maxwell's equations. Methods include finite difference time domain (FDTD) method, finite element time method (FETD), spectral element time domain (SETD) method, pseudospectral time domain (PSTD) method, finite integration technique (FIT), and multi-resolution time domain (MRTD) method. As the FDTD is the mostly commonly used time domain technique in commercial tools (FIT may also be considered as a special FDTD method), it will be discussed in more detail in section "Finite Difference Time Domain Method."
  - Integral equation solvers: Time domain integral equation methods solve the surface or volume integral equations derived from Maxwell's equations and Green's function for the background medium in time domain. Unfortunately, despite vigorous research, time domain integral equation

DOI 10.1007/978-981-4560-75-7_5-1

solvers are relatively expensive in terms of computer memory and computation time and apparently have still yet to be incorporated into a commercial simulator. Thus, time domain integral equation methods will not be discussed in detail in this chapter.

Each method has its distinct advantages and disadvantages and no one single solver is perfect for all problems, so usually multiple numerical methods may be utilized in a commercial simulation tool to fully exploit the advantages of different methods.

Basics of Electromagnetics Theory

This section summarizes the fundamental concepts and equations in electromagnetics, including Maxwell's equations and boundary conditions. The materials below in this section are the condensed theory explained by Liu (2014).

Transient Electromagnetic Fields

Maxwell's equations describe the relation between electromagnetic sources and the fields. In electromagnetic simulations, one is interested in finding the following electric field intensity **E** and magnetic field intensity **H** from governing Maxwell's equations:

$$\nabla \times \mathbf{E} = -\mathbf{M}_i - \mathbf{M}_c - \frac{\partial \mathbf{B}}{\partial t} \tag{1}$$

$$\nabla \times \mathbf{H} = -\mathbf{J}_i - \mathbf{J}_c - \frac{\partial \mathbf{D}}{\partial t} \tag{2}$$

$$\nabla \cdot \mathbf{D} = \rho_{ei} + \rho_{ec} \tag{3}$$

$$\nabla \cdot \mathbf{B} = \rho_{mi} + \rho_{mc} \tag{4}$$

where the imposed electromagnetic sources include electric current density $\mathbf{J}_i$ and magnetic current density $\mathbf{M}_i$.

In addition, the continuity equations governing the conservation law of charges are

$$\nabla \cdot \mathbf{J}_{i,c} = -\frac{\partial \rho_{ei,c}}{\partial t} \tag{5}$$

$$\nabla \cdot \mathbf{M}_{i,c} = -\frac{\partial \rho_{mi,c}}{\partial t} \tag{6}$$

respectively, for the impressed sources and the conduction sources. Note that the total free electric and magnetic charge densities are

$$\rho_e = \rho_{ei} + \rho_{ec}, \quad \rho_m = \rho_{mi} + \rho_{mc} \tag{7}$$

where ρ_{ei} (ρ_{mi}) and ρ_{ec} (ρ_{mc}) are the electric (magnetic) charge densities due to the imposed and conduction electric (magnetic) current, respectively. Note that in this chapter, the magnetic sources are

DOI 10.1007/978-981-4560-75-7_5-1

included even though so far no isolated magnetic charges have been found. The reason for the introduction of these sources is for symmetry between electric and magnetic quantities and for the future possibility of finding magnetic monopoles. Moreover, it is often convenient to use equivalent magnetic current sources in solving antenna problem, for example, aperture antennas; under such circumstances, magnetic sources are typically more convenient in computation.

Additional equations are provided by constitutive relations. For anisotropic media, the relations between flux densities and field intensities are

$$\mathbf{D} = \overline{\overline{\epsilon}}\,\mathbf{E} = \epsilon_0 \overline{\overline{\epsilon}}_r\ \mathbf{E}, \tag{8}$$

$$\mathbf{B} = \overline{\overline{\mu}}\,\mathbf{H} = \mu_0 \overline{\overline{\mu}}_r\,\mathbf{H}. \tag{9}$$

where ϵ_0 and μ_0 are the permittivity and permeability of the vacuum, while $\overline{\overline{\epsilon}}$ and $\overline{\overline{\epsilon}}_r$ are the permittivity and relative permittivity tensors of the medium, and $\overline{\overline{\mu}}$ and $\overline{\overline{\mu}}_r$ are the permeability and relative permeability tensor of the medium, respectively. In general for a linear medium, the 3×3 permittivity and permeability tensors $\overline{\overline{\epsilon}}$ and $\overline{\overline{\mu}}$ are independent of the field strength. The conduction current densities are related to the field intensities by Ohm's law:

$$\mathbf{J}_c = \overline{\overline{\sigma}}_e \mathbf{E}, \tag{10}$$

$$\mathbf{M}_c = \overline{\overline{\sigma}}_m \mathbf{H}, \tag{11}$$

where σ_e (siemens/meter) and σ_m (ohms/m) are called the electric and magnetic conductivities, respectively. In general, electromagnetic media can be inhomogeneous, that is, all parameters $\overline{\overline{\epsilon}}, \overline{\overline{\mu}}, \overline{\overline{\sigma}}_e$, and $\overline{\overline{\sigma}}_m$ may be functions of position. Furthermore, some media can be dispersive, i.e., the parameters are functions of frequency.

Given the impressed electric and magnetic sources $\mathbf{J}_i\ \rho_{ei}$, $\mathbf{M}_i$, ρ_{mi}, the unknown quantities to be solved are $\mathbf{E}$, $\mathbf{H}$, $\mathbf{D}$, $\mathbf{B}$, $\mathbf{J}_c$, $\mathbf{M}_c$, ρ_{ec}, and ρ_{mc}. Equations 8, 9, 10, and 11 together with Eqs. 1, 2, 5, and 6 provide 20 scalar equations, from which all 20 unknowns can be solved. With the above constitutive relations, Maxwell's equation can now be written as

$$\nabla \times \mathbf{E} = -\mathbf{M}_i - \overline{\overline{\sigma}}_m \mathbf{H} - \overline{\overline{\mu}} \frac{\partial \mathbf{H}}{\partial t} \tag{12}$$

$$\nabla \times \mathbf{H} = \mathbf{J}_i + \overline{\overline{\sigma}}_e \mathbf{E} + \overline{\overline{\epsilon}} \frac{\partial \mathbf{E}}{\partial t} \tag{13}$$

$$\nabla \cdot \overline{\overline{\epsilon}}\mathbf{E} = \rho_{ei} + \rho_{ec} \tag{14}$$

$$\nabla \cdot \overline{\overline{\mu}}\mathbf{H} = \rho_{mi} + \rho_{mc} \tag{15}$$

In most cases only the isotropic media will be treated in this chapter, where $\overline{\overline{\epsilon}} = \epsilon \overline{\overline{I}}$, $\overline{\overline{\mu}} = \mu \overline{\overline{I}}$, $\overline{\overline{\sigma}}_e = \sigma_e \overline{\overline{I}}$, $\overline{\overline{\sigma}}_m = \sigma_m \overline{\overline{I}}$ where ϵ and μ are the scalar permittivity and permeability and $\overline{\overline{I}}$ is the identity tensor.

Time-Harmonic Electromagnetic Fields

Time-harmonic EM fields are those with a sinusoidal variation in time, for example,

DOI 10.1007/978-981-4560-75-7_5-1

$$\mathbf{E}(\mathbf{r},t) = \hat{x}E_x(\mathbf{r})\cos\left[\omega t + \phi_x(\mathbf{r})\right] + \hat{y}E_y(\mathbf{r})\cos\left[\omega t + \phi_y(\mathbf{r})\right] + \hat{z}E_z(\mathbf{r})\cos\left[\omega t + \phi_z(\mathbf{r})\right] \tag{16}$$

where the constant ω is the angular frequency. To simplify the treatment of time-harmonic fields, one can introduce the phasor notation $\tilde{\mathbf{E}}(\mathbf{r})$ of the corresponding instantaneous field $\mathbf{E}(\mathbf{r}, t)$. The relation between the instantaneous and phasor expressions is

$$\mathbf{E}(\mathbf{r},t) = \Re e\left[\tilde{\mathbf{E}}(\mathbf{r})e^{j\omega t}\right] \tag{17}$$

In general, the phasor expression

$$\tilde{\mathbf{E}}(\mathbf{r}) = \hat{x}\tilde{E}_x(\mathbf{r})\exp[j\phi_x(\mathbf{r})] + \hat{y}\tilde{E}_y(\mathbf{r})\exp\left[j\phi_y(\mathbf{r})\right] + \hat{z}\tilde{E}_z(\mathbf{r})\exp[j\phi_z(\mathbf{r})]$$

is complex, even though the instantaneous expression $\mathbf{E}(\mathbf{r}, t)$ is always real.

With this relation between the instantaneous and phasor expressions of the field, the time derivative $\frac{\partial}{\partial t}$ on the instantaneous field corresponds to $j\omega$ times the phasor expression of the field. Therefore, one can derive Maxwell's equations for the phasor form of electromagnetic fields as

$$\nabla \times \tilde{\mathbf{E}} = -\tilde{\mathbf{M}}_i - \overline{\overline{\sigma}}_m\tilde{\mathbf{H}} - j\omega\overline{\overline{\mu}}\tilde{\mathbf{H}} \tag{18}$$

$$\nabla \times \tilde{\mathbf{H}} = \tilde{\mathbf{J}}_i + \overline{\overline{\sigma}}_e\tilde{\mathbf{E}} + j\omega\overline{\overline{\epsilon}}\tilde{\mathbf{E}} \tag{19}$$

$$\nabla \cdot \epsilon\tilde{\mathbf{E}} = \widetilde{\rho}_{ei} + \widetilde{\rho}_{ec} \tag{20}$$

$$\nabla \cdot \mu\tilde{\mathbf{H}} = \widetilde{\rho}_{mi} + \widetilde{\rho}_{mc} \tag{21}$$

The continuity equations are

$$\nabla \cdot \sigma_e\tilde{\mathbf{E}} = -j\omega\widetilde{\rho}_{ec}, \tag{22}$$

$$\nabla \cdot \sigma_m\tilde{\mathbf{H}} = -j\omega\widetilde{\rho}_{mc}. \tag{23}$$

By introducing complex permittivity and permittivity,

$$\widetilde{\overline{\overline{\epsilon}}} = \overline{\overline{\epsilon}} - \frac{j\overline{\overline{\sigma}}_e}{\omega} \tag{24}$$

$$\widetilde{\overline{\overline{\mu}}} = \overline{\overline{\mu}} - \frac{j\overline{\overline{\sigma}}_m}{\omega} \tag{25}$$

Equations 18, 19, 20, 21, 22, and 23 can be rewritten compactly as

$$\nabla \times \tilde{\mathbf{E}} = -j\omega\widetilde{\overline{\overline{\mu}}}\tilde{\mathbf{H}} - \tilde{\mathbf{M}}_i \tag{26}$$

DOI 10.1007/978-981-4560-75-7_5-1

$$\nabla \times \tilde{\mathbf{H}} = j\omega\tilde{\bar{\bar{\epsilon}}}\tilde{\mathbf{E}} + \tilde{\mathbf{J}}_i \tag{27}$$

$$\nabla \cdot \tilde{\bar{\bar{\epsilon}}}\tilde{\mathbf{E}} = \tilde{\rho}_{ei} \tag{28}$$

$$\nabla \cdot \tilde{\bar{\bar{\mu}}}\tilde{\mathbf{H}} = \tilde{\rho}_{mi} \tag{29}$$

For simplicity, in the following discussions, the tilde on phasor quantities and the complex permittivity and permeability will be dropped, keeping in mind that all these quantities are in general complex.

Electric and Magnetic Vector Potentials

In the linear regime, material properties of an isotropic medium μ_r, ϵ_r, σ_e, and σ_m are independent of the field strength, and then Maxwell's equations constitute a linear system. Then the principle of linear superposition applies, i.e., electromagnetic fields can be written as the superposition of the fields ($\mathbf{E}_A$, $\mathbf{H}_A$) generated by $\mathbf{J}_i$ and ($\mathbf{E}_F$, $\mathbf{H}_F$) generated by $\mathbf{M}_i$:

$$\mathbf{E} = \mathbf{E}_A + \mathbf{E}_F, \quad \mathbf{H} = \mathbf{H}_A + \mathbf{H}_F \tag{30}$$

Because of the principle of superposition, one can solve ($\mathbf{E}_A$, $\mathbf{H}_A$) and ($\mathbf{E}_F$, $\mathbf{H}_F$) separately and then sum them up to obtain the total fields.

In particular, ($\mathbf{E}_A$, $\mathbf{H}_A$) due to $\mathbf{J}_i$ can be determined from the vector potential $\mathbf{A}$. In a homogeneous medium, $\mathbf{A}$ is governed by the vector Helmholtz equation

$$\nabla^2\mathbf{A} + \omega^2\mu\epsilon\mathbf{A} = -\mu\mathbf{J}_i \tag{31}$$

subject to a gauge condition because the divergence of $\mathbf{A}$ needs to specified to be unique. One may choose the so-called Lorenz condition (gauge)

$$\nabla \cdot \mathbf{A} = j\omega\mu\epsilon\phi_e = 0 \tag{32}$$

as the condition for its divergence, where ϕ_e is the electric scalar potential. Alternatively, Coulomb's gauge ($\nabla \cdot \mathbf{A} = 0$) can be chosen for the divergence.

Similarly, ($\mathbf{E}_F$, $\mathbf{H}_F$) generated by $\mathbf{M}_i$ can be obtained by the electric vector potential $\mathbf{F}$. This vector potential satisfies the vector Helmholtz equation

$$\nabla^2\mathbf{F} + k^2\mathbf{F} = -\epsilon\mathbf{M}_i \tag{33}$$

subject to the Lorenz condition

$$\nabla \cdot \mathbf{F} + j\omega\mu\epsilon\phi_m = 0 \tag{34}$$

where ϕ_m is the magnetic scalar potential.

Because of the principle of superposition, the total fields due to both electric and magnetic sources are

DOI 10.1007/978-981-4560-75-7_5-1

$$\begin{aligned}\mathbf{E} &= \mathbf{E}_A + \mathbf{E}_F = -j\omega\left[\mathbf{A} + \frac{1}{k^2}\nabla(\nabla\cdot\mathbf{A})\right] - \frac{1}{\epsilon}\nabla\times\mathbf{F} \\ \mathbf{H} &= \mathbf{H}_A + \mathbf{H}_F = -j\omega\left[\mathbf{F} + \frac{1}{k^2}\nabla(\nabla\cdot\mathbf{F})\right] + \frac{1}{\mu}\nabla\times\mathbf{A}\end{aligned} \tag{35}$$

where $k^2 = \omega^2\mu\epsilon$.

By the principle of superposition, if the electric current density is $\mathbf{J}_i(\mathbf{r})$ in a homogeneous medium, the magnetic vector potential is

$$\mathbf{A}(\mathbf{r}) = \mu\int_V \mathbf{J}(\mathbf{r}')g(\mathbf{r},\mathbf{r}')d\mathbf{r}' = \mu\mathbf{J}\bigotimes g \tag{36}$$

where

$$g(\mathbf{r},\mathbf{r}') = \frac{e^{-jk|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \tag{37}$$

is called scalar Green's function for the homogeneous medium and $\bigotimes$ denotes the three-dimensional spatial convolution. Similarly, the electric vector potential due to an arbitrary volume magnetic current source $\mathbf{M}(\mathbf{r})$ is given by

$$\mathbf{F}(\mathbf{r}) = \epsilon\int_V \mathbf{M}(\mathbf{r}')g(\mathbf{r},\mathbf{r}')d\mathbf{r}' = \epsilon\mathbf{M}\bigotimes g \tag{38}$$

which is the solution to Eq. 33.

Lumped Ports and Wave Ports

In antenna simulation, one needs to apply the signal to the antenna, usually through either a lumped port (also known as a discrete port) or a wave port.

A **lumped port** consists of a voltage (or current) source and an internal impedance and is connected to the computational mesh (or grid) as a lumped element. Usually the internal impedance is chosen as the standard characteristic impedance of the transmission line (such as a coaxial cable) connected to the antenna, for example, 50 Ω. Through the field-circuit coupling, one can solve for the current (or voltage) at the lumped port as well as the electromagnetic field excited in the whole computation domain. Then the input impedance and scattering parameters can be calculated.

A **wave port** is usually a more realistic representation of the antenna source when it is connected to a transmission line or a waveguide. Its input signal is represented as one (usually the fundamental) waveguide mode of the transmission line or waveguide. The waveguide mode pattern and its corresponding propagation constant as a function of frequency are obtained by solving an eigenvalue problem of the corresponding infinitely long waveguide, over the desired frequency band. When one of these incident waveguide modes (usually the fundamental mode) encounters the antenna structure, it will be partially scattered back to the waveguide and partially transmitted out through the antenna. The normalized scattered field versus the incident field gives the scattering parameter S_{11}.

DOI 10.1007/978-981-4560-75-7_5-1

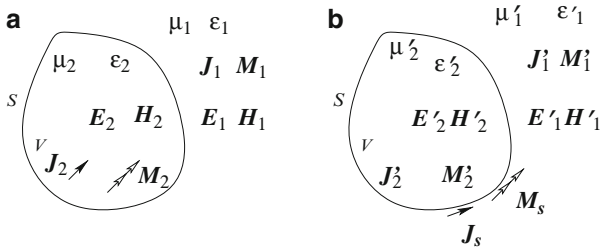


Fig. 1 Surface equivalence theorem. (**a**) The original problem. (**b**) The equivalent problem with surface sources

Near-Field and Far-Field Computation and Radiation Parameters

The lumped port or wave port inputs signal with excite electromagnetic waves in the antenna structure and gives rise to radiation of electromagnetic fields. The fields on the antenna outer surface or on a virtual box enclosing the antenna are known as the near fields. They can be obtained by solving Maxwell's equations discussed in sections.

Once the near fields are obtained, the electromagnetic fields at any location outside the antenna can be calculated by using background Green's functions. In particular, antenna engineers need to know the fields in the far zone to obtain the radiation pattern, gain, and antenna efficiency.

In order to obtain the near fields, Maxwell's equations are solved by various numerical methods in commercial antenna design tools. Described below are three commonly used methods, namely, method of moments (MOM), finite element method (FEM), and finite difference time domain (FDTD) method, as described in more detail by Liu (2015).

Method of Moments for Surface Integral Equations

Surface Equivalence Principle

The surface equivalence theorem is based on the uniqueness theorem: The field in a lossy region V is uniquely determined by the sources inside V plus the tangential electric field components specified over the surface S enclosing V (or the tangential magnetic field components or a combination of tangential **E** and **H**).

In Fig. 1, (a) is equivalent to (b) **outside** volume V if the surface sources are chosen as

$$\mathbf{J}_s = \hat{n} \times \left(\mathbf{H}_1' - \mathbf{H}_2'\right), \quad \mathbf{M}_s = -\hat{n} \times \left(\mathbf{E}_1' - \mathbf{E}_2'\right) \tag{39}$$

Note that inside and outside V, both the medium properties (μ_1', ϵ_1'), (μ_2', ϵ_2') and the fields $\left(\mathbf{E}_1' - \mathbf{H}_1'\right)$, $\left(\mathbf{E}_2' - \mathbf{H}_2'\right)$ can be chosen arbitrarily as long as (a) they satisfy Maxwell's equations in the respective new media by the new sources $\left(\mathbf{J}_1' - \mathbf{M}_1'\right)$ and $\left(\mathbf{J}_2' - \mathbf{M}_2'\right)$ and (b) they satisfy the boundary conditions on the surface S. In practice, obviously, one should make choices such that some fields are relevant to that in the original problem (otherwise, the solution to the equivalent problem will be useless).

By this surface equivalence principle, if the interested fields are $\left(\mathbf{E}_1' - \mathbf{H}_1'\right)$ outside V, the original problem can be solved if one can solve the equivalent problem (b). Here are several possible choices of the medium and fields inside V that can simplify the solution of (b).

If one chooses $\mu_1' = \mu_2' = \mu_1$, $\epsilon_1' = \epsilon_2' = \epsilon_1$, $\mathbf{E}_2' = 0$, $\mathbf{H}_2' = 0$, $\mathbf{J}_1' = \mathbf{J}_1$, $\mathbf{J}_2' = 0$, $\mathbf{M}_1' = \mathbf{M}_1$, and $\mathbf{M}_2' = 0$, then the equivalent currents become

$$\mathbf{J}_s = \hat{n} \times \mathbf{H}_1, \quad \mathbf{M}_s = -\hat{n} \times \mathbf{E}_1 \tag{40}$$

DOI 10.1007/978-981-4560-75-7_5-1

which radiate in the homogeneous medium (μ_1, ϵ_1). Therefore, the vector potentials due to these sources can be found through Green's function:

$$\mathbf{A}^{\text{sct}}(\mathbf{r}) = \mu_1 \int_S \mathbf{J}_s(\mathbf{r}') g_1(k_1 R) ds' \tag{41}$$

$$\mathbf{F}^{\text{sct}}(\mathbf{r}) = \epsilon_1 \int_S \mathbf{M}_s(\mathbf{r}') g_1(k_1 R) ds' \tag{42}$$

where $R = |\mathrm{r} - \mathrm{r}'|$ and $g_1(k_1 R) = \frac{e^{-jk_1 R}}{4\pi R}$ is Green's function in medium 1. However, again the sources in (Eq. 40) are unknown to be solved by integral equations or by some approximations. Here these surface currents will be solved using the surface integral equations described below.

Surface Integral Equations

To find the equivalent surface currents, one needs to form surface integral equations. This can be done with the above surface equivalence principle applied to the **exterior problem** and the **interior problem**, as detailed below.

The Exterior Problem: Using the above equivalence principle to the exterior problem, one has ($\mathbf{J}_{s1}$, $\mathbf{M}_{s1}$) on the surface S radiating in a homogeneous medium 1 to produce the scattered fields outside the object, where

$$\mathbf{J}_{s1} = \hat{n} \times \mathbf{H}_1 = \hat{n} \times \mathbf{H}_1^{\text{inc}} + \hat{n} \times \mathbf{H}_1^{\text{sct}} \tag{43}$$

$$\mathbf{M}_{s1} = -\hat{n} \times \mathbf{E}_1 = -\hat{n} \times \mathbf{E}_1^{\text{inc}} - \hat{n} \times \mathbf{E}_1^{\text{sct}} \tag{44}$$

where $\mathbf{E}_1^{\text{s}}$ and $\mathbf{H}_1^{\text{s}}$ represent the scattered electric and magnetic fields in the homogeneous medium 1.

The Interior Problem: Similarly, using the above surface equivalence principle to the interior problem, one has ($\mathbf{J}_{s2}$, $\mathbf{M}_{s2}$) on the surface S radiating in a homogeneous medium 2 to produce the scattered fields inside the object, where

$$-\mathbf{J}_{s2} = \hat{n} \times \mathbf{H}_2 = \hat{n} \times \left[\mathbf{H}_2^{\text{inc}} + \mathbf{H}_2^{\text{sct}}\right] \tag{45}$$

$$-\mathbf{M}_{s2} = -\hat{n} \times \mathbf{E}_2 = -\hat{n} \times \left[\mathbf{E}_2^{\text{inc}} + \mathbf{E}_2^{\text{sct}}\right] \tag{46}$$

where $\mathbf{E}_2^{\text{sct}}$ and $\mathbf{H}_2^{\text{sct}}$ represent the scattered electric and magnetic fields in the homogeneous medium 2.

If the object is dielectric, the tangential components of the electric and magnetic fields must be continuous across the interface S, or

$$\mathbf{J}_{s1} = -\mathbf{J}_{s2} \equiv \mathbf{J}_s, \quad \mathbf{M}_{s1} = -\mathbf{M}_{s2} \equiv \mathbf{M}_s \tag{47}$$

Hence, using the formulas for the scattered fields, one arrives at the surface EFIE

$$\hat{n} \times \mathbf{E}_1^{\text{inc}} = -\mathbf{M}_s - \hat{n} \times \left\{-j\omega\mu_1 \left[\mathbf{J}_s \bigotimes g_1 + \tfrac{\nabla}{k_1^2}(\nabla_s \cdot \mathbf{J}_s \bigotimes g_1)\right] - \nabla \times [\mathbf{M}_s \bigotimes g_1]\right\}_{S_+}$$
$$\hat{n} \times \mathbf{E}_2^{\text{inc}} = -\mathbf{M}_s - \hat{n} \times \left\{\left[-j\omega\mu_2 \left[\mathbf{J}_s \bigotimes g_2 + \tfrac{\nabla}{k_2^2}(\nabla_s \cdot \mathbf{J}_s \bigotimes g_2)\right] - \nabla \times [\mathbf{M}_s \bigotimes g_2]\right\}_{S_-}$$

and the surface MFIE

$$\hat{n} \times \mathbf{H}_1^{\text{inc}} = \mathbf{J}_s - \hat{n} \times \left\{ \left[-j\omega\epsilon_1 \left[\mathbf{M}_s \otimes g_1 + \frac{\nabla}{k_1^2} (\nabla_s \cdot \mathbf{M}_s \otimes g_1) \right] + \nabla \times [\mathbf{J}_s \otimes g_1] \right\}_{S_+}$$

$$-\hat{n} \times \mathbf{H}_2^{\text{inc}} = -\mathbf{J}_s - \hat{n} \times \left\{ \left[-j\omega\epsilon_2 \left[\mathbf{M}_s \otimes g_2 + \frac{\nabla}{k_2^2} (\nabla_s \cdot \mathbf{M}_s \otimes g_2) \right] - \nabla \times [\mathbf{J}_s \otimes g_2] \right\}_{S_-}$$

Note that $\mathbf{E}_2^{\text{inc}}$ and $\mathbf{H}_2^{\text{inc}}$ are the incident fields from medium 2 when there is a source inside. Therefore, unlike the incident fields in medium 1, this can be only due to a finite source but cannot be a plane wave.

The above equations only consider the observation point at $S_{\pm}$, i.e., infinitely close to the object surface. If the observation point is placed exactly on the surface, however, the terms involving the gradient of Green's function will have singularity; furthermore, the contributions of the surface convolution integral need to consider the local solid angle. Taking these into consideration, when the observation point is exactly on the object surface S, the EFIEs and MFIEs can be modified as

$$\begin{bmatrix} \mu_{r1}\mathcal{L}_1 & \mathcal{K}_1 + \mathcal{I}_1 \\ \mu_{r2}\mathcal{L}_2 & \mathcal{K}_2 + \mathcal{I}_2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{J}}_s \\ \mathbf{M}_s \end{bmatrix} = -\begin{bmatrix} \hat{n}_1 \times \mathbf{E}_1^{\text{inc}} \\ \hat{n}_2 \times \mathbf{E}_2^{\text{inc}} \end{bmatrix} \tag{48}$$

$$\begin{bmatrix} -(\mathcal{I}_1 + \mathcal{K}_1) & \epsilon_{r1}\mathcal{L}_1 \\ -(\mathcal{I}_2 + \mathcal{K}_2) & \epsilon_{r2}\mathcal{L}_2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{J}}_s \\ \mathbf{M}_s \end{bmatrix} = -\begin{bmatrix} \hat{n}_1 \times \tilde{\mathbf{H}}_1^{\text{inc}} \\ \hat{n}_2 \times \tilde{\mathbf{H}}_2^{\text{inc}} \end{bmatrix} \tag{49}$$

where $\hat{n}_1$ and $\hat{n}_2 = \hat{n}$ are the outward unit normal directions for medium 1 and medium 2, respectively, and the normalized $\tilde{\mathbf{J}}_s = \eta_0 \mathbf{J}_s$ and $\tilde{\mathbf{H}}^{\text{inc}} = \eta_0 \mathbf{H}^{\text{inc}}$ are used here. For $i = 1, 2$, one has

$$\mathcal{I}_i[\mathbf{f}] = (\hat{n}_i \cdot \hat{n}) \frac{\Omega_i}{4\pi} [\mathbf{f}] \tag{50}$$

$$\mathcal{L}_i[\mathbf{f}] = jk_0 \hat{n} \times \left[\int_S g_i(\mathbf{r}, \mathbf{r}') \mathbf{f}(\mathbf{r}') ds' + \frac{\nabla}{k_i^2} \int_S g_i(\mathbf{r}, \mathbf{r}') \nabla_s' \cdot \mathbf{f}(\mathbf{r}') ds' \right] \tag{51}$$

$$\mathcal{K}_i[\mathbf{f}] = \hat{n} \times \fint_s \nabla g_i(\mathbf{r}, \mathbf{r}') \times \mathbf{f}(\mathbf{r}') ds' \tag{52}$$

where $\mathrm{f} = \tilde{\mathbf{J}}_s$, M_s, $\hat{n} = \hat{n}_2 = -\hat{n}_1$ is the object's unit outward normal, $\fint$ represents the Cauchy principal integral, and Ω_i is the internal solid angle in region i; note that $\Omega_2 = \Omega$. Note that a Cauchy principal integral is one that does not include the contribution of the singular point at $\mathbf{r} = \mathbf{r}'$ in the above equations, which avoids the singularity of the derivative of Green's function.

For an electric impedance boundary, since $\mathbf{M} = -\hat{n} \times \widetilde{\eta}_e \tilde{\mathbf{J}}_s$ on the surface S, only the first equations above are needed. Thus, the EFIE for an electric impedance boundary is

$$[\mu_{ri}\mathcal{L}_i + (\mathcal{I}_i + \mathcal{K}_i)\hat{n} \times \widetilde{\eta}_e] \left[\tilde{\mathbf{J}}_s \right] = -\hat{n}_i \times \mathbf{E}^{\text{inc}} \tag{53}$$

and the MFIE for an impedance object is

$$[(\mathcal{I}_i + \mathcal{K}_i) - \epsilon_{ri}\mathcal{L}_i\hat{n} \times \hat{\eta}_e]\left[\tilde{\mathbf{J}}_s\right] = \hat{n}_i \times \eta_0 \mathbf{H}_i^{\text{inc}} \tag{54}$$

When $\widetilde{\eta}_e = 0$, the object becomes a PEC. Note that while the EFIE is applicable to both closed PEC objects and PEC shells, the MFIE only applies to closed PEC objects. This is because for a PEC shell, the electric current density on the shell is no longer given by $\mathbf{J}_s = \hat{n} \times \mathbf{H}$, but by $\mathbf{J}_s = \hat{n} \times (\mathbf{H}_2 - \mathbf{H}_1)$ where $\mathbf{H}_1$ and $\mathbf{H}_2$ are the magnetic fields on both sides of the shell.

Once the unknown current densities are obtained, the vector potentials $\mathbf{A}$ and $\mathbf{F}$ can be obtained through appropriate Green's functions

$$\mathbf{A}_i^{\text{sct}} = \mu_i \int_S ds' \tilde{\mathbf{J}}_s(\mathbf{r}') g_i(\mathbf{r} - \mathbf{r}'), \qquad \mathbf{F}_i^{\text{sct}} = \epsilon_i \int_S ds' \mathbf{M}_s(\mathbf{r}') g_i(\mathbf{r} - \mathbf{r}') \tag{55}$$

where $g_{1,\ 2}(\mathbf{r})$ are scalar Green's functions for the homogeneous media in the outer ($i = 1$) and inner ($i = 2$) regions, respectively.

Alternatively, the field everywhere in the ith region ($i = 1, 2$) can be found by

$$\begin{aligned} \frac{\Omega_i}{4\pi}\mathbf{E}(\mathbf{r}) &= \mathbf{E}_i^{\text{inc}}(\mathbf{r}) + \int_S \nabla_{gi}(\mathbf{r}, \mathbf{r}') \times \mathbf{M}_i(\mathbf{r}') ds' \\ &+ jk_0\eta_0\mu_{ri}\int_S \left[g_i(\mathbf{r}, \mathbf{r}')\mathbf{J}_i(\mathbf{r}') + \frac{1}{k_i^2}\nabla g_i(\mathbf{r}, \mathbf{r}')\nabla_s' \cdot \mathbf{J}_i(\mathbf{r}') \right] ds' \end{aligned} \tag{56}$$

$$\begin{aligned} \frac{\Omega_i}{4\pi}\mathbf{H}(\mathbf{r}) &= \mathbf{H}_i^{\text{inc}}(\mathbf{r}) + \int_S \nabla_{gi}(\mathbf{r}, \mathbf{r}') \times \mathbf{J}_i(\mathbf{r}') ds' \\ &+ jk_0\eta_0\epsilon_{ri}\int_S \left[g_i(\mathbf{r}, \mathbf{r}')\mathbf{M}_i(\mathbf{r}') + \frac{1}{k_i^2}\nabla g_i(\mathbf{r}, \mathbf{r}')\nabla_s' \cdot \mathbf{M}_i(\mathbf{r}') \right] ds' \end{aligned} \tag{57}$$

where $\mathbf{J}_i = -(\hat{n}_i \cdot \hat{n})\tilde{\mathbf{J}}_s$ and $\mathbf{M}_i = -(\hat{n}_i \cdot \hat{n})\mathbf{M}_s$ because of the normal directions being opposite for inner and outer regions. Note that in the above, if the observation point is away from the surface of the object, $\Omega_i = 4\pi$; when the observation point is on the surface S, only for the tangential components of the field, one uses Ω_i for the left-hand sides, but 4π for Ω_i for the normal components of $\mathbf{E}$ and $\mathbf{H}$.

Combined Field Integral Equations

Combined integral equations are used to overcome the internal resonance problem: when the operating frequency is near the object's internal resonance frequencies, the tangential electric and/or magnetic field on the object surface is not unique, as one may add any linear combination of the eigenmodes to the solution and the integral equation would still be satisfied. This is because the current due to an eigenmode of the cavity does not radiate.

The combined field integral equations can be obtained from Eqs. 48 and 49 by taking a linear combination of the two equations as

DOI 10.1007/978-981-4560-75-7_5-1

$$\begin{bmatrix} \alpha\mu_{r1}\mathcal{L}_1 - (1-\alpha)(\mathcal{I}_1 + \mathcal{K}_1) & \alpha(\mathcal{I}_1 + \mathcal{K}_1) + (1-\alpha)\epsilon_{r1}\mathcal{L}_1 \\ \alpha\mu_{r2}\mathcal{L}_2 - (1-\alpha)(\mathcal{I}_2 + \mathcal{K}_2) & \alpha(\mathcal{I}_2 + \mathcal{K}_2) + (1-\alpha)\epsilon_{r2}\mathcal{L}_2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{J}}_s \\ \mathbf{M}_s \end{bmatrix} = -\begin{bmatrix} \hat{n}_1 \times \left[\alpha\mathbf{E}_1^{\text{inc}} + (1-\alpha)\tilde{\mathbf{H}}_1^{\text{inc}}\right] \\ \hat{n}_2 \times \left[\alpha\mathbf{E}_2^{\text{inc}} + (1-\alpha)\tilde{\mathbf{H}}_2^{\text{inc}}\right] \end{bmatrix} \tag{58}$$

where $0 \leq \alpha \leq 1$ is the factor of the EFIE.

Alternatively, one can add the two equations in Eq. 48 to form one EFIE, and the equations in Eq. 49 to form one MFIE to yield the following CFIE:

$$\begin{bmatrix} \mu_{r1}\mathcal{L}_1 + \mu_{r2}\mathcal{L}_2 & \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{K}_1 + \mathcal{K}_2 \\ \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{K}_1 + \mathcal{K}_2 & -\epsilon_{r1}\mathcal{L}_1 - \epsilon_{r2}\mathcal{L}_2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{J}}_s \\ \mathbf{M}_s \end{bmatrix} = -\begin{bmatrix} \hat{n} \times \left(\mathbf{E}_1^{\text{inc}} - \mathbf{E}_2^{\text{inc}}\right) \\ \hat{n}_2 \times \left(\tilde{\mathbf{H}}_2^{\text{inc}} - \tilde{\mathbf{H}}_1^{\text{inc}}\right) \end{bmatrix} \tag{59}$$

Note that in the special case where the surface is smooth, one has $\mathcal{I}_1 - \mathcal{I}_2 = -\frac{\Omega}{4\pi}$ in the above equations. The integral equation in Eq. 59 is also known as the PMCHWT (Poggio-Miller-Chang-Harrington-Wu-Tsai) formulation based on the works by Poggio and Miller (1973), Chang and Harrington (1977), and Wu and Tsai (1977).

Method of Moments (MOM)

The surface integral equations in Eqs. 48 and 49 are referred as the SIEs in strong form. Denote $\tilde{\mathbf{J}}_s = \eta_0 \mathbf{J}_s$. Expand $\tilde{\mathbf{J}}_s$ and $\mathbf{M}_s$ in terms of divergence-conforming basis functions $\mathbf{f}_n^J$ and $\mathbf{f}_n^M$:

$$\tilde{\mathbf{J}}_s = \sum_{n=1}^{N^J} J_n \mathbf{f}_n^J(\mathbf{r}), \quad \mathbf{M}_s = \sum_{n=1}^{N^M} M_n \mathbf{f}_n^M(\mathbf{r}) \tag{60}$$

and substitute them into Eqs. 48 and 49. Let the testing functions for $\tilde{\mathbf{J}}_s$ and $\mathbf{M}_s$ be $\mathbf{w}_m^J$ and $\mathbf{w}_m^M$, respectively. By testing Eq. 48 with the testing functions $\hat{n} \times \mathbf{w}_m^J$ and $\hat{n} \times \mathbf{w}_m^M$

$$\int_S ds \begin{bmatrix} \hat{n} \times \mathbf{w}_m^J & 0 \\ 0 & \hat{n} \times \mathbf{w}_m^J \end{bmatrix} \cdot \begin{bmatrix} \mu_{r1}\mathcal{L}_1 & \mathcal{I}_1 + \mathcal{K}_1 \\ \mu_{r2}\mathcal{L}_2 & \mathcal{I}_2 + \mathcal{K}_2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{J}}_s \\ \mathbf{M}_s \end{bmatrix} = \int_S ds \begin{bmatrix} \hat{n} \times \mathbf{w}_m^J & 0 \\ 0 & \hat{n} \times \mathbf{w}_m^J \end{bmatrix} \cdot \begin{bmatrix} \hat{n} \times \mathbf{E}_1^{\text{inc}} \\ -\hat{n} \times \mathbf{E}_2^{\text{inc}} \end{bmatrix}$$

one has

$$\begin{bmatrix} \mathbf{Z}_1^J & \mathbf{Z}_1^M \\ \mathbf{Z}_2^J & \mathbf{Z}_2^M \end{bmatrix} \begin{bmatrix} \mathbf{J} \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{V}_1^E \\ \mathbf{V}_2^E \end{bmatrix} \tag{61}$$

where $\mathbf{J}$ and $\mathbf{M}$ are vectors containing the unknown expansion coefficients in Eq. 60 for $i = 1, 2$,

DOI 10.1007/978-981-4560-75-7_5-1

$$\begin{aligned}\mathbf{Z}_i^J &\equiv \int_S ds\left(\hat{n}\times\mathbf{w}_m^J\right)\cdot\mu_{ri}\mathcal{L}_i\left[\mathbf{f}_n^J\right]\\ &= jk_0\mu_{ri}\int_S ds\int_S g_i(\mathbf{r},\mathbf{r}')\mathbf{w}_m^J(\mathbf{r})\cdot\mathbf{f}_n^J(\mathbf{r}')ds'\\ &-\frac{jk_0\mu_{ri}}{k_i^2}\int_S ds\int_S g_i(\mathbf{r},\mathbf{r}')\nabla_s\cdot\mathbf{w}_m^J(\mathbf{r})\nabla_s'\cdot\mathbf{f}_n^J(\mathbf{r}')ds'\end{aligned} \tag{62}$$

$$\begin{aligned}\mathbf{Z}_i^M &\equiv \int_S ds\left(\hat{n}\times\mathbf{w}_m^J\right)\cdot\left[\mathcal{K}_i+\mathcal{I}_i\right]\left[\mathbf{f}_n^M\right]\\ &= (-1)^i\int_S ds\frac{\Omega_i(\mathbf{r})}{4\pi}\left[\hat{n}\times\mathbf{w}_m^J(\mathbf{r})\right]\cdot\mathbf{f}_n^M(\mathbf{r})\\ &+\int_S ds\fint\mathbf{w}_m^J(\mathbf{r})\cdot\left[\nabla g_i(\mathbf{r},\mathbf{r}')\times\mathbf{f}_n^M(\mathbf{r}')\right]ds'\end{aligned} \tag{63}$$

Note that integration by parts is used in the second terms of $\mathbf{Z}_1^J$ and $\mathbf{Z}_2^J$. Similarly, the excitation vectors are given by

$$\begin{aligned}\mathbf{V}_i^E &= -\int_S ds\left(\hat{n}\times\mathbf{w}_m^J\right)\cdot\left(\hat{n}_i\times\mathbf{E}_i^{\text{inc}}\right)\\ &(-1)^{i+1}\int_S ds\mathbf{w}_m^J(\mathbf{r})\cdot\mathbf{E}_i^{\text{inc}}(\mathbf{r})\end{aligned} \tag{64}$$

Note that in the above, the identity $(\hat{n}\times\mathbf{w})\cdot(\hat{n}\times\mathbf{a}) = \mathbf{a}\cdot\mathbf{w}$ if $\hat{n}\cdot\mathbf{w} = 0$ is used. Equation 61 is the MOM matrix equation for the surface integral equation for the general case.

By the same token, the MFIEs in Eq. 49 can be written in a matrix form using the MOM

$$\begin{bmatrix}\mathbf{Z}_1^J & \mathbf{Z}_1^M\\ \mathbf{Z}_2^J & \mathbf{Z}_1^M\end{bmatrix}\begin{bmatrix}\mathbf{J}\\ \mathbf{M}\end{bmatrix} = \begin{bmatrix}\mathbf{V}_1^H\\ \mathbf{V}_2^H\end{bmatrix} \tag{65}$$

where for $i = 1, 2$,

$$\begin{aligned}\mathbf{Z}_i^M &\equiv \int_S ds\left(\hat{n}\times\mathbf{w}_m^M\right)\cdot\epsilon_{ri}\mathcal{L}_i\left[\mathbf{f}_n^M\right]\\ &= jk_0\epsilon_{ri}\int_S ds\int_S g_i(\mathbf{r},\mathbf{r}')\mathbf{w}_m^M(\mathbf{r})\cdot\mathbf{f}_n^M(\mathbf{r}')ds'\\ &-\frac{jk_0\epsilon_{ri}}{k_i^2}\int_S ds\int_S g_i(\mathbf{r},\mathbf{r}')\nabla_s\cdot\mathbf{w}_m^M(\mathbf{r})\nabla_s'\cdot\mathbf{f}_n^M(\mathbf{r}')ds'\end{aligned} \tag{66}$$

DOI 10.1007/978-981-4560-75-7_5-1

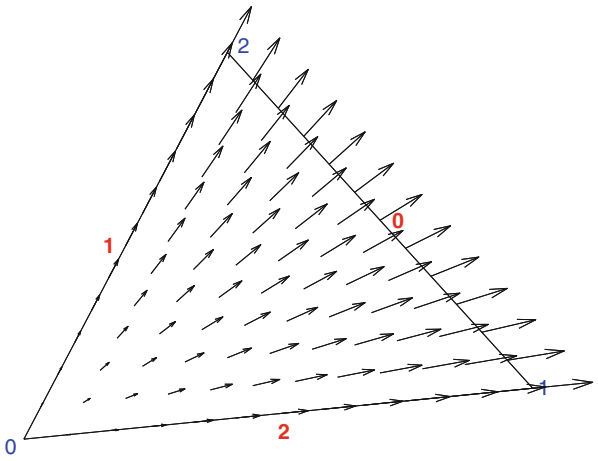


Fig. 2 Divergence-conforming basis function $\mathbf{B}_0$ associated with edge 0 (from node 1 to node 2) on a triangle. This is also known as the constant-normal/linear tangential (*CN/LT*) basis function. In computational electromagnetics, this is also better known as the Rao-Wilton-Glisson (*RWG*) function. The other two basis functions on this triangle, $\mathbf{B}_1$ and $\mathbf{B}_2$, are similar and thus not shown here

$$\begin{aligned}\mathbf{Z}_i^J \equiv & -\int_S ds\left(\hat{n}\times\mathbf{w}_m^J\right)\cdot\left[\mathcal{I}_i+\mathcal{K}_i\right]\left[\mathbf{f}_n^J\right] \\ = & (-1)^{i+1}\int_S ds\frac{\Omega_i(\mathbf{r})}{4\pi}\left[\hat{n}\times\mathbf{w}_m^M(\mathbf{r})\right]\cdot\mathbf{f}_n^J(\mathbf{r}) \\ & -\int_S ds\,{-\!\!\!\!\!\!\int}_S\mathbf{w}_m^M(\mathbf{r})\cdot\left[\nabla g_i(\mathbf{r},\mathbf{r}')\times\mathbf{f}_n^J(\mathbf{r}')\right]ds'\end{aligned} \tag{67}$$

Note that the integration by parts has been used in the second terms of $\mathbf{Z}_1^M$ and $\mathbf{Z}_2^M$. Similarly, the excitation vectors are given by

$$\begin{aligned}\mathbf{V}_i^H &= -\int_S ds\left(\hat{n}\times\mathbf{w}_m^M\right)\cdot\left(\hat{n}_i\times\tilde{\mathbf{H}}_i^{\text{inc}}\right) \\ &= (-1)^{i+1}\int_S ds\mathbf{w}_m^M(\mathbf{r})\cdot\tilde{\mathbf{H}}_i^{\text{inc}}(\mathbf{r})\end{aligned} \tag{68}$$

Divergence-Conforming Basis Functions

A Straight Triangle Element

When solving for $\mathbf{J}_s$ and $\mathbf{M}_s$ in the surface integral equations, the normal component should be continuous between adjacent elements. Vector basis functions satisfying this condition are known as divergence-conforming basis functions.

For a straight triangle element, there are three mixed-order divergence conformation basis functions. These are the so-called CN/LT basis functions, because their normal component is constant along an edge, while the tangential component with respect to an edge changes linearly within an element. These CN/LT functions are also known as the triangular rooftop or Rao-Wilton-Glisson (RWG) functions. The *i*th divergence-conforming basis corresponding to the *i*th edge of a triangle on the *xy* plane can be written as

DOI 10.1007/978-981-4560-75-7_5-1

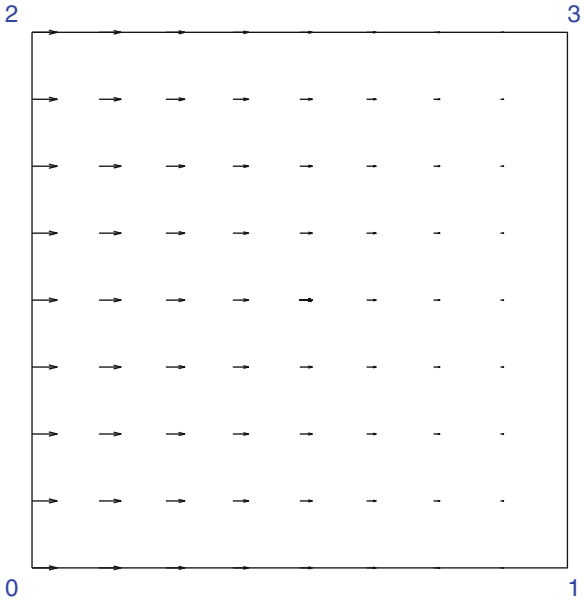


Fig. 3 Divergence-conforming basis function (CN/LT element) $\mathbf{B}_{02}^{(\xi,1)}$ in a reference square element, where subscript 02 indicates that the face is formed by vertices 0 and 2. The other three basis functions associated with the four faces, $\mathbf{B}_{01}$, $\mathbf{B}_{23}$, and $\mathbf{B}_{13}$, are similar except they are associated with the other three faces

$$\mathbf{B}_i = \omega_i \hat{z} \times (s_{i+2}\nabla s_{i+1} - s_{i+2}\nabla s_{i+2}), \quad i = 0, 1, 2 \tag{69}$$

where s_i (i = 0,1,2) are the simplex coordinates of the three vertices of the triangle. The divergence-conforming basis function $\mathbf{B}_0$ for a triangle is shown in Fig. 2.

A Reference Square Element

A curved quadrilateral element can be mapped into a reference square element by curvilinear mapping. On square reference element, the Mth order divergence-conforming basis functions are

$$\begin{aligned} \mathbf{B}^{(\xi,M)} &= \hat{\xi}\phi_i^{(M)}(\xi)\phi_j^{(M-1)}(\eta), \quad i = 0, \cdots, M; \qquad j = 1, \cdots, M-1 \\ \mathbf{B}^{(\eta,M)} &= \hat{\eta}\phi_i^{(M-1)}(\xi)\phi_j^{(M)}(\eta), \quad i = 0, \cdots, M-1; \quad j = 0, \cdots, M \end{aligned} \tag{70}$$

where $\phi_i^{(M)}(\xi)$ is the Mth mixed-order scalar basis function. For example, the 1st mixed-order rooftop functions are

$$\begin{aligned} \mathbf{B}^{(\xi,1)} &= \hat{\xi}p(\eta)t(\xi) \\ \mathbf{B}^{(\eta,1)} &= \hat{\eta}t(\eta)p(\xi) \end{aligned} \tag{71}$$

where $p(\xi)$ is a pulse basis function and $t(\xi)$ is a triangular basis function. One such 1st mixed-order rooftop function, $\mathbf{B}^{(\xi,\,1)}$, is shown in Fig. 3.

Coordinate Transformation for Divergence-Conforming Basis Functions

The above divergence-conforming basis functions have been given on a straight triangle and a square element. These can be referred to as the reference elements. In practice, however, one requires the elements to be able to conform with the problem geometry; thus, curved triangle and quadrilateral

DOI 10.1007/978-981-4560-75-7_5-1

elements are needed. The basis functions in such elements can be derived from those in the reference elements.

Contravariant Transformation: For vector integral equations, the unknowns are the electric and magnetic current densities ($\mathbf{J}_s$ and $\mathbf{M}_s$). Since these vectors have normal component continuity across adjacent elements, one needs to use the contravariant transformation

$$\mathbf{B} = \frac{1}{\det(\mathbf{J})}\mathbf{J}^T\hat{\mathbf{B}} \tag{72}$$

where $\hat{\mathbf{B}}$ is its representation in the reference element where it has normal continuity (contravariant components), or equivalently

$$\begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} = \frac{1}{\det(\mathbf{J})}\mathbf{J}^T \begin{bmatrix} \hat{B}_{\xi'} \\ \hat{B}_{\eta'} \\ \hat{B}_{\zeta'} \end{bmatrix} \tag{73}$$

Here the vector $\hat{\mathbf{B}}$ in the reference domain has been written in its contravariant components

$$\hat{\mathbf{B}} = \left(\hat{\mathbf{B}}\cdot\hat{\xi}'\right)\hat{\xi} + \left(\hat{\mathbf{B}}\cdot\hat{\eta}'\right)\hat{\eta} + \left(\hat{\mathbf{B}}\cdot\hat{\zeta}'\right)\hat{\zeta} = \hat{B}_{\xi'}\hat{\xi} + \hat{B}_{\eta'}\hat{\eta} + \hat{B}_{\zeta'}\hat{\zeta} \tag{74}$$

Examples of such field vectors are electric flux density **D** and magnetic flux density **B**. The divergence of such a vector can be written as

$$\nabla\cdot\mathbf{B} = \frac{1}{\det(\mathbf{J})}\hat{\nabla}\cdot\hat{\mathbf{B}} \tag{75}$$

where $\hat{\nabla}\cdot\hat{\mathbf{B}}$ represents the divergence in the reference element.

The surface integral equations are often applied to curved surfaces. Thus, one needs to map a curved element in three dimensions into a flat surface in two dimensions where the reference 2-D elements (e.g., triangle or square) are located. If the 2-D reference coordinates are (ξ, n), the mapping is

$$\begin{bmatrix} \dfrac{\partial}{\partial\xi} \\ \dfrac{\partial}{\partial\eta} \end{bmatrix} = \mathbf{J}\begin{bmatrix} \dfrac{\partial}{\partial x} \\ \dfrac{\partial}{\partial y} \\ \dfrac{\partial}{\partial z} \end{bmatrix}, \quad \mathbf{J} = \begin{bmatrix} \dfrac{\partial x}{\partial\xi} & \dfrac{\partial y}{\partial\xi} & \dfrac{\partial z}{\partial\xi} \\ \dfrac{\partial x}{\partial\eta} & \dfrac{\partial y}{\partial\eta} & \dfrac{\partial z}{\partial\eta} \end{bmatrix} \tag{76}$$

where the Jacobian matrix **J** is a 2 × 3 matrix and can be obtained easily by Lagrange interpolation functions (scalar basis functions) $\{\phi_n(\xi, \eta)\}$:

DOI 10.1007/978-981-4560-75-7_5-1

$$\begin{aligned} x &= \sum_n x_n \phi_n(\xi, \eta) \\ y &= \sum_n y_n \phi_n(\xi, \eta) \\ z &= \sum_n z_n \phi_n(\xi, \eta) \end{aligned} \tag{77}$$

where (x_n, y_n, z_n) are the locations of the control points. Under this transformation, the differential surface is

$$ds = |\mathbf{J}| d\xi d\eta \tag{78}$$

where $|\mathbf{J}|$ is not exactly the determinant of the Jacobian matrix as this Jacobian matrix is not square; instead, it is given by

$$|\mathbf{J}| = \sqrt{\left(\frac{\partial y}{\partial \xi}\frac{\partial z}{\partial \eta} - \frac{\partial z}{\partial \xi}\frac{\partial y}{\partial \eta}\right)^2 + \left(\frac{\partial z}{\partial \xi}\frac{\partial x}{\partial \eta} - \frac{\partial x}{\partial \xi}\frac{\partial z}{\partial \eta}\right)^2 + \left(\frac{\partial x}{\partial \xi}\frac{\partial y}{\partial \eta} - \frac{\partial y}{\partial \xi}\frac{\partial x}{\partial \eta}\right)^2} \tag{79}$$

The contravariant transformation for the divergence-conforming vector is

$$\begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} = \frac{1}{|\mathbf{J}|}\mathbf{J}^T \begin{bmatrix} \hat{B}_{\xi'} \\ \hat{B}_{\eta'} \end{bmatrix}, \quad \text{or } \mathbf{B} = \frac{1}{|\mathbf{J}|}\mathbf{J}^T \hat{\mathbf{B}} \tag{80}$$

The divergence of this vector is

$$\nabla_s \cdot \mathbf{B} = \frac{1}{|\mathbf{J}|}\hat{\nabla} \cdot \hat{\mathbf{B}} = \frac{1}{|\mathbf{J}|}\left(\frac{\partial \hat{B}_{\xi'}}{\partial \xi} + \frac{\partial \hat{B}_{\eta'}}{\partial \eta}\right) \tag{81}$$

The above expressions can be substituted into Eqs. 66 and 67 to obtain the impedance matrix in the MOM.

MOM Solution and MLFMA Acceleration

In the traditional MOM, matrix Eqs. 61 and 65 can be rewritten as

$$\mathbf{ZI} = \mathbf{V} \tag{82}$$

where $\mathbf{Z}$ is called the impedance matrix and $\mathbf{I}$ contains all the unknown expansion coefficients.

Traditionally, in the MOM the matrix Eq. 82 is solved either directly by methods such as Gauss elimination or LU decomposition or iteratively by Krylov subspace methods (Harrington 1968). However, as these impedance matrices are dense and complex, solving these matrix equations is very expensive. If the total number of unknowns in the system is N, a direct method costs (N^2) in memory and $O(N^3)$ in CPU time, which are prohibitive for large-scale problems.

The most remarkable breakthrough in integral equation solvers is the development of the fast multipole method (Rokhlin 1983; Coifman et al. 1993) and multilevel fast multipole algorithm (MLFMA, also called MLFMM) (Song and Chew 1995). It has been developed to reduce the CPU and memory requirements for the iterative solver from $O(N^2)$ to $O(N \log N)$, where N is the total number of unknowns. For more details of the MLFMA, the reader is referred to Chew et al. (2001). The MLFMA has been utilized in design tools such as FEKO.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_5-1

Finite Element Method

The three-dimensional antenna design problem can be solved by the frequency domain finite element method (FEM) through the weak form formulation. The FEM has the excellent capability of modeling arbitrary structures with its unstructured mesh and can also treat material anisotropy and dispersion easily (Jin 1993; Volakis et al. 1998). Here the goal for the FEM is to solve Maxwell's equations in a 3-D anisotropic medium. The vector Helmholtz equation can be derived from Maxwell's curl equations (Faraday's law and Ampere's law) as

$$-\nabla \times \left(\overline{\overline{\mu}}_r^{-1} \nabla \times \mathbf{E}\right) + k_0^2 \overline{\overline{\epsilon}}_r \mathbf{E} = \mathbf{S}_e \tag{83}$$

where the source terms are

$$\mathbf{S}_e = j\omega\mu_0 \mathbf{J} + \nabla \times \left(\overline{\overline{\mu}}_r^{-1} \mathbf{M}\right)$$

and the anisotropic and potentially lossy medium is characterized by the complex relative permittivity and permeability tensors $\overline{\overline{\epsilon}}_r$ and $\overline{\overline{\mu}}_r$, respectively.

Weak Form Equation

One can derive the weak form of Maxwell's equations for anisotropic media by testing Eq. 83 with a vector testing function $\mathbf{w}_m$:

$$\begin{aligned} &\int_V \left[-(\nabla \times \mathbf{w}_m) \cdot \overline{\overline{\mu}}_r^{-1} (\nabla \times \mathbf{E}) + k_0^2 \mathbf{w}_m \cdot \overline{\overline{\epsilon}}_r \mathbf{E}\right] dv \\ &= \int_S \mathbf{w}_m \cdot \hat{n} \times \overline{\overline{\mu}}_r^{-1} (\nabla \times \mathbf{E}) ds + \int_V \mathbf{w}_m \cdot \mathbf{S}_e dv = -j\omega\mu_0 \int_S \mathbf{w}_m \cdot (\hat{n} \times \mathbf{H}) ds + \int_V \mathbf{w}_m \cdot \mathbf{S}_e dv \end{aligned} \tag{84}$$

These equations must be solved in conjunction with the outer boundary conditions.

Note that for both PEC and PMC outer boundary conditions, the surface integral terms in the above two equations are equal to zero.

For an unbounded problem, exact or approximate boundary conditions can be used at the outer boundary to mimic a radiation boundary condition. In recent years, the perfectly matched layer (PML) has become a powerful absorbing boundary condition. If PML is used an outer absorbing medium (Teixeira and Chew 1998), the permittivity and permeability tensors must be modified as

$$\overline{\overline{\epsilon}}_r' = [\det(\mathbf{S})]^{-1} \mathbf{S} \overline{\overline{\epsilon}}_r \mathbf{S}, \quad \overline{\overline{\mu}}_r' = [\det(\mathbf{S})]^{-1} \mathbf{S} \overline{\overline{\mu}}_r \mathbf{S} \tag{85}$$

where

$$\mathbf{S} = \text{diag}\left[\frac{1}{e_x}, \frac{1}{e_y}, \frac{1}{e_z}\right] \tag{86}$$

and $e_\eta = a_\eta - j\omega_\eta/\omega$ is the complex stretching factor in η direction ($\eta = x, y, z$) and a_η is the scaling factor, while ω is the attenuation factor. For the special case of the PML for an isotropic medium, one has

DOI 10.1007/978-981-4560-75-7_5-1

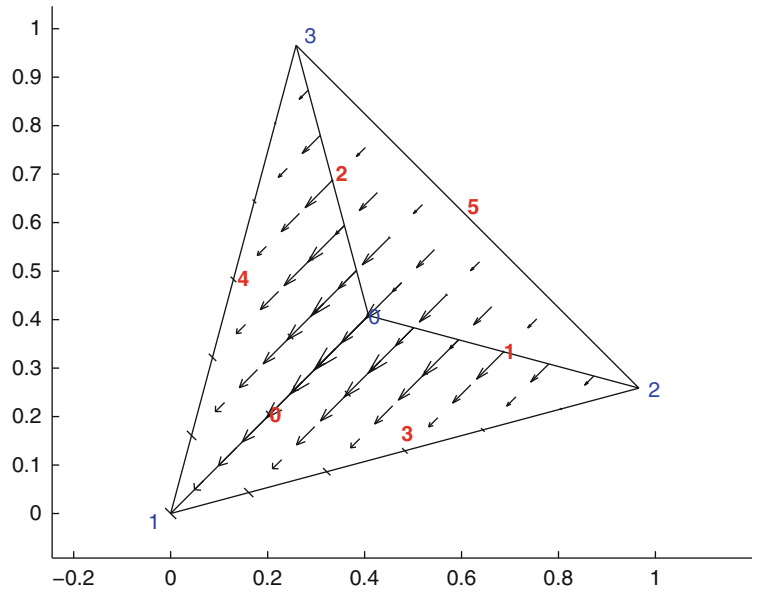


Fig. 4 CT/LN edge elements $\mathbf{B}_{ij}$ in a tetrahedron

$$\overline{\overline{\epsilon}}_r' = \epsilon_r \Lambda, \overline{\overline{\mu}}_r' = \mu_r \Lambda, \Lambda = \text{Diag}\left(\frac{e_y e_z}{e_x}, \frac{e_x e_z}{e_y}, \frac{e_x e_y}{e_z}\right) \tag{87}$$

The PML can be backed at the outer boundary by a perfect electric conductor or by a perfect magnetic conductor, making boundary integral term disappear in the weak form equation.

Curl-Conforming Vector Basis Functions

Edge Elements for Straight Tetrahedrons

For the vector electromagnetic problem, if scalar basis functions are used to expand the unknown field **E** above, they will produce spurious modes. Curl-conforming vector basis functions must be used. The most commonly used basis functions in 3-D electromagnetic problems in FEM are the mixed-order curl-conforming vector basis functions developed by Nedelec in 1980. These basis functions are commonly known as edge elements because they are based on the edges of the element, and the tangential components of these basis functions are one order lower than the normal component.

CT/LN (constant tangential/linear normal) edge elements are the lowest mixed-order edge elements for a tetrahedron in 3-D (Peterson et al. 1997). There are six such basis functions, each corresponding to one edge of the tetrahedron. The basis function for the edge *ij* (from node *i* to node *j*) is

$$\mathbf{B}_{ij} = w_{ij}\left(s_i \nabla s_j - s_j \nabla s_i\right) \tag{88}$$

where w_{ij} is the edge length between nodes *i* and *j* and s_{i} and s_{j} are the simplex coordinates of these nodes in the tetrahedron. The curl of this basis is $\nabla \times \mathbf{B}_{ij} =$ constant within the element; the divergence of this basis $\nabla \cdot \mathbf{B}_{ij} = 0$ within the element, but it is nonzero at the edges because of the discontinuity of its normal component at the edges.

The CT/LN edge elements for a tetrahedron are shown in Fig. 4.

The LT/QN (linear tangential/quadratic normal) edge elements are the next higher mixed-order basis functions for a tetrahedron. They have linear tangential component and quadratic normal component

DOI 10.1007/978-981-4560-75-7_5-1

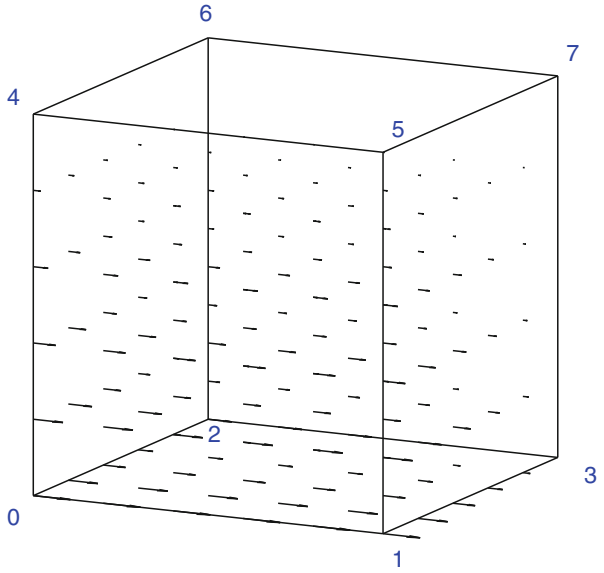


Fig. 5 CT/LN edge element $\mathbf{B}_{\xi,\,000}$ in a reference cube, where $\hat{\xi}$ is along the direction of edge 01

within an element. There are 20 such edge elements for one tetrahedron. If the four nodes of the tetrahedron are labeled as nodes 1, 2, 3, and 4, then these 20 LT/QN edge elements contain:

- 12 linear tangential functions (two per edge)

$$\mathbf{B}_{ij}^{(1)} = s_i \nabla s_j, \quad i \neq j \tag{89}$$

  where $ij = 12, 21, 13, 31, 14, 41, 23, 32, 24, 42, 34, 43$ and
- 8 quadratic normal functions (two per face)

$$\mathbf{B}_{ijk}^{(2)} = s_i s_j \nabla s_k - s_i s_k \nabla s_j, \quad i \neq j \neq k \tag{90}$$

  where $(i, j, k) = (1, 2, 3)$ and $(2, 3, 1)$ for face 123; $(1, 2, 4)$ and $(2, 4, 1)$ for face 124; $(2, 3, 4)$ and $(3, 4, 2)$ for face 234; and $(1, 3, 4)$ and $(3, 4, 1)$ for face 134. Note that these are unnormalized basis functions, i.e., their tangential components are not equal to 1 at the locations corresponding to edges.

Edge Elements for a Cube

The Mth mixed-order curl-conforming basis functions in a reference cubic element are

$$\begin{aligned}
\mathbf{B}_{\xi,\,ijk}^{(M)} &= \hat{\xi}\phi_i^{(M-1)}(\xi)\phi_j^{(M)}(\eta)\phi_k^{(M)}(\zeta), \\
&\quad i = 0, \cdots, M-1; (j,k) = 0, \cdots, M \\
\mathbf{B}_{\eta,\,ijk}^{(M)} &= \hat{\eta}\phi_i^{(M)}(\xi)\phi_j^{(M-1)}(\eta)\phi_k^{(M)}(\zeta), \\
&\quad j = 0, \cdots, M-1; (i,k) = 0, \cdots, M \\
\mathbf{B}_{\zeta,\,ijk}^{(M)} &= \hat{\zeta}\phi_i^{(M)}(\xi)\phi_j^{(M)}(\eta)\phi_k^{(M-1)}(\zeta), \\
&\quad k = 0, \cdots, M-1; (i,j) = 0, \cdots, M
\end{aligned} \tag{91}$$

The edge elements for a hexahedron are shown in Fig. 5.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_5-1

Mapping Curl-Conforming Basis Functions to Nonstandard Elements

The edge elements shown above are for standard straight tetrahedron and cube elements. In practice, one would need to be able to apply these elements to nonstandard elements such as curved tetrahedron elements and straight and curved hexahedron elements. One needs to ensure that not only the geometry is properly mapped but also the basis functions satisfy the continuity of tangential components.

If the coordinate mapping from (x, y, z) coordinates to (ξ, η, ζ) coordinates is defined by

$$\begin{bmatrix} \frac{\partial}{\partial \xi} \\ \frac{\partial}{\partial \eta} \\ \frac{\partial}{\partial \zeta} \end{bmatrix} = \mathbf{J} \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{bmatrix}, \quad \mathbf{J} = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} & \frac{\partial z}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} & \frac{\partial z}{\partial \eta} \\ \frac{\partial x}{\partial \zeta} & \frac{\partial y}{\partial \zeta} & \frac{\partial z}{\partial \zeta} \end{bmatrix} \tag{92}$$

where $\mathbf{J}$ is the Jacobian matrix, one can define the base vectors in the (ξ, η, ζ) coordinates as

$$\begin{bmatrix} \hat{\xi} \\ \hat{\eta} \\ \hat{\zeta} \end{bmatrix} = \mathbf{J} \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} \tag{93}$$

In addition, one can define the reciprocal base vectors as

$$\begin{bmatrix} \hat{\xi}' \\ \hat{\eta}' \\ \hat{\zeta}' \end{bmatrix} = \begin{bmatrix} \nabla\xi \\ \nabla\eta \\ \nabla\zeta \end{bmatrix} = \begin{bmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial \xi}{\partial y} & \frac{\partial \xi}{\partial z} \\ \frac{\partial \eta}{\partial x} & \frac{\partial \eta}{\partial y} & \frac{\partial \eta}{\partial z} \\ \frac{\partial \zeta}{\partial x} & \frac{\partial \zeta}{\partial y} & \frac{\partial \zeta}{\partial z} \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} = \frac{1}{\det(\mathbf{J})} \begin{bmatrix} \hat{\eta} \times \hat{\zeta} \\ \hat{\zeta} \times \hat{\xi} \\ \hat{\xi} \times \hat{\eta} \end{bmatrix} \tag{94}$$

Similarly, it can be shown that

$$\begin{bmatrix} \hat{\xi} \\ \hat{\eta} \\ \hat{\zeta} \end{bmatrix} = \det(\mathbf{J}) \begin{bmatrix} \hat{\eta}' \times \hat{\zeta}' \\ \hat{\zeta}' \times \hat{\xi}' \\ \hat{\xi}' \times \hat{\eta} \end{bmatrix} \tag{95}$$

The base vectors and reciprocal base vectors satisfy the orthogonality

$$\hat{\xi} \cdot \hat{\xi}' = \hat{\eta} \cdot \hat{\eta}' = \hat{\zeta} \cdot \hat{\zeta}' = 1 \tag{96}$$

$$\hat{\xi} \cdot \hat{\eta}' = \hat{\xi} \cdot \hat{\zeta}' = \hat{\eta} \cdot \hat{\zeta}' = 0 \tag{97}$$

Obviously,

$$\hat{\xi} \cdot \left(\hat{\eta} \times \hat{\zeta} \right) = \frac{1}{\hat{\xi}' \cdot \left(\hat{\eta} \times \hat{\zeta}' \right)} = \det(\mathbf{J}) \tag{98}$$

DOI 10.1007/978-981-4560-75-7_5-1

Under the above coordinate transformation, a vector **V** may be represented by its covariant components

$$\mathbf{V} = \left(\mathbf{V}\cdot\hat{\xi}\right)\hat{\xi}' + (\mathbf{V}\cdot\hat{\eta})\hat{\eta}' + \left(\mathbf{V}\cdot\hat{\zeta}\right)\hat{\zeta}' = \mathbf{V}_{\xi}\hat{\xi}' + \mathbf{V}_{\eta}\hat{\eta}' + \mathbf{V}_{\zeta}\hat{\zeta}' \tag{99}$$

or by its contravariant components

$$\mathbf{V} = \left(\mathbf{V}\cdot\hat{\xi}'\right)\hat{\xi} + (\mathbf{V}\cdot\hat{\eta}')\hat{\eta} + \left(\mathbf{V}\cdot\hat{\zeta}'\right)\hat{\zeta} = \mathbf{V}_{\xi'}\hat{\xi} + \mathbf{V}_{\eta'}\hat{\eta} + \mathbf{V}_{\zeta'}\hat{\zeta} \tag{100}$$

Covariant Mapping: To preserve the tangential continuity of a vector **E** after coordinate mapping, one should use the covariant mapping as

$$\mathbf{E} = \mathrm{J}^{-1}\hat{\mathbf{E}} \tag{101}$$

where $\hat{\mathbf{E}}$ is its representation in the reference element where it has tangential continuity. This equation can be written more explicitly as

$$\begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = \mathbf{J}^{-1}\begin{bmatrix} E_\xi \\ E_\eta \\ E_\zeta \end{bmatrix} \tag{102}$$

Examples of such field vectors are electric field **E** and magnetic field **H**.

Contravariant Mapping: On the other hand, for a field vector **B** that satisfies normal component continuity, one should use the contravariant mapping

$$\mathbf{B} = \frac{1}{\det(\mathbf{J})}\mathbf{J}^T\hat{\mathbf{B}} \tag{103}$$

where $\hat{\mathbf{B}}$ is its representation in the reference element where it has normal continuity (contravariant components), or equivalently

$$\begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} = \frac{1}{\det(\mathbf{J})}\mathbf{J}^T\begin{bmatrix} B_{\xi'} \\ B_{\eta'} \\ B_{\zeta'} \end{bmatrix} \tag{104}$$

Examples of such field vectors are electric flux density **D** and magnetic flux density **B**.

The curl of a covariant vector is a contravariant vector. Therefore, the transformation of the curl operation of a covariant vector is

$$\nabla\times\mathbf{E} = \frac{1}{\det(\mathbf{J})}\mathbf{J}^T\hat{\nabla}\times\hat{\mathbf{E}} \tag{105}$$

where $\hat{\nabla}\times\hat{\mathbf{E}}$ is the curl in the reference (ξ, η, ζ) domain. All these transformations can be used in the weak form Helmholtz equation to obtain the system matrix equation, as described below.

DOI 10.1007/978-981-4560-75-7_5-1

Solution of FEM Matrix Equation

Now consider a simple case where the outer boundary is PEC or PMC, or an unbounded problem truncated by PML backed by PEC or PMC. In this case, the surface integral term in Eq. 84 vanishes. Thus, the weak form Eq. 84 becomes

$$-\int_V (\nabla \times \mathbf{w}_m)\overline{\overline{\mu}}_r^{-1} \cdot (\nabla \times \mathbf{E})dv + k_0^2 \int_V \mathbf{w}_m \overline{\overline{\epsilon}}_r \cdot \mathbf{E}dv = \int_V \mathbf{w}_m \cdot \mathbf{S}_e dv \tag{106}$$

The unknown electric field is expanded as

$$\mathbf{E}(\mathbf{r}) \approx \sum_{n=1}^{N} e_n \mathbf{f}_n(x, y, z) \tag{107}$$

where $\mathbf{f}_n$ is the curl-conforming basis functions. Furthermore, one chooses $\mathbf{w}_m = \mathbf{f}_m$ under the Galerkin method. Then the weak form Eq. 106 is converted into a matrix equation

$$\mathbf{ZI} = \mathbf{V} \tag{108}$$

where

$$Z_{mn} = -\int_V \nabla \times \mathbf{f}_m \cdot \overline{\overline{\mu}}_r^{-1} \nabla \times \mathbf{f}_n dv + k_0^2 \int_V \mathbf{f}_m \cdot \overline{\overline{\epsilon}}_r \mathbf{f}_n dv \tag{109}$$

and $\mathbf{V}$ is the excitation vector, while $\mathbf{I}$ contains all the unknown expansion coefficients for $\mathbf{E}$.

The FEM system equation in Eq. 108 is a sparse matrix because low-order basis functions only have interactions with their adjacent elements. For a smaller system, this equation can be solved by a direct matrix inversion method such as Gaussian elimination of LU decomposition. However, if the number of degrees of freedom N is too large for a direct solver, a preconditioned iterative solution is usually performed. Recently, domain decomposition methods have gained much attention (Vouvakis et al. 2006; Zhao et al. 2007). The domain decomposition method has also started to be utilized by commercial tools such as HFSS.

Finite Difference Time Domain Method

For wideband simulations, a time domain solution is often preferred because one can obtain all results within that bandwidth with just one single simulation. The goal for a time domain method is to solve for the transient electromagnetic fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{H}(\mathbf{r}, t)$ as functions of space and time.

In this section, the finite difference time domain (FDTD) method (Yee 1966; Taflove and Hagness 2005) is present as it is arguably the most commonly used time domain methods in computational electromagnetics. The key in the FDTD method is the use of a staggered grid in both time domain and spatial domain, so that the central differencing scheme can be applied to both spatial and temporal derivatives.

Time domain Maxwell's curl equations, i.e., Faraday's law and Ampere's law,

DOI 10.1007/978-981-4560-75-7_5-1

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} - \sigma_m \mathbf{H} - \mathbf{M} \tag{110}$$

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t} + \sigma_e \mathbf{E} + \mathbf{J} \tag{111}$$

are solved in the FDTD method.

Time Domain PML Equations

For electromagnetic wave propagation and scattering in an unbounded medium, however, one needs to apply the radiation boundary condition to the above equations. Recently, the most popular radiation boundary condition is achieved by the perfectly matched layer (PML) originally proposed by Berenger (Berenger 1994; Chew and Weedon 1994). However, that PML formulation is based on the split-field components and is shown to be only weakly well posed. Here the strongly well-posed PML formulation by Fan and Liu (2003) is used.

In the strongly well-posed PML, Faraday's law and Ampere's law are modified in the PML in a compact form:

$$\nabla \times \tilde{\mathbf{E}} = -\mu \frac{\partial \tilde{\mathbf{H}}}{\partial t} - (\sigma_m + \mu \mathbf{\Lambda}_1)\tilde{\mathbf{H}} - \mathrm{M} - (\sigma_m \mathbf{\Lambda}_1 + \mu \mathbf{\Lambda}_2)\mathbf{H}^{(1)} - \sigma_m \mathbf{\Lambda}_3 \mathbf{H}^{(2)} \tag{112}$$

$$\nabla \times \tilde{\mathbf{H}} = \epsilon \frac{\partial \tilde{\mathbf{E}}}{\partial t} + (\sigma_e + \epsilon \mathbf{\Lambda}_1)\tilde{\mathbf{E}} + \mathbf{J} + (\sigma_e \mathbf{\Lambda}_1 + \epsilon \mathbf{\Lambda}_2)\mathbf{E}^{(1)} + \sigma_e \mathbf{\Lambda}_3 \mathbf{E}^{(2)} \tag{113}$$

where $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{H}}$ are the modified electric field and magnetic field inside the PML, while the field variables with superscripts (1) and (2) denote the first and second time integration of the corresponding fields. In the physical (non-PML) domain, $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{H}}$ reduce to $\mathbf{E}$ and $\mathbf{H}$, respectively, and no auxiliary fields are needed. Moreover, it is assumed that the sources $\mathbf{J}$ and $\mathbf{M}$ only exist in the physical domain. The PML auxiliary time-integrated fields are governed by ordinary partial differential equations:

$$\frac{\partial \mathbf{E}^{(1)}}{\partial t} = \tilde{\mathbf{E}} - \mathbf{\Lambda}_0 \mathbf{E}^{(1)} \tag{114}$$

$$\frac{\partial \mathbf{E}^{(2)}}{\partial t} = \mathbf{E}^{(1)} \tag{115}$$

$$\frac{\partial \mathbf{H}^{(1)}}{\partial t} = \tilde{\mathbf{H}} - \mathbf{\Lambda}_0 \mathbf{H}^{(1)} \tag{116}$$

$$\frac{\partial \mathbf{H}^{(2)}}{\partial t} = \mathbf{H}^{(1)} \tag{117}$$

In the above,

$$\mathbf{\Lambda}_0 = \mathrm{Diag}\left[\omega_x, \omega_y, \omega_z\right] \tag{118}$$

DOI 10.1007/978-981-4560-75-7_5-1

$$\mathbf{\Lambda}_1 = \mathrm{Diag}\left[\omega_y + \omega_z - \omega_x, \omega_x + \omega_z - \omega_y, \omega_x + \omega_y - \omega_z\right] \tag{119}$$

$$\mathbf{\Lambda}_2 = \mathrm{Diag}\left[\left(\omega_x - \omega_y\right)\left(\omega_x - \omega_z\right), \left(\omega_y - \omega_x\right)\left(\omega_y - \omega_z\right), \ \left(\omega_z - \omega_x\right)\ \left(\omega_z - \omega_y\right)\right] \tag{120}$$

$$\mathbf{\Lambda}_3 = \mathrm{Diag}\left[\omega_y\omega_z, \omega_z\omega_x, \omega_x\omega_y\right] \tag{121}$$

Equations 112, 113, 114, 115, 116, and 117 constitute the unsplit-field, strongly well-posed PML equations, with some auxiliary ordinary differential equations for $\mathbf{E}^{(1)}$, $\mathbf{E}^{(2)}$, $\mathbf{H}^{(1)}$, and $\mathbf{H}^{(2)}$. Note that except for the lower-order terms, Eqs. 112, and 113 are identical to original 3-D Maxwell's equations; furthermore, the lower-order terms require little additional computational cost since they are obtained by the ordinary differential equations.

In the above, ω_x, ω_y, and ω_z are the PML attenuation coefficients along the x, y, and z directions and are usually chosen to have a quadratic profile, increasing from the inner PML boundary to the outer boundary (Liu 1997, 1999).

Finite Difference Temporal Discretization

The 3-D FDTD method discretizes the fields with staggered grids in both temporal and spatial dimensions. In the temporal dimension, for example, the electric field and magnetic current density are defined at the integer time steps

$$\tilde{\mathbf{E}}^n = \mathbf{E}(\mathbf{r}, t = n\Delta t), \quad \mathrm{M}^n = \mathrm{M}(\mathbf{r}, t = n\Delta t), \tag{122}$$

while the magnetic field and electric current density are defined at the half-integer time steps

$$\mathbf{H}^{n+1/2} = \mathbf{H}(\mathbf{r}, t = (n + 1/2)\Delta t), \quad \mathbf{J}^{n+1/2} = \mathbf{J}(\mathbf{r}, t = (n + 1/2)\Delta t) \tag{123}$$

The central finite difference approximation in time gives

$$\frac{\partial f(\mathbf{r}, t)}{\partial t} = \frac{f(\mathbf{r}, t + \Delta t/2) - f(\mathbf{r}, t - \Delta t/2)}{\Delta t} + O\left(\Delta t^2\right) \tag{124}$$

where $O(\Delta t^2)$ indicates that the error term is of second order with Δt, as can be proved by using Taylor expansion. Similarly, if one uses the forward or backward finite differencing schemes, the error will be only first order. Using Eq. 124 in Eqs. 111 and 110, at $t = (n + 1/2)\Delta t$, one has

$$\epsilon(\mathbf{r})\frac{\mathbf{E}^{n+1} - \mathbf{E}^n}{\Delta t} + \sigma_e(\mathbf{r})\frac{\mathbf{E}^{n+1} - \mathbf{E}^n}{2} \approx \left(\nabla \times \mathbf{H}^{n+1/2} - \mathbf{J}^{n+1/2}\right) \tag{125}$$

Similarly, at $t = n\Delta t$, one has

$$\mu(\mathbf{r})\frac{\mathbf{H}^{n+1/2} - \mathbf{H}^{n-1/2}}{\Delta t} + \sigma_m(\mathbf{r})\frac{\mathbf{H}^{n+1/2} - \mathbf{H}^{n-1/2}}{2} \approx \left(\nabla \times \mathbf{E}^n - \mathbf{M}^n\right) \tag{126}$$

Hence,

DOI 10.1007/978-981-4560-75-7_5-1

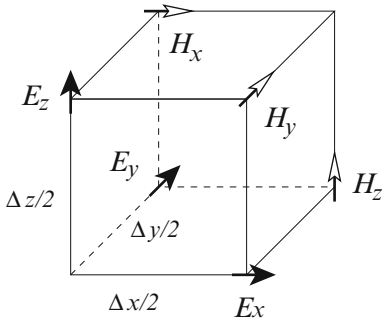


Fig. 6 The relative locations of field components in a 1/8 cell of the Yee staggered grid

$$\mathbf{H}^{n+\frac{1}{2}} = \frac{\mu - \Delta t\sigma_m/2}{\epsilon + \Delta t\sigma_m/2}\mathbf{H}^{n-\frac{1}{2}} - \frac{\Delta t}{\mu + \Delta t\sigma_m/2}\left[\nabla \times \mathbf{E}^n + \mathbf{M}^n\right] \tag{127}$$

$$\mathbf{E}^{n+1} = \frac{\epsilon - \Delta t\sigma_e/2}{\epsilon + \Delta t\sigma_e/2}\mathbf{E}^n + \frac{\Delta t}{\epsilon + \Delta t\sigma_e/2}\left[\nabla \times \mathbf{H}^{n+1/2} + \mathbf{J}^{n+1/2}\right] \tag{128}$$

This is a leap-frog system:

$$\left(\mathbf{H}^{-12}, \mathbf{E}^0\right) \rightarrow \mathbf{H}^{1/2} \rightarrow \mathbf{E}^1 \rightarrow \mathbf{H}^{3/2} \rightarrow \mathbf{E}^2 \rightarrow \mathbf{H}^{5/2} \cdots \tag{129}$$

Through the time stepping, $\mathbf{H}^{n\ +\ 1/2}$ and $\mathbf{E}^{n+1}$ can be found for all time steps provided that $\nabla \times \mathbf{E}$ and $\nabla \times \mathbf{H}$ can be obtained for the earlier time steps. The curl operations in the above are obtained by the spatial central differencing scheme below.

Finite Difference Spatial Discretization

Similar to the above staggered temporal grid where **E** and **H** are staggered by half a time step, one adopts the Yee grid for the spatial discretization so that the central differencing scheme can be applied. The computational domain of dimensions $L_x \times L_y \times L_z (x \in [x_{\min}, x_{\max} = x_{\min} + L_x]), y \in \left[y_{\min}, y_{max} = y_{\min} + L_y\right]$, $z \in \left[z_{\min}, z_{\max} = z_{\min} + L_z\right]\big)$ is discretized uniformly by $N_x \times N_y \times N_z$ cells, each with dimensions $\Delta x \times \Delta y \times \Delta z$, where $\Delta x = L_x/N_x$, $\Delta y = L_y/N_y$, and $\Delta z = L_z/N_z$ are the cell size along the x, y, and z directions, respectively. The integer grid point locations are

$$\begin{aligned} x_i &= x_{\min} + (i-1)\Delta x, \quad i = 1, \cdots, N_x + 1 \\ y_i &= y_{\min} + (j-1)\Delta y, \quad j = 1, \cdots, N_y + 1 \\ z_k &= z_{\min} + (k-1)\Delta z, \quad k = 1, \cdots, N_z + 1 \end{aligned} \tag{130}$$

while the half-integer grid points are at

$$\begin{aligned} x_{i+\frac{1}{2}} &= x_{\min} + \left(i - \frac{1}{2}\right)\Delta x, \quad i = 1, \cdots, N_x \\ y_{j+\frac{1}{2}} &= y_{\min} + \left(j - \frac{1}{2}\right)\Delta y, \quad j = 1, \cdots, N_y \\ z_{k+\frac{1}{2}} &= z_{\min} + \left(k - \frac{1}{2}\right)\Delta z, \quad k = 1, \cdots, N_z \end{aligned} \tag{131}$$

Furthermore, the temporal grid is also staggered, with the integer temporal grid points at

$$t_n = n\Delta t, \quad n = 1, \cdots, Nt \tag{132}$$

while the half-integer temporal grid points are at

$$t_{n+\frac{1}{2}} = \left(n + \frac{1}{2}\right)\Delta t, \quad n = -1, \cdots, Nt \tag{133}$$

Different components of the electromagnetic fields are located at different locations in this Yee grid, as shown in Fig. 6.

Specifically, one sets the tangential electric field components at the edge centers of a unit cell,

$$\begin{aligned} E^n_{x,\,i+1/2,j,k} &= E_x\left(x_{i+\frac{1}{2}}, y_j, z_k; t_n\right) \\ E^n_{y,\,i,j+1/2,k} &= E_y\left(x_i, y_{j+\frac{1}{2}}, z_k; t_n\right) \\ E^n_{z,\,i,j,\,k+1/2} &= E_z\left(x_i, y_j, z_{k+\frac{1}{2}}; t_n\right) \end{aligned} \tag{134}$$

and the same locations for **J**:

$$\begin{aligned} J^{n+\frac{1}{2}}_{x,\,i+1/2,j,k} &= J_x\left(x_{i+\frac{1}{2}}, y_j, z_k; t_{n+\frac{1}{2}}\right) \\ J^{n+\frac{1}{2}}_{y,\,i,j+1/2,k} &= J_y\left(x_i, y_{j+\frac{1}{2}}, z_k; t_{n+\frac{1}{2}}\right) \\ J^{n+\frac{1}{2}}_{z,\,i,j,\,k+1/2} &= J_z\left(x_i, y_j, z_{k+\frac{1}{2}}; t_{n+\frac{1}{2}}\right) \end{aligned} \tag{135}$$

The normal magnetic field components are set at the face centers of the unit cell,

$$\begin{aligned} H^{n+\frac{1}{2}}_{x,\,i,j+\frac{1}{2},k+\frac{1}{2}} &= H_x\left(x_i, y_{j+\frac{1}{2}}, z_{k+\frac{1}{2}}; t_{n+\frac{1}{2}}\right) \\ H^{n+\frac{1}{2}}_{y,\,i+\frac{1}{2},j,k+\frac{1}{2}} &= H_y\left(x_{i+\frac{1}{2}}, y_j, z_{k+\frac{1}{2}}; t_{n+\frac{1}{2}}\right) \\ H^{n+\frac{1}{2}}_{z,\,i+\frac{1}{2},j+\frac{1}{2},k} &= H_z\left(x_{i+\frac{1}{2}}, y_{j+\frac{1}{2}}, z_k; t_{n+\frac{1}{2}}\right) \end{aligned} \tag{136}$$

and the same locations for **M**:

$$\begin{aligned} M^n_{x,\,i,j+\frac{1}{2},k+\frac{1}{2}} &= M_x\left(x_i, y_{j+\frac{1}{2}}, z_{k+\frac{1}{2}}; t_n\right) \\ M^n_{y,\,i+\frac{1}{2},j,k+\frac{1}{2}} &= M_y\left(x_{i+\frac{1}{2}}, y_j, z_{k+\frac{1}{2}}; t_n\right) \\ M^n_{z,\,i+\frac{1}{2},j+\frac{1}{2},k} &= M_z\left(x_{i+\frac{1}{2}}, y_{j+\frac{1}{2}}, z_k; t_n\right) \end{aligned} \tag{137}$$

Note that in the above, the superscripts have been used to denote the temporal locations according to the central differencing scheme in temporal derivatives.

Then using the central differencing scheme (124) yields, for example,

DOI 10.1007/978-981-4560-75-7_5-1

$$\left.\frac{\partial H_z^{n+\frac{1}{2}}}{\partial y}\right|_{i+\frac{1}{2},j,k} \approx \frac{1}{\Delta y}\left[H_z^{n+\frac{1}{2}}\left(i+\frac{1}{2},j+\frac{1}{2},k\right) - H_z^{n+\frac{1}{2}}\left(i+\frac{1}{2},j-\frac{1}{2},k\right)\right] \equiv \frac{1}{\Delta y} Dy\left[H_z^{n+\frac{1}{2}}\right]\left(i+\frac{1}{2},j,k\right) \tag{138}$$

where the difference operator D_ξ ($\xi = x, y, z$) operating on a function $f_\eta^d(\eta = x, y, z;\ \ d = n \text{ or } n+\frac{1}{2})$ at spatial temporal grid point d is defined at the spatial grid point (a, b, c) as

$$D_\xi\left[f_\eta^d\right]_{a,b,c} = f^d_{\eta,(a,b,c)+\hat{\eta}\frac{1}{2}} - f^d_{\eta,(a,b,c)-\hat{\eta}\frac{1}{2}}, \quad \xi,\eta = x,y,z \tag{139}$$

Note that in the above expression, d, a, b, c are integers or half-integers depending of the field components.

Outer Boundary Conditions

Note that in the above updating equations, boundary conditions are needed. If a computational domain of $L_x \times L_y \times L_z$ is uniformly divided into $N_x \times N_y \times N_z$ cells, with $\Delta x = L_x/N_x$, $\Delta y = L_y/N_y$, and $\Delta z = L_z/N_z$, the number of field components in the whole domain is

$$\begin{aligned} E_x:&\quad N_x\left(N_y+1\right)\left(N_z+1\right)\\ E_y:&\quad \left(N_x+1\right)N_y\left(N_z+1\right)\\ E_z:&\quad \left(N_x+1\right)\left(N_y+1\right)N_z\\ H_x:&\quad \left(N_x+1\right)N_yN_z\\ H_y:&\quad N_x\left(N_y+1\right)N_z\\ H_z:&\quad N_xN_y\left(N_z+1\right)\end{aligned}$$

where it is assumed that the outer boundaries have the tangential **E** and normal **H** components. These boundary field components cannot be obtained by the time-stepping equations above, so they need to be determined by boundary conditions.

PEC Boundary Conditions

When the PEC boundary condition are required on the outer boundaries, it is convenient to make the tangential components of the electric field to be located at the outer boundaries, as discussed above where the tangential electric field components are located at the edge centers of a unit cell.

For example, PEC at $x = x_{min}$ and $x = x_{max}$ planes can be simply applied by setting

$$\begin{gathered} E_{y,1,j,k} = E_{y,N_x+1,j,k} = E_{z,1,j,k} = E_{z,N_x+1,j,k} = 0\\ H_{x,1,j,k} = H_{x,N_x+1,j,k} = 0\end{gathered} \tag{140}$$

In this case, one does not need to update these field components if they have been set zero at the beginning of the program.

Another boundary condition is the impedance boundary condition at one or all outer boundaries.

PMC Boundary Conditions

For PMC boundary conditions, it is more convenient to set the tangential components of the magnetic field at the outer boundary. This will require to shift the cell half a cell or, equivalently, exchange **E** and **H** in the above field locations. Then, one can apply the boundary conditions easily by setting

DOI 10.1007/978-981-4560-75-7_5-1

$$H_{y,1,j,k} = H_{y,N_x+1,j,k} = H_{z,1,j,k} = H_{z,N_x+1,j,k} = 0$$
$$E_{x,1,j,k} = E_{x,N_x+1,j,k} = 0 \tag{141}$$

Radiation Boundary Conditions

For unbounded media, one has to use a radiation condition to simulate the outgoing scattered fields so that there is negligible reflection from the truncated boundary at the computational edge. Several approximate radiation boundary conditions are possible:

- **Transmitting boundary condition:** Field extrapolation by using various transmitting boundary conditions, for example, Liao's, Lindman's, and Mur's absorbing boundary conditions.
- **Sponge boundary condition:** Tapered absorber outside the computational domain. This consists of a layer of lossy material of finite thickness outside the interested domain; the outer boundary conditions of this absorber are usually the PEC or PMC boundary conditions. This boundary condition is also known as the sponge boundary condition.
- **PML absorber:** In this case, the perfectly matched layer is used as a special absorber that provides highly effective attenuation to outgoing waves without giving rise to noticeable reflections.

In the following subsection, the PML absorbing boundary condition will be presented in the updating equations in the FDTD method.

The PML Updating Equations

The strongly well-posed PML for 3-D lossy media formulated by Fan and Liu (2003) is used here as presented in Eqs. 112, 113, 114, 115, 116, and 117. The discretization of these equations is similar to the regular Maxwell's equations, with the only difference being the additional ordinary differential equations in Eqs. 114, 115, 116, and 117. With the staggered grid discretization in temporal and spatial derivatives, one can obtain the field updating equations for the magnetic field:

$$\begin{aligned}\tilde{H}^{n+\frac{1}{2}}_{x,i,j+\frac{1}{2},k+\frac{1}{2}} &= B_{x1}\tilde{H}^{n-\frac{1}{2}}_{x,i,j+\frac{1}{2},k+\frac{1}{2}} - B_{x2}D_y\left[\tilde{E}^n_z\right]_{i,j+\frac{1}{2},k+\frac{1}{2}} \\ &\quad + B_{x3}D_z\left[\tilde{E}^n_y\right]_{i,j+\frac{1}{2},k+\frac{1}{2}} - B_{x4}H^{(1),n}_{x,i,j+\frac{1}{2},k+\frac{1}{2}} \\ &\quad - B_{x5}H^{(2),n}_{x,i,j+\frac{1}{2},k+\frac{1}{2}} - B_{x0}M^n_{x,i,j+\frac{1}{2},k+\frac{1}{2}}\end{aligned} \tag{142}$$

$$\begin{aligned}\tilde{H}^{n+\frac{1}{2}}_{y,i+\frac{1}{2},j,k+\frac{1}{2}} &= B_{y1}\tilde{H}^{n-\frac{1}{2}}_{y,i+\frac{1}{2},j,k+\frac{1}{2}} - B_{y2}D_z\left[\tilde{E}^n_x\right]_{i+\frac{1}{2},j,k+\frac{1}{2}} \\ &\quad + B_{y3}D_x\left[\tilde{E}^n_z\right]_{i+\frac{1}{2},j,k+\frac{1}{2}} - B_{y4}H^{(1),n}_{y,i+\frac{1}{2},j,k+\frac{1}{2}} \\ &\quad - B_{y5}H^{(2),n}_{y,i+\frac{1}{2},j,k+\frac{1}{2}} - B_{y0}M^n_{y,i+\frac{1}{2},j,k+\frac{1}{2}}\end{aligned} \tag{143}$$

$$\begin{aligned}\tilde{H}^{n+\frac{1}{2}}_{z,i+\frac{1}{2},j+\frac{1}{2},k} &= B_{z1}\tilde{H}^{n-\frac{1}{2}}_{z,i+\frac{1}{2},j+\frac{1}{2},k} - B_{z2}D_x\left[\tilde{E}^n_y\right]_{i+\frac{1}{2},j+\frac{1}{2},k} \\ &\quad + B_{z3}D_y\left[\tilde{E}^n_x\right]_{i+\frac{1}{2},j+\frac{1}{2},k} - B_{z4}H^{(1),n}_{z,i+\frac{1}{2},j+\frac{1}{2},k} \\ &\quad - B_{z5}H^{(2),n}_{z,i+\frac{1}{2},j+\frac{1}{2},k} - B_{z0}M^n_{z,i+\frac{1}{2},j+\frac{1}{2},k}\end{aligned} \tag{144}$$

where the finite difference coefficients are

DOI 10.1007/978-981-4560-75-7_5-1

$$B_{x0} = \left.\frac{2\Delta t}{2\mu + \Delta t\left(\sigma_m + \omega_{yzx}\mu\right)}\right|_{i,j+\frac{1}{2},k+\frac{1}{2}} \tag{145}$$

$$B_{x1} = \left.\frac{2\mu - \Delta t\left(\sigma_m + \omega_{yxz}\mu\right)}{2\mu + \Delta t\left(\sigma_m + \omega_{yzx}\mu\right)}\right|_{i,j+\frac{1}{2},k+\frac{1}{2}} \tag{146}$$

$$B_{x2} = B_{x0}/\Delta y, \quad B_{x3} = B_{x0}/\Delta z \tag{147}$$

$$B_{x4} = \left.\frac{2\Delta t\left(\sigma_m\omega_{yzx} + \phi_{xy}\phi_{xz}\mu\right)}{2\mu + \Delta t\left(\sigma_m + \omega_{yzx}\mu\right)}\right|_{i,j+\frac{1}{2},k+\frac{1}{2}} \tag{148}$$

$$B_{x5} = \left.\frac{2\Delta t\sigma_m\omega_y\omega_z}{2\mu + \Delta t\left(\sigma_m + \omega_{yzx}\mu\right)}\right|_{i,j+\frac{1}{2},k+\frac{1}{2}} \tag{149}$$

$$B_{y0} = \left.\frac{2\Delta t}{2\mu + \Delta t\left(\sigma_m + \omega_{xzy}\mu\right)}\right|_{i+\frac{1}{2},j,k+\frac{1}{2}} \tag{150}$$

$$B_{y1} = \left.\frac{2\mu - \Delta t\left(\sigma_m + \omega_{xzy}\mu\right)}{2\mu + \Delta t\left(\sigma_m + \omega_{xzy}\mu\right)}\right|_{i+\frac{1}{2},j,k+\frac{1}{2}} \tag{151}$$

$$B_{y2} = B_{y0}/\Delta z, \quad B_{y3} = B_{y0}/\Delta x \tag{152}$$

$$B_{y4} = \left.\frac{2\Delta t\left(\sigma_m\omega_{xzy} + \phi_{yx}\phi_{yz}\mu\right)}{2\mu + \Delta t\left(\sigma_m + \omega_{xzy}\mu\right)}\right|_{i+\frac{1}{2},j,k+\frac{1}{2}} \tag{153}$$

$$B_{y5} = \left.\frac{2\Delta t\sigma_m\omega_x\omega_z}{2\mu + \Delta t\left(\sigma_m + \omega_{xzy}\mu\right)}\right|_{i+\frac{1}{2},j,k+\frac{1}{2}} \tag{154}$$

$$B_{z0} = \left.\frac{2\Delta t}{2\mu + \Delta t\left(\sigma_m + \omega_{xzy}\mu\right)}\right|_{i+\frac{1}{2},j+\frac{1}{2},k} \tag{155}$$

$$B_{z1} = \left.\frac{2\mu - \Delta t\left(\sigma_m + \omega_{xyz}\mu\right)}{2\mu + \Delta t\left(\sigma_m + \omega_{xyz}\mu\right)}\right|_{i+\frac{1}{2},j+\frac{1}{2},k} \tag{156}$$

DOI 10.1007/978-981-4560-75-7_5-1

$$B_{z2} = B_{x0}/\Delta x, \quad B_{z3} = B_{z0}/\Delta y \tag{157}$$

$$B_{z4} = \left.\frac{2\Delta t\left(\sigma_m \omega_{xyz} + \phi_{zx}\phi_{zy}\mu\right)}{2\mu + \Delta t\left(\sigma_m + \omega_{xyz}\mu\right)}\right|_{i+\frac{1}{2},j+\frac{1}{2},k} \tag{158}$$

$$B_{z5} = \left.\frac{2\Delta t\sigma_m \omega_x \omega_y}{2\mu + \Delta t\left(\sigma_m + \omega_{xyz}\mu\right)}\right|_{i+\frac{1}{2},j+\frac{1}{2},k} \tag{159}$$

Similarly, the updating equations for electric field components are

$$\begin{aligned}\tilde{E}^{n+1}_{x,i+\frac{1}{2},j,k} = {} & A_{x1}\tilde{E}^{n}_{x,i+\frac{1}{2},j,k} + A_{x2}D_y\left[\tilde{H}_z^{n+\frac{1}{2}}\right]_{i+\frac{1}{2},j,k} \\ & - A_{x3}D_z\left[\tilde{H}_y^{n+\frac{1}{2}}\right]_{i+\frac{1}{2},j,k} - A_{x4}E^{(1),n+\frac{1}{2}}_{x,i+\frac{1}{2},j,k} \\ & - A_{x5}E^{(2),n+\frac{1}{2}}_{x,i+\frac{1}{2},j,k} - A_{x0}J^{n+\frac{1}{2}}_{x,i+\frac{1}{2},j,k}\end{aligned} \tag{160}$$

$$\begin{aligned}\tilde{E}^{n+1}_{y,i,j+\frac{1}{2},k} = {} & A_{y1}\tilde{E}^{n}_{y,i,j+\frac{1}{2},k} + A_{y2}D_z\left[\tilde{H}_x^{n+\frac{1}{2}}\right]_{i,j+\frac{1}{2},k} \\ & - A_{y3}D_x\left[\tilde{H}_z^{n+\frac{1}{2}}\right]_{i,j+\frac{1}{2},k} - A_{y4}E^{(1),n+\frac{1}{2}}_{y,i,j+\frac{1}{2},k} \\ & - A_{y5}E^{(2),n+\frac{1}{2}}_{y,i,j+\frac{1}{2},k} - A_{y0}J^{n+\frac{1}{2}}_{y,i,j+\frac{1}{2},k}\end{aligned} \tag{161}$$

$$\begin{aligned}\tilde{E}^{n+1}_{z,i,j,k+\frac{1}{2}} = {} & A_{z1}\tilde{E}^{n}_{z,i,j,k+\frac{1}{2}} + A_{z2}D_x\left[\tilde{H}_y^{n+\frac{1}{2}}\right]_{i,j,k+\frac{1}{2}} \\ & - A_{z3}D_y\left[\tilde{H}_x^{n+\frac{1}{2}}\right]_{i,j,k+\frac{1}{2}} - A_{z4}E^{(1),n+\frac{1}{2}}_{z,i,j,k+\frac{1}{2}} \\ & - A_{z5}E^{(2),n+\frac{1}{2}}_{z,i,j,k+\frac{1}{2}} - A_{z0}J^{n+\frac{1}{2}}_{z,i,j,k+\frac{1}{2}}\end{aligned} \tag{162}$$

where the finite difference coefficients are

$$A_{x0} = \left.\frac{2\Delta t}{2\epsilon + \Delta t\left(\sigma_e + \omega_{yzx}\epsilon\right)}\right|_{i+\frac{1}{2},j,k} \tag{163}$$

$$A_{x1} = \left.\frac{2\epsilon - \Delta t\left(\sigma_e + \omega_{yzx}\epsilon\right)}{2\epsilon + \Delta t\left(\sigma_e + \omega_{yzx}\epsilon\right)}\right|_{i+\frac{1}{2},j,k} \tag{164}$$

$$A_{x2} = A_{x0}/\Delta y, \quad A_{x3} = A_{x0}/\Delta z \tag{165}$$

$$A_{x4} = \left.\frac{2\Delta t\left(\sigma_e \omega_{yzx} + \phi_{xy}\phi_{xz}\epsilon\right)}{2\epsilon + \Delta t\left(\sigma_e + \omega_{yzx}\epsilon\right)}\right|_{i+\frac{1}{2},j,k} \tag{166}$$

$$A_{x5} = \left.\frac{2\Delta t\sigma_e \omega_y \omega_z}{2\epsilon + \Delta t\left(\sigma_e + \omega_{yzx}\epsilon\right)}\right|_{i+\frac{1}{2},j,k} \tag{167}$$

$$A_{y0} = \left.\frac{2\Delta t}{2\epsilon + \Delta t\left(\sigma_e + \omega_{xzy}\epsilon\right)}\right|_{i,j+\frac{1}{2},k} \tag{168}$$

$$A_{y1} = \left.\frac{2\epsilon - \Delta t\left(\sigma_e + \omega_{xzy}\epsilon\right)}{2\epsilon + \Delta t\left(\sigma_e + \omega_{xzy}\epsilon\right)}\right|_{i,j+\frac{1}{2},k} \tag{169}$$

$$A_{y2} = A_{y0}/\Delta z, \quad A_{y3} = A_{y0}/\Delta x \tag{170}$$

$$A_{y4} = \left.\frac{2\Delta t\left(\sigma_e \omega_{xzy} + \phi_{yx}\phi_{yz}\epsilon\right)}{2\epsilon + \Delta t\left(\sigma_e + \omega_{xzy}\epsilon\right)}\right|_{i,j+\frac{1}{2},k} \tag{171}$$

$$A_{y5} = \left.\frac{2\Delta t\sigma_e \omega_x \omega_z}{2\epsilon + \Delta t\left(\sigma_e + \omega_{xzy}\epsilon\right)}\right|_{i,j+\frac{1}{2},k} \tag{172}$$

$$A_{z0} = \left.\frac{2\Delta t}{2\epsilon + \Delta t\left(\sigma_e + \omega_{xyz}\epsilon\right)}\right|_{i,j,k+\frac{1}{2}} \tag{173}$$

$$A_{z1} = \left.\frac{2\epsilon - \Delta t\left(\sigma_e + \omega_{xyz}\epsilon\right)}{2\epsilon + \Delta t\left(\sigma_e + \omega_{xyz}\epsilon\right)}\right|_{i,j,k+\frac{1}{2}} \tag{174}$$

$$A_{z2} = A_{z0}/\Delta x, \quad A_{z3} = A_{z0}/\Delta y \tag{175}$$

$$A_{z4} = \left.\frac{2\Delta t\left(\sigma_e \omega_{xyz} + \phi_{zx}\phi_{zy}\epsilon\right)}{2\epsilon + \Delta t\left(\sigma_e + \omega_{xyz}\epsilon\right)}\right|_{i,j,k+\frac{1}{2}} \tag{176}$$

$$A_{z5} = \left.\frac{2\Delta t\sigma_e \omega_x \omega_y}{2\epsilon + \Delta t\left(\sigma_e + \omega_{xyz}\epsilon\right)}\right|_{i,j,k+\frac{1}{2}} \tag{177}$$

In the above, if the field components are in the PML region, the above PML equations require the time-integrated electric field components:

DOI 10.1007/978-981-4560-75-7_5-1

$$E_x^{(2),n+\frac{1}{2}} = E_x^{(2),n-\frac{1}{2}} + \Delta t E_x^{(1),n} = E_x^{(2),n-\frac{1}{2}} + \frac{\Delta t}{2}\left[E_x^{(1),n+\frac{1}{2}} + E_x^{(1),n-\frac{1}{2}}\right] = E_x^{(2),n-\frac{1}{2}} + \Delta t\left[\frac{2}{2+\omega_x\Delta t}E_x^{(1),n-\frac{1}{2}} + \frac{\Delta t}{2+\omega_x\Delta t}\tilde{E}_x^n\right] \tag{178}$$

$$E_y^{(2),n+\frac{1}{2}} = E_y^{(2),n-\frac{1}{2}} + \Delta t E_y^{(1),n} = E_y^{(2),n-\frac{1}{2}} + \frac{\Delta t}{2}\left[E_y^{(1),n+\frac{1}{2}} + E_y^{(1),n-\frac{1}{2}}\right] = E_y^{(2),n-\frac{1}{2}} + \Delta t\left[\frac{2}{2+\omega_y\Delta t}E_y^{(1),n-\frac{1}{2}} + \frac{\Delta t}{2+\omega_y\Delta t}\tilde{E}_y^n\right] \tag{179}$$

$$E_z^{(2),n+\frac{1}{2}} = E_z^{(2),n-\frac{1}{2}} + \Delta t E_z^{(1),n} = E_z^{(2),n-\frac{1}{2}} + \frac{\Delta t}{2}\left[E_z^{(1),n+\frac{1}{2}} + E_z^{(1),n-\frac{1}{2}}\right] = E_z^{(2),n-\frac{1}{2}} + \Delta t\left[\frac{2}{2+\omega_z\Delta t}E_z^{(1),n-\frac{1}{2}} + \frac{\Delta t}{2+\omega_z\Delta t}\tilde{E}_z^n\right] \tag{180}$$

$$E_x^{(1),n+\frac{1}{2}} = \frac{2-\omega_x\Delta t}{2+\omega_x\Delta t}E_x^{(1),n-\frac{1}{2}} + \frac{2\Delta t}{2+\omega_x\Delta t}\tilde{E}_x^n \tag{181}$$

$$E_y^{(1),n+\frac{1}{2}} = \frac{2-\omega_y\Delta t}{2+\omega_y\Delta t}E_y^{(1),n-\frac{1}{2}} + \frac{2\Delta t}{2+\omega_y\Delta t}\tilde{E}_y^n \tag{182}$$

$$E_z^{(1),n+\frac{1}{2}} = \frac{2-\omega_z\Delta t}{2+\omega_z\Delta t}E_z^{(1),n-\frac{1}{2}} + \frac{2\Delta t}{2+\omega_z\Delta t}\tilde{E}_z^n \tag{183}$$

where the second-order time-integrated fields are updated first because this avoids storing two steps of the first-order time-integrated fields.

Similarly, for the time-integrated magnetic field components, one has

$$H_x^{(2),n} = H_x^{(2),n-1} + \Delta t H_x^{(1),n-\frac{1}{2}} = H_x^{(2),n-1} + \frac{\Delta t}{2}\left[H_x^{(1),n} + H_x^{(1),n-1}\right] = H_x^{(2),n-1} + \Delta t\left[\frac{2}{2+\omega_x\Delta t}H_x^{(1),n-1} + \frac{\Delta t}{2+\omega_x\Delta t}\tilde{H}_x^{n-\frac{1}{2}}\right] \tag{184}$$

$$H_y^{(2),n} = H_y^{(2),n-1} + \Delta t H_y^{(1),n-\frac{1}{2}} = H_y^{(2),n-1} + \frac{\Delta t}{2}\left[H_y^{(1),n} + H_y^{(1),n-1}\right] = H_y^{(2),n-1} + \Delta t\left[\frac{2}{2+\omega_y\Delta t}H_y^{(1),n-1} + \frac{\Delta t}{2+\omega_y\Delta t}\tilde{H}_y^{n-\frac{1}{2}}\right] \tag{185}$$

$$H_z^{(2),n} = H_z^{(2),n-1} + \Delta t H_z^{(1),n-\frac{1}{2}} = H_z^{(2),n-1} + \frac{\Delta t}{2}\left[H_z^{(1),n} + H_z^{(1),n-1}\right] = H_z^{(2),n-1} + \Delta t\left[\frac{2}{2+\omega_z\Delta t}H_z^{(1),n-1} + \frac{\Delta t}{2+\omega_z\Delta t}\tilde{H}_z^{n-\frac{1}{2}}\right] \tag{186}$$

DOI 10.1007/978-981-4560-75-7_5-1

$$H_x^{(1),n} = \frac{2-\omega_x\Delta t}{2+\omega_x\Delta t} H_x^{(1),n-1} + \frac{2\Delta t}{2+\omega_x\Delta t}\tilde{H}_x^{n-\frac{1}{2}} \tag{187}$$

$$H_y^{(1),n} = \frac{2-\omega_y\Delta t}{2+\omega_y\Delta t} H_y^{(1),n-1} + \frac{2\Delta t}{2+\omega_y\Delta t}\tilde{H}_y^{n-\frac{1}{2}} \tag{188}$$

$$H_z^{(1),n} = \frac{2-\omega_z\Delta t}{2+\omega_z\Delta t} H_z^{(1),n-1} + \frac{2\Delta t}{2+\omega_z\Delta t}\tilde{H}_z^{n-\frac{1}{2}} \tag{189}$$

The above equations give all field components for all time steps if the outer boundary conditions are known (usually it is the PEC boundary condition outside the PML). The procedure for the 3-D FDTD method is as follows:

Step 0: For time step $n = 0$, initialize all field components so that they satisfy the initial conditions for $\tilde{\mathbf{H}}^{-\frac{1}{2}}$, $\tilde{E}^0$, $\mathbf{H}^{(2),-1}$, $\mathbf{H}^{(1),-1}$, $\mathbf{E}^{(2),-\frac{1}{2}}$, and $\mathbf{E}^{(1),-\frac{1}{2}}$. For non-PML regions (i.e., the physical domain), the time-integrated fields are not needed. (Most of the time these are zero before $t = 0$.) For a required time window t_{max}, find $N_t = ceil(t_{max}/\Delta t)$, where Δt should be smaller than the maximum time step allowed by the CFL stability condition.

Step 1: Update the time-integrated magnetic fields $\mathbf{H}^{(2),n}$ and then $\mathbf{H}^{(1),n}$ inside the PML.

Step 2: Update the magnetic field components $\tilde{\mathbf{H}}^{n+\frac{1}{2}}$ inside PML and $\tilde{\mathbf{H}}^{n+\frac{1}{2}}$ in the physical domain besides the outer boundary (these can use the same array). Apply boundary conditions for the magnetic field components on the outer boundary; for PML backed by a PEC outer boundary, this is not needed explicitly because the zero normal magnetic field on PEC is already set in the initial condition.

Step 3: Update the time-integrated electric field $\mathbf{E}^{(2),n+\frac{1}{2}}$ and then $\mathbf{E}^{(1),n+\frac{1}{2}}$ inside the PML.

Step 4: Update the electric field component $\mathbf{E}^{n+1}$ everywhere except for the outer boundary. Apply boundary conditions for the electric field components on the outer boundary; for PML backed by a PEC outer boundary, this is not needed explicitly because the zero tangential electric field is already set in the initial condition.

Step 5: Increase the time step index n by 1. If $n > N_t$, then exit; otherwise, go back to Step 1 above.

Note that in the non-PML (physical) region, there is no need to include the time-integrated field components.

Accuracy and Stability Conditions

As discussed above, the FDTD method uses a regular (structured) grid where the grid lines are parallel to the coordinate axes. For objects having curved surfaces, therefore, there will be a so-called staircasing error due to the staircased approximation. As a result, the FDTD has lower accuracy than the finite element method in terms of geometrical representation. Typically, one can use the following guideline to choose the discretization step:

$$\Delta x \le \frac{\lambda_{\min}}{N_{PPW}} \tag{190}$$

where $\lambda_{\min}$ is the minimum wavelength in the whole domain at the highest frequency of interest and N_{PPW} denotes the sampling density in terms of the number of points per minimum wavelength. Typically, one

needs to choose $N_{PPW} = 10$ to 20 to ensure good accuracy. However, this is only a rough guideline, as the numerical dispersion error increases linearly with the number of time steps.

Given a fixed spatial discretization, the maximum time step increment has to satisfy the following condition:

$$\Delta t \leq \frac{\Delta x}{\sqrt{D} c_{\max}} \tag{191}$$

in order for the FDTD method to be stable, where c_{max} is the maximum speed of light $c_{\max} = c/\sqrt{\min(\mu_r \epsilon_r)}$ and D is the dimensionality; for example, $D = 3$ for 3-D problems.

Techniques to Improve the FDTD Method

In a homogeneous medium or a smooth medium where the material properties change continuously with space, all field components are continuous. Therefore, the central finite difference scheme used in the FDTD method will give the second-order accuracy.

On the other hand, if the material is discontinuous inside the computational domain, the FDTD method will only have the first-order accuracy. Furthermore, for curve geometries, the FDTD method will introduce the staircasing approximation error, further decreasing the accuracy.

Despite the simplicity and robustness of the FDTD method, its lower accuracy greatly limits its applications. Over the years, many techniques have been proposed to improve the FDTD method, especially to remove the first-order accuracy and the staircasing approximation. Below only the enlarged cell technique to improve the FDTD method will be discussed.

The Enlarged Cell Technique (ECT)

A conformal finite difference time domain (CFDTD) method has been put forward by Dey and Mittra (Dey and Mittra 1997; Yu and Mittra 2000) to accurately model curved perfectly conducting objects. The method deforms the grid locally to accommodate the curvature of the PEC surface and, therefore, avoids the significant staircase error of the conventional FDTD method in approximating boundaries.

In the CFDTD method, however, the deformed grid may lead to a numerical instability due to the existence of some small-area cells. To overcome this problem, previous methods either remove the problematic small cells by assuming them totally inside the conducting bodies or reduce the time step (Dey and Mittra 1997) to obtain a stable solution. These will either decrease the accuracy or increase the run time.

The ECT was developed by Xiao and Liu (2004, 2008) to preserve the accuracy of the CFDTD method without affecting its efficiency from the FDTD method. The most important step in the ECT is to enlarge the unstable small cells into the adjacent cells so that the scheme remains stable, yet without affecting the local accuracy because it makes the electromotive force conservative.

An obvious advantage of the ECT is that it is more efficient than the conventional CFDTD method because it can obtain a stable solution without time step increment reduction. Another advantage is found by recalling that the FDTD algorithm usually has a smaller dispersion error for a larger time step. So the ECT is even more accurate than the conventional CFDTD method by allowing a larger time step. Such an ECT method has been utilized in commercial tool Wavenology.

Toward the Multiscale DGTD Method

The predominant time domain commercial tools utilize FDTD (or FIT) methods, which are limited to single-scale problems. In order to solve multiscale problems with both electrically coarse but large regions as well as electrically fine details, it is desirable to use different techniques for different regions.

DOI 10.1007/978-981-4560-75-7_5-1

The discontinuous Galerkin time domain (DGTD) method provides such a possibility. It uses the discontinuous Galerkin method; thus, different regions can have different and non-conformal meshes (or grids), thereby allowing different schemes to be used to their advantages. The finite element time domain (FETD), finite difference time domain (FDTD), and spectral element time domain (SETD) methods have been combined in the DGTD framework to solve multiscale problems (Chen and Liu 2013). Furthermore, a nonlinear circuit solver based on SPICE has been combined into this DGTD method to allow true circuit-field interactions in some commercial tools (e.g., Wavenology).

Conclusion

Commercial antenna design tools have seen rapid development over the last two decades. Although most tools use the traditional finite element method, method of moments, and finite difference time domain method, there is a noticeable trend toward solving multiscale problems by developing new technologies that can allow multi-solver integration. This is expected to greatly impact the antenna design optimization and the integration of antennas with sensor networks.

Cross-References

- ▶ Historic Review of Maxwell's Equations
- ▶ Numerical Modeling in Antenna Engineering
- ▶ Optimization Methods in Antenna Engineering

References

Berenger J-P (1994) A perfectly matched layer for the absorption of electromagnetic waves. J Comput Phy 114:185–200

Chang Y, Harrington R (1977) A surface formulation for characteristic modes of material bodies. IEEE Trans Antennas Propag 25:789–795

Chen J, Liu QH (2013) Discontinuous Galerkin time-domain methods for multiscale electromagnetic simulations: a review. Proc IEEE 101:242–254

Chew WC, Weedon WH (1994) A 3D perfectly matched medium from modified Maxwell's equations with stretched coordinates. Microwave Opt Tech Letters 7:599–604

Chew WC, Michielssen E, Song JM, Jin JM (2001) Fast and efficient algorithms in computational electromagnetics. Artech House Inc, Boston

Coifman R, Rokhlin V, Wandzura S (1993) The fast multipole method for the wave equation: a pedestrian prescription. IEEE Antennas Propag Mag 35:7–12

Dey S, Mittra R (1997) A locally conformal finite-difference time-domain (FDTD) algorithm for modeling three-dimensional perfectly conducting objects. IEEE Microwave Guided Wave Letters 7:273–275

Fan GX, Liu QH (2003) A strongly well-posed PML in lossy media. IEEE Antennas Wireless Propag Letters 2:97–100

Harrington RF (1968) Field Computation by moment methods. MacMillan, New York

Jin J (1993) The finite element method in electromagnetics. Wiley, New York

DOI 10.1007/978-981-4560-75-7_5-1

Liu QH (1997) A FDTD algorithm with perfectly matched layers for conductive media. Microw Opt Technol Lett 14:134–137

Liu QH (1999) PML and PSTD algorithm for arbitrary lossy anisotropic media. IEEE Microw Guid Wave Lett 9:48–50

Liu QH (2014) Electromagnetics theory. Class notes for ECE 571, Duke University

Liu QH (2015) Computational electromagnetics. Class notes for ECE 577, Duke University

NEC (2015) www.nec2.org. Accessed on 04 Feb 2015

Peterson AF, Ray SL, Mittra R (1997) Computational methods for electromagnetics. IEEE Press, New York

Poggio AJ, Miller EK (1973) Integral equations solutions in three dimensional scattering problems. In: Mittra R (ed) Computer techniques for electromagnetics. Pergamon, New York

Rokhlin V (1983) Rapid solution of integral equations of classical potential theory. J Comput Phy 60:187–207

Song JM, Chew WC (1995) Multilevel fast-multipole algorithm for solving combined field integral equations of electromagnetic scattering. Microw Opt Technol Lett 10:14–19

Taflove A, Hagness SC (2005) Computational electrodynamics: the finite-difference time-domain method, 3rd edn. Artech House, Boston

Teixeira FL, Chew WC (1998) General closed-form PML constitutive tensors to match arbitrary bianisotropic and dispersive linear media. IEEE Microw Guid Wave Lett 8:223–225

Volakis JL, Chatterjee A, Kempel LC (1998) Finite element method for electromagnetics. IEEE Press

Vouvakis MN, Cendes Z, Lee JF (2006) A FEM domain decomposition method for photonic and electromagnetic band gap structures. IEEE Trans Antennas Propag 54:721–733

Wu TK, Tsai LL (1977) Scattering from arbitrarily-shaped lossy dielectric bodies of revolution. Radio Sci 12:709–718

Xiao T, Liu QH (2004) Enlarged cells for the conformal FDTD method to avoid the time step reduction. IEEE Microw Wireless Compon Lett 14:551–553

Xiao T, Liu QH (2008) A 3-D enlarged cell technique (ECT) for the conformal FDTD method. IEEE Trans Antennas Propag 56:765–773

Yee K (1966) Numerical solutions of initial boundary value problems involving Maxwell's equations in isotropic media. IEEE Trans Antennas Propag 14:302–307

Yu W, Mittra R (2000) A conformal FDTD algorithm for modeling perfectly conducting objects with curve-shaped surfaces and edges. Microw Opt Technol Lett 27:136–138

Zhao K, Rawat V, Lee SC, Lee JF (2007) A domain decomposition method with nonconformal meshes for finite periodic and semi-periodic structures. IEEE Trans Antennas Propag 55:2559–2570

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_6-2

Numerical Modeling in Antenna Engineering

W. C. Chew[a]*, L. J. Jiang[b], S. Sun[b], W. E. I. Sha[b], Q. I. Dai[a], M. Fallahpour[d] and Y. M. Wu[c]
[a]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, China
[b]Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, China
[c]Key Laboratory for Information Science of Electromagnetic Waves (MoE), School of Information Science and Technology, Fudan University, Shanghai, China
[d]Department of Electrical Engineering, Stanford University, Stanford, CA, USA

Abstract

The principal computational electromagnetics techniques for solving antenna problems are reviewed. An introduction is given on a historical review of how antenna problems were solved in the past. The call for precise solutions calls for the use of numerical methods as found in computational electromagnetics. A brief introduction on differential equation solutions and integral solutions is given. The Green's function concept is introduced to facilitate the formulation of integral equations. Numerical methods and fast algorithms to solve these equations are discussed.

Then an overview of how electromagnetic theory relates to circuit theory is presented. Then the concept of partial element equivalence circuit is introduced to facilitate solutions to more complex problems. In antenna technology, one invariably has to have a good combined understanding of the wave theory and circuit theory.

Next, the discussion on the computation of electromagnetic solutions in the "twilight zone" where circuit theory meets wave theory was presented. Solutions valid for the wave physics regime often become unstable facing low-frequency catastrophe when the frequency is low.

Due to advances in nanofabrication technology, antennas can be made in the optical frequency regime. But their full understanding requires the full solutions of Maxwell's equations. Also, many models, such as perfect electric conductors, which are valid at microwave frequency, are not valid at optical frequency. Hence, many antenna concepts need rethinking in the optical regime.

Next, an emerging area of the use of eigenanalysis methods for antenna design is discussed. This can be the characteristic mode analysis or the natural mode analysis. These analysis methods offer new physical insight not possible by conventional numerical methods.

Then the discussion on the use of the domain decomposition method to solve highly complex and multi-scale antenna structures is given. Antennas, due to the need to interface with the circuit theory, often have structures ranging from a fraction of a wavelength to a tiny fraction of a wavelength. This poses a new computational challenge that can be overcome by the domain decomposition method.

Many antenna designs in the high-frequency regime or the ray optics regime are guided by ray physics and the adjoining mathematics. These mathematical techniques are often highly complex due to the rich physics that come with ray optics. The discussion on the use of these new mathematical techniques to reduce computational workload and offering new physical insight is given.

A conclusion section is given to summarize this chapter and allude to future directions.

*Email: w-chew@uiuc.edu

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_6-2

Keywords

Computational electromagnetics; Antenna modeling; Fast algorithms; Circuit theory; Wave theory; PEEC; Low-frequency breakdown; Calderon preconditioning; Nanoantennas; Design rules; Gain; Directivity; Characteristic mode analysis; Normal mode expansion; Domain decomposition; Equivalence principle algorithm; High-frequency asymptotics; Numerical steepest descent path

Introduction

Maxwell's equations, completed by James Clerk Maxwell in 1864, have been guiding electromagnetic theory for over 150 years now (Maxwell 1865). Solutions to Maxwell's equations have offered physical insight and guidance for myriads of electromagnetics technologies including antenna design. In the beginning, it was the use of analytic solutions. For instance, the solution of a radiating Hertzian dipole offers insight on the physics of its radiation. It provides physical insight on its radiation pattern, the near field and far field of the dipole, its radiation resistance, its directivity, and its gain. These are all figures of merit for an antenna designer.

Due to the complex shapes of antennas, closed-form solutions are quickly exhausted. Then approximate solutions are sought, and they have been used to gain physical insight into the antenna design. For instance, the half-wave dipole has no closed-form solution, but physical intuition tells that the current distribution is close to being sinusoidal. Hence, many figures of merit of a half-wave dipole can be calculated based on approximate models.

Another area of antenna design where approximation methods reign superior is in high-frequency antennas or reflector antennas. Due to the large radius of curvature of the antennas, quasi-optical techniques such as the physical optics approximation can be used to ascertain the radiation field of the antenna. Ray optics can be used to gain physical insight. Edge diffraction from the edge of the parabolic reflector can be estimated by using the canonical solution of the Sommerfeld half-plane (Sommerfeld 1896). Uniform asymptotic theory and uniform theory of diffraction are both used to refine the solutions (Keller 1962; Kouyoumjian and Pathak 1974; Lee and Deschamps 1976; Hansen 1981). The idea of approximate methods to gain physical insight prevails up to the late 1970s and early 1980s even for microstrip antenna designs (Lo et al. 1979; Chew and Kong 1981).

Even though modeling and simulation are important, physical insight is equally important. The most famous of this is the design of the Yagi-Uda antenna, proposed by Professor Yagi and his student, Uda, in 1926 (Yagi and Uda 2014). They had the insight that an array of dipoles can act like a waveguide. The simple dipole, made of a length of straight wire, is also called a linear antenna. When this linear antenna is slightly less than a half wavelength, it is capacitive, and an array of them can act like a waveguide. Hence, the dipole array, put in front of a half-wave dipole, can help guide the energy of the radiation field in the forward direction, enhancing its radiation in that direction and, hence, its directivity. When the linear antenna is slightly larger than a half wavelength, it becomes inductive and ceases to be a waveguide. Such a linear antenna is used as a reflector in the Yagi-Uda array.

When numerical methods became popular, integral equation methods were designed to solve linear antennas more accurately. Examples of such integral equation are the Hallén integral equation and Pocklington integral equation (Balanis 2012b). Approximate numerical method together with semi-analytic methods was used to solve these integral equations. Semi-analytic methods were popular because of the slowness of computers back then (King 1956).

DOI 10.1007/978-981-4560-75-7_6-2

Computers became powerful very quickly with the advancement of the electronic computers. Differential equation solvers as well as integral equation solvers were developed to solve Maxwell's equations. In differential equation solvers, the unknowns are the field. Hence, the unknown counts are spread over volumetric space and, therefore, scale as the volume of the space of L^3 where L is the typical length of the simulation region. Moreover, in order to simulate an infinite region, absorbing boundary conditions are needed at the boundary of the simulation domain.

In integral equation solvers, the unknowns are the sources or the currents on the antenna. In many cases, these unknowns reside on the surface of the antenna structure, and thus, they live in a 2D manifold. In the cases of antennas that can be modeled by wire structures, the unknowns reside on a 1D manifold. Consequently, integral equation solvers became popular because the unknown count can be smaller. Very soon, integral equations were solved with the method of moments and applied to linear antennas. Linear antennas were popular because the unknowns were functions of a one-dimensional variable, and henceforth, they live in a one-dimensional space. Usually, many practical problems can be solved with a small number of unknowns and therefore needing the solution of small matrix systems.

As the antenna structure becomes more complex, more unknowns are needed. Early complex antenna structures were modeled with wire structures as in the numerical electromagnetic code (NEC). Very soon, surface patch models were developed and popularized (Burke and Poggio 1981; Rao et al. 1982).

A drawback of the integral equation solver is the memory usage. The matrix system is often dense and hence consumes $O(N^2)$ memory usage. For antennas that can be modeled as surface patches, $N \sim O((kL)^2)$ where L is the typical dimension of the antenna. Henceforth, memory usage scales as $(kL)^4$ for them. When used to solve linear antennas, $N \sim O(kL)$ and therefore the memory usage scales as $(kL)^2$.

On the other hand, differential equation solvers are easier to implement even though unknown counts are larger. Moreover, the innate matrix system that is related to a differential equation is inherently sparse. Therefore, when combined with iterative solvers, the matrix system can be efficiently stored with $O(N)$ memory where N is the dimension of the matrix and, hence, the number of unknowns. Furthermore, each matrix-vector product can be effected in $O(N)$ operations. Consequently, the iterative solver can be effected with $O(N)$ operations per iteration. Typically, the memory usage and CPU time usage scale as $(kL)^3$ when grid dispersion error is ignored. Hence, it could be more efficient than integral equation solvers.

A powerful differential equation solver is the finite-difference time-domain method invented by Yee (1966) and popularized by Taflove (1995). In this method, the construction of the matrix system is not needed, and a matrix-vector product is effected by some rather simple operations. Since only the unknown fields at different time steps need to be stored, the memory and CPU time usage are both $O(N)$. Even though FDTD is solved with time stepping, it is equivalent to iteration in iterative solvers. Iterative solvers are generally applied to frequency-domain solutions.

Due to advancements of differential equation solvers, their efficiency, and their simplicity, they have become rather popular. They can be used to solve problems with tens of thousands of unknowns easily and, hence, many practical real-world problems.

In recent years, however, fast integral equation solvers have been developed (Rokhlin 1990; Coifman et al. 1993; Chew et al. 2001). In these solvers, iterative methods are used to solve integral equations without the need to construct the matrix system. Thus, the matrix is never stored and only the action of the matrix on a vector needs to be stored. Henceforth, memory requirements of the fast integral equation solvers are reduced to $O(N \log N)$ for dynamic problems while to $O(N)$ to static or low-frequency problems. Moreover, the matrix-vector product can be effected in $O(N \log N)$ operations. So, the memory usage for an antenna that can be modeled by surface currents scales as $(kL)^2$ and the operations per iteration scale as $(kL)^2 \log kL$, greatly reducing the memory and CPU usage compared to differential equation solvers.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_6-2

Differential Equations-Maxwell's Equations

Maxwell's equations establish that the theory of electricity and magnetism and theory of optics are actually one and the same theory. Both of them have waves propagating at the velocity of light. The velocity of light, and astronomical number, was established by Roemer via astronomical observation long ago (in year 1676).

Maxwell's equations in the time domain are completely described by the four equations:

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{1}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{2}$$

$$\nabla \cdot \mathbf{D} = \varrho \tag{3}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{4}$$

When Maxwell's equations were first derived by Maxwell, they were not as elegant and succinct as the above. It was the work of Oliver Heaviside who casts them into the above form (Heaviside 1888).

For time-varying problems, the third and the fourth equations are derivable from the first two by taking their divergence. Hence, only the first two equations can be considered as independent, and only them need to be solved for time-varying problems.

Since there are four unknowns **E**, **H**, **D**, and **B** with two independent equations, the constitutive relations are the addition equations in order to obtain four equations and four unknowns, viz., for isotropic media,

$$\mathbf{D} = \epsilon \mathbf{E} \tag{5}$$

$$\mathbf{B} = \mu \mathbf{H} \tag{6}$$

The above constitutive relations are sufficient to describe most materials at microwave frequencies. For some materials such as ferrites, the above will have to be replaced with the constitutive relations for anisotropic media.

The Yee algorithm will solve the first two of the above first-order partial differential equations directly. But many algorithms will first convert the above into one second-order partial differential equation. By assuming $\exp(-i\omega t)$ dependence, $\partial/\partial t$ can be replaced by $-i\omega$ in the above equations. Then by dividing the first equation by ϵ, taking its curl, and then replacing the resultant right-hand side with the second Maxwell's equations, one arrives at

$$\nabla \times \epsilon^{-1} \nabla \times \mathbf{H} - \omega^2 \mu \mathbf{H} = \nabla \times \epsilon^{-1} \mathbf{J} \tag{7}$$

Similarly, one has

$$\nabla \times \mu^{-1} \nabla \times \mathbf{E} - \omega^2 \epsilon \mathbf{E} = i\omega \mathbf{J} \tag{8}$$

DOI 10.1007/978-981-4560-75-7_6-2

One needs only solve one of the two equations above since they are derivable from each other. When the above equation is solved by a numerical method, the field **H** or **E** is treated as the unknown, and the source is assumed known.

Scalar and Dyadic Green's Function

The Green's function is useful because it expresses the field in terms of the source of a system. It is the point source response of the partial differential equation or the field generated due to a point source. When a medium is homogeneous, such a response or Green's function can be found easily. It is well known that when the Helmholtz equation is driven by a point source, the response is the scalar Green's function, namely,

$$\nabla^2 g(\mathbf{r}, \mathbf{r}') + k^2 g(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}') \tag{9}$$

and $g(\mathbf{r}, \mathbf{r}') = \exp(ik|\mathbf{r} - \mathbf{r}'|)/(4\pi|\mathbf{r} - \mathbf{r}'|)$ and $k = \omega\sqrt{\mu\epsilon} = \omega/c$. Hence, when the medium is homogeneous, the above vector wave equation can be written as

$$\nabla\nabla \cdot \mathbf{E} - \nabla \cdot \nabla\mathbf{E} - k^2\mathbf{E} = i\omega\mu\mathbf{J} \tag{10}$$

where the identity has been used that $\nabla \times \nabla \times \mathbf{E} = \nabla\nabla \cdot \mathbf{E} - \nabla \cdot \nabla\mathbf{E}$. Furthermore, by noting that $\nabla \cdot \mathbf{E} = \varrho/\epsilon$, the above becomes

$$\nabla^2\mathbf{E} + k^2\mathbf{E} = -i\omega\mu\mathbf{J} + \nabla\varrho/\epsilon \tag{11}$$

By using the principle of linear superposition and the solution to Eq. 9, the solution to the above can be written as

$$\mathbf{E}(\mathbf{r}) = i\omega\mu\int_V d\mathbf{r}' g(\mathbf{r}, \mathbf{r}')\mathbf{J}(\mathbf{r}') - \int_V d\mathbf{r}' g(\mathbf{r}, \mathbf{r}')\nabla'\varrho(\mathbf{r}')/\epsilon \tag{12}$$

$$= i\omega\mu\int_V d\mathbf{r}' g(\mathbf{r}, \mathbf{r}')\mathbf{J}(\mathbf{r}') - \frac{1}{i\omega\epsilon}\nabla\int_V d\mathbf{r}' g(\mathbf{r}, \mathbf{r}')\nabla' \cdot \mathbf{J}(\mathbf{r}') \tag{13}$$

The above is reminiscent of that

$$\mathbf{E} = i\omega\mathbf{A} - \nabla\phi \tag{14}$$

One can also symbolically write that

$$\mathbf{E}(\mathbf{r}) = i\omega\mu\int_V d\mathbf{r}'\overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}(\mathbf{r}') \tag{15}$$

where the action of the dyadic Green's function on the current **J** plus the integral is to effect the integral in Eq. 12 above.

Integral Equation of Excitation

An antenna is usually made of metal and excited by a source. The source generates an incident field on the antenna. At microwave frequencies, metallic structures can be approximated by a perfect electric

DOI 10.1007/978-981-4560-75-7_6-2

conductor (PEC) where the tangential component of the electric field is zero. Otherwise, there would be infinite surface current flowing on the surface of the PEC structure.

As a result, a surface current is induced on the antenna so as to produce a field that cancels the incident field on the surface of the antenna structure. This physical concept can be represented by the following integral equation:

$$-\hat{t} \cdot \mathbf{E}_{inc}(\mathbf{r}) = i\omega\mu\hat{t} \cdot \int_S dS' \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}(\mathbf{r}'), \quad \mathbf{r} \in S \tag{16}$$

Source Excitation Methods

Reflector antennas are usually driven by a feed which is usually a horn antenna. The modeling of the feed can be done by the equivalence principle. For smaller antennas at lower frequencies, oftentimes, an antenna is driven by a current source or voltage source at its terminal. The equivalence principle can be used to obtain the equivalence problem that needs to be solved before approximations are made.

Equivalence Principle

The equivalence principle can be derived mathematically by invoking Green's theorem for scalar waves or vector Green's theorem for electromagnetic waves. But it can be argued physically or by the Gedankenexperiment. Imaging a set of sources generates **E** and **H** fields in free space. One defines a surface S on which one will impress the surface electric current $\mathbf{J}_s = \hat{n} \times \mathbf{H}$ and surface magnetic current $\mathbf{M}_s = \mathbf{E} \times \hat{n}$. One assumes that the fields $\mathbf{E} = \mathbf{H} = 0$ inside S and that the original fields are unchanged outside S. These fields definitely satisfy the jump conditions induced by $\mathbf{J}_s$ and $\mathbf{M}_s$ at the interface. Furthermore, the fields outside satisfy the boundary conditions (including the radiation condition at infinity) on the surface S, and hence, by the uniqueness theorem, they are the only solution (Figs. 1, 2, 3, 4, and 5).

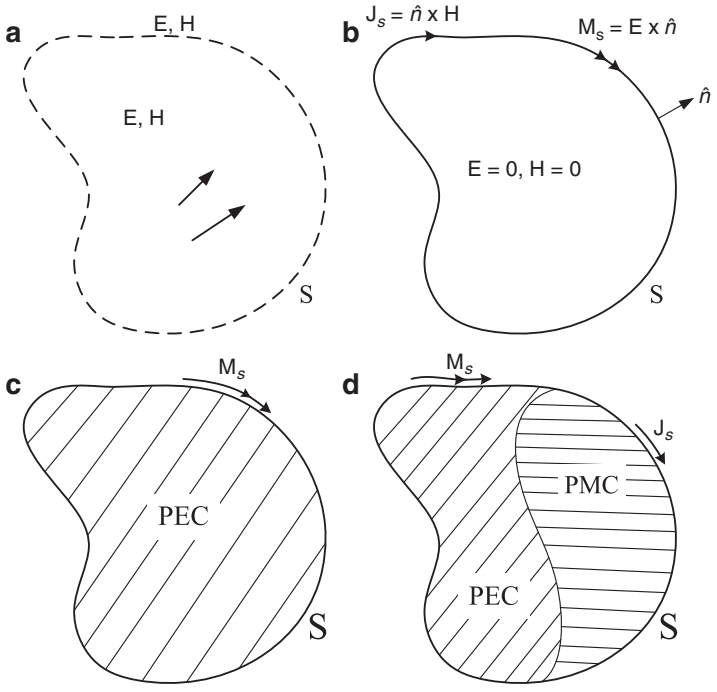


Fig. 1 By the equivalence principle, the four cases (**a**–**c**) and (**d**) generate equivalent fields outside the surface S

DOI 10.1007/978-981-4560-75-7_6-2

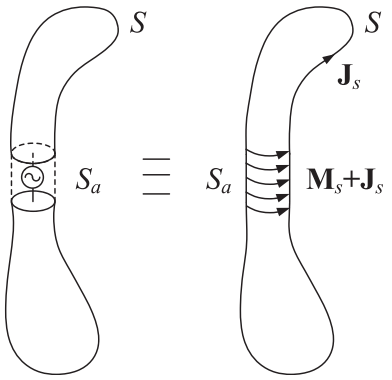


Fig. 2 For a voltage source driving an antenna at the gap, the equivalence principle can be invoked to make the case on the right to be equivalent to the case on the *left*. When the gap is filled with PEC, the electric current can be removed, with only the magnetic current impressed on it. When the gap is filled with PMC, the magnetic current can be removed with the electric current impressed on the PMC surface

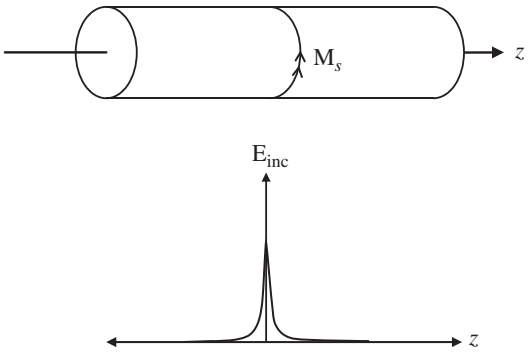


Fig. 3 An antenna driven by a very thin magnetic current ribbon. The incident field generated by the magnetic current ribbon can be approximated by a delta function

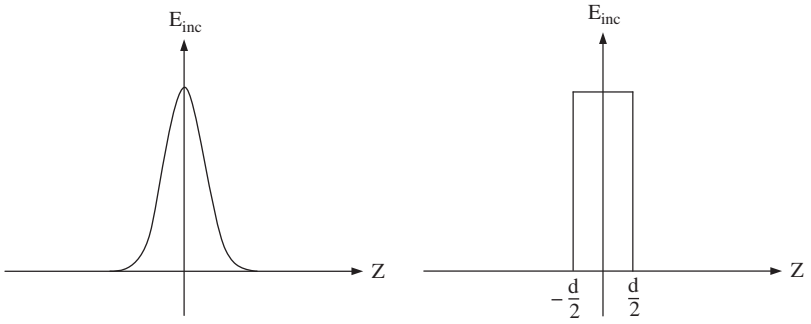


Fig. 4 The plot of the incident field generated by a finite-width magnetic ribbon. The incident field generated by the finite-width magnetic current ribbon can be approximated by a pulse function

Since the fields are zero inside S, one can insert a PEC inside S without affecting the fields outside. When the PEC fills the whole of the volume inside S, then the surface current $\mathbf{J}_s$ cannot radiate, as it will induce an image current that is negative of itself. Consequently, the magnetic current $\mathbf{M}_s$ alone, when impressed on the PEC surface, will radiate the same fields outside. A similar argument can be made when the volume is filled with PMC or a mixture thereof.

DOI 10.1007/978-981-4560-75-7_6-2

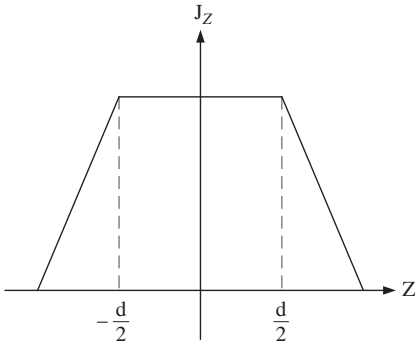


Fig. 5 The antenna can also be excited by a current source. The current source can be represented by one basis function with known and fixed amplitude

With the above equivalence principle established, one can use that to derive an equivalence model for an antenna driven by a source at its gap. First, by wrapping a surface S on the antenna snugly, one can represent the fields outside the antenna with impressed currents $\mathbf{J}_s$ and $\mathbf{M}_s$ per the above discussion. On the PEC surface of the antenna, $\mathbf{M}_s = 0$ since $\hat{n} \times \mathbf{E} = 0$. The only nonzero surface currents are $\mathbf{J}_s$ and $\mathbf{M}_s$ at the gap region. Now, one can fill the gap with PEC, and again, $\mathbf{J}_s$ does not radiate. Hence, only $\mathbf{M}_s$ impressed on the surface associated with the gap region is needed to produce the equivalent fields outside the antenna.

The impressed magnetic current can be modeled by a magnetic ribbon current. This magnetic current can be used to generate the incident field on the antenna, which will in turn induce a current on it. What is to be noted is that the magnitude of the magnetic current is related to the voltage across the gap. When the magnetic current ribbon is very thin, it produces a field that is sharply peaked at the location just beneath the ribbon. This sharply peaked function can be approximated by a delta function. Therefore, one can approximate the incident field just as a delta function. This is the well-known delta-gap model.

When the magnetic ribbon has a finite width, the incident field it generates will have a Gaussian-like pulse shape. This can be used as the incident field on the antenna. This Gaussian pulse will induce a voltage across the gap. But for simplicity, one can approximate the Gaussian pulse incident field with a square pulse with the same area. It will induce the same voltage drop across the gap (Lo et al. 2013).

To realize a magnetic ribbon source in the real world, one can use a small toroid as shown in Fig. 6. As is well known, a small loop antenna is like a magnetic dipole, and the toroid can be thought of as a string of magnetic dipoles forming a circle emulating a magnetic current loop. The current induced on the antenna wire is a transformer action, and the antenna wire is the secondary winding of the transformer. Such toroidal antennas have been used in well-logging exploration tools (Clark 1985).

Alternatively, one can fill the gap region with a perfect magnetic conductor (PMC) and impress only the electric current $\mathbf{J}_s$ as the equivalent current on the gap surface. One can solve this excitation problem to obtain $\hat{n} \times \mathbf{E}$ on the surface of the PMC gap and hence the voltage. With this voltage, one can calculate the input impedance of the antenna.

But a gap filled with a PMC material is an idealization of an open circuit where no current can flow. As an approximation, one can replace the gap with one filled with free space again and drive the antenna with a current source at the gap. The resultant voltage at the gap can then be calculated to obtain the input impedance. This is the current source driven antenna excitation model.

Subspace Projection Method

Integral equations and differential equations previously derived represent linear operator equations. One can describe a general linear problem with a linear operator equation as

DOI 10.1007/978-981-4560-75-7_6-2

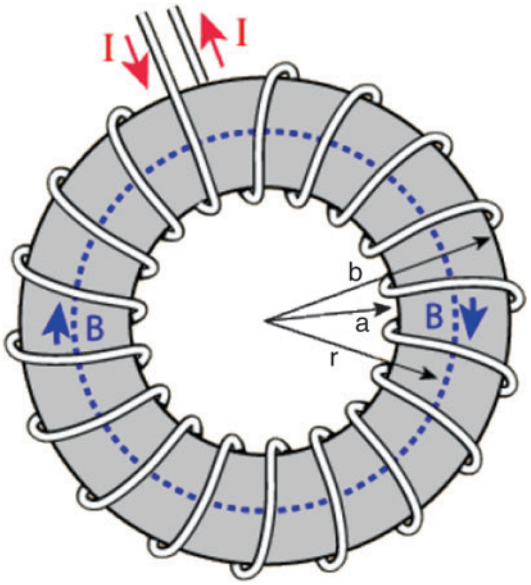


Fig. 6 A magnetic ribbon source can be approximated by a small toroid. The figure is from hyperphysics.phy-astr.gsu.edu/hbase/magnetic/toroid.html

$$\mathcal{L}f = h \tag{17}$$

where $\mathcal{L}$ represents a linear operator. This linear operator represents a differential operator in the case of differential equations, while it represents an integral operator for integral equations. In mathematical notation, one often writes $\mathcal{L} : V \rightarrow W$, meaning that it is a linear operator that maps elements of the vector space V to the elements of the vector space W, where V is known as the domain space and W the range space of the operator $\mathcal{L}$.

One illustrates the subspace projection method with scalar equations as the mathematical concept does not differ greatly when vector equations are used. For the differential operator, an explicit form may be

$$\left(\nabla^2 + k^2\right) f(\mathbf{r}) = h(\mathbf{r}) \tag{18}$$

which is the Helmholtz wave equation or the frequency-domain version of the wave equation. A boundary condition together with the radiation condition or a loss condition has to be stipulated in order to make the solution to the above equation unique (Chew 1990).

For the integral equation, one can consider a scalar integral equation and an explicit form is given:

$$\int d\mathbf{r}' g(\mathbf{r}, \mathbf{r}') f(\mathbf{r}') = h(\mathbf{r}) \tag{19}$$

where $g(\mathbf{r}, \mathbf{r}') = \exp(ik|\mathbf{r} - \mathbf{r}'|)/(4\pi|\mathbf{r} - \mathbf{r}'|)$ is the Green's function which is also called the kernel.

These equations are not amenable to computation. They are equations in the infinite-dimensional Hilbert space with uncountably infinite indices. To make them computable, one can find a subspace in which an approximate solution of the above equations can be found and then project the solution into this subspace. This process can be methodically described as follows.

One chooses a basis set with N basis functions f_n, $n = 1, \ldots, N$ that spans the subspace that can approximate the domain space of the operator. One expands the unknown f in terms of the basis functions in this subspace, namely,

DOI 10.1007/978-981-4560-75-7_6-2

$$f(\mathbf{r}) \doteq \sum_{n=1}^{N} a_n f_n(\mathbf{r}) \tag{20}$$

The linear operator equation can then be written as

$$\sum_{n=1}^{N} a_n \mathcal{L} f_n \doteq h \tag{21}$$

The above equation can only be approximately satisfied since one has picked a finite number of basis functions. Moreover, the solution to Eq. 21 is still untenable. To make it more easily solvable, one converts Eq. 21 into a matrix equation by weighting or testing the equation with $w_m(\mathbf{r})$, $m = 1, \ldots, N$. Consequently, one has

$$\sum_{n=1}^{N} a_n \langle w_m, \mathcal{L} f_n \rangle = \langle w_m, h \rangle, \quad m = 1, \cdots, N \tag{22}$$

where the inner product is defined as

$$\langle u, v \rangle = \int u(\mathbf{r}) v(\mathbf{r}) d\mathbf{r} \tag{23}$$

The integral is over the support of the functions u and v. When they are defined over a volume (surface), one has a volume (surface) integral.

The above Eq. 22 implies that it can be exactly satisfied if one projects the Eq. 21 onto a subspace spanned by $w_m(\mathbf{r})$, $m = 1, \ldots, N$.

Recent research has shown that if the testing functions w_m are chosen from the dual space of the range space, the above can be a good approximation or matrix representation of the original operator equation (Cools et al. 2009). The dual space can be larger or smaller than the range space.

Equation 22 now constitutes a matrix equation:

$$\overline{\mathbf{L}} \cdot \mathbf{a} = \mathbf{h} \tag{24}$$

$$\left(\overline{\mathbf{L}}\right)_{mn} = \langle w_m, \mathcal{L} f_n \rangle \tag{25}$$

$$(\mathbf{a})_n = a_n \tag{26}$$

$$(\mathbf{h})_m = \langle w_m, h \rangle \tag{27}$$

$$(\mathbf{a}) = \left(\overline{\mathbf{L}}\right)^{-1} \cdot \mathbf{h} \tag{28}$$

The matrix equation can be solved by various means. In the above, the choice of the testing function is determined by the auxiliary equation (Chew 1990). The testing function should be chosen so that the auxiliary equation is solved accurately, whereas the choice of the expansion or basis function is to solve the primary equation accurately. The auxiliary equation is the adjoint or the transpose of the original equation where the range space and the domain space are swapped. It can be shown that the left domain

DOI 10.1007/978-981-4560-75-7_6-2

space of an operator is the dual space of the range space. Hence, if the testing functions solve the auxiliary equation well, they approximate the left domain space of the original operator, which is the dual space of the range space of the original equation, well. This is in agreement with the aforementioned recent findings.

The above procedure of converting the operator equation into a matrix equation is the underpinning method behind the finite-element method or the method of moments. They are variously known as Galerkin's method, Petrov–Galerkin's method, method of weighted residuals, collocation method, and point matching method (Chew 1990, Chap. 5). But they all can be regarded as subspace projection methods.

When $\mathcal{L}$ is a differential operator, the matrix $\overline{\mathbf{L}}$ is sparse because a differential operator is a local operator. However, when $\mathcal{L}$ is an integral operator, the matrix $\overline{\mathbf{L}}$ is dense. In differential equations, the unknown is the field that permeates all of space. Hence, the unknown count is usually large in differential equation solvers. But they come with sparse matrix systems that are cheaper to solve and store.

For integral equations, the unknown is the induced current on the antenna structure. Hence, the unknown count is smaller since the current resides only on or in a structure with finite support. But the ensuing matrix system is dense and is hard to store. Moreover, the matrix assembly of forming the matrix elements

$$\langle w_m, \mathcal{L} f_n \rangle \tag{29}$$

is tedious since the operator involves singularity that has to be evaluated with care (Graglia 1993).

Iterative Methods

When the unknown count is large, the matrix $\overline{\mathbf{L}}$ is never generated. Instead, iterative, matrix-free methods in numerical linear algebra are used to solve the matrix equation instead. Iterative methods can be made matrix-free because one needs only to write a code to produce the result

$$\overline{\mathbf{L}} \cdot \mathbf{a} \tag{30}$$

namely, the result of the action of a matrix on a vector, but not the matrix itself. For sparse matrices, this action can be effected with $O(N)$ operations. For dense matrices, it costs $O(N^2)$ to effect this action. However, various fast solvers have been developed that allow this action to be effected in $O(N)$ or $O(N \log N)$ operations (Chew et al. 2001, ref. therein).

When it comes to solving a matrix equation iteratively, numerical linear algebra plays an important role (Trefethen et al. 1997; Chew et al. 2009). The condition number and the distribution of the eigenvalues of the matrix system determine the convergence rate of iterative methods. The Krylov subspace method is a popular way to understand the convergence rate of iterative methods. In this method, one finds the solution to the matrix equation by finding the best-fit solution in a subspace called the Krylov subspace

$$\mathcal{K}^K\left(\overline{\mathbf{L}}, \mathbf{r}_0\right) = \left\{\mathbf{r}_0, \overline{\mathbf{A}} \cdot \mathbf{r}_0, \overline{\mathbf{A}}^2 \cdot \mathbf{r}_0, \cdots, \overline{\mathbf{A}}^{K-1} \cdot \mathbf{r}_0\right\} \tag{31}$$

where $\mathbf{r}_0 = \mathrm{h} - \overline{\mathbf{L}} \cdot \mathrm{a}_0$ where a_0 is the initial guess to the solution. The above Krylov subspace can be generated by performing $K-1$ matrix-vector product with the matrix $\overline{\mathbf{L}}$ on the vector $\mathbf{r}_0$. The method finds the optimal solution at the K-th iteration, $\mathbf{a}_K$, by letting it be $\mathbf{a}_K = \mathbf{a}_0 + \mathbf{z}_K$, such that $\mathbf{z}_K \in \mathcal{K}^K\left(\overline{\mathbf{L}}, \mathbf{r}_0\right)$. By so doing, the residual at the K-th iteration is

DOI 10.1007/978-981-4560-75-7_6-2

$$\begin{aligned}\mathbf{r}_K &= \mathbf{h} - \overline{\mathbf{L}} \cdot \mathbf{a}_K \\ &= \mathbf{r}_0 - \overline{\mathbf{L}} \cdot \mathbf{z}_K \in \left\{\mathbf{r}_0, \overline{\mathbf{L}} \cdot \mathbf{r}_0, \overline{\mathbf{L}}^2 \cdot \mathbf{r}_0, \cdots, \overline{\mathbf{L}}^K \cdot \mathbf{r}_0\right\} \\ &\in \mathcal{K}^{K+1}\left(\overline{\mathbf{L}}, \mathbf{r}_0\right)\end{aligned} \tag{32}$$

In other words, the residual error that is to be minimized can be written as

$$\begin{aligned}\mathbf{r}_K &= \mathbf{r}_0 + \alpha_1 \overline{\mathbf{L}} \cdot \mathbf{r}_0 + \alpha_2 \overline{\mathbf{L}}^2 \cdot \mathbf{r}_0 + \cdots + \alpha_K \overline{\mathbf{L}}^K \cdot \mathbf{r}_0 \\ &= \sum_{k=0}^{K} \alpha_k \overline{\mathbf{L}}^k \cdot \mathbf{r}_0 = P_K^o\left(\overline{\mathbf{L}}\right) \cdot \mathbf{r}_0\end{aligned} \tag{33}$$

where

$$P_K^o(x) = 1 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_K x^K$$

is an optimal polynomial with $\alpha_0 = 1$ and the coefficients $\alpha_1, \ldots, \alpha_K$ are chosen to minimize $\mathbf{r}_K$.

To understand the convergence of the above matrix polynomial, one expands r_0 in terms of the left and right eigenvectors of $\overline{\mathbf{L}}$ or (Chew et al. 2009, ref. therein)

$$\mathbf{r}_0 = \sum_{n=1}^{N} \mathbf{v}_n \left(\mathbf{w}_n^t \cdot \mathbf{r}_0\right) \tag{34}$$

where $\mathbf{v}_n$ is the right eigenvector, while $\mathbf{w}_n$ is the left eigenvector. They can be shown to share the same set of eigenvalues and are mutually orthogonal. That is, $\mathbf{w}_n \cdot \mathbf{v}_{n'} = \delta_{nn'}$. Therefore,

$$\overline{\mathbf{L}} \cdot \mathbf{v}_n = \lambda_n \mathbf{v}_n \tag{35}$$

$$\mathbf{w}_n^t \cdot \overline{\mathbf{L}} = \lambda_n \mathbf{w}_n^t \tag{36}$$

where the eigenvalue λ_n can be complex. One can easily show that

$$\overline{\mathbf{L}}^k \cdot \mathbf{r}_0 = \sum_{n=1}^{N} \lambda_n^k \mathbf{v}_n \left(\mathbf{w}_n^t \cdot \mathbf{r}_0\right) \tag{37}$$

By substituting the above into Eq. 33, one has

$$\mathbf{r}_K = \sum_{n=1}^{N} \sum_{k=0}^{K} \alpha_k \lambda_n^k \mathbf{v}_n \left(\mathbf{w}_n^t \cdot \mathbf{r}_0\right) = \sum_{n=1}^{N} P_K^o(\lambda_n) \mathbf{v}_n \left(\mathbf{w}_n^t \cdot \mathbf{r}_0\right) \tag{38}$$

$$\mathbf{r}_K = \sum_{k=0}^{K} \overline{\mathbf{L}}^k \cdot \mathbf{r}_0 = \sum_{n=1}^{N} P_K^o(\lambda_n) \mathbf{v}_n \left(\mathbf{w}_n^t \cdot \mathbf{r}_0\right) \tag{39}$$

$$\mathbf{r}_0 = g - \overline{\mathbf{L}} \cdot \mathbf{a}_0 \tag{40}$$

$$\mathbf{r}_1 = \mathbf{r}_0 + \alpha_1' \overline{\mathbf{L}} \cdot \mathbf{r}_0 \tag{41}$$

DOI 10.1007/978-981-4560-75-7_6-2

It is seen that the residual error in the K-th iteration is proportional to the value of the optimal polynomial $P_K^o(\lambda_n)$ at the N eigenvalues λ_n. If $K = N$, a polynomial can be found such that $P_K^o(\lambda_n)$ is exactly zero at all these eigenvalues, meaning that the residual error is zero. When K is less than N, if the eigenvalues are clustered together on the complex plane, the residual error can still be made small. Since the polynomial has value such that $P_K^o(0) = 1$, if there are many eigenvalues near to the origin, it is difficult to fit the polynomial so that it is small close to the origin. Hence, having eigenvalues close to the origin is bad for convergence, since it is hard to find the fitting polynomial. By the same token, if the eigenvalues are spread widely over the complex plane, it is hard to find a fitting polynomial as well. All these cases give rise to a bad condition number of the matrix system which is defined as $|\lambda_{max}|/|\lambda_{min}|$, the ratio of the largest eigenvalue to the smallest eigenvalue. Therefore, an ill-conditioned matrix has bad convergence when solved with iterative solvers.

Fast Algorithm for Integral Equations

Due to the increased workload of numerical methods when the antenna structure becomes electrically large (large compared to wavelength), there has been much interest in fast methods to solve the ensuing matrix equations derived from Maxwell's equations.

Differential equation solvers naturally give rise to sparse matrices. The solution process can be made matrix-free easily. The downside is the existence of the grid dispersion error whose deleterious effect increases with the size of the problem (Lee and Cangellaris 1992). In this case, the effort has been in reducing the unknown counts as the unknowns are fields that permeate all of space pervasively. Also, accuracy improvement is necessary to mitigate grid dispersion errors. A higher-order method is used to reduce unknown counts but with the peril of reducing the sparsity of the matrix. Also, the unknown count in differential equation solvers grows with the volume of the simulation region and, hence, suffers from the cruelty of dimensionality.

Methods to invert the finite-element matrix directly have been studied extensively. Because of the sparsity of the FEM matrix for differential equations, it can be directly inverted in $N^{1.5}$ in 2D and N^2 in 3D by the nested dissection ordering method (Axelsson and Barker 1984). Moreover, for domains where the shape is oblong, the frontal method proves popular in inverting the finite-element matrix.

Another hot area of research in differential equation solvers is the design of absorbing boundary conditions (ABCs) (Chew 1990, Chap. 4). Berenger's perfectly matched layers (PMLs) (Berenger 1994) and coordinate-stretching PML (Chew and Weedon 1994) have become highly popular among numericists in this area. Coordinate-stretching PML draws inspirations from the area of matched asymptotics where coordinate stretching is used to emphasize a certain physics of the problem. Furthermore, an anisotropic-medium PML has been developed (Sacks et al. 1995). This topic has spurred the interest of many researchers.

Most ABCs are not perfect; a rigorous, perfect ABC is actually a boundary integral equation truncation of a differential equation solution domain solved by FEM or FDM (Jin 2002; Volakis et al. 1998; Zhu and Cangellaris 2006). This has been avoided in the past because boundary integral equations give rise to dense matrix systems that are expensive to solve and store. But the advent of fast algorithms has changed the landscape (Chew et al. 2001; Peng et al. 2011). Boundary integral equations accelerated by fast algorithms have been used to reduce the domain size of finite-element methods since they act as rigorous absorbing boundary conditions.

Cruelty of Computational Complexity

Integral equations were difficult and expensive to solve in the past. They usually give rise to a dense matrix system requiring $O(N^3)$ computer time to solve and $O(N^2)$ memory requirements. These computational complexities are just too unwieldy for large problems. But advances in fast methods have

DOI 10.1007/978-981-4560-75-7_6-2

eliminated this bottleneck by reducing the size of α in the exponent in N^{α} in these computational complexity scalings.

For iterative solvers, a sleuth of method can reduce the computational time complexity to $O(N \log N)$. Moreover, many of these methods can be made matrix-free so that only the unknown vector and the diagonal part of the matrix corresponding to near-neighbor interactions need to be stored. Hence, the memory requirements can be reduced to $O(N)$. There are essentially three popular methods to speed up the solutions of integral equations:

- Fast Fourier transform-based techniques: In this technique, the integral operator is cast into a form that resembles a convolutional integral as much as possible. Then FFT is used to expedite the convolution, enabling its evaluation of matrix-vector products in $O(N \log N)$ operations. For surface structures or sparsely packed structures where a lot of zero padding in the FFT is needed, the complexity is worse and is not optimal (Borup and Gandhi 1984).
- Matrix compression-based techniques: The MOM matrix that follows from solving integral equations is low rank. The reasons for low rank are two: over-discretization of mesh density and far interactions between currents on the object. The Nyquist sampling theorem necessitates the discretization of at least two points per wavelength in order to capture the oscillatory nature of the currents on an object. But oftentimes, discretization far above the Nyquist sampling rate is used. In this case, redundancies are created in the unknown counts, and the rank of the ensuing matrix system is lower rank than the unknown counts. Such matrix systems can be easily compressed using matrix compression techniques, such as wavelets (Wagner et al. 1993, ref. therein), adaptive cross approximations, simple fast multipole, etc. (Zhao et al. 2004, ref. therein). Interaction matrices with low ranks due to far interactions are harder to compress. They cannot be compressed beyond the Nyquist barrier (Wagner et al. 1993), even though ray physics-based methods have been used to compress them further beyond the Nyquist barrier. When wave physics is involved, the only viable way to compress the matrices for far interaction efficiently is the multipole-based methods (Chew et al. 2001).
- Multipole-based methods: The simple fast multipole method can expedite the solution involving circuit physics or Laplace's solutions very easily. Extension of such an algorithm for wave physics cases has been proposed, but its verbatim use for wave physics cases does not work. The only viable method of expediting the wave physics case is the multilevel fast multipole algorithm (Chew et al. 2001), where anterpolation and interpolation between levels are added. The algorithm is a tree-based algorithm. The matrices for far interactions are analytically diagonalized on paper rather than by algebraic or numerical means. The number of diagonal elements needed is proportional to the rank of the matrices. The reason is that the ranks of the matrices increase with the group size compared to wavelength. Near the upper levels of the inverted tree, the ranks of the matrices are higher, while near the lower levels, their ranks are lower.

To be of the varying ranks of these matrices, their dimensions are different for different levels even if they are diagonalized. Hence, anterpolation and interpolation are needed between levels due to their different ranks, very much in the spirit of the multilevel multigrid schemes. With this augmentation, the multilevel fast multipole algorithm provides optimal complexity of performing a matrix-vector product with $O(N \log N)$ complexity. It is with this algorithm that wave physics problems with tens of millions to hundreds of millions and over three billion unknowns have been solved (Michiels et al. 2014).

The key to the multilevel fast multipole algorithm is the factorization of the matrix element L_{ij} when elements i and j are far apart. Then one can express (Chew et al. 2001).

$$L_{ij} = \tilde{\mathbf{V}}^{t}_{f,i,i_1} \cdot \bar{\mathbf{I}}^{t}_{1} \cdot \bar{\bar{\beta}}_{i_1,i_2} \cdot \bar{\mathbf{I}}^{t}_{2} \cdot \cdots \bar{\bar{\beta}}_{i_N,L} \cdot \bar{\bar{\alpha}}_{LL'} \cdot \bar{\bar{\beta}}_{L',j_N} \cdots \bar{\mathbf{I}}_{2} \cdot \bar{\bar{\beta}}_{j_2,j_1} \cdot \bar{\mathbf{I}}_{1} \cdot \tilde{\mathbf{V}}_{s,j_1,j} \tag{42}$$

DOI 10.1007/978-981-4560-75-7_6-2

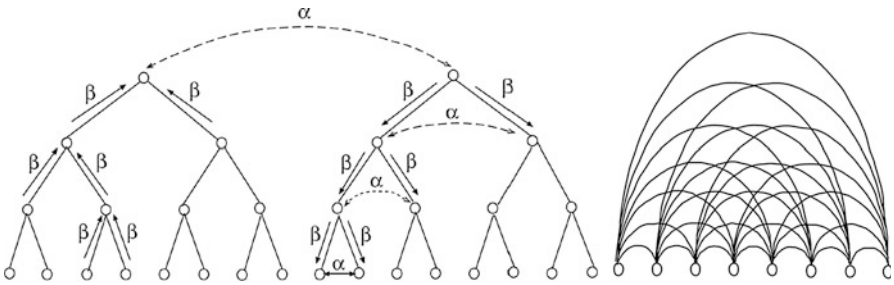


Fig. 7 The factorization of the matrix L_{ij} in Eq. 42 allows the matrix-vector product to emulate a telephone network connection (*left*) (Chew et al. 2001). The direct line connections between telephones (*right*) requires a large number of lines

The factorization allows a matrix-vector product to emulate a multilevel telephone network connection as shown in Fig. 7, where the number of telephone lines can be greatly reduced compared to a direct telephone connection. In the above, the $\widetilde{\overline{\alpha}}$ and $\widetilde{\overline{\beta}}$ matrices are diagonal, while the $\bar{\mathbf{I}}$ matrices are quasi-diagonal interpolation and anterpolation matrices that are not square. Hence, the storage requirements of the factorized matrices are small, and they can be reused in a tree algorithm. There are log N levels in this multilevel algorithm. Each level requires $O(N)$ in workload as well as memory requirements. Hence, the CPU and memory complexity are both $O(N \log N)$. Other factorizations do not lead to $O(N \log N)$ complexity.

The Circuit Models of Electromagnetic Structures

When electromagnetic waves are guided between electronic components through interconnects such as coaxial lines, strip lines, or other type of waveguides, a circuit is formed (Pozar 2011; Celik et al. 2002). The electronic circuit system is the static or quasi-static approximation of electrodynamics. Both fundamental circuit principles, Kirchhoff's current law (KCL) and Kirchhoff's voltage law (KVL), can be derived from Maxwell's equations (Pillegi et al. 1995; Balanis 2012a). Antennas can be considered as terminating devices in the circuit system for matching the impedance of the circuit to that of the propagation channel within a certain frequency band. However, unlike most quasi-static or static circuit components, the antenna's working principle has to depend on the dynamic field. Due to the displacement current, the electric field and the magnetic field couple to each other, which enables wave propagation (Chew 1990). It means that the reliable antenna modeling methods have to be based on the full-wave analysis. However, the circuit model of radiating structures can help to integrate their designs with other circuit subsystems to characterize and optimize the system performance. Hence, parameters such as the input impedance and return loss are popularly used. In practice, because a lot of antennas behave as the low Q resonating tank, many of their designs also rely on circuit interpretations for the performance optimization. Hence, circuit modeling methodologies for distributive electromagnetic structures will be the focus in this part.

Basic Circuit Laws from Maxwell's Equations

Resistance

For a current filament in a segment of the metallic wire with the finite conductivity σ, if the internal electric field $\mathbf{E}$ is uniform, the conducted current due to $\mathbf{E}$ is

$$\mathbf{J} = \sigma \mathbf{E} \tag{43}$$

DOI 10.1007/978-981-4560-75-7_6-2

Both current density and field are vectors in the same direction for isotropic media. For the current component flowing along the x direction in the Cartesian coordinate system, it is only related to the x component of **E**

$$J_x = \sigma E_x \tag{44}$$

For the static field, the electric field is equal to the negative gradient of the potential. If the conductor cross-section area that is orthogonal to the x direction is ΔS and J_x is assumed to be constant over a very small length ΔL_x, one can derive the current through the conductor along the x direction from Eq. 44

$$I_x = \frac{\Delta V_x}{R_x} \tag{45}$$

where ΔV_x is the potential (voltage) drop along the **x** direction and R_x is

$$R_x = \frac{\Delta L_x}{\sigma \Delta S} = \frac{\rho \Delta L_x}{\Delta S} \tag{46}$$

where ρ is the conventional resistivity. Hence, under the static approximation, Eq. 44 is the circuit Ohm's law in the **x** direction while Eq. 43 is the circuit Ohm's law in all directions. For electrodynamics, the electric field is not a simple gradient of the scalar potential. The current distribution could change rapidly as a function of space. Hence, Ohm's circuit law can only be applied for electrically small parts. For electrically large circuit systems, the full circuit model truly based on electrodynamics will be very complicated.

Inductance

Inductance characterizes the ability of a conductor in creating a voltage in either itself or other conductors through the current flowing through it. Based on Faraday's law, this definition is usually connected with looped conductors (Pozar 2011; Balanis 2012a). However, with the help from auxiliary potentials, one can define the partial inductance that is also consistent with the regular inductance definition. The relationship between the induced voltage v, current i, and inductance L is defined as

$$v = L \frac{\partial i}{\partial t} \tag{47}$$

If the time-dependent voltage v is measured on the current loop itself, L is called the self-inductance. If it is measured on another loop, L is called the mutual inductance.

It is also convenient to define the inductance through the magnetic flux when the dimension of the circuit system is electrically small. If ψ is the total magnetic flux through the loop generated by a current i, one has

$$\psi = Li \tag{48}$$

Assume that a loop is made of the perfect electric conductor and its cross section is zero. If the operating frequency is so low that the loop is electrically small, then the current on the loop has almost zero phase change. However, for the high-frequency case, the current changes rapidly. The conventional loop-based inductance will become improper for the physical interpretation and modeling.

DOI 10.1007/978-981-4560-75-7_6-2

Capacitance

The capacitance is the measure of the charge capacity of a conductor. It usually needs to be relative to a certain reference voltage or potential. If it is relative to infinity, it is the self-capacitance. If it is relative to another conductor, it is the mutual capacitance with that conductor. A classic definition of the capacitance is

$$C = \frac{Q}{V} \tag{49}$$

where Q is the net charge on the conductor while V is the bias DC voltage on the conductor relative to its reference (the potential of its reference is usually considered as 0). This definition is in a dual relationship with Eq. 48. Similar to Eq. 47, another definition can be written as (Balanis 2012a)

$$i = C\frac{\partial v}{\partial t} \tag{50}$$

Hence, the capacitance causes the branch currents between the conductor and its reference. Under the dynamic situation, the voltage variation at different locations of the conductor could be highly different and the potential itself is a limited term in describing the dynamic field. Hence, the practical capacitance definition for the full-wave case is more complicated than the static case.

Generally, the capacitance is less sensitive to frequencies since most conductors being used are of high conductivities. Hence, the static capacitance approximation based on Eq. 49 works very well in many practical modeling methods.

KVL

Based on the integral form of Faraday's law, the overall voltage drop along a closed loop can be written as

$$\sum_c v_c = \oint_C d1 \cdot \mathbf{E} = -\frac{\partial \psi}{\partial t} \tag{51}$$

where the total magnetic flux is $\psi = \int_S ds \cdot \mathbf{B}$, the subscript c could be R, L, or C, v_R means the voltage drop due to resistors in the loop, v_L means the voltage drop due to inductors in the loop, and v_C means the voltage drop due to capacitors in the loop. Using Eq. 48, one has

$$\sum_c v_c = -L_s \frac{\partial_i}{\partial t} \tag{52}$$

or

$$\sum_c v_c + L_s \frac{\partial_i}{\partial t} = 0 \tag{53}$$

Here, L_s is the parasitic loop self-inductance when all lumped components are shorted. Sometimes it is also called the strayed inductance (Balanis 2012a). When the circuit dimension is electrically small, L_s is a very small value. For example, a circular loop with a 1 cm radius and 1 mm wire cross-section radius has a

DOI 10.1007/978-981-4560-75-7_6-2

self-inductance of 2.99×10^{-8}H. And it is already a very big circuit loop. Hence, if one ignores the parasitic voltage introduced by L_s under the low-frequency assumption, Eq. 53 becomes the standard Kirchhoff's voltage law (KVL):

$$\sum_c v_c = 0 \tag{54}$$

But if the frequency is very high, the voltage introduced by the self-inductance cannot be ignored. Conventional KVL will not be accurate anymore. Equation 53 has to be used to consider the effect of L_s.

KCL

Based on the continuity equation, a circuit node enclosed by a small closed volume V satisfies

$$\oint_S d\mathbf{S} \cdot \mathbf{J} + \frac{\partial}{\partial t} \int_V dr\rho = 0 \tag{55}$$

where ρ is the volumetric charge density while $\mathbf{J}$ is the current density. Under the low-frequency approximation, currents only flow through metal wires. The closed surface integral of the current density becomes the summation of branch currents. Further, by Eqs. 49 and 50, one has

$$\sum_b I_b + C_n \frac{\partial v_n}{\partial t} = 0 \tag{56}$$

This equation is also for low frequencies (Balanis 2012a). The subscript b refers to all branches. Here, v_n is the node voltage while C_n is the total parasitic self-capacitance of all branches connected to the node relative to the zero potential ground at infinity. It can be computed by excluding all lumped components inside $V$and computing the total self-capacitance of all conductors. At low frequencies, C_n is so small that it could be approximated as zero. Then the standard Kirchhoff's current law (KCL) is achieved:

$$\sum_b I_b = 0 \tag{57}$$

At high frequencies, the parasitic current introduced by the self-capacitance cannot be ignored. The conventional KCL based on the pure circuit theory will not be accurate anymore.

Partial Element Equivalent Circuit Method

The partial element equivalent circuit (PEEC) method was developed by Albert Ruehli to solve parasitic coupling problems by building the equivalence circuit based on the integral equation (Ruehli 1972, 1974; Ruehli and Brennan 1973). It was initiated for static or quasi-static applications with no retardation (delay) considered. Later it was extended to the full-wave case where the retardation is included. PEEC is a very convenient bridging method between electromagnetics and circuit theories. It is broadly used by many electronic design automation (EDA) softwares. It also greatly helps the modeling of resonating structures such as antennas. Even though the PEEC method solves the problem in the circuit solver, its formulation actually starts from the integral equation (Chew 1990). The modified nodal analysis (MNA)-based circuit solvers (Ho et al. 1975) basically consider the electric field and magnetic field decoupling effect happening at the low-frequency regime. Hence, it involves both KCL and KVL in the formulation. Some low-frequency full-wave integral equation simulation methods have also been developed based on

DOI 10.1007/978-981-4560-75-7_6-2

the same physical insight, such as the augmented electric field integral equation method (A-EFIE) (Qian and Chew 2009).

Partial Inductance

Assume that there are two loops i and j with currents I_i and I_j (Ruehli 1972). For low frequencies, both I_i and I_j are approximated to be constant over each loop. Then the mutual inductance L_{ij} between two loops is

$$L_{ij} = \frac{\psi_{ij}}{I_j}, \tag{58}$$

where ψ_{ij} is the magnetic flux through loop i due to the current in loop j. It can be further written as

$$\psi_{ij} = \int_{S_i} d\mathbf{S}_i \cdot \mathbf{B}_{ij} = \frac{1}{a_i} \int_{a_i} da_i \oint_i d\mathbf{l}_i \cdot \mathbf{A}_{ij}, \tag{59}$$

where S_i is the loop surface of loop i, $\mathbf{B}_{ij}$ is the magnetic flux density, $\mathbf{A}_{ij}$ is the magnetic vector potential at loop i due to the current in loop j, and a_i is the cross-section area of the loop i wire. The vector potential $\mathbf{A}_{ij}$ is formulated through an integral using the homogeneous medium Green's function as its kernel:

$$\mathbf{A}_{ij} = \oint_{\text{loop}_j} d\mathbf{r}' \frac{e^{-jk|\mathbf{r}_i - \mathbf{r}'|}}{4\pi|\mathbf{r}_i - \mathbf{r}'|} \mu \mathbf{J}_j(\mathbf{r}'). \tag{60}$$

Assuming that the working frequency is low and cross sections of wires are always small, then the current is almost uniform along the loop and at the cross section. Hence, one has

$$\mathbf{A}_{ij} = \frac{\mu}{4\pi} \frac{I_j}{a_j} \int_{a_j} da_j \oint_j d\mathbf{l}_j \left[\frac{1}{r_{ij}}\right], \tag{61}$$

where I_j is the loop j's current from the common circuit point of view. Using Eqs. 58 and 59, one has

$$L_{ij} = \frac{1}{a_i a_j} \frac{\mu}{4\pi} \oint_i \oint_j \int_{a_i} \int_{a_j} \left[d\mathbf{l}_i \cdot d\mathbf{l}_j da_i da_j \frac{1}{r_{ij}} \right]. \tag{62}$$

If a_i and a_j are small (current filaments), it can be further simplified to be independent from the cross section

$$L_{ij} = \frac{\mu}{4\pi} \oint_i \oint_j \left[d\mathbf{l}_i \cdot d\mathbf{l}_j \frac{1}{r_{ij}} \right]. \tag{63}$$

If the wire cross section is not small, the wire has to be split along the longitudinal direction into many current filaments. Each current filament is assumed to have a constant current at its cross section. Then Eq. 63 depicts the result of a group of inductors connected in a parallel fashion.

With the thin- and short-wire assumption, if one divides each loop into small segments, the mutual inductance is the summation of partial inductive contributions. For example, if loop i is partitioned into K pieces while loop j is divided into M pieces, then

DOI 10.1007/978-981-4560-75-7_6-2

$$L_{ij} = \sum_{k=1}^{K} \sum_{m=1}^{M} \frac{\mu}{4\pi} \frac{1}{a_{ik} a_{jm}} \int_{l_{ik}} \int_{a_{ik}} \int_{l_{jm}} \int_{a_{jm}} da_{ik} da_{jm} \frac{d\mathbf{l}_{ik} \cdot d\mathbf{l}_{jm}}{r_{km}}, \tag{64}$$

where subscript *ik* means the *k*-th current filament on loop *i*, *jm* means the *m*-th current filament on loop *j*, and *p* means *partial*. Here, a_{ik} is the cross-section area of the *k*-th current filament on loop *i* and l_{ik} is the *k*-th current filament of loop *i*. Other symbols are defined in a similar way. If one defines the kernel of the double summations in Eq. 64 to be the partial inductance $L_{p_{ij}}^{km}$, one has

$$L_{p_{ij}}^{km} = \frac{\mu}{4\pi} \frac{1}{a_{ik} a_{jm}} \int_{l_{ik}} \int_{a_{ik}} \int_{l_{jm}} \int_{a_{jm}} da_{ik} da_{jm} \frac{\left| d\mathbf{l}_{ik} \cdot d\mathbf{l}_{jm} \right|}{r_{km}}. \tag{65}$$

Then the total inductance is the summation of partial inductances

$$L_{ij} = \sum_{k} \sum_{m} S_{ij}^{km} L_{p_{ij}}^{km}, \tag{66}$$

where S_{ij}^{km} is the sign of the dot project of $d\mathbf{l}_{ik} \cdot d\mathbf{l}_{jm}$. Hence, it is either +1 or −1. If one ignores the integration over cross sections, the partial inductance is simplified to

$$L_{p_{ij}}^{km} = \frac{\mu}{4\pi} \int_{l_{ik}} \int_{l_{jm}} \frac{\left| d\mathbf{l}_{ik} \cdot d\mathbf{l}_{jm} \right|}{r_{km}} \tag{67}$$

The physical meaning of $L_{p_{ij}}^{km}$ is very intriguing. If one places both current filaments l_{ik} and l_{jm} on the same plane, as shown in Fig. 8, the projection lines from ending points of current *ik* to *jm* will form an

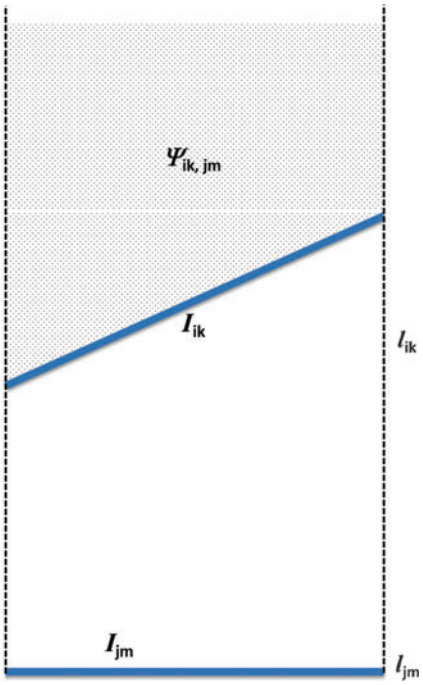


Fig. 8 The magnetic flux loop for the partial inductance

DOI 10.1007/978-981-4560-75-7_6-2

infinite loop with current *ik* and the infinity. Since the integration kernel (static Green's function) is zero at infinity, l_{jm} is orthogonal to both projection lines, and $\int_{l_{ik}}$ can be written as a loop integral in the partial inductance definition (Eq. 62) (if one ignores the absolute value operation $|\cdot|$ in the kernel). Hence,

$$L_{p_{ij}}^{km} = \frac{\mu}{4\pi}\oint_{l_{ik}} d\mathbf{l}_{ik} \cdot \int_{l_{jm}} \frac{d\mathbf{l}_{jm}}{r_{km}} = \oint_{l_{ik}} d\mathbf{l}_{ik} \cdot \frac{\mathbf{A}_{i,km}}{I_{jm}} = \frac{1}{I_{jm}} \int_{S_{ik}} d\mathbf{S}.\mathbf{B}_{i,km}. \tag{68}$$

where $\mathbf{A}_{i,km}$ is the magnetic vector potential defined in Eq. 60 generated inside the infinite loop by the current filament l_{jm}. Hence, even though the formulation of the partial inductance does not have a complete loop, it actually uses the loop formed with infinity through current project lines,

$$L_{p_{ij}}^{km} = \frac{\psi_{ik,jm}}{I_{jm}}. \tag{69}$$

The above definitions and properties are also applicable to cases when $i = j$ and $k = m$. It means that they are valid for self-inductances too. Using the mutual partial inductance as the reference, it can be understood that the self-partial inductance corresponds to the magnetic flux through a loop formed by the infinity, two orthogonal projection lines, and the current filament itself.

If two current filaments are both on the same rectangular loop and parallel to each other, they will have opposite current flow directions, as shown in Fig. 9. Then in Eq. 66 $S_{ii}^{km} = -1$ is for the mutual partial inductance $L_{p_{ii}}^{km}$ and $S_{ii}^{mm} = +1$ is for the self-partial inductance $L_{p_{ii}}^{km}$. Hence, the magnetic flux area from l_{ik} to infinity is completely canceled from the area from l_{im}. Hence, only the flux area in between the two current filaments is kept. When this area is closed by another pair of vertical current filaments, it forms a complete loop and the loop area will directly determine the total inductance, which is consistent with the conventional circuit concept.

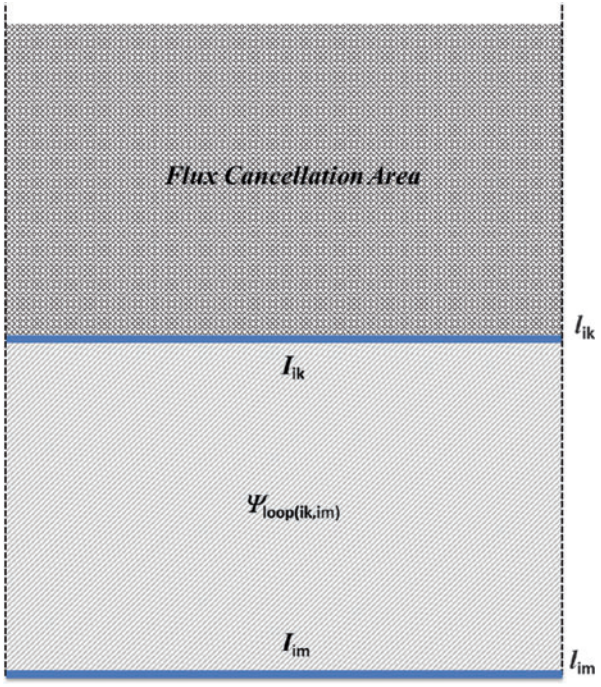


Fig. 9 The magnetic flux loop for the loop based on partial inductances

DOI 10.1007/978-981-4560-75-7_6-2

PEEC Formulation

Assume that circuit conducting interconnects are partitioned into N_l current filaments (Ruehli et al. 2003, 2013; Rong and Cangellaris 2001). If the conductivity is finite, its time-domain electric field for one wire segment is

$$\mathbf{E}(\mathbf{r},t) = \frac{\mathbf{J}(\mathbf{r},t)}{\sigma} + \frac{\partial \mathbf{A}(\mathbf{r},t)}{\partial t} + \nabla\phi(\mathbf{r},t) \tag{70}$$

The vector potential **A** is contributed from all current filaments including itself. Hence, it can be written as

$$\mathbf{A}(\mathbf{r},t) = \sum_{n_l=1}^{N_L} \frac{\mu}{4\pi} \int_{V_{n_l}} dr' G(\mathbf{r},\mathbf{r}') \mathbf{J}_{n_l}(\mathbf{r}',t') \tag{71}$$

where $\mathbf{J}_{nl}$ is the current density vector on the n_l-th current filament and $G(\mathbf{r}, \mathbf{r}')$ is the Green's function

$$G(\mathbf{r},\mathbf{r}') = \frac{1}{|\mathbf{r}-\mathbf{r}'|}. \tag{72}$$

t' is the source time with the retardation

$$t' = t - \frac{|\mathbf{r}-\mathbf{r}'|}{c/\sqrt{\mu_r \epsilon_r}}. \tag{73}$$

The scalar potential ϕ is contributed from the charge distribution. Since good conductors are usually used in the design, there is almost no charge inside the conductor. Hence, all charges are on the conductor surface. Assuming that all surfaces are partitioned into N_C nonoverlapping patches and the charge density on each patch is generally defined as ρ, the scalar potential is written as

$$\phi(\mathbf{r},t) = \sum_{n_c=1}^{N_C} \frac{1}{4\pi\epsilon} \int_{S_{n_c}} d\mathbf{r}' G(\mathbf{r},\mathbf{r}') \rho_{n_c}(\mathbf{r}',t') \tag{74}$$

Hence, Eq. 70 can be written as

$$\begin{aligned} \mathbf{E}(\mathbf{r},t) \quad &= \frac{\mathbf{J}(\mathbf{r},t)}{\sigma} + \sum_{n_l=1}^{N_L} \frac{\mu}{4\pi} \int_{V_{n_l}} d\mathbf{r}' G(\mathbf{r},\mathbf{r}') \frac{\partial \mathbf{J}_{n_l}(\mathbf{r}',t')}{\partial t} \\ &+ \sum_{n_c=1}^{N_C} \frac{1}{4\pi\epsilon} \nabla \int_{S_{n_c}} dS' G(\mathbf{r},\mathbf{r}') \rho_{n_c}(\mathbf{r}',t') \end{aligned} \tag{75}$$

For the m-th current filament, it is either inside or on the surface of the conductor. Hence, the electric field along its filament direction is always zero if one does Galerkin's testing along the current filament. Hence, one can have

DOI 10.1007/978-981-4560-75-7_6-2

$$\begin{aligned}&\frac{1}{\sigma}\int_{V_m} d\mathbf{r}_m J_m(\mathbf{r},t)\\&+\sum_{n_l=1}^{N_L}\frac{\mu}{4\pi}\int_{V_m}\int_{Vn_l} d\mathbf{r}_m d\mathbf{r}'_{n_l} G(\mathbf{r},\mathbf{r}')\frac{\partial J_{n_l}(t'_m)}{\partial t}S_{m,n_l}\\&+\sum_{n_c=1}^{N_C}\frac{1}{4\pi\epsilon}\int_{V_m} d\mathbf{r}_m\mathbf{u}_m\cdot\nabla\int_{S_{n_c}} ds'_{n_c} G(\mathbf{r},\mathbf{r}')\rho_{n_c}(\mathbf{r}',t')=0.\end{aligned} \tag{76}$$

Its first term corresponds to the resistive voltage drop

$$V_{R_m}=\frac{1}{\sigma}\int_{V_m} d\mathbf{r}_m J_m(\mathbf{r},t)=\frac{I_m}{\sigma}\int_{l_m} dl_m=\frac{l_m}{\sigma}I_m=R_m I_m. \tag{77}$$

where l_m is the longitudinal length of the m-th current filament. Hence, the first term shows the ohmic voltage drop due to the lossy conductor.

The second term corresponds to the inductance effect. Here, $S_{m,\,nl}$ is the vector dot between the testing current direction and the source current direction. If one assumes all filaments are along the Cartesian coordinate axes, $S_{m,n_l}=+1,-1$, or 0 depending on their relative directions. Using the partial inductance concept introduced in the previous subsection, one has

$$V_{L_m}=\sum_{n_l=1}^{N_L}\frac{\mu}{4\pi}\frac{1}{a_m a_{n_l}}\int_{V_m}\int_{V_{n_l}} d\mathbf{r}_m d\mathbf{r}'_{n_l} G(\mathbf{r},\mathbf{r}')\frac{\partial J_{n_l}(t'_m)}{\partial t}S_{m,n_l} \tag{78}$$

$$=\sum_{n_l=1}^{N_L} L_p^{mn_l}\frac{d}{dt}I_{n_l}(t'_m). \tag{79}$$

Hence, the second term is the voltage drop due to partial inductances. It shall be noted that there are two types of inductances contributed to V_{Lm}: one is the self-partial inductance and another one is the mutual partial inductance. Because there are many current filaments in reality, mutual partial inductances are much more complicated than the self-partial inductance.

The third term corresponds to the capacitive coupling. It could be represented by potentials or capacitances. It uses the central difference to approximate the gradient where $\mathbf{u}_m$ is the unit vector direction of the mth current flow. Assume that u_m represents x, y or z direction, then

$$\begin{aligned}V_{C_m}&=\sum_{n_c=1}^{N_C}\frac{a_m}{4\pi\epsilon}\left[\int_{S_{n_c}} dS' G(\mathbf{r}_{m+},\mathbf{r}')\rho_{n_c}(\mathbf{r}',t')-\int_{S_{n_c}} dS' G(\mathbf{r}_{m-},\mathbf{r}')\rho_{n_c}(\mathbf{r}',t')\right]\\&=\sum_{n_c=1}^{N_C}\frac{1}{4\pi\epsilon}\rho_{n_c}\cdot\left[\int_{S_{n_c}} dS' G(\mathbf{r}_{m+},\mathbf{r}')-\int_{S_{n_c}} dS' G(\mathbf{r}_{m-},\mathbf{r}')\right]\\&=\sum_{n_c=1}^{N_C} Q_{n_c}\left[pp^{+}_{i(m,n_c)}-pp^{-}_{i(m,n_c)}\right].\end{aligned} \tag{80}$$

DOI 10.1007/978-981-4560-75-7_6-2

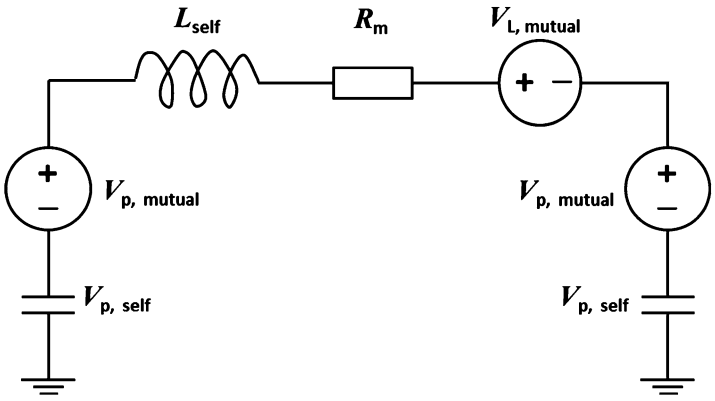


Fig. 10 The equivalent circuit of the PEEC model

where

$$pp^{\pm}_{i(m,n_c)} = \frac{1}{4\pi\epsilon}\frac{1}{a_{n_c}}\int_{S_{nc}} dS' G(\mathbf{r}_{m\pm}, \mathbf{r}'). \tag{81}$$

Here, the position $\mathbf{r}_{m+}$ and $\mathbf{r}_{m-}$ are the two end points of the m-th current filament. The current filaments' terminating nodes are located at the centers of piecewise constant charge density patches. This is very similar to the cell structure in the FDTD algorithm (Chew 1990). The third term therefore is interpreted as the charge-controlled voltage source (CCVC) in the equivalent circuit. By considering all three terms and their relationships with other parts of the system, an equivalent circuit can be easily built, as shown in Fig. 10.

The L_{self} is the self-partial inductance, R_m is the ohmic loss due to the finite conductivity, and $V_{i,\text{mutual}}$ is the voltage generated by other currents from other places through partial mutual inductances. Here, $V_{p,\text{mutual}}$ is the voltage generated by other branch currents from other places through partial mutual potentials, and $V_{p,\text{self}}$ is the voltage generated by the self-capacitive coupling or the self-partial potential coefficient.

Through this approach, the distributive structure can be modeled using a circuit. Since each electrically small piece could contribute to the mutual partial inductance and capacitively coupled potentials, the resultant circuit is a densely coupled one. If this circuit is solved by the circuit solver SPICE (Nagel and Pederson 1973; Warwick 2009), it results in a dense numerical matrix. Its solving process will be very expensive. Many methods could be employed to further solve this problem. It shall be noted that conventional SPICE solvers do not handle retardation. It means that the true full-wave circuit models cannot be solved by SPICE directly. Hence, certain modifications to SPICE solvers are necessary for handling circuit models based on the PEEC method.

Low-Frequency Computational Electromagnetics

An electromagnetic field can be obtained by solving the wave equations with enforced boundary condition. Because of the linearity of problems, the solution can be obtained in the frequency domain by Fourier transforming the equations in time, which theoretically covers the frequency range from static to optical regime. Generally speaking, the problems in solving the wave equations in the frequency domain can be classified into three regimes, which are defined as low-frequency regime, mid-frequency regime, and high-frequency regime (Chew et al. 2008). For low frequency, the typical electrical size (L) is much less than wavelength (λ), i.e., $L \ll \lambda$. At low frequency, the electromagnetic field physics is close to

DOI 10.1007/978-981-4560-75-7_6-2

that of the static field, where the magnetic field and the electric field are decoupled from each other. Hence, the electromagnetic wave equations reduce to the equation of electrostatics and the equation of magnetostatics that are completely decoupled from each other (Zhao and Chew 2000). Furthermore, the electrostatics equation or Poisson's equation governs the capacitive circuit physics, while the magnetostatics equation or Laplace's equation governs the inductive circuit physics. In the mid-frequency regime, the wavelength is comparable with the object size, $L \approx \lambda$. In this regime, the electric and magnetic fields are tightly coupled together. Thus, the electromagnetic fields become waves and are oscillatory as is typical of wave physics. In the high-frequency regime, the wavelength is much less than the object size, $L \gg \lambda$. The plane waves of the ray physics thus become dominant. In this section, our discussion will be focused on the low-frequency computation electromagnetics, which becomes more and more important in recent interdisciplinary research.

Understanding of Low-Frequency Region

Low-frequency computational electromagnetics has been studied for many years (Chew et al. 2008; Zhao and Chew 2000). Theoretically, the current or charge at static limit does not radiate but generates the field which is inversely proportional to the square of the distance from the source. For the on-chip-level antenna design, the typical size of the interconnects and devices is usually so much smaller than the wavelength (nanometer to micrometer levels in the standard CMOS process).

It implies that the surrounding field is primarily quasi-static; that is why circuit designers can utilize the lumped-circuit model instead of using the distributed-circuit model for the passive components, such as the interconnects, the inductors, and the capacitors. Hence, in order to capture the circuit physics with a universal solver, the well-developed full-wave solvers have to be revisited and corrected at low-frequency limit.

When $\omega \rightarrow 0$, the four Maxwell equations can be reduced into two parts: the magnetostatic part:

$$\nabla \times \mathbf{H} = \mathbf{J} \tag{82}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{83}$$

and the electrostatic part:

$$\nabla \times \mathbf{E} = 0 \tag{84}$$

$$\nabla \cdot \mathbf{D} = \rho = \lim \frac{\nabla \cdot \mathbf{J}}{i\omega} \tag{85}$$

Note that the current $\mathbf{J}$ in the electrostatic part must be of the order of ω in order to keep the right-hand side bounded. According to the Helmholtz decomposition (Collins 1966), the total current $\mathbf{J}$ naturally decomposes into a solenoidal (divergence-free) part and an irrotational (curl-free) part as

$$\mathbf{J} = \mathbf{J}_{sol} + \mathbf{J}_{irr} \tag{86}$$

where the solenoidal part $\mathbf{J}_{sol}$ represents eddy currents that produce primarily the magnetic field and the irrotational part $\mathbf{J}_{irr}$ represents charge currents that produce primarily the electric field at low frequencies. Both of them are equally important in capturing the inductive and capacitive circuit physics. When $\omega \rightarrow 0$, $\mathbf{J}_{irr}$ becomes much smaller than $\mathbf{J}_{sol}$; hence, the former usually cannot be captured correctly in the numerical calculation, due to the limit of computer precision. This is also known as the low-frequency inaccuracy problem (Chew et al. 2008). The problem in the electric field integral equation (EFIE) and its

DOI 10.1007/978-981-4560-75-7_6-2

solutions will be discussed in the following sections. On the other hand, the similar low-frequency problem also happened in the time-domain EFIE (TD-EFIE) (Andriulli et al. 2009), while the low-frequency inaccuracy problem happened in the magnetic field integral equation (MFIE) (Zhang et al. 2003).

Low-Frequency Problems in EFIE Formulation

Due to the decoupling between the electric field and magnetic field, there is a low-frequency breakdown problem in the EFIE formulation. The reasons for this breakdown can be classified into two catalogs: the frequency breakdown and the mesh density breakdown. The frequency breakdown occurs because the contribution of the vector potential is swamped by that of the scalar potential at low frequencies. Moreover, it will evoke the null space of the divergence operator in the scalar potential term and make the MoM matrix become extremely ill conditioned. Basically, the traditional EFIE operator can be written in its mixed potential form as

$$\mathcal{T}(\boldsymbol{J}) = \mathcal{T}_s(\boldsymbol{J}) + \mathcal{T}_h(\boldsymbol{J}) \tag{87}$$

in which the smoothing $(\mathcal{T}_s)$ and hypersingular $(\mathcal{T}_h)$ terms are defined as

$$\mathcal{T}_s(\boldsymbol{J}) = i\omega\mu\hat{\boldsymbol{n}}_r \times \int_{S'} dS' g(\boldsymbol{r}, \boldsymbol{r}')\boldsymbol{J} \tag{88}$$

$$\mathcal{T}_h(\boldsymbol{J}) = -\frac{1}{i\omega\varepsilon}\hat{\boldsymbol{n}}_r \times \nabla \int_{S'} dS' g(\boldsymbol{r}, \boldsymbol{r}')\nabla' \cdot \boldsymbol{J} \tag{89}$$

where

$$g(\boldsymbol{r}, \boldsymbol{r}') = \frac{e^{ik_0|r-r'|}}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|} \tag{90}$$

is the free-space Green's function, k_0 is the wave number in the free space, ε and μ are the relative permeability and permittivity, and $\boldsymbol{J}$ is the surface current on an arbitrarily shaped PEC surface S' whose outward pointing unit normal at $\boldsymbol{r}$ is denoted by $\hat{n}_r$. The domain of both $\mathcal{T}_s$ and $\mathcal{T}_h$ is the surface S'. Here, $\mathcal{T}_s$ is referred to as a smoothing operator because the range space of the operator is smoothed by one order relative to its domain, while the hypersingular operator $\mathcal{T}_h$ lowers the smoothness of the domain space.

When $\omega \to 0$, the hypersingular term $\mathcal{T}_h$ which is $O(\omega^{-1})$ dominates over the smoothing term $\mathcal{T}_s$ which is $O(\omega^1)$. As mentioned above, because of the existence of the divergence operator in Eq. 89 and $\nabla' \cdot \boldsymbol{J}_{sol} = 0$, $\mathcal{T}_h$ has a null space. Thus, the ill-conditioned $\mathcal{T}$ behaves like a first-kind operator between the solenoidal and irrotational subspaces (Adams 2004). This makes the impedance matrix nearly singular and unsolvable at low frequencies (Zhao and Chew 2000). This is the so-called low-frequency breakdown problem for the EFIE operator.

Low-Frequency Remedies

In past decades, many research efforts have been carried out to avoid the imbalance inherent in the traditional EFIE operator. The most popular one is the loop-tree or loop-star decomposition, which separates the electrostatic and magnetostatic physics at low frequencies (Zhao and Chew 2000; Vecchi 1999; Lee et al. 2003). By introducing the charge in the unknown list and enforcing the current continuity constraint, the augmented version of EFIE was proposed, which is known as A-EFIE (Qian and Chew

DOI 10.1007/978-981-4560-75-7_6-2

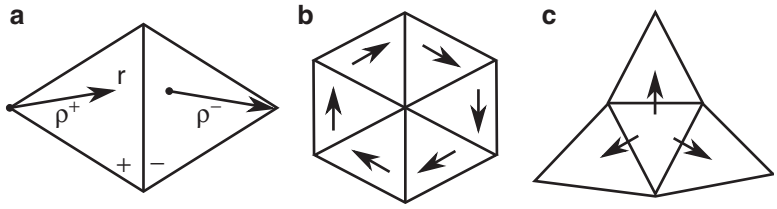


Fig. 11 Representations of (**a**) RWG, (**b**) loop, and (**c**) star basis functions

2008). The frequency scaling can be normalized in a balanced manner, thus remedying the low-frequency breakdown problem. In addition, the success of the developed Calderón multiplicative preconditioned EFIE (CMP-EFIE) method (Hsiao and Kleinman 1997; Adams 2004) has also made the original EFIE into a second-kind Fredholm integral equation operator, which is the so-called self-regularizing property of the $\mathcal{T}$ operator and immune to mesh density breakdown. Unfortunately, although stable at low frequencies, A-EFIE and CMP-EFIE all suffer from the low-frequency inaccuracy problem for the solved currents. As a remedy, the perturbation method has been applied to capture the accurate currents at different frequency orders (Qian and Chew 2010; Sun et al. 2013b).

Loop-Tree Decomposition

Instead of using the popular Rao–Wilton–Glisson (RWG) basis function, one can decompose the RWG basis function into the loop basis function whose members have zero divergence and the tree basis (or star basis) function whose members have nonzero divergence, as shown in Fig. 11. Here, $\rho^{\pm}(\mathbf{r})$ is the vector from the point $\mathbf{r}$ to the apex of the respective triangles, and $\pm$ is the support of the respective triangles. This is known as the loop-tree or loop-star decomposition, which is a quasi-Helmholtz decomposition because the tree or the star expansion functions are not curl-free. In some work based on the finite-element method, this is also known as the tree-cotree decomposition (Manges and Cendes 1995). Here, one shall call the space spanned by the RWG basis function as the RWG space, the subspace spanned by the tree (or star) basis function as the tree (or star) space, and that spanned by the loop basis function as the loop space. Therefore, the total surface current can be expanded as

$$\mathbf{J}(r') = \sum_{n=1}^{N_L} I_{Ln}\mathbf{J}_{Ln}(r') + \sum_{n=1}^{N_C} I_{Cn}\mathbf{J}_{Cn}(r') \tag{91}$$

where $\mathbf{J}_{Ln}(\mathbf{r}')$ is the divergence-free loop basis function such that $\nabla \cdot \mathbf{J}_{Ln}(\mathbf{r}') = 0$ and $\mathbf{J}_{Cn}(\mathbf{r}')$ is the nondivergence-free tree (or a star) basis function such that $\nabla \cdot \mathbf{J}_{Cn}(\mathbf{r}') \neq 0$, and it is used to model the charge in the system. The above can be written in the matrix form as

$$\mathbf{J}(r') = \mathbf{J}_L^t(r') \cdot \mathbf{I}_L + \mathbf{J}_C^t(r') \cdot \mathbf{I}_C \tag{92}$$

where the four column vectors in the right-hand side contain the expansion and coefficient elements in Eq. 91. By substituting the above into the EFIE and testing the result with the same set of the functions as in Galerkin's method, one has a matrix equation as

$$\begin{bmatrix} \overline{\mathbf{Z}}_{LL} & \overline{\mathbf{Z}}_{LC} \\ \overline{\mathbf{Z}}_{CL} & \overline{\mathbf{Z}}_{CC} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{I}_L \\ \mathbf{I}_C \end{bmatrix} = \begin{bmatrix} \mathbf{V}_L \\ \mathbf{V}_C \end{bmatrix} \tag{93}$$

where

DOI 10.1007/978-981-4560-75-7_6-2

$$\mathbf{V}_L = -\left\langle \mathbf{J}_L(r), \mathbf{E}^{inc}(r) \right\rangle \tag{94}$$

$$\mathbf{V}_C = -\left\langle \mathbf{J}_C(r), \mathbf{E}^{inc}(r) \right\rangle \tag{95}$$

$$\overline{\mathbf{Z}}_{LL} = i\omega\mu \left\langle \mathbf{J}_L(r), g(r,r'), \mathbf{J}_L^t(r') \right\rangle \tag{96}$$

$$\overline{\mathbf{Z}}_{LC} = i\omega\mu \left\langle \mathbf{J}_L(r), g(r,r'), \mathbf{J}_C^t(r') \right\rangle \tag{97}$$

$$\overline{\mathbf{Z}}_{CL} = i\omega\mu \left\langle \mathbf{J}_C(r), g(r,r'), \mathbf{J}_L^t(r') \right\rangle = \overline{\mathbf{Z}}^t{}_{LC} \tag{98}$$

$$\begin{aligned} \overline{\mathbf{Z}}_{CC} &= i\omega\mu \left\langle \mathbf{J}_C(r), g(r,r'), \mathbf{J}_C^t(r') \right\rangle \\ &- \frac{i}{\omega\varepsilon} \left\langle \nabla . \mathbf{J}_C(r), g(r.r'), \nabla' \cdot \mathbf{J}_C^t(r') \right\rangle \end{aligned} \tag{99}$$

In order to solve the system equation when $\omega \to 0$, the frequency normalization should be implemented as detailed in (Zhao and Chew 2000). It implies that even after the loop-tree decomposition, the matrix system is still ill conditioned because of the imbalance of the matrix elements, which is the so-called frequency-scaling behavior. Recently, the application of the loop-star/tree decomposition with the magnetic field and the Calderón preconditioned electric field integral operators is also studied (Andriulli 2012).

A-EFIE

Another remedy to the low-frequency breakdown problem is the A-EFIE, which avoids the imbalance inherent in the conventional EFIE by introducing the charge as the additional unknown. The same notations as those used in (Qian and Chew 2008) are used here. By introducing the current continuity condition, one can obtain the A-EFIE system as

$$\begin{bmatrix} \overline{\mathbf{V}} & \overline{\mathbf{D}}^T \cdot \overline{\mathbf{P}} \\ \overline{\mathbf{D}} & k_0^2 \overline{\mathbf{I}} \end{bmatrix} \cdot \begin{bmatrix} ik_0 \boldsymbol{j} \\ c_0 \rho \end{bmatrix} = \begin{bmatrix} \eta_0^{-1} \mathbf{b} \\ 0 \end{bmatrix} \tag{100}$$

where $\overline{\mathbf{V}}$ and $\overline{\mathbf{P}}$ are the vector potential and scalar charge matrices, respectively, $\overline{\mathbf{I}} \in \mathbf{R}^{p \times e}$ is an identity matrix, $\boldsymbol{j} \in \mathbf{C}^{e \times 1}$ represents the vector of current coefficient, $\rho \in \mathbf{C}^{p \times 1}$ represents the vector of charge coefficient, and $b \in \mathbf{C}^{e \times 1}$ denotes the excitation. The superscripts p and e indicate the number of patches and edges, respectively. Mathematically, it has a generalized saddle point form and can be reduced to a standard saddle point form in the low-frequency regime. Physically, this new equation system can also be understood as a complete Kirchhoff circuit law, which is solvable and stable with low-frequency circuit physics. It means that the first equation in Eq. 100 represents the Kirchhoff voltage law, while the second equation represents the Kirchhoff current law.

At very low frequencies, charge neutrality still causes rank deficiency of the A-EFIE system in Eq. 100. A simple method has been proposed to remedy this issue by dropping one charge unknown for each spanning tree, which results the A-EFIE system as

$$\begin{bmatrix} \overline{\mathbf{V}} & \overline{\mathbf{D}}^T \cdot \overline{\mathbf{P}} \cdot \overline{\mathbf{B}} \\ \overline{\mathbf{F}} \cdot \overline{\mathbf{D}} & k_0^2 \overline{\mathbf{I}}_r \end{bmatrix} \cdot \begin{bmatrix} ik_0 \boldsymbol{j} \\ c_0 \rho_r \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix} \tag{101}$$

DOI 10.1007/978-981-4560-75-7_6-2

where the backward matrix $\overline{\mathbf{B}}$ and forward matrix $\overline{\mathbf{F}}$ are defined to map the charge unknowns back and forth, respectively, and ρ_r denotes the charge of unknown vectors. Then, one has

$$\rho_r = \overline{\mathbf{F}} \cdot \rho, \quad \rho = \overline{\mathbf{B}} \cdot \rho_r \tag{102}$$

However, the saddle point problems always suffer from the poor spectral property due to the zero diagonal block. Hence, the constraint preconditioner is usually required in order to regularize the saddle point matrix for easier solution. Since this preconditioning matrix should be inverted first, a highly sparse matrix is preferred and is given by

$$\overline{M} = \begin{bmatrix} \overline{\mathbf{V}}_d & \overline{\mathbf{D}}_r{}^T \cdot \overline{\mathbf{P}}_{rd} \\ \overline{\mathbf{D}}_r & k_0^2 \overline{\mathbf{I}}_r \end{bmatrix} \tag{103}$$

which can be solved by a fast direct solver like the multifrontal method. It implies that it will incur more computational cost in addition to the more unknowns in comparison with the traditional EFIE system. More details about this constraint preconditioning technique can be found in Benzi et al. (2005). Alternatively, the Calderón multiplicative preconditioner can be further implemented on A-EFIE, which does not need inverse preconditioning and avoids the saddle point problems as well (Sun et al. 2013a).

Calderon Preconditioning

Another popular remedy is to utilize the Calderón identity, which can be expressed as (Adams 2004).

$$\mathcal{T}^2(\boldsymbol{J}) = -\frac{\boldsymbol{J}}{4} + \mathcal{K}^2(\boldsymbol{J}) \tag{104}$$

It can be used to conduct a new operator by projecting the operator $\mathcal{T}$ of Eq. (87) into itself. In the above, $\mathcal{K}$ is the magnetic field integral equation (MFIE) operator (Chew et al. 2008)

$$\mathcal{K}(\boldsymbol{J}) = \hat{\boldsymbol{n}}_{\boldsymbol{r}} \times \nabla \times \int_{S'} dS' g(\boldsymbol{r}, \boldsymbol{r}') \boldsymbol{J}. \tag{105}$$

It is important to notice from Eq. 104 that the composite operator $\mathcal{T}^2$ is actually a second-kind integral operator with its spectrum accumulating at -0.25. In other words, the $\mathcal{T}$ operator can be utilized to precondition itself, the so-called "self-regularizing property".

A multiplicative form of the Calderón multiplicative preconditioner (CMP) (Andriulli et al. 2008) was developed based on the div- and quasi-curl-conforming basis function, called the Buffa–Christiansen (BC) basis function (Buffa and Christiansen 2007). This preconditioning is straightforward to implement and easily integrated into existing MoM codes based on RWG basis functions. For the discretization of composite operator $\mathcal{T}^2$, the inner $\mathcal{T}$ operator is discretized by using div-conforming RWG basis function (source) and $\hat{n}_r \times$ RWG basis function, while the outer $\mathcal{T}$ operator is discretized by using div- and quasi-curl-conforming BC basis function (source) and $\hat{n}_r \times$ BC basis function. Therefore, the resulting Gram matrix is a mixed curl- and quasi-curl-conforming matrix, which is highly sparse and invertible. Subsequently, the CMP is applied on the combined field integral equation (CFIE) formulation for PEC objects (Bağci et al. 2009) and also for the single-source integral equations (Valdés et al. 2011).

It should also be noted that the BC basis function represents a subset of the functions proposed by Chen and Wilton in 1990 (Chen and Wilton 1990), which was named the dual basis. In other words, the idea of

DOI 10.1007/978-981-4560-75-7_6-2

these two basis functions is identical. Both of these two basis functions are a linear combination of RWG basis functions on the barycentrically refined triangles within a polygon pair and have the same dual basis property, i.e., approximately orthogonal to the original RWG basis function (Tong et al. 2009). The reason why the BC basis function has received more attention in the electromagnetics community is because of its success in the CMP (Andriulli et al. 2008), where the well-conditioned nature of the Gram matrices linking the BC basis functions to $\hat{n}_r \times$ RWG basis functions is ensured. Hence, one should rightfully call them the Chen–Wilton–Buffa–Christiansen (CWBC) basis function.

According to the decomposition in Eq. 87, the composite operator $\mathcal{T}^2$ can be further expanded as

$$\mathcal{T}^2 = \mathcal{T}_s^2 + \mathcal{T}_s\mathcal{T}_h + \mathcal{T}_h\mathcal{T}_s + \mathcal{T}_h^2 \tag{106}$$

where

$$\mathcal{T}_s^2(\boldsymbol{J}) = -\omega^2\mu^2\hat{\boldsymbol{n}}_{\boldsymbol{r}} \times \int_{S'} dS' g(\boldsymbol{r},\boldsymbol{r}') \left[\hat{\boldsymbol{n}}'_{\boldsymbol{r}} \times \int_{S''} dS'' g(\boldsymbol{r}',\boldsymbol{r}'')\boldsymbol{J}\right] \tag{107}$$

$$\mathcal{T}_s\mathcal{T}_h(\boldsymbol{J}) = -\hat{\boldsymbol{n}}_{\boldsymbol{r}} \times \int_{S'} dS' g(\boldsymbol{r},\boldsymbol{r}') \left[\hat{\boldsymbol{n}}'_{\boldsymbol{r}} \times \nabla' \int_{S''} dS'' g(\boldsymbol{r}',\boldsymbol{r}'')\nabla''\cdot\boldsymbol{J}\right] \tag{108}$$

$$\mathcal{T}_h\mathcal{T}_s(\boldsymbol{J}) = -\hat{\boldsymbol{n}}_{\boldsymbol{r}} \times \nabla \int_{S'} dS' g(\boldsymbol{r},\boldsymbol{r}')\nabla'\cdot \left[\hat{\boldsymbol{n}}'_{\boldsymbol{r}} \times \int_{S''} dS'' g(\boldsymbol{r}',\boldsymbol{r}'')\boldsymbol{J}\right] \tag{109}$$

$$\mathcal{T}_h^2(\boldsymbol{J}) = -\frac{1}{\omega^2\varepsilon^2}\hat{\boldsymbol{n}}_{\boldsymbol{r}} \times \nabla \int_{S'} dS' g(\boldsymbol{r},\boldsymbol{r}')\nabla'\cdot \left[\hat{\boldsymbol{n}}'_{\boldsymbol{r}} \times \nabla' \int_{S''} dS'' g(\boldsymbol{r}',\boldsymbol{r}'')\nabla''\cdot\boldsymbol{J}\right]. \tag{110}$$

Recalling the surface Helmholtz decomposition of the surface current $\boldsymbol{J}$ (Hsiao and Kleinman 1997),

$$\boldsymbol{J} = \nabla_S\phi + \hat{\boldsymbol{n}}_{\boldsymbol{r}} \times \nabla_S\psi \tag{111}$$

where ϕ and ψ are scalar functions defined on S. The first term of Eq. 111 is purely irrotational (curl-free) while the second term is purely solenoidal (divergence-free). It is easy to show that (Adams 2004)

$$\nabla_S \cdot (\mathcal{T}_h(\boldsymbol{J})) = \nabla_S \cdot (\hat{\boldsymbol{n}}_{\boldsymbol{r}} \times \nabla\Phi) = 0. \tag{112}$$

where $\nabla_S\cdot$ is the surface divergence operator. As a result, the square of the hypersingular term $\left(\mathcal{T}_h^2\right)$ in Eq. 110 is identically zero. Consequently, the decomposed operator $\mathcal{T}^2$ in Eq. 106 can be rewritten as

$$\mathcal{T}^2 = \mathcal{T}_s^2 + \mathcal{T}_h\mathcal{T}_s + \mathcal{T}_s\mathcal{T}_h. \tag{113}$$

Notice that from Eqs. 107 to 109, the $\mathcal{T}_s^2 \sim O(\omega^2)$, while the $\mathcal{T}_s\mathcal{T}_h + \mathcal{T}_h\mathcal{T}_s \sim O(\omega^0)$, when $\omega \to 0$. In other words, the $\mathcal{T}_s\mathcal{T}_h$ and $\mathcal{T}_h\mathcal{T}_s$ are frequency invariant, so that these two terms behave like the identity operator. Meanwhile, $\mathcal{T}_s^2$ behaves like a compact operator and approaches zero when $\omega \to 0$. Therefore, the total operator $\mathcal{T}^2$ can be considered as an identity operator plus a compact operator, making it a well-conditioned second-kind Fredholm integral operator immune to low-frequency breakdown. Meanwhile, $\mathcal{T}_h^2 \sim O(\omega^{-2})$ implies that the square of the hypersingular term has to be set to zero; otherwise it will

DOI 10.1007/978-981-4560-75-7_6-2

swamp the contributions from the other three terms at low frequencies. However, if $\boldsymbol{J}$ is of the order smaller than ω^0 and the higher-order current is also important for certain problems, the decomposed $\mathcal{T}^2$ (without $\mathcal{T}_h^2$) still decreases with frequency. It does not cause the breakdown of the MoM but causes the inaccuracy of the electric current at low frequencies.

Perturbation Method

The low-frequency inaccuracy problem happens in many formulations such as EFIE, MFIE, and CMP-EFIE. As detailed in (Zhang et al. 2003; Qian and Chew 2010), although the formulations are stable at extreme low frequencies, the current unknowns cannot be computed accurately. To remedy this inaccuracy problem, a perturbation method introduced in this section will be discussed based on the expanded CMP-EFIE formulation in Eq. 113 (Sun et al. 2013b).

By expanding the Green's function with Taylor series as

$$g(\boldsymbol{r},\boldsymbol{r}') \approx \frac{1}{4\pi R}\left[1 + ik_0R + \frac{1}{2}(ik_0R)^2\right] \tag{114}$$

the sub-matrices can be expanded with respect to a small parameter $\delta = ik_0$. Then, one has

$$\left[\overline{\boldsymbol{Z}}_p^q\right] = \left[\overline{\boldsymbol{Z}}_p^{q(0)}\right] + \delta\left[\overline{\boldsymbol{Z}}_p^{q(1)}\right] + \delta^2\left[\overline{\boldsymbol{Z}}_p^{q(2)}\right] + O\left(\delta^3\right) \tag{115}$$

where the impedance matrices for the smoothing term are given by

$$\left[\overline{\boldsymbol{Z}}_p^{s(0)}\right]_{mn} = \frac{\mu}{4\pi}\int_{S_m} dS\boldsymbol{f}_{pm}(\boldsymbol{r})\cdot\int_{S_n} dS'\frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|}\boldsymbol{f}_{pn}(\boldsymbol{r}') \tag{116}$$

$$\left[\overline{\boldsymbol{Z}}_p^{s(1)}\right]_{mn} = \frac{\mu}{4\pi}\int_{S_m} dS\boldsymbol{f}_{pm}(\boldsymbol{r})\cdot\int_{S_n} dS'\boldsymbol{f}_{pn}(\boldsymbol{r}') \tag{117}$$

$$\left[\overline{\boldsymbol{Z}}_p^{s(2)}\right]_{mn} = \frac{\mu}{8\pi}\int_{S_m} dS\boldsymbol{f}_{pm}(\boldsymbol{r})\cdot\int_{S_n} dS'|\boldsymbol{r}-\boldsymbol{r}'|\boldsymbol{f}_{pn}(\boldsymbol{r}') \tag{118}$$

and the impedance matrices for the hypersingular term are given by

$$\left[\overline{\boldsymbol{Z}}_p^{h(0)}\right]_{mn} = \frac{1}{4\pi\varepsilon}\nabla\int_{S_m} dS\boldsymbol{f}_{pm}(\boldsymbol{r})\cdot\int_{S_n} dS'\frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|}\nabla'\boldsymbol{f}_{pn}(\boldsymbol{r}') \tag{119}$$

$$\left[\overline{\boldsymbol{Z}}_p^{h(1)}\right]_{mn} = \frac{1}{4\pi\varepsilon}\nabla\int_{S_m} dS\boldsymbol{f}_{pm}(\boldsymbol{r})\cdot\int_{S_n} dS'\nabla'\cdot\boldsymbol{f}_{pn}(\boldsymbol{r}') \tag{120}$$

$$\left[\overline{\boldsymbol{Z}}_p^{h(2)}\right]_{mn} = \frac{1}{8\pi\varepsilon}\nabla\int_{S_m} dS\boldsymbol{f}_{pm}(\boldsymbol{r})\cdot\int_{S_n} dS'|\boldsymbol{r}-\boldsymbol{r}'|\nabla'\cdot\boldsymbol{f}_{pn}(\boldsymbol{r}'). \tag{121}$$

At low frequencies, $\overline{Z}_p^q \approx \overline{Z}_p^{q(0)}$, because the zeroth-order of the Green's function in Eq. 114 is equal to its original form of Eq. 90.

DOI 10.1007/978-981-4560-75-7_6-2

For the current and excitation vectors, the same notations as those in Qian and Chew (2010) are used as

$$ik_0\boldsymbol{j} = \boldsymbol{j}^{(0)} + \delta\boldsymbol{j}^{(1)} + \delta^2\boldsymbol{j}^{(2)} + O\left(\delta^3\right) \tag{122}$$

$$\boldsymbol{b} = \boldsymbol{b}^{(0)} + \delta\boldsymbol{b}^{(1)} + \delta^2\boldsymbol{b}^{(2)} + O\left(\delta^3\right) \tag{123}$$

where

$$\left[\boldsymbol{b}^{(0)}\right]_m = -\eta^{-1}\int_{S_m} dS\boldsymbol{f}_{RWGm}(r)\cdot\boldsymbol{E}_0 \tag{124}$$

$$\left[\boldsymbol{b}^{(1)}\right]_m = -\eta^{-1}\int_{S_m} dS\boldsymbol{f}_{RWGm}(r)\cdot\boldsymbol{E}_0\left(\hat{k}_{inc}\cdot\tilde{\boldsymbol{r}}\right) \tag{125}$$

$$\left[\boldsymbol{b}^{(2)}\right]_m = -\frac{1}{2}\eta^{-1}\int_{S_m} dS\boldsymbol{f}_{RWGm}(r)\cdot\boldsymbol{E}_0\left(\hat{k}_{inc}\cdot\tilde{\boldsymbol{r}}\right)^2 \tag{126}$$

and $\hat{k}_{inc}$ is the unit vector of the incident direction and the tilde above $\boldsymbol{r}$ indicates the normalization by l, where l is a typical length scale.

Substituting them into Eq. 113 and matching the coefficients of like powers of δ, a recurrent system of equations for the current functions $\boldsymbol{j}$ is obtained. Firstly, matching the zeroth order of δ gives the lowest-order equation as

$$\left[\overline{\boldsymbol{Z}}^{s(0)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\overline{\boldsymbol{Z}}^{h(0)}_{RW\,G} + \overline{\boldsymbol{Z}}^{h(0)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\overline{\boldsymbol{Z}}^{s(0)}_{RW\,G}\right]\cdot\boldsymbol{j}^{(0)} = \overline{\boldsymbol{Z}}^{h(0)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\boldsymbol{b}^{(0)} \tag{127}$$

It can be noticed that the impedance matrix is equal to the three-term CMP-EFIE without $\mathcal{T}_h^2$ in Eq. 113 in static regime, where the contributions from the $\mathcal{T}_h^2$ of $O(\omega^2)$ disappear. That means the resultant matrix is similar to the form of an identity operator as the aforementioned analysis. Hence, it has a good spectral property at low frequencies and its iterative solution converges well. For the inductive circuit problem like an inductor, the loop current is inversely proportional to frequency, so that the leading term of current is ω^{-1}. According to Eq. 122, the inductive current $ik_0\boldsymbol{j}$ would be of this zeroth-order solution.

For the plane-wave scattering problem, the leading term of the current $\boldsymbol{j}$ is of the order of ω^0. Hence, $\boldsymbol{j}^{(0)} = 0$ and one can obtain the first- and second-order equations by matching the first and second order of δ as

$$\left[\overline{\boldsymbol{Z}}^{s(0)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\overline{\boldsymbol{Z}}^{h(0)}_{RW\,G} + \overline{\boldsymbol{Z}}^{h(0)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\overline{\boldsymbol{Z}}^{s(0)}_{RW\,G}\right]\cdot\boldsymbol{j}^{(1)} = \overline{\boldsymbol{Z}}^{h(1)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\boldsymbol{b}^{(0)} + \overline{\boldsymbol{Z}}^{h(0)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\boldsymbol{b}^{(1)} \tag{128}$$

and

$$\begin{aligned}\left[\overline{\boldsymbol{Z}}^{s(0)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\overline{\boldsymbol{Z}}^{h(0)}_{RW\,G} + \overline{\boldsymbol{Z}}^{h(0)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\overline{\boldsymbol{Z}}^{s(0)}_{RW\,G}\right]\cdot\boldsymbol{j}^{(2)} &= \overline{\boldsymbol{Z}}^{s(0)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\boldsymbol{b}^{(0)} + \overline{\boldsymbol{Z}}^{h(1)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\boldsymbol{b}^{(1)} + \overline{\boldsymbol{Z}}^{h(0)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\boldsymbol{b}^{(2)} \\ &\quad - \left[\overline{\boldsymbol{Z}}^{s(0)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\overline{\boldsymbol{Z}}^{h(1)}_{RW\,G} + \overline{\boldsymbol{Z}}^{h(0)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\overline{\boldsymbol{Z}}^{s(1)}_{RWG}\right]\cdot\boldsymbol{j}^{(1)} \\ &\quad - \left[\overline{\boldsymbol{Z}}^{s(1)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\overline{\boldsymbol{Z}}^{h(0)}_{RW\,G} + \overline{\boldsymbol{Z}}^{h(1)}_{CW\,BC}\overline{\boldsymbol{G}}_m^{-1}\overline{\boldsymbol{Z}}^{s(0)}_{RW\,G}\right]\cdot\boldsymbol{j}^{(1)}\end{aligned} \tag{129}$$

DOI 10.1007/978-981-4560-75-7_6-2

Once the currents in the first three orders are obtained by solving (127, 128, and 129), the far-field results at very low frequencies can be obtained accurately.

It is important to notice that the impedance matrices on the left-hand side of Eqs. 127, 128, and 129 have both contributions from the smoothing and hypersingular terms, thus avoiding the null-space problem and breakdown at low frequencies. Similar to the augmented EFIE formulations, the decomposed CMP-EFIE operator is balanced at low frequencies after setting $\mathcal{T}_h^2 = 0$.

In particular, it can be observed that the leading term of the current $\boldsymbol{j}$ is of the order of ω^1 for the capacitive or differential circuit problems. It implies that the leading term of $ik_0\boldsymbol{j}$ is of the order of ω^2. Hence, the zeroth-order current $\boldsymbol{j}^{(0)}$ in Eq. 122, which is of the order of ω^0, and the first-order current $\boldsymbol{j}^{(1)}$ should be zero. For the delta-gap voltage source excitation, there is only the zeroth order $\boldsymbol{b}^{(0)}$ left in the right-hand side of the system. By utilizing this natural characteristic, the solution complexity of the matrix systems in Eqs. 127, 128, and 129 can be reduced significantly as

$$\left[\overline{\boldsymbol{Z}}_{CW\,BC}^{s(0)}\overline{\boldsymbol{G}}_m^{-1}\overline{\boldsymbol{Z}}_{RW\,G}^{h(0)} + \overline{\boldsymbol{Z}}_{CW\,BC}^{h(0)}\overline{\boldsymbol{G}}_m^{-1}\overline{\boldsymbol{Z}}_{RW\,G}^{s(0)}\right]\cdot\boldsymbol{j}^{(2)} = \overline{\boldsymbol{Z}}_{CW\,BC}^{s(0)}\overline{\boldsymbol{G}}_m^{-1}\boldsymbol{b}^{(0)} \tag{130}$$

where the second term in the left-hand side is usually much smaller than the first term, thus leading to a simpler formulation for the capacitive problems (Liu et al. 2014).

Optical Nanoantennas

A microwave or radiowave antenna is an electrical device which converts electric power into radiowaves, and vice versa. As analogues of microwave and radiowave antennas, optical nanoantennas (see Fig. 12) convert freely propagating optical radiation into localized electromagnetic energy, and vice versa (Novotny and Hulst 2011). They are indispensable building blocks for manipulating and controlling light–matter interaction at subwavelength scales. Nanoantennas have broad applications (see Fig. 13) in photodetection, solar energy, light emission, sensing, microscopy, and (surface-enhanced Raman) spectroscopy (Bharadwaj et al. 2009; Giannini et al. 2011; Biagioni et al. 2012; Krasnok et al. 2013). Currently, the tools of nanoscience and nanotechnology, such as focused ion beam milling (Muhlschlegel et al. 2005), electron-beam lithography (Kinkhabwala et al. 2009), and self-assembly schemes (Kalkbrenner et al. 2005), enable fabrication of nanoantennas down to nanoscales.

Theoretical Model

Computational electromagnetics (Chew et al. 2000; Jin 2014; Taflove and Hagness 2005), which is used for modeling the interaction of electromagnetic fields with physical objects and surrounding environment, plays an important role in characterizing and optimizing the optical design of nanoantennas. A rigorous, fast, and efficient solution to Maxwell's equations facilitates understanding underlying working principles, reducing experimental costs, and accelerating research and development periods. With the aid of state-of-the-art methods, critical physical parameters in the design of nanoantennas can be illustrated for observation and analyzed for optimization. It is highly desirable to know the strengths and weaknesses of various theoretical methods in modeling nanoantennas. Different from classical antennas, rigorous full-wave simulation of nanoantennas faces the following new challenges: (1) accurate near-field calculation; (2) computing numerical dyadic Green's functions in inhomogeneous electromagnetic environment; (3) complex multilayered, periodic, random, or composite structures; (4) highly dispersive and lossy materials (not perfect electric conductors); (5) strong evanescent wave couplings and mode hybridizations; and (6) multi-scale, nonlinear, and multiphysics effects.

DOI 10.1007/978-981-4560-75-7_6-2

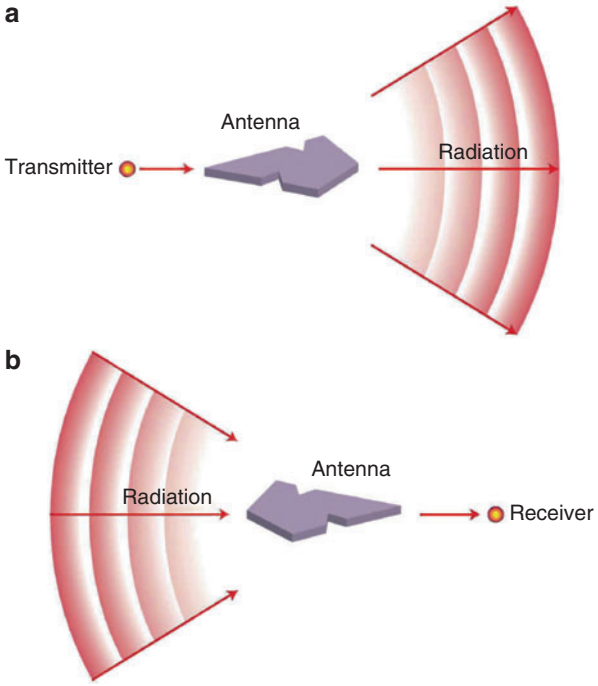


Fig. 12 Schematic patterns of nanoantennas. (**a**) Transmitting nanoantenna. (**b**) Receiving nanoantenna. *Arrows* indicate the direction of energy flow. The two configurations are related by the principle of reciprocity (The figure is taken from Novotny and Hulst (2011))

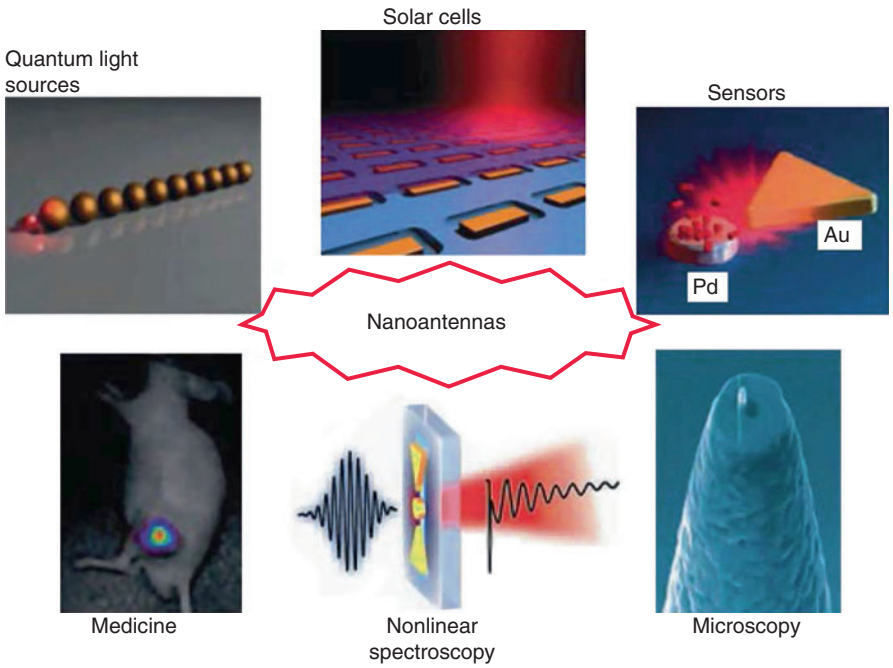


Fig. 13 Modern applications of nanoantennas (The figure is taken from Krasnok et al. (2013))

Time-Domain Methods Versus Frequency-Domain Methods

Most optical materials are dispersive; therefore, a recursive convolution method (Luebbers et al. 1990) or a piecewise linear recursive convolution method (Kelley and Luebbers 1996) must be adopted for time-

DOI 10.1007/978-981-4560-75-7_6-2

domain methods. For noble metals with plasmonic effects in the visible light range, such as silver and gold, the complex refractive index has to be described by a large number of summation terms in the Lorentz–Drude model leading to a long calculation time. However, for frequency-domain methods, an experimentally tabulated refractive index of the dispersive material is incorporated directly. Another difficulty in time-domain methods is the treatment of periodic boundary conditions particularly for the oblique incidence caused by the anticausal property of the Floquet theorem. Hence, the ability of frequency-domain methods to handle the case of oblique incidence is clearly an important advantage over time-domain methods (Veysoglu et al. 1993). Moreover, time-domain methods suffer from numerical dispersion and stability problems in contrast to frequency-domain methods. This drawback becomes serious if a three-dimensional large-scale nanoantenna structure is investigated. A significant merit associated with a time-domain method is a broadband simulation at the frequencies of interest. The frequency-domain method can employ parallel computing technique to circumvent the problem. Compared to frequency-domain methods, time-domain methods cannot obtain accurate eigenvalue solutions but save considerable computer resources. Additionally, time-domain methods are easier and more convenient to model nonlinear and multiphysics effects in nanoantennas.

Integral Equation Methods Versus Differential Equation Methods

Differential equation methods involving finite-difference and finite-element algorithms (Jin 2014; Taflove and Hagness 2005) can treat a variety of inhomogeneous boundary conditions conveniently. The methods have a powerful ability to model a complex nanoantenna structure. The resultant matrix by differential equation methods is sparse due to the "local" nature of differential operators of Maxwell's equations or wave equations. The method consumes the memory cost of $O(N)$ and complexity of $O(N)$ per matrix-vector multiplication in the Krylov subspace iteration algorithm (Vandervorst 1992). Moreover, multifrontal or multigrid methods (Davis and Duff 1997) can speed up the solution process of the differential equations. To simulate the interaction between light and nanoantennas, an efficient absorption boundary condition (Berenger 1994; Chew and Weedon 1994) and additional volume grids enclosing the nanoantenna have to be adopted.

In comparison with differential equation methods, integral equation methods (Chew 1999, 2008) connect field components to equivalent currents by using "global" integral operators represented with dyadic Green's functions. As a result, integral equation methods always guarantee higher accuracy but lead to full dense matrix. Fortunately, matrix-free fast algorithms (Chew et al. 2000), such as fast Fourier transform (Catedra et al. 1989) and fast multipole methods (Greengard and Rokhlin 1987; Song et al. 1997), can significantly reduce computer resources required by solving the dense matrix. Thanks to the Green's tensor, the integral equation methods automatically satisfy the radiation boundary condition but need singularity treatments. In particular, the surface integral equation method having a unique feature of surface triangulation produces much smaller unknowns. However, the method can only be employed to analyze a homogeneous or piecewise-homogeneous structure. For an arbitrary inhomogeneity or complex environment encountered in nanoantennas, the near-field calculation by the surface integral equation method is hard to implement.

Mode-Matching Methods

The mode-matching method (Chew 1999; Bienstman and Baets 2001) is a commonly used technique for the formulation of optical problems, especially for structures consisting of two or more separated regions. It is based on expanding the fields and matching them at the boundaries of different regions and thus lends itself naturally to the analysis of some nanoantenna structures. The most representative of mode-matching methods involves rigorous coupled-wave analysis (Moharam et al. 1995), scattering matrix method (Yonekura et al. 1999), plane-wave expansion method (Johnson and Joannopoulos 2001), and T-matrix

DOI 10.1007/978-981-4560-75-7_6-2

method (Mishchenko et al. 2010). Using cheap computer resources, these methods are particularly useful in characterizing the optical response of periodic or multilayered nanoantennas, Yagi-Uda nanoantennas, and nanoantenna array. Absorption and radiation (scattering) characteristics of nanoantennas can be obtained with a program. However, mode-matching methods are not well suited for characterizing plasmonic effects because a large quantity of modes are required to describe the plasmon coupling and hybridization.

Optical Nanoantennas Versus Classical Antennas

This section introduces the basic principles and physical parameters that govern the operation of nanoantennas. The differences between classical antennas and optical nanoantennas are summarized.

Source and Receiver

Classical antennas use current (voltage) and electrical load as a source and receiver, respectively. However, the source and receiver of nanoantennas become quantum objects such as molecules, atoms, ions, or quantum dots. At this length scale, the light–matter interaction becomes quantized (Novotny and Hulst 2011).

Resonant Transducer

The design of classical antennas relates to the wavelength λ of incident or radiative electromagnetic waves (Novotny and Hulst 2011). For example, a half-wave antenna has a length L of λ/2, and a Yagi-Uda antenna has fixed separations between elements at fractions of λ. At microwave and radiowave regimes, metals can be regarded as perfect electric conductors (PECs). Thus, the scaling law was adopted to design antennas from one frequency to another. However, the scaling breaks down at optical frequencies, where metals (gold, silver, etc.) are highly lossy and dispersive materials and the skin depth of metals becomes comparable to the nanoantenna size. The permittivity of metals at optical regime can be expressed by a superposition of the Lorentz–Drude model (Maier 2007). Although an effective wavelength (Novotny and Hulst 2011) or a low-frequency circuit theory has been utilized (Alu and Engheta 2008) to simplify the nanoantenna design, a rigorous solution to Maxwell's equations is essential to capture the wave interaction between nanoantennas and sources/receivers. Particularly, the size of a resonant nanoantenna is much smaller than the wavelength of radiation, which breaks the half-wavelength limit.

Absorption, Scattering, and Extinction Cross Sections

The far-field radiation pattern adopted in classical antennas is still very useful to describe the angular response of nanoantenna radiation or scattering. However, the total absorption and scattering properties of a nanoantenna are easier to measure and calculate for optimized designs.

Given a plane-wave excitation (laser excitation can be well approximated as a plane-wave excitation in nanoantenna experiments), the scattering cross section, which estimates the total scattering strength of a nanoantenna as a function of frequency, is defined as (Tsang et al. 2000)

$$\sigma_s = \frac{\int_S \Re e\left[\frac{1}{2}\mathbf{E}_{sca} \times \mathbf{H}^*_{sca}\right] \cdot d\mathbf{S}}{|\mathbf{S}_{inc}|} \tag{131}$$

where * denotes the complex conjugation, S is an arbitrary surface enclosing the nanoantenna, $\mathbf{S}_{inc} = \frac{1}{2}\left[\mathbf{E}_{inc} \times \left(\mathbf{H}^*_{inc}\right)\right]$ is the incident energy flux, and $\mathbf{E}_{sca}$ and $\mathbf{H}_{sca}$ are the scattered electric and

DOI 10.1007/978-981-4560-75-7_6-2

magnetic fields, respectively. Considering the total power absorbed by a nanoantenna, absorption cross section is of the form (Tsang et al. 2000)

$$\sigma_a = -\frac{\int_S \frac{1}{2}\Re e[\mathbf{E}\times\mathbf{H}^*]\cdot d\mathbf{S}}{|\mathbf{S}_{inc}|} = \frac{\int_V k_0\epsilon_r''(\mathbf{r})|\mathbf{E}(\mathbf{r})|^2 dV}{|\mathbf{E}_{inc}|^2} \tag{132}$$

where ϵ_r'' is the imaginary part of relative permittivity and k_0 is the wave number of free space (background medium). The extinction cross section describes the intrinsic losses of a nanoantenna system including both absorption loss and scattering (or leaky) loss. Hence, the sum of the scattering and absorption cross sections is the extinction cross-section σ_e

$$\sigma_e = \sigma_a + \sigma_s \tag{133}$$

Moreover, another mathematical expression for the extinction cross section is to employ the optical theorem at the far-field limit (Tsang et al. 2000)

$$\sigma_e = \frac{4\pi}{k_0}\Im m\left[\mathbf{e}_{inc}\cdot\overline{\mathbf{F}}\cdot\mathbf{e}_{inc}\right] \tag{134}$$

and

$$\mathbf{E}_{sca} = \frac{\exp(ik_0 r)}{r}\overline{\mathbf{F}}\cdot\mathbf{E}_{inc}, \quad r\to\infty \tag{135}$$

where i is the imaginary unit and $\mathbf{e}_{inc}$ is the polarization unit vector of the incident electric field $\mathbf{E}_{inc}$. The absorption, scattering, and extinction cross sections can be obtained by a postprocessing procedure after rigorously solving Maxwell's equations (He et al. 2012).

Directivity, Gain, and Reciprocity

Nanoantenna gain is a product of radiation efficiency η_r and directivity D, i.e., $G = \eta_r D$, which is the same as classical antennas. The radiation efficiency is defined as the ratio of the far-field power P_r radiated by the emitter with a nanoantenna to the total delivered power, i.e., $\eta_r = P_r/(P_r + P_l)$, where P_l is the total ohmic loss by nanoantenna absorption. This is also identical to classical antennas. However, considering internal losses of the emitter itself, an internal efficiency is defined as (Krasnok et al. 2013)

$$\eta_{in} = \frac{P_0}{P_0 + P_{0l}} \tag{136}$$

where P_{0l} and P_0 are the internal losses and the power radiated by the emitter in the absence of the nanoantenna, respectively. Finally, the radiation efficiency is expressed as

$$\eta_r = \frac{P_r}{P_r + P_l + P_0(1-\eta_{in})/\eta_{in}} \tag{137}$$

From the above equations, the introduction of nanoantennas could significantly improve the radiation efficiency of a poor emitter ($\eta_{in} \ll 1$). Because of a resonant enhancement, the radiated power P_r with the

DOI 10.1007/978-981-4560-75-7_6-2

nanoantenna is much larger than the radiated power P_0 without the nanoantenna. Similar to classical antennas, reciprocity holds true for nanoantennas if the medium is the reciprocal medium $\left(\overline{\epsilon}_r = \overline{\epsilon}_r^t, \overline{\mu}_r = \overline{\mu}_r^t\right)$. Considering two nanoantenna systems with different sources $\mathbf{J}_1$ and $\mathbf{J}_2$, which respectively produce the fields $\mathbf{E}_1$ and $\mathbf{E}_2$, then $\int_V \mathbf{E}_1 \cdot \mathbf{J}_2 dV = \int_V \mathbf{E}_2 \cdot \mathbf{J}_1 dV$.

Input Impedance Versus Local Density of States

For classical antennas, guided structures (transmission line, coplanar waveguide, microstrip line, etc.) are utilized to control the amplitude and phase of feeding source. The input impedance of antennas should be matched with the characteristic impedance of guided structures to minimize the reflection. However, the guided structure is not needed for nanoantennas if the source is a quantum emitter or their clusters, which are located around the feeding gap (element) of nanoantennas. In this situation, the impedance matching concept is not applicable to the nanoantenna design. Typically, a focused laser beam is used to excite or feed nanoantennas. The excitation field to quantum emitters is enhanced by the nanoantenna. Meanwhile, radiation field from quantum emitters is again enhanced by the nanoantenna. For surface-enhanced Raman scattering (SERS), the enhancement factor is written as (Giannini et al. 2011)

$$G_{SERS} = \frac{\left|\mathbf{E}^p(\mathbf{r}_0, \omega_i)\right|^2}{\left|\mathbf{E}^p_{inc}(\mathbf{r}_0, \omega_i)\right|^2} \cdot \frac{\left|\mathbf{E}^d(\mathbf{r}_0, \omega_r)\right|^2}{\left|\mathbf{E}^d_{inc}(\mathbf{r}_0, \omega_r)\right|^2} \tag{138}$$

where $\mathbf{r}_0$ is the emitter point and $|\mathbf{E}|^2/|\mathbf{E}_{inc}|^2$ are due to local field enhancements. ω_i and ω_r are the angular frequencies respectively for the incident (pump) laser beam and Raman-shifted radiation field. The enhancement factor can be easily obtained by a near-field simulation of nanoantennas. It should be noted that $\mathbf{E}^p_{inc}(\mathbf{r}_0, \omega_i)$ is the incident E-field from the laser beam (that can be regarded as a plane wave) while $\mathbf{E}^d_{inc}(\mathbf{r}_0, \omega_r)$ is the incident E-field from the emitter (that can be regarded as a Hertzian dipole).

To let quantum emitters efficiently radiate electromagnetic waves, photon local density of states (LDOS) at the emitter point or spontaneous emission (decay) rate of the emitter should be enhanced. The spontaneous emission rate counts the number of spontaneously emitted photons per unit time. The LDOS counts the number of electromagnetic modes at the emitter point. Each electromagnetic mode can be taken as a decay channel. The more decay channels there are, the easier it is for an excited atom to emit photons via returning to its ground state (Novotny and Hulst 2011). Therefore, more decay channels mean a larger spontaneous emission rate. Nanoantennas could significantly boost spontaneous emissions (named Purcell effect) due to the highly localized near-field enhancement induced by plasmonic effects. For example, the gap plasmonic mode supported at metallic bow-tie antennas enhances the single-molecule fluorescence (Kinkhabwala et al. 2009).

The LDOS in reciprocal, lossless, inhomogeneous, and anisotropic medium can be connected to dyadic Green's functions or Green's tensor

$$\rho(\mathbf{r}_0, \omega_0) = \frac{k_0^2}{\pi\omega_0} \mathrm{Tr}\left\{\overline{\epsilon}_r^{1/2} \cdot \Im m\left[\overline{\mathbf{G}}^e(\mathbf{r}_0, \mathbf{r}_0; \omega_0)\right] \cdot \overline{\epsilon}_r^{1/2} + \overline{\mu}_r^{1/2} \cdot \Im m\left[\overline{\mathbf{G}}^m(\mathbf{r}_0, \mathbf{r}_0; \omega_0)\right] . \overline{\mu}_r^{1/2}\right\} \tag{139}$$

where $\overline{\mathbf{G}}^e$ and $\overline{\mathbf{G}}^m$ are the electric and magnetic dyadic Green's functions in inhomogeneous medium. $\overline{\epsilon}_r$ and $\overline{\mu}_r$ are the relative permittivity and relative permeability tensors. For lossless, isotropic, inhomogeneous, and nonmagnetic medium, the above can be simplified to (Novotny and Hecht 2006)

DOI 10.1007/978-981-4560-75-7_6-2

$$\begin{aligned}\rho(\mathbf{r}_0,\omega_0) &= \sum_{\mathrm{k}} |\mathrm{u}_{\mathrm{k}}(\mathbf{r}_0)|^2 \delta(\omega_{\mathrm{k}}-\omega_0) \\ &= \frac{2\omega_0}{\pi c^2}\mathrm{Tr}\left\{\Im\mathrm{m}\left[\overline{\mathbf{G}}^{\mathrm{e}}(\mathbf{r}_0,\mathbf{r}_0;\omega_0)\right]\right\}\end{aligned} \tag{140}$$

where $c = 1/\sqrt{\mu\epsilon}$ is the speed of light in the medium. Here, $\mathbf{u_k}$ are the eigenmodes of E-field in inhomogeneous medium satisfying an orthogonal relation $\int_v \mathbf{u}_{\mathrm{k}'}^{\dagger}(\mathbf{r})\epsilon_r(\mathbf{r})\mathbf{u}_{\mathrm{k}}(\mathbf{r})d\mathbf{r} = \delta_{k',k}$. The LDOS can be separated into radiative and non-radiative parts, which are connected to radiation efficiency in Eq. 137. According to Fermi's golden rule (Novotny and Hecht 2006), the spontaneous emission rate γ_{se} is proportional to the LDOS

$$\gamma_{se} = \frac{\pi\omega_0}{3\hbar\epsilon_0}|\mathbf{p}|^2\rho(\mathbf{r}_0,\omega_0) \tag{141}$$

where $\mathbf{p}$ is the electric dipole moment of the atomic (molecular) transition.

Computational electromagnetics plays an important role in obtaining the Green's tensor in arbitrary inhomogeneous environment. The Green's tensor in inhomogeneous medium can be found by solving the following vector wave equations:

$$\nabla\times\overline{\mu}^{-1}\cdot\nabla\times\overline{\mathbf{G}}^{\mathrm{e}} - \omega_0^2\overline{\epsilon}\cdot\overline{\mathbf{G}}^{\mathrm{e}} = \overline{\mathbf{I}}\delta(\mathbf{r}-\mathbf{r}') \tag{142}$$

$$\nabla\times\overline{\epsilon}^{-1}\cdot\nabla\times\overline{\mathbf{G}}^{m} - \omega_0^2\overline{\mu}\cdot\overline{\mathbf{G}}^{m} = \nabla\times\left[\overline{\epsilon}^{-1}\cdot\overline{\mathbf{I}}\delta(\mathbf{r}-\mathbf{r}')\right] \tag{143}$$

A finite-difference method has been developed to discretize the delta function (monopole source) and derivative of delta function (dipole source) (Qiao et al. 2011). Taking a 2D line source as an example, one gets

$$\delta\left(x-m'\Delta_x, y-n'\Delta_y\right)|_{x=m'\Delta_x, y=n'\Delta_y} \approx \frac{1}{\Delta_x\Delta_y} \tag{144}$$

$$\begin{aligned}\frac{\partial\delta\left(x-m'\Delta_x, y-n'\Delta_y\right)}{\partial y} &\approx -\frac{\delta\left(x-m'\Delta_x, y-(n'+0.5)\Delta_y\right)}{\Delta_y} \\ &\quad + \frac{\delta\left(x-m'\Delta_x, y-(n'+0.5)\Delta_y\right)}{\Delta_y}\end{aligned} \tag{145}$$

where Δ_x and Δ_y are the spatial steps along the x and y directions and (m', n') is the grid index corresponding to the line source point. Using the above techniques and the finite-difference implementation, the numerical dyadic Green's functions can be found accordingly. Recently, using multilayered Green's functions and integral equation approach, the numerical Green's tensor around an arbitrarily shaped metallic nanostructure deposited on a multilayered structure has been calculated (Chen et al. 2012b).

Weak Coupling and Strong Coupling Regimes

Consider a nanoantenna operating at a single resonant mode $\mathbf{u_k}$, such as a metallic dipole antenna. In this situation, spontaneous emission rate in Eq. 141 can be rewritten as (Andreani et al. 1999)

DOI 10.1007/978-981-4560-75-7_6-2

$$\gamma_{se} = \frac{2}{3\hbar\epsilon_0}|\mathbf{p}|^2\frac{Q}{V} \tag{146}$$

where Q is the quality factor of the resonant antenna and $V = 1/|\mathbf{u_k}(\mathbf{r}_0)|^2$ is the mode volume. According to Jaynes–Cummings model, strong and weak coupling regimes can be distinguished by using the atom-field coupling constant (Andreani et al. 1999)

$$\kappa = \sqrt{\frac{|\mathbf{p}|^2\omega}{2\hbar\epsilon_0 V}} \tag{147}$$

Weak coupling obeys the condition $\kappa \ll \gamma_{se}$. In this irreversible regime, the effect of vacuum fluctuation fields on the emitter can be treated as a perturbation. Due to large radiation and absorption losses, spontaneously emitted photons cannot react back upon the emitter. Strong coupling obeys the condition $\kappa \gg \gamma_{se}$. In this reversible regime, photons remain in the cavity for a long time and are absorbed and re-emitted many times, which is called vacuum Rabi oscillation. Strong coupling allows the electron and photon entangled to be a polariton, which is a quasiparticle. The strong coupling regime is quite important to cavity quantum electrodynamics and quantum computing. In Eqs. 146 and 147, the important characteristic parameters Q and V can be rigorously calculated by an eigenvalue analysis. Calculation of the mode volume V is trivial if the eigenmode $\mathbf{u_k}$ around the (resonant) working frequency ω of the nanoantenna has been obtained. The quality factor can be extracted from corresponding complex eigenfrequency via the equation $Q = \Re e(\omega_k)/[2\Im m(\omega_k)]$. Several fast and universal eigenvalue solvers have been developed for an arbitrary electromagnetic system (Dai et al. 2012, 2013, 2014; Sha et al. 2014).

Design Rules for Optical Nanoantennas

Broadband Nanoantennas

A broadband nanoantenna can be an absorber for light harvesting or be a (near-field) concentrator for light enhancement (Atwater and Polman 2010). It also can be a far-field reflector for light trapping (Yu et al. 2010). Tapered nonresonant structures (Maksymov et al. 2011) support propagating (not standing) waves and thus a broadband response. Alternatively, the multisized nanoantenna (Cui et al. 2011) is another way to obtain a broadband absorption or scattering. The mode coupling (hybridization) between individual resonators induces a spectral overlap between eigenmodes (Ye and He 2010). Meanwhile, high-order modes could be excited via a strong mode coupling or symmetry breaking (Giannini et al. 2011). According to the coupled-mode theory, a single-mode resonator will be operated at the critical coupling regime if the radiation loss is balanced with the absorption loss (Yu et al. 2010), which is similar to a maximum power transfer condition in electrical engineering. Under this condition, a maximum absorption can be achieved but with a narrow bandwidth. Additionally, a broadband reflector can be realized by resonators operated at the over-coupling regime (Yu et al. 2010), where the radiation loss is much larger than the absorption (ohmic) loss.

Wavelength-Selective Nanoantennas

Wavelength selectivity by nanoantennas is required to resolve and match vibrational modes of a target molecule. Radiation and ohmic losses can be reduced to construct a narrow-band resonator with a high-quality factor. A defect in a periodic structure can induce the wavelength selectivity as well (Joannopoulos

DOI 10.1007/978-981-4560-75-7_6-2

et al. 2008). The wavelength selectivity has been realized by a single dielectric sphere, where the constructive and destructive interferences between the broad dipole eigenmode (first term of the spherical harmonics in Mie series) and the narrow quadrupole eigenmode (second term of the spherical harmonics in Mie series) induce an extraordinary Fano resonance with an asymmetric and narrow spectral line (Liu et al. 2012). The Fano spectral line is fundamentally different from the Lorentz spectral line achieved by a single-mode resonator.

Directional Nanoantennas

Directional nanoantennas are useful for far-field detection and sensing. A high directivity can be obtained by the following rules: (1) far-field interference by using an antenna array (Liu et al. 2011); (2) near-field interference by using the reflector, such as Yagi-Uda antennas (Taminiau et al. 2008); (3) using optical couplers, such as gratings, Fabry–Perot directive antenna, etc. (Martin-Moreno et al. 2003); and (4) using high-order modes and mode interference, such as a single sphere antenna, whispering gallery antenna, etc. (Gerard et al. 2009; Liu et al. 2012; Staude et al. 2013).

Metallic Nanoantennas Versus Dielectric Nanoantennas

Metallic nanoantennas have an unprecedented ability to concentrate and confine electromagnetic fields at a deep subwavelength scale. Dielectric nanoantennas (Pellegrini et al. 2009) have a larger geometric size but with a lower ohmic loss. Dielectric nanoantennas could easily excite high-order eigenmodes or exhibit both electric and magnetic resonances (Gerard et al. 2009; Staude et al. 2013) and thus show a strong scattering capability. Also, they can be semiconducting materials themselves and are more compatible to optoelectronic devices (Chen et al. 2012a). Dielectric nanoantennas can be integrated with metallic ones to achieve better performances (Devilez et al. 2010).

Nonlinear Optical Nanoantennas

Metals at optical regime have nonlinear susceptibility. Strong electromagnetic fields in plasmonic nanoantennas significantly enhance nonlinear processes including second harmonic generation, Kerr effect, four-wave mixing, etc. (Kauranen and Zayats 2012). Far-field detection of high-order harmonic generation depends on (1) pump field intensity at the frequency of fundamental mode, (2) spatial overlap between fundamental eigenmode and high-order harmonic eigenmode, and (3) outcoupling (far-field radiation) strength of high-order harmonic eigenmode. There are many designs in literatures to enhance the nonlinear effects of nanoantennas (Kauranen and Zayats 2012). The nonlinear response of nanoantennas can be rigorously modeled with the self-consistent solution to coupled Maxwell's equations and hydrodynamic equation (Corvi and Schaich 1986). Using a surface susceptibility tensor, a surface integral equation approach has been proposed to model the second harmonic generation from symmetric metallic objects, where the surface contribution is much larger than the bulk one (Makitalo et al. 2011; Xiong et al. 2014). It is worthy of mentioning that symmetry breaking is an efficient way to enhance the second harmonic generation from plasmonic nanoantennas (Pu et al. 2010).

Control of Light Angular Momentum by Nanoantennas

Similar to an electron, a photon has spin and orbital angular momenta, which are respectively defined as (Bliokh et al. 2013)

$$\mathbf{P}_s = \frac{1}{8\omega}\nabla \times \Im m\left[\frac{\epsilon_0}{\mu_r}(\mathbf{E}^* \times \mathbf{E}) + \frac{\mu_0}{\epsilon_r}(\mathbf{H}^* \times \mathbf{H})\right] \tag{148}$$

DOI 10.1007/978-981-4560-75-7_6-2

$$\mathbf{P}_o = \frac{1}{4\omega}\Im m\left[\frac{\epsilon_0}{\mu_r}(\mathbf{E}^* \cdot \nabla \mathbf{E}) + \frac{\mu_0}{\epsilon_r}(\mathbf{H}^* \cdot \nabla \mathbf{H})\right] \tag{149}$$

The summation of the spin and orbital angular momenta is equal to the electromagnetic momentum density

$$\mathbf{P} = \mathbf{P}_s + \mathbf{P}_0 = \frac{\mathrm{Re}(\mathbf{E}^* \times \mathbf{H})}{2c^2} \tag{150}$$

Moreover, the time-averaged helicity density, which characterizes the difference between the number of right-hand and left-hand circularly polarized photons, is defined as (Bliokh et al. 2013)

$$h = \frac{-1}{2\omega c_0}\Im m\,(\mathbf{E}^* \cdot \mathbf{H}) \tag{151}$$

From Eqs. 148 and 151, linear polarized light has zero spin angular momentum and zero helicity. The left-handed and right-handed circular polarizations have different spin angular momentum and helicity. From Eq. 149, the orbital angular momentum of a light beam depends on the field spatial distribution (not on the polarization). It can be split into an internal and an external orbital angular momenta. The internal orbital angular momentum is an origin-independent angular momentum of a light beam that can be associated with a helical or twisted wavefront. The external orbital angular momentum is the origin-dependent (center of the beam-dependent) angular momentum. The total angular momentum is conserved if light goes through a rotationally symmetric nanostructure. To rotate circularly polarized light, chiral metamaterials as building blocks have been constructed via mirror symmetry breaking (Kwon et al. 2008; Ye and He 2010; Ye et al. 2010). Consequently, if a rotationally symmetric nanoantenna constructed by chiral metamaterials converts the right-hand circularly polarized light to the left-hand circularly polarized light, both spin and orbital angular momenta of light will be changed based on the total angular momenta conservation. Additionally, all the above physical quantities including spin and orbital angular momenta and helicity density can be computed by classical computational electromagnetics.

Tunable Active Nanoantennas

Optical responses of nanoantennas involving resonance frequency, bandwidth, radiation pattern, near-field distribution, etc., can be dynamically tunable and switchable by using optical, electrical, and mechanical ways. Plasmonic effects could induce a strong optical force to reconfigure the geometry of nanoantennas (Bonakdar et al. 2012) or enhance nonlinear (Kerr) effects to modify refractive indices of antennas or loads (Chen and Alu 2010). Permittivity of liquid crystals (Berthelot et al. 2009) and graphene (Yao et al. 2013; Ren et al. 2013) can be tunable with electrostatic fields. Hence, they have been introduced to electrostatically tune optical responses of nanoantennas. Nanomechanical control of an optical antenna was also reported in literatures (Merlein et al. 2008).

Modal Analysis for Antenna Design

Modal analysis, as the preoccupation of many researchers, has been successfully adopted in the investigation of conducting enclosures such as waveguides and resonators at different frequency bands (Collin

1991; Van Bladel 1985). Recently, characteristic mode analysis gains popularity among antenna engineers as it provides a systematic approach for antenna design which is no longer experience or intuition based. Initiated by Garbacz and refined by Harrington and Mautz (Garbacz and Turpin 1971; Harrington and Mautz 1971), characteristic mode theory was popularized by the work of Cabedo-Fabres within the last decade (Cabedo-Fabres et al. 2007). On the other hand, the well-documented literature on natural mode analysis mainly focuses on bounded or semi-bounded systems (even dielectric resonator antennas are normally enclosed by impenetrable boundaries when their modal shapes are sought for). Recent effort addresses bounded and unbounded problems using a general framework based on natural mode expansion, which offers useful physical insight into antenna operation, as well as an alternative of CMA for modal design and engineering (Dai et al. 2012, 2014).

Characteristic Mode Analysis

Considering the electric field integral equation (EFIE) for a perfect electric conductor (PEC) object, the characteristic mode analysis (CMA), formulated on top of the method of moments (MoM), aims to obtain the characteristic current modes $\mathbf{J}_n$ supported on the PEC surface S by solving the generalized eigenvalue equation

$$\overline{\mathbf{Z}} \cdot \mathbf{J}_n = v_n \overline{\mathbf{R}} \cdot \mathbf{J}_n \tag{152}$$

or

$$\overline{\mathbf{X}} \cdot \mathbf{J}_n = \lambda_n \overline{\mathbf{R}} \cdot \mathbf{J}_n, \tag{153}$$

where $\overline{\mathbf{Z}} = \overline{\mathbf{R}} + i\overline{\mathbf{X}}$, and $v_n = 1 + i\lambda_n$. Meanwhile, $\overline{\mathbf{Z}}$ is the impedance matrix for a given frequency, whose entries are calculated as

$$\begin{aligned} \left(\overline{\mathbf{Z}}\right)_{mn} &= i\omega\mu \left\langle \mathbf{f}_m, \overline{\mathbf{G}}, \mathbf{f}_n \right\rangle_S \\ &= i\omega\mu \int_s \mathrm{d}\mathbf{r}\, \mathbf{f}_m(\mathbf{r}) \cdot \int_s \overline{\mathbf{G}}, (\mathbf{r}, \mathbf{r}') \cdot \mathbf{f}_n(\mathbf{r}') \mathrm{d}\mathbf{r}', \end{aligned} \tag{154}$$

where $\overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}')$ is the dyadic Green's function and $\mathbf{f}_n(\mathbf{r})$ are RWG basis functions.

Since $\overline{\mathbf{R}}$ and $\overline{\mathbf{X}}$ are dense matrices, the use of fast algorithms for full matrix-vector product is indispensable to solve for λ_n and $\mathbf{J}_n$ iteratively. Otherwise, CMA can only be used to analyze small objects. The characteristic value (eigenvalue) λ_n is important as its magnitude is proportional to the object's reactive power. When $\lambda_n = 0$, the correspondent characteristic mode $\mathbf{J}_n$ is at resonance which is efficient in receiving or radiating energy. When $\lambda_n > 0$ ($\lambda_n < 0$), mode $\mathbf{J}_n$ is storing electric (magnetic) energy. The characteristic angle can be calculated as $\phi_n = 180^\circ - \arctan(\lambda_n)$. Hence, mode $\mathbf{J}_n$ is at resonance when $\phi_n = 180^\circ$ and stores energy only when $\phi_n = 90^\circ$ or 270°.

When an incident field $\mathbf{E}_{inc}$ exists, the induced current on the PEC object can be calculated by solving

$$\overline{\mathbf{Z}} \cdot \mathbf{I}_{ind} = \mathbf{V} \tag{155}$$

where

$$(\mathbf{V})_n = -\langle \mathbf{f}_n, \mathbf{E}_{inc} \rangle_S. \tag{156}$$

DOI 10.1007/978-981-4560-75-7_6-2

Since the orthogonal property of characteristic modes holds for $\lambda_m \neq \lambda_n$

$$\mathbf{J}_m^T \cdot \overline{\mathbf{R}} \cdot \mathbf{J}_n = 0, \tag{157}$$

one can easily expand $\mathbf{I}_{ind}$ in terms of $\mathbf{J}_n$ as

$$\mathbf{I}_{ind} = \sum_n \alpha_n \mathbf{J}_n \tag{158}$$

where the modal expansion coefficients are

$$\alpha_n = \frac{1}{1 + i\lambda_n} \frac{\mathbf{J}_n^T \cdot \mathbf{V}}{\mathbf{J}_n^T \cdot \overline{\mathbf{R}} \cdot \mathbf{J}_n} \tag{159}$$

Natural Mode Expansion

Natural mode analysis, when formulated in terms of integral equations, differs from CMA in that it is closely related to the problem of seeking for $\mathbf{J}$ such that $\overline{\mathbf{Z}} \cdot \mathbf{J} = 0$. This requires one to find natural resonances of an open system or, equivalently, roots of $\det\left[\overline{\mathbf{Z}}(\omega_n)\right] = 0$ on the complex plane. On the other hand, CMA should always yield real λ_n as a function of given frequencies along the real axis, while the correspondent $\mathbf{J}_n$ can never satisfy $\overline{\mathbf{Z}} \cdot \mathbf{J}_n = 0$.

It is quite challenging to find natural modes with integral equation solvers as the complexity of calculating $\det\left[\overline{\mathbf{Z}}\right]$ is $O(N^3)$ where N is the number of unknowns. In the singularity expansion-based studies, natural modes are extracted from the measured data rather than being directly calculated (Baum et al. 1991). However, natural mode analysis can be easily carried out if it is formulated in terms of differential equations.

Bounded Case

Consider an arbitrary inhomogeneity (where $\overline{\epsilon}$ and $\overline{\mu}$ are tensors) bounded in domain V by impenetrable boundaries at surface S. The electric field $\mathbf{E}_n$ and magnetic field $\mathbf{H}_n$ of a natural mode are governed by the linear eigenvalue equations

$$\nabla \times \overline{\mu}^{-1} \cdot \nabla \times \mathbf{E}_n = \omega_n^2 \overline{\epsilon} \cdot \mathbf{E}_n \tag{160}$$

and

$$\nabla \times \overline{\epsilon}^{-1} \cdot \nabla \times \mathbf{H}_n = \omega_n^2 \overline{\mu} \cdot \mathbf{H}_n, \tag{161}$$

respectively. On the surface S, the field satisfies $\hat{n} \times \mathbf{E}_n = 0$ or $\hat{n} \times \mathbf{H}_n = 0$ for PEC or PMC boundary, respectively, where $\hat{n}$ is the unit normal vector of S.

Taking the case of $\mathbf{E}_n$ as an example, similar analysis applies to the case of $\mathbf{H}_n$. When the medium is lossless, namely, $\overline{\epsilon} = \overline{\epsilon}^\dagger$ and $\overline{\mu} = \overline{\mu}^\dagger$, or reciprocal, namely, $\overline{\epsilon} = \overline{\epsilon}^T$ and $\overline{\mu} = \overline{\mu}^T$, one can show that for $\omega_m^2 \neq \omega_n^2$,

$$\left\langle \mathbf{E}_m^\gamma, \overline{\epsilon} \cdot \mathbf{E}_n \right\rangle_V = \int_V \mathbf{E}_m^\gamma(\mathbf{r}) \cdot \overline{\epsilon} \cdot \mathbf{E}_n(\mathbf{r})\, d\mathbf{r} = 0, \tag{162}$$

DOI 10.1007/978-981-4560-75-7_6-2

where $\gamma = *$ for the lossless case and omitted for the reciprocal case. Nontrivial degenerate modes satisfying $\omega_m^2 = \omega_n^2 \neq 0$ can be orthogonalized by applying the Gram–Schmidt process.

The operator $\left(\nabla \times \overline{\mu}^{-1} \cdot \nabla \times\right)$ has a null space spanned by countably infinite irrotational eigenmodes $\mathbf{E}_q$ associated with zero eigenvalues. They are orthogonal to nontrivial modes $\mathbf{E}_n$ and are expressible as $\mathbf{E}_q = \nabla \psi_q$, where ψ_q are solutions of eigenvalue equation

$$\nabla \cdot \overline{\epsilon} \cdot \nabla \psi_q(\mathbf{r}) = k_q^2 \psi_q(\mathbf{r}) \tag{163}$$

associated with eigenvalues k_q^2. Under prescribed boundary conditions (Dirichlet or Neumann) of a bounded V, for $k_p^2 \neq k_q^2$, the orthogonality property $\left\langle \psi_p^{\gamma}, \psi_p \right\rangle_V = 0$ or $\left\langle \mathbf{E}_p^{\gamma}, \overline{\epsilon} \cdot \mathbf{E}_q \right\rangle_V = 0$ holds, where γ is defined for the lossless case and reciprocal case, accordingly.

By using the eigenbasis $\mathbf{E}_n$ including $\mathbf{E}_q$, the solution of Eq. 8 in V can be expanded as

$$\mathbf{E}(\mathbf{r}) = \sum_n^{\infty} \alpha_n \mathbf{E}_n(\mathbf{r}) \tag{164}$$

where modal expansion coefficients α_n can be obtained using the modal orthogonality as

$$\alpha_n = \frac{1}{\omega_n^2 - \omega^2} \frac{\left\langle \mathbf{E}_n^{\gamma}, i\omega \mathbf{J} \right\rangle_V}{\left\langle \mathbf{E}_n^{\gamma}, \overline{\epsilon} \cdot \mathbf{E}_n \right\rangle_V} \tag{165}$$

In low-frequency applications, $\mathbf{E}_q$ are crucial in expanding the field. If the source $\mathbf{J}$ is not divergence-free, $\mathbf{E}_q$ may need to be included in the modal expansion solution.

When the inhomogeneity is lossy, namely, $\overline{\epsilon} \neq \overline{\epsilon}^{\dagger}$ or $\overline{\mu} \neq \overline{\mu}^{\dagger}$, or nonreciprocal, namely, $\overline{\epsilon} \neq \overline{\epsilon}^{T}$ or $\overline{\mu} \neq \overline{\mu}^{T}$, Eq. 162 is no longer valid. An approach for constructing orthogonal eigenmodes is to introduce an auxiliary system to the original one as follows (Chen and Lien 1980):

$$\Delta \times \left[\overline{\mu}^{-1}\right]^{\nu} \cdot \nabla \times \mathbf{E}_n^a = \left(\omega_n^2\right)^{\gamma} \overline{\epsilon}^{\nu} \cdot \mathbf{E}_n^a \tag{166}$$

where $\nu = \dagger$ and $\nu = T$ for lossy and nonreciprocal cases, respectively, and γ is defined the same as before. Equation 166 should be constructed in the same domain V as in Eq. 160, and the auxiliary fields $\mathbf{E}_n^a$ should satisfy the same boundary conditions as in Eq. 160 but with different material media. Therefore, one can obtain the nondegenerate orthogonality property as

$$\left\langle \mathbf{E}_m^{a\gamma}, \overline{\epsilon} \cdot \mathbf{E}_n \right\rangle_V = 0. \tag{167}$$

Similar to the original problem, degenerate curl-free modes of the auxiliary system satisfy $\mathbf{E}_q^a = \nabla \psi_q^a$ where ψ_q^a are eigensolutions of

$$\nabla \cdot \overline{\epsilon} \cdot \nabla \psi_q^a(\mathbf{r}) = \left(k_q^2\right)^{\gamma} \psi_q^a(\mathbf{r}) \tag{168}$$

Again, for $k_p^2 \neq k_q^2$, one has $\left\langle \mathbf{E}_p^{a\gamma}, \overline{\epsilon} \cdot \mathbf{E}_q \right\rangle_V = 0$. Therefore, the field excited by any current $\mathbf{J}$ can be expanded in terms of the complete set (assumed to be) of eigenbasis $\mathbf{E}_n$ where α_n are in a simple form of

DOI 10.1007/978-981-4560-75-7_6-2

$$\alpha_n = \frac{1}{\omega_n^2 - \omega^2} \frac{\left\langle \mathbf{E}_n^{a\gamma}, i\omega\mathbf{J} \right\rangle_V}{\left\langle \mathbf{E}_n^{a\gamma}, \overline{\epsilon} \cdot \mathbf{E}_n \right\rangle_V}. \tag{169}$$

Unbounded Case

A straightforward extension of natural mode expansion (NME) to unbounded problems is to let S tend to infinity. A small loss has to be introduced to the medium in V, which is equivalent to requiring the wave to be outgoing at infinity. The small loss also guarantees the uniqueness of the field solution due to a given source (Chew 1990).

Based on the modal behavior, natural modes can be approximately classified into trapped modes and exterior modes. The former resonate due to the inhomogeneity with most energy confined, while the latter resonate between the inhomogeneity and the far boundaries. Note that the energy confinement for a trapped mode in 3D problems is never perfect, unless the inhomogeneity is a perfectly enclosed cavity or a dielectric with an infinitely high refractive index. In certain geometries, some trapped modes are coupled to external radiation, which are also regarded as leaky modes.

As S approaches infinity, the spectra of trapped modes slightly broaden while the modal shapes remain almost unchanged. On the other hand, exterior modes are not "immune" to external variations. When the boundaries expand, the distribution of exterior modes on the complex ω plane becomes increasingly dense, which eventually yields a continuum. Thus, one may intuitively write the modal expansion solution of unbounded fields in the form of

$$\mathbf{E}(\mathbf{r}) = \sum_{n=1}^{N} \alpha_n \mathbf{E}_n(\mathbf{r}) + \int \mathrm{d}\omega' \alpha(\omega') \mathbf{E}(\omega'\mathbf{r}). \tag{170}$$

The discrete summation in Eq. 170 corresponds to trapped modes, while the integral corresponds to continuous exterior modes where $\alpha(\omega')$ should include the density of states.

In numerical studies, one can replace PEC/PMC boundaries with PMLs (Berenger 1994; Chew and Weedon 1994) to emulate unbounded media. When PMLs are implemented, the eigenspectra become discrete from which trapped modes can be easily distinguished from exterior ones since the quality factors

$$Q = \left| \frac{\Re e[\omega_n]}{2\Im m[\omega_n]} \right| \tag{171}$$

of the former are obviously larger. Moreover, exterior modes in this case include resonances inside PMLs which have large $\Im m[\omega_n]$. Using PMLs enables one to correctly capture the physics of trapped modes which are important in characterizing the field behavior of an excited system. Therefore, the model order can be greatly reduced through NME using only a few important modes.

The governing Eqs. 160 and 161 can be discretized with conventional differential equation solvers, such as FDM and FEM. The resultant generalized eigenvalue equations only have sparse matrices and therefore can be easily solved with iterative eigensolvers, such as Lanczos and Arnoldi methods (Lehoucq and Sorensen 1996). Shift-invert techniques can be used to calculate interior spectrum of the system, from which one may extract trapped modes that are physically important. For example, the −1st-, 0th-, and +1st-order modes of a composite right-/left-handed (CRLH) waveguide slot are found as shown in Fig. 14a–c. Consider an open-ended CRLH waveguide slot antenna (Fig. 15a) (Dong and Itoh 2010), where the substrate has a thickness of $h = 1.27$ mm and a relative permittivity of Eq. 44. Parameters in Fig. 15a are set as $\omega_1 = 0.3$ mm, $\omega_2 = 0.45$ mm, $\omega_3 = 9.6$ mm, and $\omega_s = 3.9$ mm. Figure 15b shows 50 natural frequencies searched around 11 GHz. One can easily identify five trapped modes as they have

DOI 10.1007/978-981-4560-75-7_6-2

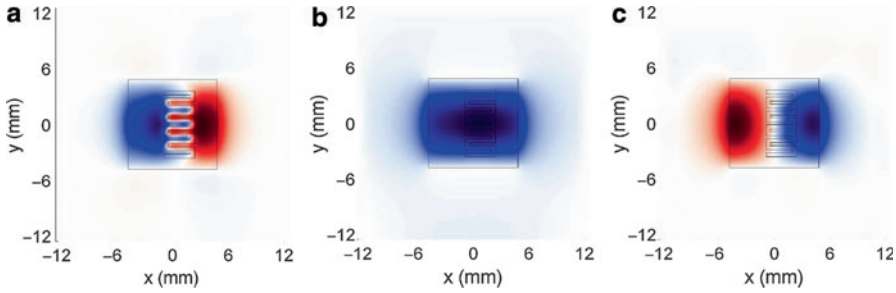


Fig. 14 Simulated modal shapes $\Re e[E_z]$ on an xy plane for several trapped modes: (**a**) −1st-order mode. (**b**) 0th-order mode. (**c**) +1st-order mode

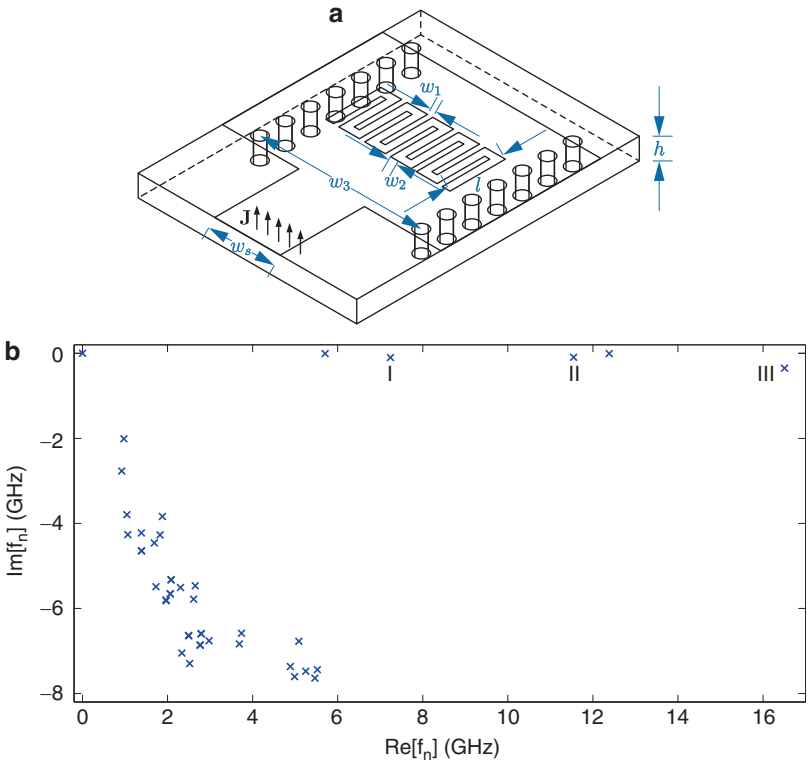


Fig. 15 (**a**) Open-ended CRLH-SIW slot antenna. (**b**) Computed complex eigenfrequencies around 11 GHz

small $\Im m[f_n]$, among which I, II, and III are −1st-, 0th-, and +1st-order resonances, respectively. The rest with large $\Im m[f_n]$ are exterior modes, which are well separated from trapped ones due to the existence of PMLs.

When the antenna is excited by a z-polarized current sheet operating at 5.9 GHz, the excited field $\Im m[E_z^{\mathrm{dir}}]$ can be approximately expanded using five trapped modes as good agreement is observed between Fig. 16a, b. More quantitatively, comparisons of the directly calculated field and the expanded fields with different numbers of trapped modes along $y = 0.36$ mm are given in Fig. 16c. The relative error is shown in Fig. 16d with respect to the number of modes used to expand the field. The small error near the source (Fig. 16c) may be reduced by taking the irrotational modes into account as the source is not divergence-free. A reduced modal approximation of the excited field in a properly operating antenna is achieved since only a few trapped modes dominate the expanded solution. Provided that trapped modes can be solved for by the shift-invert technique in a preprocessing stage, NME is well suited for a fast wideband field

DOI 10.1007/978-981-4560-75-7_6-2

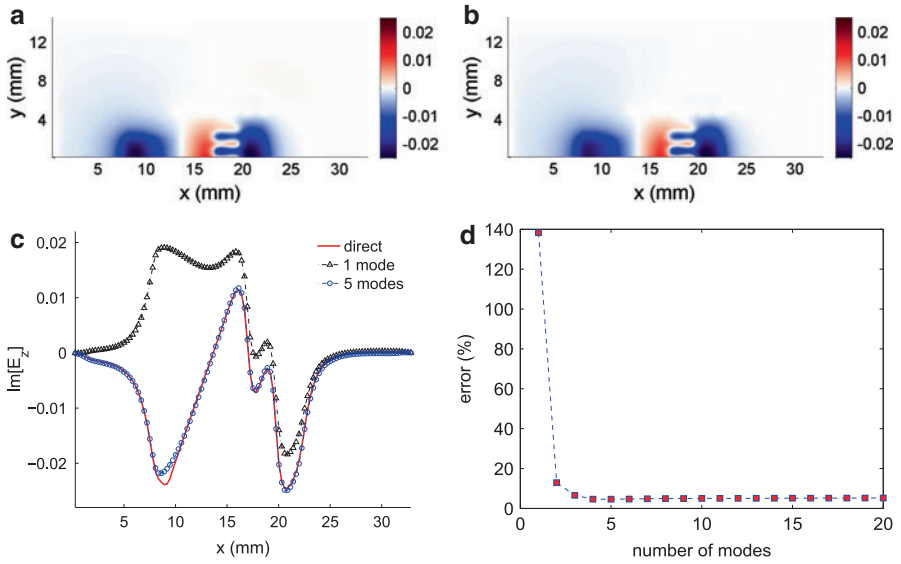


Fig. 16 A current sheet operating at 5.9 GHz is introduced to the CRLH waveguide slot antenna. $\mathcal{I}m[E_z]$ of (**a**) directly calculated field on xy plane at $z = 0.635$ mm. (**b**) Expanded field using trapped modes on the same plane. (**c**) Comparisons of directly calculated and expanded $\mathcal{I}m[E_z]$ along an observation line $y = 0.36$ mm. (**d**) Relative error with respect to the number of modes used to expand the field

approximation and for problems with different source configurations since the field reconstruction with a few modes consumes negligible CPU time.

Incorporation of PMLs in natural mode analysis leads to linear eigenvalue problems since PMLs behave as a wideband absorber for electromagnetic waves, whose working frequency can always be predetermined. This greatly simplifies the calculation of natural modes and renders NME a much simpler expression than that in Baum et al. (1991). Besides indicating modal radiation ability by modal quality factors, natural mode analysis also provides guidance on source placement for the excitation of desired modes. Furthermore, NME offers more flexibility in modeling complicated antennas than CMA where dielectric parts are included.

Equivalence Principle Algorithm for Antennas

The lack of analytical methods in terms of closed-form formulas or approximate asymptotic expressions has been a real motivation for antenna engineers to invoke computational electromagnetic methods to analyze and design antennas. Although the currently available numerical methods can address many problems in antenna structures modeling, there are other issues mostly related to multi-scale problems which cannot be properly addressed by these methods (Li and Chew 2007).

Multi-scale Problems

In general, traditional modeling techniques mainly focus on one scale (e.g., macroscopic or microscopic scales) and the physics change significantly over these scales. For example, when the macroscopic scale (or macroscale) is used, only the effect of the larger scales (sizes) is of interest and the effect of the smaller scales is modeled by some constitutive relations (Weinan 2012). It is known that macroscopic scale is efficient and microscopic scale is accurate. Hence, simultaneous incorporation of models at different scales helps to exploit the advantages offered by all of those scales. This so-called multi-scale concept for the electromagnetic modeling has been the research focus in the past two decades. A typical multi-scale

DOI 10.1007/978-981-4560-75-7_6-2

electromagnetic problem can be raised because of the nature of the electromagnetic fields. In fact, electromagnetic physics can be divided into: circuit physics, wave physics, and ray physics. From a numerical point of view, to model a multi-scale problem where both low-frequency circuit physics and high-frequency wave physics coexist, disparate meshing is not avoidable. This leads to a large number of unknowns reducing the convergence rate of the iterative solvers. As an example, consider a small patch antenna mounted on roof of a car. The antenna dimensions are in centimeters while the dimensions of the car are in meters. These are examples of multi-scale problems. To address these problems, a few ideas have been suggested where the domain decomposition method (DDM), the most popular one, will be discussed next.

Domain Decomposition Method

The main idea of domain decomposition method relies on dividing a larger solution domain into smaller nonoverlapping subdomains. Each of these individual subdomains can be solved independently while the interaction between adjacent subdomains is included through interface boundary conditions to guarantee the unique solution. In fact, the solutions of each subdomain are stitched together using some transmission condition (e.g., Robin-type transmission condition) and iterative algorithm to derive the solution of the large problem (Lu et al. 2008). In this manner, the solution of each subdomain is independent of the other subdomains. Hence, the meshing strategy can be optimally selected for each subdomain. The freedom to select the mesh strategy helps to facilitate the parallelization of the solution and its reuse and also to improve the condition number of the matrix system. These advantages have encouraged the combination of DDM with other electromagnetic solvers such as finite-element method and finite-difference frequency-domain method. Then, the domain decomposition method was used for solving three-dimensional large finite periodic problems such as photonic or electromagnetic bandgap structures and frequency-selective surfaces. In addition, relatively large dielectric aperture array, patch antenna array, and slotted waveguide array were successfully analyzed using DDM. To extend the application of DDM for integral equation-based solvers, the equivalence principle algorithm (EPA) was introduced by Chew (Chew and Lu 1993; Lu and Chew 1995).

Equivalence Principle Algorithm

Based on EPA, the entire solution domain is decomposed into many subdomains by defining appropriate closed surfaces which are called equivalence surfaces. The subdomains are enclosed by the equivalence surfaces and communicate mutually only through the closed surfaces. Consider a simple scenario where there are only two objects, namely, Object1 and Object2, and they are enclosed by two separate surfaces, namely, S_1 and S_2, respectively (Fig. 17). The entire EPA procedure for this case can be cast into four steps:

1. Outside-in propagation
   Initially, the collective effect of every other exterior sources is modeled by equivalent magnetic and/or electric currents on the equivalence surface S_1. These equivalent sources generate the original fields inside the S_1 and zero field outside. These equivalent currents, called incident currents are denoted as $\mathbf{J}_1^{inc}$ and $\mathbf{M}_1^{inc}$.
2. Solving for the induced current on enclosed object
   Having the incident field on Object1 produced by the equivalence currents and by invoking a solution method (e.g., MoM) the induced currents on Object1 denoted as $\mathbf{J}_1$ and $\mathbf{M}_1$, can be found.
3. Inside-out propagation
   Having found the currents on Object1, the scattered field distribution can be readily calculated anywhere inside or on the surface of S_1. Then, the equivalence principle can be used to calculate the

DOI 10.1007/978-981-4560-75-7_6-2

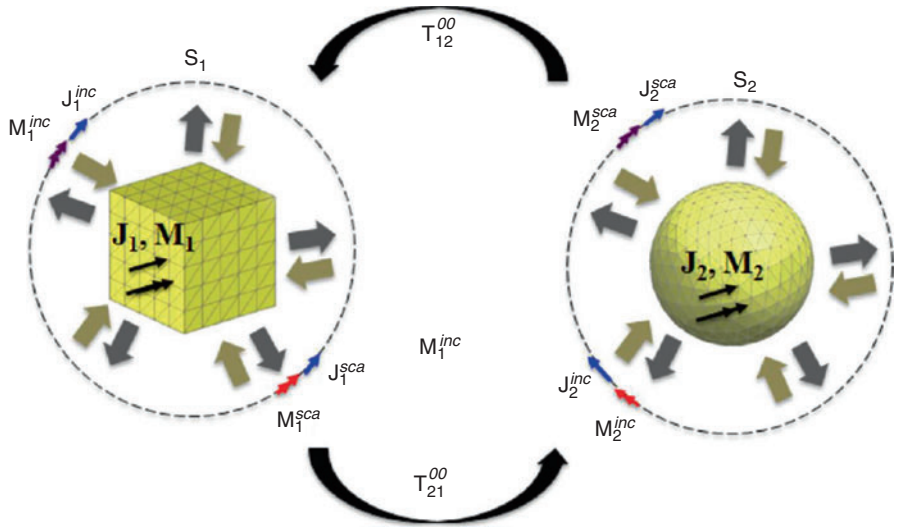


Fig. 17 A simple two-object scenario to explain the equivalence principle algorithm

equivalent currents on the surface (by defining the region of interest to be outside of S_1). These equivalent currents, called scattered currents are denoted as $\mathbf{J}_1^{sca}$ and $\mathbf{M}_1^{sca}$.

A similar procedure is applied to the other defined equivalence surfaces and the enclosed objects in a parallel manner. However, the procedure will not capture the whole physics until the interaction between the surfaces is also considered.

4. Interaction between equivalence surfaces

The interaction between S_1 and other equivalence surfaces (i.e., S_2 in this case) is modeled by introduction of an operator called the translation operator. Its main function is to transfer the effect of the equivalent currents from one equivalence surface to another equivalence surface. In contrast with domain decomposition methods for differential equation-based solvers, the subdomains do not have to touch each other (i.e., white space between the subdomains is allowed).

The entire EPA procedure for the two-object two-equivalence surface scenario can be described through the following equations:

$$\begin{bmatrix} \mathbf{J}_m^{sca} \\ \mathbf{M}_m^{sca} \end{bmatrix} = \begin{bmatrix} -\hat{n} \times \mathcal{K}_H^s \\ -\hat{n} \times \mathcal{L}_E^s \end{bmatrix} \cdot \left[\mathcal{L}_{\text{solver}}^{-1}\right] \cdot \begin{bmatrix} -\mathcal{L}_E^s & -\mathcal{K}_H^s \end{bmatrix} \cdot \begin{bmatrix} \mathbf{J}_m^{inc} \\ \mathbf{M}_m^{inc} \end{bmatrix} = S_{mm} \cdot \begin{bmatrix} \mathbf{J}_m^{inc} \\ \mathbf{M}_m^{inc} \end{bmatrix} \tag{172}$$

$$\begin{bmatrix} \mathbf{J}_m^{sca} \\ \mathbf{M}_m^{sca} \end{bmatrix} = \begin{bmatrix} -\hat{n} \times \mathcal{K}_H^s & -\hat{n} \times \mathcal{L}_H^s \\ -\hat{n} \times \mathcal{L}_E^s & -\hat{n} \times \mathcal{K}_E^s \end{bmatrix} \cdot \begin{bmatrix} \mathbf{J}_n^{sca} \\ \mathbf{M}_n^{sca} \end{bmatrix} = T_{mn}^{oo} \cdot \begin{bmatrix} \mathbf{J}_n^{sca} \\ \mathbf{M}_n^{sca} \end{bmatrix} \tag{173}$$

where $\mathcal{L}_E$, $\mathcal{K}_E$, $\mathcal{L}_H$, and $\mathcal{K}_H$ are integral operators and they are well defined in the literature (Li and Chew 2007). Also $\mathcal{L}_{\text{solver}}^{-1}$ represents the current solver part. Moreover, S_{mm} represents scattering operator for equivalence surface S_m and T_{mn}^{oo} shows the translation operator between surface m and n where $m, n = \{1, 2\}$. Combination of these two equations results in a system of equations as

$$\begin{bmatrix} \mathbf{J}_1^{sca} \\ \mathbf{M}_1^{sca} \end{bmatrix} - S_{11} \cdot T_{12}^{oo} \cdot \begin{bmatrix} \mathbf{J}_2^{sca} \\ \mathbf{M}_2^{sca} \end{bmatrix} = S_{11} \cdot \begin{bmatrix} \mathbf{J}_1^{inc} \\ \mathbf{M}_1^{inc} \end{bmatrix} \tag{174}$$

DOI 10.1007/978-981-4560-75-7_6-2

$$-S_{22}\cdot T_{21}^{oo}\cdot\begin{bmatrix}\mathbf{J}_1^{sca}\\ \mathbf{M}_1^{sca}\end{bmatrix}+\begin{bmatrix}\mathbf{J}_2^{sca}\\ \mathbf{M}_2^{sca}\end{bmatrix}=S_{22}\cdot\begin{bmatrix}\mathbf{J}_2^{inc}\\ \mathbf{M}_2^{inc}\end{bmatrix} \tag{175}$$

By solving this system of equations, the unknown scattered equivalent currents on equivalence surfaces will be calculated. Similar equations can be derived for more objects with more equivalence surfaces.

EPA for Connected Regions

The EPA procedure works well for objects which are not intercepted by the equivalence surface. However, in the antenna applications and many other electromagnetic problems, unconnected (isolated) region models are seldom used and one region may be connected with some other regions through wires, lumped elements, switches, and ground plane. In this case, the original continuous currents will be cut by the equivalence surface which leads into electric charge accumulation on both sides at the connection. This charge accumulation and current discontinuity result in a singular field generation which will induce a singular equivalence current on the intercepting equivalence surfaces and affects the accuracy of the results. Moreover, having objects very close to the equivalence surface requires computationally costly singularity subtraction and near-field treatments. In the literature, three methods, namely, singularity matching approach, fictitious junction basis method, and the tap basis scheme have been suggested to retain the current continuity and avoid the singularity. Tap basis scheme has been shown to model the continuous current accurately and no singular integral will be involved (Li and Chew 2007, 2008).

Tap Basis Scheme

For scenarios where the enclosed objects by an equivalence surface are connected to other objects enclosed by other equivalence surfaces, a tap basis scheme has been introduced in (Li and Chew 2007, 2008) to keep the current flowing smoothly in and out of the equivalence surfaces. Assume a metallic strip which is divided into two sections by defining two-equivalence surfaces (Fig. 18a). The strip is then divided into three regions by defining a transition region or tap basis region (Fig. 18b). To account for the tap basis, Eqs. 174 and 175 should be modified as

$$\begin{bmatrix}\mathbf{J}_1^{sca}\\ \mathbf{M}_1^{sca}\\ \mathbf{t}_{t1}\end{bmatrix}-S_{11}\cdot\begin{bmatrix}T_{12}^{oo}\cdot\begin{bmatrix}\mathbf{J}_2^{sca}\\ \mathbf{M}_2^{sca}\end{bmatrix}\\ -\mathcal{L}_{tt}^{-1}\cdot(\mathbf{t}_{t1}+\mathbf{t}_{t2})\end{bmatrix}=S_{11}\cdot\begin{bmatrix}\mathbf{J}_1^{inc}\\ \mathbf{M}_1^{inc}\\ -\mathcal{L}_{tt}\cdot\mathbf{E}_t^{i}\end{bmatrix} \tag{176}$$

$$\begin{bmatrix}\mathbf{J}_2^{sca}\\ \mathbf{M}_2^{sca}\\ \mathbf{t}_{t2}\end{bmatrix}-S_{22}\cdot\begin{bmatrix}T_{21}^{oo}\cdot\begin{bmatrix}\mathbf{J}_1^{sca}\\ \mathbf{M}_1^{sca}\end{bmatrix}\\ -\mathcal{L}_{tt}^{-1}\cdot(\mathbf{t}_{t1}+\mathbf{t}_{t2})\end{bmatrix}=S_{22}\cdot\begin{bmatrix}\mathbf{J}_2^{inc}\\ \mathbf{M}_2^{inc}\\ -\mathcal{L}_{tt}\cdot\mathbf{E}_t^{i}\end{bmatrix} \tag{177}$$

where $\mathcal{L}_{tt}^{-1}$ and $\mathbf{E}_t^i$ are the current solver operator and the incident field on the tap basis part, respectively. Moreover, $\mathbf{t}_{ti}=\mathcal{L}_{ti}\cdot\mathbf{J}_i$ represents the induced field by current on $i-th$ metallic object (i.e., $\mathbf{J}_i$) on the tap region. Expressions Eqs. 176 and 177 constitute six equations for six sets of unknowns [$\mathbf{J}_1^{sca}$, $\mathbf{M}_1^{sca}$, $\mathbf{t}_{t1}$, $\mathbf{J}_2^{sca}$, $\mathbf{M}_2^{sca}$, $\mathbf{t}_{t2}$]. The equations can be set up for the N object case. If no tap basis is involved with the object, the bottom rows of the equations disappear, and one reverts back to the ordinary EPA (Eqs. 174 and 175).

Antenna Modeling Using EPA in Literature

In the past two decades, the equivalence principle algorithm has been successfully used to solve multi-scale electromagnetic problems including antenna problems. Some of the reported works in the literature

DOI 10.1007/978-981-4560-75-7_6-2

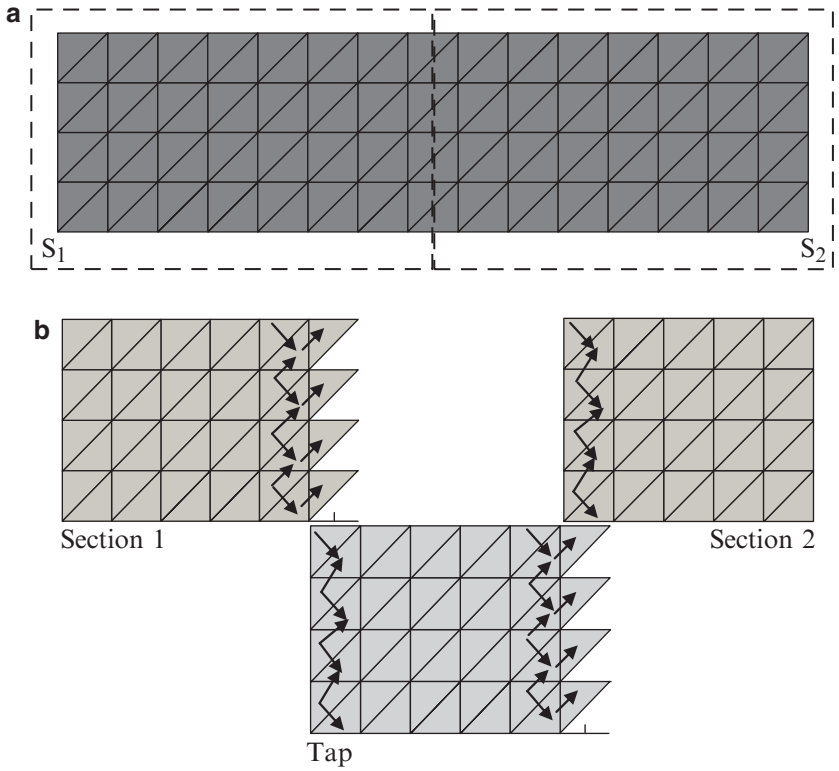


Fig. 18 Metallic strip. (**a**) Divided into two sections by two-equivalence surfaces, (**b**) introduction of tap basis to keep the current continuity

for analyzing antennas using EPA are summarized in Tables 1 and 2. The applications of EPA to analyze array antennas are listed in Table 1. Array of monopoles, U-shaped slot microstrips, and Yagi antennas have been successfully modeled and solved using EPA. In Table 2, applications of EPA for some other special cases are listed where EPA has been applied for analyzing an antenna mounted on the roof of a car, reconfigurable pixelled antennas, and a bow-tie antenna. Overall, the application of EPA helped significantly reduce the computational burden and accelerate the calculations.

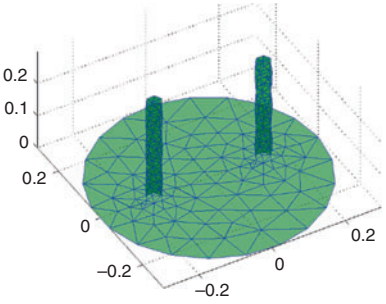
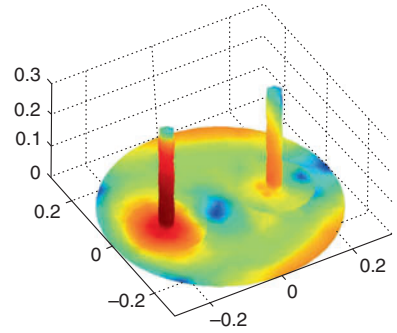
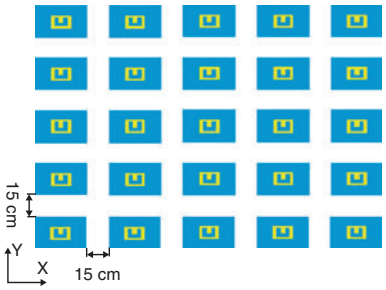
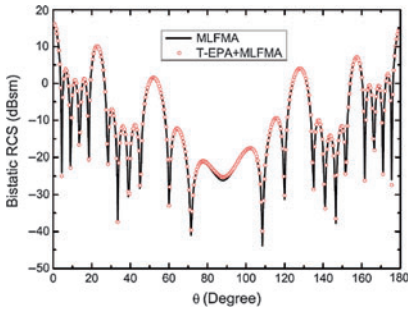
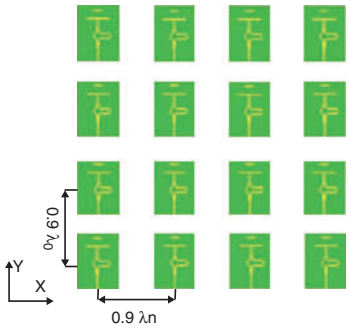
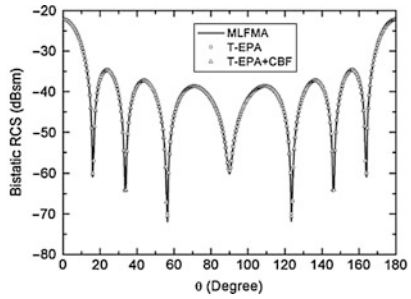
Recently, several ideas have been suggested to improve the accuracy and convergence rate of EPA. Most of these works try to address problems associated with the equivalence surfaces. Application of multilevel fast multipole algorithm to calculate the interactions between the equivalence surfaces significantly accelerated the whole EPA procedure. Also, in Shao et al. (2014), utilization of grid-robust higher-order vector basis (GRHOVB) instead of curvilinear RWG function for the equivalence surfaces has been proposed to improve the accuracy of EAP. In Shao et al. (2012), instead of solving for both electric and magnetic currents on the equivalence surfaces, the concept of generalized impedance boundary condition (GIBC) has been utilized to define a relationship between these currents and reduce the number of unknowns. Despite all of these endeavors, EPA is still an open area for further research. The goal of the future works is to improve the accuracy and the speed of convergence as well as modifying EPA for newly emerging applications such as solar cells, implantable devices, and nanoscale structures.

High-Frequency Techniques for Antennas

For antennas with a reflector or a platform that is large compared to the wavelength, high-frequency techniques can be used to simulate electromagnetic scattered fields very efficiently. They are important in

DOI 10.1007/978-981-4560-75-7_6-2

Table 1 Application of EPA to model array antennas (Li and Chew 2007; Shao et al. 2011, 2012, 2013)

Antenna type	Desired output
Two monopoles on ground plane	Current distribution
	
Array of microstrip antennas	Bistatic radar cross section (RCS)
	
Array of Yagi antennas	Bistatic RCS
	

three aspects (Samii and Israel 1980; Yaghjian 1984; Chou et al. 2005). First, high-frequency methods are much faster than full-wave methods on calculating electromagnetic scattered fields from electrically large antennas. Second, high-frequency methods could provide physical insights for the understanding of the radiation mechanisms of antennas. Typical high-frequency insights lead to wave physical critical points: the shadow and reflection boundaries, the caustics from surfaces of antennas, etc. Third, for complicated antenna structures, high-frequency methods have potential to be combined with the full-wave method well to solve radiation problems. Specifically, wave fields in electrically large parts exhibit the high-frequency ray physics phenomenon. While in electrically small and moderate parts, wave fields have circuit and wave physics behaviors, respectively. Thus, once the hybrid solvers in terms of high-frequency methods and full-wave solvers are developed, the efficiency of the electromagnetic wave simulation could be greatly enhanced.

DOI 10.1007/978-981-4560-75-7_6-2

Table 2 Application of EPA to solve antenna problems (Li and Chew 2008; Fallahpour et al. 2014; Shao et al. 2014)

Antenna type	Desired output
Antenna mounted on roof of a car	Radiation pattern
Reconfigurable pixelled antenna	RCS
Bow tie antenna on PEC plate	Radiation pattern

When the product of the wave frequency k and the diameter of the antenna a becomes tens to thousands, the electromagnetic waves assume the high-frequency localization phenomenon. In this sense, the high-frequency scattered fields have the high-frequency wave physics phenomenon and ray physics. Thus, high-frequency wave fields propagate like rays. With the aid of the high-frequency ray physics theory, various high-frequency methods were developed in the last century. For instance, for calculating the reflected and transmitted wave fields, the high-frequency methods contain the geometric optics (GO) method (Ling et al. 1989) and the physical optics (PO) method (Macdonald 1913). While for computing the diffracted wave fields in Fig. 19a, these methods correspond to the geometry theory of

DOI 10.1007/978-981-4560-75-7_6-2

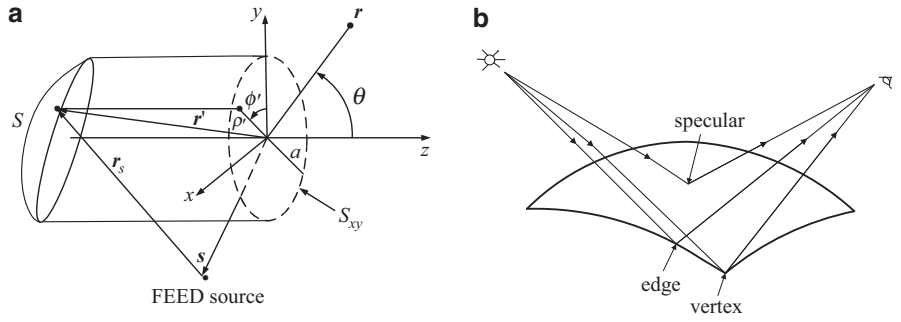


Fig. 19 (**a**) A shaped reflector antenna system. (**b**) High-frequency ray fields from the scatterer, including the reflected fields from the specular points and the diffracted fields from the edge and vertex

diffraction (GTD) (Keller 1962; Hansen 1981), uniform theory of diffraction (UTD) (Kouyoumjian and Pathak 1974), uniform asymptotic diffraction (UAT) (Lee and Deschamps 1976), physical theory of diffraction (PTD) (Ufimtsev 2005), and incremental length diffraction coefficients (ILDC) (Yaghjian et al. 1996). In the following, these high-frequency methods are introduced in an elegant way. Furthermore, fast solvers for the high-frequency scattered fields are addressed, including the high-frequency asymptotic method (HFA) (Conde et al. 2001; Carluccio et al. 2010) and the numerical steepest descent path method (NSDP) (Wu et al. 2012a, b, 2013a, b, c, 2014).

Radiation from Antennas

For the reflector antenna system with a location of the feed as shown in Fig. 19a, the electric scattered far fields could be expressed as

$$\mathbf{E}^{(s)}(\mathbf{r}) \approx \frac{ike^{ikr}}{4\pi r}\sqrt{\frac{\mu}{\epsilon}}\left(\bar{\mathbf{I}} - \hat{r}\hat{r}\right) \cdot \int_S dr' e^{-ik\hat{r}\cdot\mathbf{r}'}\mathbf{J}(\mathbf{r}') \tag{178}$$

Here, $\mathbf{J}(\mathbf{r}')$ is the induced electric current on the surface of the antenna, S, and $\mathbf{r}$ and $\mathbf{r}'$ denote the observation and integration source points. The projection aperture of S in the xy plane is the circular region, S_{xy}, with radius a as shown in Fig. 19a. Then, the radiation integrals $\mathbf{F}(\mathbf{r})$ could be obtained from Eq. 178, with the formulation

$$\mathbf{F}(\mathbf{r}) = \int_S d\mathbf{r}' e^{-ik\hat{r}\cdot\mathbf{r}'}\mathbf{J}(\mathbf{r}') = \int_0^a \int_0^{2\pi} e^{-ik\hat{r}\cdot\mathbf{r}'}\mathbf{J}(\mathbf{r}')\rho' T(\rho', \phi')d\rho' d\phi' \tag{179}$$

Here, $T(\rho', \phi')$ is the Jacobian transformation function for the antenna surface between (x, y, z) and (ρ, ϕ, z) coordinate systems.

The vector surface integrals $\mathbf{F}(\mathbf{r})$ in Eq. 179 are called the radiation integrals. Furthermore, the surface integral contains the phase information of the dyadic Green's function in the far-field region, the induced electric current, and the integration of the curvature from S.

High-Frequency Techniques

Geometric Optics Method

The ray path between any two points $\mathbf{p}_1$ and $\mathbf{p}_2$ in three-dimensional space follows Fermat's principle, that is,

DOI 10.1007/978-981-4560-75-7_6-2

$$\delta \int_{\mathbf{p}_1}^{\mathbf{p}_2} n(s)\, ds = 0 \tag{180}$$

Here, $n(s)$ is the refraction index of the medium and δ is the variation operator in variational calculus. Equation 180 expresses that the ray follows the path that minimizes the optical distances between $\mathbf{p}_1$ and $\mathbf{p}_2$. On invoking Fermat's principle in Eq. 180, the reflected rays could be determined from the incident field. This method is called the geometrical optics (GO) method. Furthermore, based on the ray tube theory and conservation of energy flux for light intensity, the reflected field $\mathbf{E}^{(r)}(\mathbf{r})$ at the distance s from the reflection point $\mathbf{p}_r$ of the scatterer could be expressed as

$$\mathbf{E}^{(r)}(\mathbf{r}) = \mathbf{E}^{(i)}(\mathbf{p}_r) \cdot \overline{\mathbf{R}} \sqrt{\frac{\rho_1(\mathbf{p}_r)\rho_2(\mathbf{p}_r)}{(\rho_1(\mathbf{p}_r)+s)\ (\rho_2(\mathbf{p}_r)+s)}} e^{iks} \tag{181}$$

Here, $\overline{\mathbf{R}}$ is the dyadic reflection coefficient, $\rho_1(\mathbf{p}_r)$ and $\rho_2(\mathbf{p}_r)$ are the principle radii of S at the point $\mathbf{p}_r$ of reflection, and s is the distance from $\mathbf{p}_r$ to $\mathbf{r}$ along the direction of reflected field. For the antenna with penetrable material, the transmitted wave fields also exist. The typical GO method for calculating the wave fields from the complicated cavity structure is the shooting and bouncing ray method (SBR) (Ling et al. 1989). Actually, by adopting the SBR method, multiple scattering wave fields are captured and calculated efficiently.

The GO method for calculating the scattered field has close connection with the scattered field in Eq. 178. On invoking the high-frequency asymptotic method for Eq. 178 in section "Basic Circuit Laws from Maxwell's Equations" in the following, one can obtain that the leading term in Eq. 178 corresponds to the GO ray field at the specular reflection point.

Geometric Theory of Diffraction

When the rays come to the wedges, the diffracted wave fields and the reflection and shadow boundaries come out. In this situation, for the case that the observation point lies in the transition and deep shadow regions, the GO method fails to capture the scattered wave fields very well. To remedy this aforementioned failure, Sommerfeld addressed the early high-frequency scattering problem from the perfect conducting half-plane (Sommerfeld 1896). The scattered field was expressed in terms of the Fresnel integral. By performing the asymptotic expansion of the Fresnel integral, the scattered far field could be easily evaluated. In the high-frequency diffracted wave current like the fringe current, the shadow and reflection boundaries were clearly demonstrated. Moreover, the regions for the observation point lying around the reflection and shadow boundaries are called transition regions.

The calculation of the diffracted ray field from the perfectly conducting curved wedge was developed by Keller (1962). This method was called the geometry theory of diffraction (GTD). The diffracted wave fields from the point $\mathbf{p}_d$ along the curved wedge has the similar formulation as (Eq. 181):

$$\mathbf{E}^{(d)}(\mathbf{r}) = \mathbf{E}^{(i)}(\mathbf{p}_d) \cdot \overline{\mathbf{D}} \sqrt{\frac{\rho_d(\mathbf{p}_d)}{s(\rho_d(\mathbf{p}_d)+s)}} e^{iks} \tag{182}$$

Here, $\overline{\mathbf{D}}$ is the dyadic diffraction coefficient in (Keller 1962), $\rho_d(\mathbf{p}_d)$ is the principle radius of the curved wedge at the point $\mathbf{p}_d$ of reflection, and s is the distance from $\mathbf{p}_d$ to $\mathbf{r}$ along the direction of the reflected field.

DOI 10.1007/978-981-4560-75-7_6-2

Furthermore, the diffraction coefficients $\overline{\mathbf{D}}$ could be obtained from the asymptotic expansion of the radiation integrals of the Sommerfeld half-plane diffraction problems. Thus, the fields from the GTD method contain both reflected and transmitted wave fields from GO method and the diffracted wave fields. Therefore, compared to the GO method, the GTD method provides a relatively mature way for calculating the high-frequency wave fields.

Again, on invoking the high-frequency asymptotic method in section "Basic Circuit Laws from Maxwell's Equations" for Eq. 178, the higher-order terms correspond to the diffracted ray fields. Moreover, these obtained diffracted ray fields agree well with the results generated from the GTD method. However, at the transition regions, these GTD diffraction coefficients become infinite; thus, the GTD diffracted wave fields become singular. Therefore, when the observation point lies around the transition regions, the GTD method breaks down.

Uniform Theory of Diffraction

To remedy the failure of the GTD method at the transition region, people developed the efficient uniform theory of diffraction (UTD) (Kouyoumjian and Pathak 1974) and uniform asymptotic diffraction (UAT) methods (Lee and Deschamps 1976). With the high-frequency wave physics aspect, on invoking the asymptotic technique, the radiation integrals for the diffracted wave fields from the wedge could be expressed as the superposition of contributions from high-frequency critical points. This superposition of critical-point contributions could also be regarded as the diffracted rays with the ray physics aspect. However, when the observation points lie in transition regions, these critical points will coalesce. In this situation, the asymptotic approximation of the radiation integrals fails to approximate these integrals well.

To account for the coupling of high-frequency critical points, the uniform asymptotic approximation of radiation integral could be adopted to remedy the breakdown of the GTD method at transitions. This method was called the UTD method, which provides the UTD diffraction coefficients in terms of the UTD transition function. The UTD transition function with the Fresnel function involved is used to compensate the GTD breakdown and makes the wave fields continuous around the transition regions. Meanwhile, with the observation point far away from the transition regions, the wave field results from the UTD and GTD methods could agree well with each other. Once Eq. 178 is derived in the uniform asymptotic way, higher accuracy for the diffracted wave fields will result in near-transition regions. The UTD method follows the asymptotic of the radiation integral. Thus, the breakdown problem of the GTD method around the transition regions could be well solved by the UTD method.

Physical Optics Method

The physical optics (PO) approximation serves as an efficient approach for analyzing the scattered electromagnetic fields. Early in 1913, Macdonald introduced the PO approximation concept by approximating the induced current on the surface of the consider scatterer (Macdonald 1913). The surface PO-induced current is approximated by doing the local tangent plane approximation, that is, the integration small surface patch along the electrically large object is assumed to be locally flat and smooth under the external high-frequency condition (Ufimtsev 2008). In this sense, for the incident high-frequency electromagnetic waves on the electrically large real-world objects, the PO approximation has revealed itself as an efficient way to calculate the scattered field (Lee and Mittra 1983). However, the PO current is not correct around the shadow boundary (Ufimtsev 2005), which leads to the fact that the physical optics method does not satisfy the reciprocity theorem.

With *ka* going from tens to thousands, by making use of the local tangential plane approximation for Eq. 179, the induced current in the lit region of considered scatterer could be approximated as

DOI 10.1007/978-981-4560-75-7_6-2

$$\mathbf{J}_{PO}(\mathbf{r}) = 2\hat{n}(\mathbf{r}) \times \mathbf{H}^{(i)}(\mathbf{r}) \tag{183}$$

Equation 183 is known as the PO current. Here, $\hat{n}(\mathbf{r})$ is the outward unit normal to the surface and $\mathbf{H}^{(i)}(\mathbf{r})$ is the incident field on the reflector antenna from the feed source. Here, $\mathbf{H}^{(i)}(\mathbf{r})$ in Eq. 183 is directly incident field generated by the feed to location $\mathbf{r}$ with the formulation

$$\mathbf{H}^{(i)}(\mathbf{r}) \simeq C\frac{e^{ikr_s}}{r_s}\mathbf{g}(\theta, \varphi) \tag{184}$$

with C as a known complex constant and $\mathbf{g}(\theta, \varphi)$ as the vector radiation function from the feed source, and it is assumed to be known.

After substituting $\mathbf{J}_{PO}$ in Eq. 183 into $\mathbf{J}$ in Eq. 178, one can obtain the PO scattered field from the reflector antenna. It could be represented by three radiation surface integrals. More specifically, the phase function in Eq. 178 contains the information from the dyadic Green's function, the shape function for the surface of the reflector antenna, and the incident magnetic field from the feed source. Owing to the highly oscillatory nature of the radiation integrand in Eq. 179, the direct numerical scheme for evaluating the integral makes the computational workload increase vastly when ka is large. The problem is more serious when there are many the observation points. Therefore, one needs to seek the efficient techniques to evaluate these PO integrals (Burkholder and Lee 2005).

Physical Theory of Diffraction

To remedy the limitation of the PO current, Ufimtsev introduced the edge wave current to correct the PO current, known as the physical theory of diffraction (PTD) method (Ufimtsev 2005). The PTD current, denoted as $\mathbf{J}_{PTD}$, takes the expression

$$\mathbf{J}_{PTD}(\mathbf{r}) = \mathbf{J}_{PO}(\mathbf{r}) + \mathbf{J}_{U}(\mathbf{r}) \tag{185}$$

Here, $\mathbf{J}_U(\mathbf{r})$ is the edge wave current which is used to correct the limited accuracy of $\mathbf{J}_{PO}(\mathbf{r})$ around the shadow boundary and in the shadow regions. Actually, $\mathbf{J}_U(\mathbf{r})$ could be obtained from the scattered wave solution of the canonical perfect conducting wedge scatterer (Ufimtsev 2005). After substituting $\mathbf{J}_{PTD}(\mathbf{r})$ in Eq. 185 into Eq. 178, one can obtain the scattered fields from the PTD method. The resultant scattered fields contain the PO scattered fields and the edge diffracted wave fields, which could also be regarded as the radiation integrals with highly oscillatory integrands. Compared to the GTD and the UTD method, the PTD method gives more accurate results than the GTD and the UTD methods via the integral formulations. Meanwhile, it is time consuming to adopt the PTD method by the direct adaptive numerical scheme for these radiation integrals.

Incremental Length Diffraction Coefficient Technique

The incremental length diffraction technique introduced by Shore and Yaghjian (Hansen and Shore 1998; Yaghjian et al. 1996) provides an efficient way to calculate the high-frequency diffracted wave fields from the perfect conducting scatterer. The nonuniform (NU) diffracted wave field could be obtained by the integration of the product of the Green's function with the NU current around the surface of the scatterer, S. Furthermore, the NU-diffracted wave field could be captured by simplifying the integration domain S into $C(\mathbf{r}_s, s)$ and dl (Hansen and Shore 1998), with dl as the shadow boundary of the scatterer. In the high-frequency regime, $C(\mathbf{r}_s, s)$ conforms locally to the shape of the scatterer around the shadow boundary point $\mathbf{r}_s$. This technique for the determination of the NU-diffracted wave field is called the incremental length diffraction coefficient (ILDC) technique (Hansen and Shore 1998; Yaghjian et al. 1996).

DOI 10.1007/978-981-4560-75-7_6-2

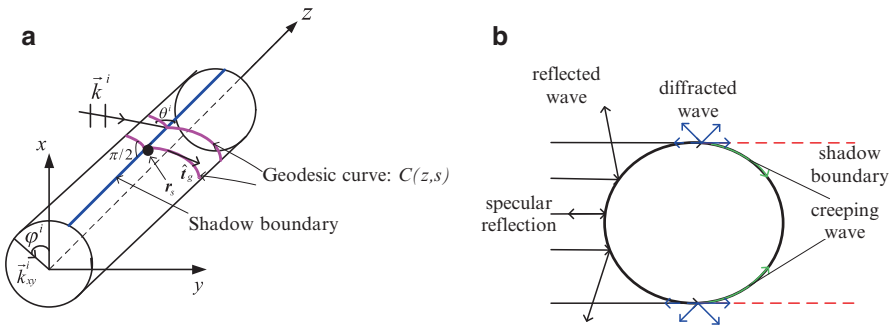


Fig. 20 (**a**) The obliquely incident plane wave impinges on the convex cylinder with finite length and the shadow boundary is shown by the *blue line* on the cylinder. The NU-Fock current in the shadow region follows the direction of the geodesic curve $C(z, s)$ by the *magenta color*. (**b**) PO current in the illuminated region and NU-Fock current around the shadow boundary and in the deep shadow region for the NU-diffracted and creeping waves

In particular, for the scatterer with convex shadow boundary, the NU-diffracted wave current is called the NU-Fock current (Fock 1946). The NU-diffracted wave field could be obtained from the integration of the product of the free-space Green's function and the NU-Fock current on S. Then, the resultant electric scattered field has the expression

$$\mathbf{E}^{(q)}(\mathbf{r}) = i\omega\mu\left(\bar{\mathbf{I}} + \frac{1}{k^2}\nabla\nabla\right) \cdot \int_S d\mathbf{r}' \frac{\mathbf{J}^{(q)}(\mathbf{r}')e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \tag{186}$$

The index q stands for the TE and TM cases (horizontal and vertical polarizations), and $\mathbf{J}^{(q)}(\mathbf{r}')$ is the NU-Fock current. In the far-field zone, Eq. 186 could be reformulated as

$$\mathbf{E}^{(q)}(\mathbf{r}) \approx \frac{ike^{ikr}}{4\pi r}\sqrt{\frac{\mu}{\epsilon}}\left(\bar{\mathbf{I}} - \hat{r}\hat{r}\right) \cdot \int_S d\mathbf{r}' e^{-ik\hat{r}\cdot\mathbf{r}'}\mathbf{J}^{(q)}(\mathbf{r}') \tag{187}$$

One can suppose that the integration surface S in Fig. 20 is given by $C(z, s) \times [z_0, z_1]$. Here, $C(z, s)$ is the integration curve and $[z_0, z_1]$ is the integration domain along the z direction. Then, with the strip of thickness dz as consideration, the far scattered field could be formulated as

$$d\mathbf{E}^{(q)}(\mathbf{r}) \approx dz\frac{ike^{ikr}}{4\pi r}\sqrt{\frac{\mu}{\epsilon}}\left(\bar{\mathbf{I}} - \hat{r}\hat{r}\right) \cdot \int_{C(z,s)} dsC'(z,s)e^{-ik\hat{r}\cdot C(z,s)}\mathbf{J}^{(q)}(C(z,s)) \tag{188}$$

In the illuminated region, the above $C(z, s)$ is normal to the shadow boundary. That is, $C(z, s)$ is the line along the $\mathbf{k}^i$ direction. In the shadow region, $C(z, s)$ is the geodesic curve which forms the constant angle with the z axis as θ^i. The above Eq. 188 is known as the ILDC formulation (Hansen and Shore 1998).

Fast Solvers for High-Frequency Scattered Fields

High-Frequency Asymptotic Method

To present the high-frequency asymptotic (HFA) method in an elegant way, one can study contour integrals of the type

$$I(k) = \int_C f(z)e^{ikg(z)}dz \tag{189}$$

DOI 10.1007/978-981-4560-75-7_6-2

Here, $\mathcal{C}$ is an integration domain in the real axis or a contour on the complex plane. Moreover, Eq. 189 could be treated as the simplification of the surface radiation integral Eq. 179 after some mathematical manipulations.

Due to the localization nature of the high-frequency wave field with the wave frequency k or the angular frequency ω approaching infinity, one can obtain an asymptotic approximation of the radiation integral in Eq. 189. This could be done by the "method of stationary phase" (or the "saddle point method"). That is, the major contribution of $I(k)$ comes from the neighborhood of the saddle point z_s satisfying $g'(z_s) = 0$. Here, one can assume that there is one saddle point z_s inside $\mathcal{C}$. Then, the leading term for the asymptotic expansion of Eq. 189 could be obtained by the expansion of $g(z)$ with the Taylor series and terminating at the second-order term, with the expression

$$I(k) \simeq f(z_s) e^{ikg(z_s)} \left(\frac{2\pi i}{kg''(z_s)} \right)^{1/2} \tag{190}$$

Furthermore, if $g''(z_s) = 0$, the above expansion Eq. 190 shall be modified with the Taylor series expansion of $g(z)$ to the third term. As is seen from Eq. 190, $I(k)$ is of $O\left(k^{-\frac{1}{2}}\right)$ and becomes increasingly accurate when k goes to infinity. The saddle point in two dimensions is called the "stationary phase point." Detailed discussion could be found in Conde et al. (2001) and Carluccio et al. (2010). By the same token, the integrand of the radiation integral in Eq. 189 is approximated by several leading terms. With the high-frequency wave physics aspect, the leading term in Eq. 190 corresponds to the GO ray fields at the specular reflection point, as shown in Fig. 19a. From Eq. 190, it is easy to see that the traditional HFA approach could provide the calculation of the radiation integral with frequency-independent workload.

Numerical Steepest Descent Path Method

Owing to the asymptotic process of the HFA method for radiation integrals in Eq. 179, the generated high-frequency scattered field results usually lose accuracy, in particular, when the parameter ka is not extremely large but lies in the high-frequency regime.

In the following, the application of the NSDP method for the PO scattered fields is considered. Since the PO scattered fields contain phase terms that have rapid variations in space, the resultant PO integrands are highly oscillatory in nature. To efficiently evaluate these surface PO integrals, the numerical steepest descent path (NSDP) method was proposed (Vico-Bondia et al. 2010; Zhang et al. 2014; (Wu et al. 2012a, b, 2013a, b, c). On invoking the NSDP method, the original PO real integration domain is deformed into the complex NSDPs on complex plane. In this manner, the PO integrands decay exponentially on these complex NSDPs, and they can be integrated with workload independent of frequency. In contrast to the HFA method, the only approximation in the NSDP method is the numerical integration of the exponential decay PO integrand. Hence, the proposed NSDP method improves the PO scattered field accuracy.

The definition of the NSDP follows that, for the given integration domain $[L_1, L_2]$ with the integrand $e^{ikg(x)}$, one can define the path functions $x = x_{L_m}(p)$, $m = 1, 2$, $p \in [0, \infty)$, satisfying three conditions (Chew 1990):

(a) $x_{L_m}(0) = L_m$, that is, the paths start at L_m.
(b) $\Re e(g(x_{L_m}(p))) = \Re e(g(x_{L_m}(0))) \equiv C$, where C is a constant.
(c) $\Im m(g(x_{L_m}(p))) = p$.

DOI 10.1007/978-981-4560-75-7_6-2

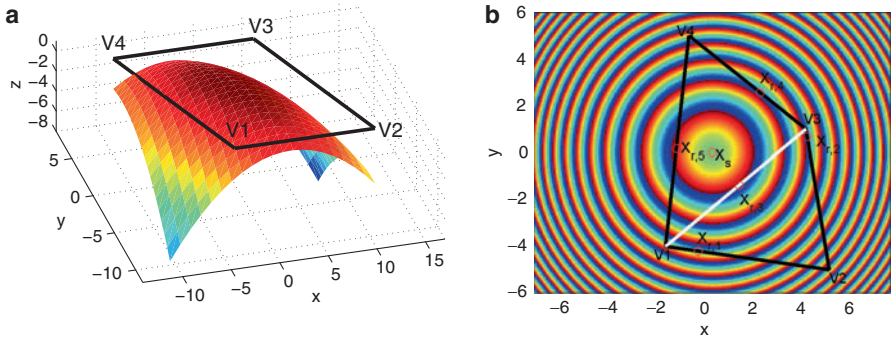


Fig. 21 (**a**) A parabolic PEC patch. (**b**) The projected $x-y$ quadrilateral domain $\mathbf{V}_1\mathbf{V}_2\mathbf{V}_3\mathbf{V}_4$ of the parabolic PEC patch

Then, after substituting $x = x_{L_m}(p)$ into the phase function $g(x)$, one sees that $e^{ikg(x_{L_m}(p))} = e^{ikg(L_m)-kp}$ decreases exponentially as $O(e^{-kp})$ when kp goes large. Hence, the path functions $x = x_{L_m}(p)$ are called the NSDPs. Through (a), (b), and (c), the NSDPs $x_{L_m}(p)$ for the end points L_m could be obtained (Chew 1990; Wu et al. 2012a, 2013a, b).

However, the phase behavior of $g(x)$ around the stationary phase point x_s is different from that around the end points. The corresponding NSDP shall also change. One notice that $g(x) - g(x_s)$ is a quadratic function around x_s after the Taylor expansion of the phase function at x_s (Chew 1990). Therefore, one can change the above condition (c) with $\Im m(g(x_{L_m}(p))) = p^2$ and define the corresponding NSDP, $x = x_0(p)$, $p \in (-\infty, \infty)$.

In the following, the application of the NSDP method for the PO scattered fields from the quadratic parabolic PEC patches as shown in Fig. 21 is considered. The projected xy quadrilateral domain is S_{xy}, denoted as $\mathbf{V}_1\mathbf{V}_2\mathbf{V}_3\mathbf{V}_4$. The edges are governed by $y = a_m x + b_m$, $m = 1, 2, 4, 5$. To capture the high-frequency critical-point phenomenon for PO scattered fields, one can consider that the PO integrand has the quadratic variation for the phase functions on the quadrilateral patches. After substituting $\mathbf{J}_{PO}$ into Eq. 178, the three PO surface integrals could be obtained. For simplicity, one can denote them as $\tilde{I}$. In the process for the reduction of the PO surface integral to several line integrals $I_2^{(a_m, b_m)}$, the complementary error function comes out in the integrand (Wu et al. 2012a, b). On invoking Stokes' phenomenon of the complementary error function, the phase functions on the edges are $g^{(m)}(x) = x^2 + (a_m x + b_m)^2$, and Stokes' lines are $y = -x - \frac{b_m}{a_m}$. Then, the NSDPs on the edges of the quadrilateral domain are (Chew 1990)

$$x_{L_m}(p) = \frac{\mathrm{sgn}(L'_m)}{\sqrt{1+a_m^2}}\sqrt{L_m'^2 + ip} + x_s, p \in [0, \infty) \tag{191}$$

$$x_{0,m}(p) = \frac{e^{\left(i\frac{\pi}{4}\right)}p}{\sqrt{1+a_m^2}} + x_s, p \in (-\infty, \infty) \tag{192}$$

where $x_s = -\frac{a_m b_m}{1+a_m^2}$ is the stationary phase point of phase function $g^{(m)}(x)$ defined on edges and $L'_m = \sqrt{1+a_m^2}(L_m - x_s)$. These NSDPs are demonstrated in Fig. 22. After changing the original PO integration path to the NSDPs and on invoking the Cauchy's integral theorem, the PO integral $\tilde{I}$ can be represented in terms of the NSDPs as

DOI 10.1007/978-981-4560-75-7_6-2

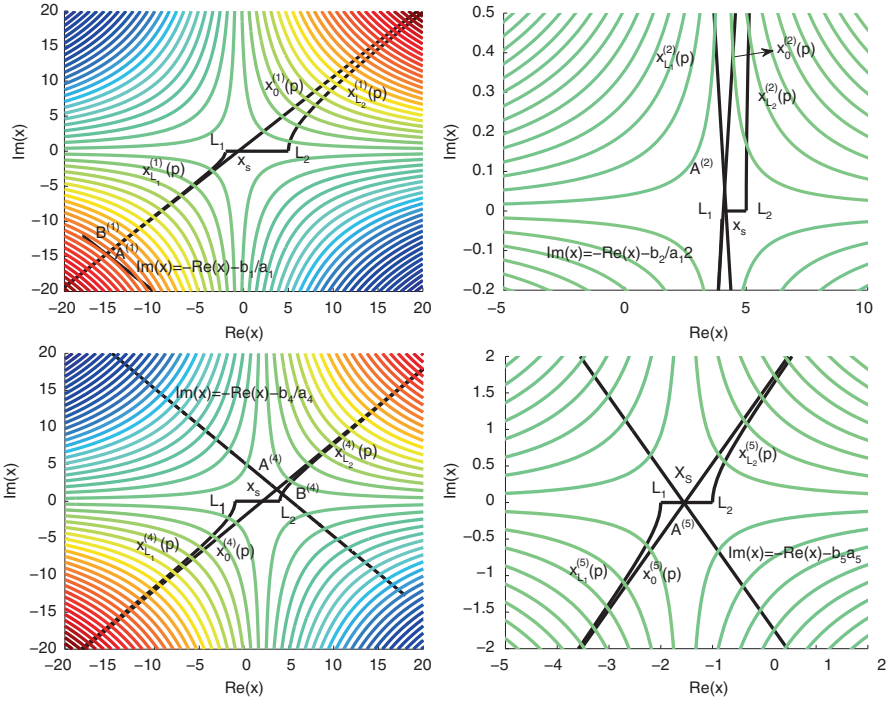


Fig. 22 The four sub-figures correspond to the NSDPs for the integrand of $I_2^{(a_m,b_m)}$, m = 1, 2, 4, 5, defined on the four edges of the quadrilateral domain $\mathbf{V}_1\mathbf{V}_2\mathbf{V}_3\mathbf{V}_4$ in Fig. 21

$$\tilde{I} = -I_2^{(a_1,b_1)} + I_2^{(a_2,b_2)} + I_2^{(a_4,b_4)} + I_2^{(a_5,b_5)} \tag{193}$$

where

$$\begin{aligned} I_2^{(a_1,b_1)} &= I_{L_1}^{(a_1,b_1)} - I_{L_2}^{(a_1,b_1)} + I_{x_s}^{(a_1,b_1)} \\ &+ K((\mathbf{V}_2(1),0)) - K((\mathbf{V}_1(1),0)) - K\left(\mathbf{B}^{(1)}\right) + K\left(\mathbf{A}^{(1)}\right) \end{aligned} \tag{194}$$

$$I_2^{(a_2,b_2)} = I_{L_1}^{(a_2,b_2)} + I_{x_s}^{(a_2,b_2)} - I_{L_2}^{(a_2,b_2)} + K((\mathbf{V}_2(1),(0)) - K\left(\mathbf{A}^{(2)}\right) \tag{195}$$

$$I_2^{(a_4,b_4)} = I_{L_1}^{(a_4,b_4)} + I_{x_s}^{(a_4,b_4)} - I_{L_2}^{(a_4,b_4)} + K\left(\mathbf{B}^{(4)}\right) - K\left(\mathbf{A}^{(4)}\right) \tag{196}$$

$$I_2^{(a_5,b_5)} = I_{L_1}^{(a_5,b_5)} + I_{x_s}^{(a_5,b_5)} - I_{L_2}^{(a_5,b_5)} + K\left(\mathbf{A}^{(5)}\right) - K((\mathbf{V}_1(1),0)) \tag{197}$$

Originally, $I_2^{(a_m,b_m)}$, m = 1, 2, 4, 5, are the highly oscillatory PO line integrals $I_2^{(a,b)}$ defined on the four edges $\overrightarrow{\mathbf{V}_1\mathbf{V}_2}$, $\overrightarrow{\mathbf{V}_2\mathbf{V}_3}$, $\overrightarrow{\mathbf{V}_3\mathbf{V}_4}$, and $\overrightarrow{\mathbf{V}_1\mathbf{V}_4}$ in Fig. 21, and in Eqs. 195, 196, and 197, $K(x)$ comes from Stokes' phenomenon of erfc(z)'s different asymptotic behaviors on different domains, which has a closed-form formula. At this point, PO integral is expressed in terms of NSDPs with an exponentially decay integrand, which could be evaluated in a frequency-independent workload.

Compared to the HFA method, one can see the advantage of the NSDP method after studying the high-frequency critical-point contributions from both methods. With the ray physics aspect, the high-frequency critical-point contributions from the stationary phase point, resonance point, and vertex point correspond

DOI 10.1007/978-981-4560-75-7_6-2

to the ray fields from the specular reflection point, the boundary specular reflection point, and the vertex point, respectively. The leading terms of the HFA approach for the high-frequency critical-point contributions including the stationary phase point, resonance point, and vertex point (Carluccio et al. 2010) are of $O\left(k^{-1}\right), O\left(k^{-\frac{3}{2}}\right), O\left(k^{-2}\right)$, respectively. On the other hand, these critical-point contributions could also be rigorously expressed via the PO integrand defined on the NSDPs. Since the process for evaluating the PO integral by the NSDP method uses less approximation, these PO integral results gain high accuracy from those produced by the HFA method. Meanwhile, the frequency-independent computation effort is achieved (Wu et al. 2012a, b, 2013a, b, c).

Conclusions

In this chapter, an overview of different computational electromagnetics (CEM) methods is given that are used to solve antenna problems. As computers become more powerful, these methods are more prevalent for antenna designs. Many software tools have been developed to enable better antenna designs. Hence, it is important for antenna designers and engineers to understand the working principles of different CEM methods or future methods that will be incorporated into these software tools. With better understanding, better antennas can be designed.

As antenna engineering often involves circuit design, it is also important that antenna designers understand how electromagnetic theory morphs into circuit theory when the feature sizes are much smaller than the wavelength. Antenna structures are designed to be of the order of wavelength or larger in order to receive the signals that are transmitted as waves. However, as these signals are received by the antennas, they invariably have to be routed through different systems using a combination of circuit and waveguide theories. It is instrumental that antenna engineers understand both the electromagnetic and circuit theory well.

As the signal is received, the signal is routed through a system where wave physics becomes circuit physics. Hence, it is also important that computation be done correctly in the "twilight" zone between the circuit theory and electromagnetic full wave theory.

Due to the advanced made in nanofabrication technology, it is now possible to make nanoantennas that are subwavelength even in the optical regime. These nanoantennas find applications in optical links in computer systems as front-end communication bus. They can provide the high data rate needed. They also can be used to enhance the optical emissions given by atoms and molecules or couple light energy to materials to enhance the generation of photoelectricity.

Many antenna designers use software to gain physical insight into the working of the antennas. It will be increasingly important that this software provides output parameters that offer a deeper insight into the workings of an antenna. This can be seen in the increasing use of characteristic mode analysis for antenna design. The future can be the use of natural mode analysis for better understanding and design of antennas.

Also, as systems become highly complex, it is prudent to decompose a larger system as a sum of smaller systems for ease of finding the solution. So the domain decomposition method comes in useful. The equivalence principle algorithm (EPA) is useful to achieve this decomposition, even though much research work is still needed.

As a final note, high-frequency scattering has always been important in design of reflector antennas. The improvement of the accuracy and speed of existing high-frequency techniques is an area that warrants further research.

General research directions are new numerical analysis tools for better antenna design and engineering. Fast CEM algorithms, robust and valid over a broad frequency range, are future goals. These future tools should handle multi-scale structures efficiently and robustly, with the least use of memory and computer

DOI 10.1007/978-981-4560-75-7_6-2

resource. Moreover, coupling to multiphysics simulation will become increasingly important as antennas are integrated into complex environments. New metrics for antenna design and performance need to be formulated as the complexity becomes higher.

References

Adams RJ (2004) Physical and analytical properties of a stabilized electric field integral equation. IEEE Trans Antennas Propag 52(2):362–372

Alu A, Engheta N (2008) Tuning the scattering response of optical nanoantennas with nanocircuit loads. Nat Photonics 2:307–310

Andreani LC, Panzarini G, Gerard JM (1999) Strong-coupling regime for quantum boxes in pillar microcavities: theory. Phys Rev B 60:13276–13279

Andriulli FP (2012) Loop-star and loop-tree decompositions: analysis and efficient algorithms. IEEE Trans Antennas Propag 60(5):2347–2356

Andriulli FP, Cools K, Bağci H, Olyslager F, Buffa A, Christiansen S, Michielssen E (2008) A multiplicative Calderón preconditioner for the electric field integral equation. IEEE Trans Antennas Propag 56(8):2398–2412

Andriulli FP, Bağci H, Vipiana F, Vecchi G, Michielssen E (2009) Analysis and regularization of the TD-EFIE low-frequency breakdown. IEEE Trans Antennas Propag 57(7):2034–2046

Atwater HA, Polman A (2010) Plasmonics for improved photovoltaic devices. Nat Mater 9:205–213

Axelsson O, Barker VA (1984) Finite element solution of boundary value problems: theory and computation. Academic Press, New York

Bağci H, Andriulli FP, Cools K, Olyslager F, Michielssen E (2009) A Calderón multiplicative preconditioner for the combined field integral equation. IEEE Trans Antennas Propag 57(10):3387–3392

Balanis CA (2012a) Advanced engineering electromagnetics, 2nd edn. Wiley, New York

Balanis CA (2012b) Antenna theory: analysis and design. Wiley, New York

Baum C, Rothwell E, Chen K-M, Nyquist D (1991) The singularity expansion method and its application to target identification. Proc IEEE 79:1481–1492

Benzi M, Golub GH, Liesen J (2005) Numerical solution of saddle point problems. Acta Numer 14:1–137

Berenger JP (1994) A perfectly matched layer for the absorption of electromagnetic waves. J Comput Phys 114:185–200

Berthelot J, Bouhelier A, Huang CJ et al (2009) Tuning of an optical dimer nanoantenna by electrically controlling its load impedance. Nano Lett 9:3914–3921

Bharadwaj P, Deutsch B, Novotny L (2009) Optical antennas. Adv Opt Photon 1:438–483

Biagioni P, Huang JS, Hecht B (2012) Nanoantennas for visible and infrared radiation. Rep Prog Phys 75:024402

Bienstman P, Baets R (2001) Optical modelling of photonic crystals and VCSELs using eigenmode expansion and perfectly matched layers. Opt Quant Electron 33:327–341

Bliokh KY, Bekshaev AY, Nori F (2013) Dual electromagnetism: helicity, spin, momentum and angular momentum. New J Phys 15:033026

Bonakdar A, Kohoutek J, Dey D, Mohseni H (2012) Optomechanical nanoantenna. Opt Lett 37:3258–3260

Borup DT, Gandhi OP (1984) Fast-Fourier transform method for calculation of SAR distributions in finely discretized inhomogeneous models of biological bodies. IEEE Trans Microw Theory Tech 32(4):355–360

DOI 10.1007/978-981-4560-75-7_6-2

Buffa A, Christiansen S (2007) A dual finite element complex on the barycentric refinement. Math Comput 76(260):1743–1769

Burke GJ, Poggio AJ (1981) "Numerical Electromagnetic Code-2," Ver. 5.7.5, Arie Voors

Burkholder RJ, Lee TH (2005) Adaptive sampling for fast physical optics numerical integration. IEEE Trans Antennas Propag 53(5):1843–1845

Cabedo-Fabres M, Antonino-Daviu E, Valero-Nogueira A, Bataller MF (2007) The theory of characteristic modes revisited: a contribution to the design of antennas for modern applications. IEEE Antennas Propag Mag 49:52–68

Carluccio G, Albani M, Pathak PH (2010) Uniform asymptotic evaluation of surface integrals with polygonal integration domains in terms of UTD transition functions. IEEE Trans Antennas Propag 58(4):1155–1163

Catedra MF, Gago E, Nuno L (1989) A numerical scheme to obtain the RCS of three-dimensional bodies of resonant size using the conjugate gradient method and the fast Fourier transform. IEEE Trans Antennas Propag 37:528–537

Celik M, Pileggi L, Odabasioglu A (2002) IC interconnect analysis. Kluwer Academic, Durdrecht

Chen PY, Alu A (2010) Optical nanoantenna arrays loaded with nonlinear materials. Phys Rev B 82:235405

Chen CH, Lien C-D (1980) The variational principle for non-self-adjoint electromagnetic problems. IEEE Trans Microw Theory Tech 28:878–886

Chen QL, Wilton DR (1990) Electromagnetic scattering by three-dimensional arbitrary complex material/conducting bodies. IEEE Int Symp Antennas Propag 2:590–593

Chen LZ, Sha WEI, Choy WCH (2012a) Light harvesting improvement of organic solar cells with self-enhanced active layer designs. Opt Express 20:8175–8185

Chen YP, Sha WEI, Choy WCH, Jiang LJ, Chew WC (2012b) Study on spontaneous emission in complex multilayered plasmonic system via surface integral equation approach with layered medium Green's function. Opt Express 20:20210–20221

Chew WC (1990) Waves and fields in inhomogeneous media. Van Nostrand Reinhold, New York, Reprinted by Piscataway, NJ: IEEE Press, 1995

Chew WC, Kong JA (1981) Asymptotic formula for the resonant frequencies of a circular microstrip antenna. J Appl Phys 52(8):5365–5369

Chew WC, Lu CC (1993) The use of Huygens equivalence principle for solving the volume integral equation of scattering. IEEE Trans Antennas Propag 41(7):897904

Chew WC, Weedon WH (1994) A 3-D perfectly matched medium from modified Maxwell's equations with stretched coordinates. Microw Opt Technol Lett 7(13):599–604

Chew WC, Jin JM, Michielssen E, Song J (2000) Fast and efficient algorithms in computational electromagnetics. Artech House, London

Chew WC, Jin JM, Michielssen E, Song JM (eds) (2001) Fast and efficient algorithms in computational electromagnetics. Artech House, Boston

Chew WC, Tong MS, Hu B (2008) Integral equation methods for electromagnetic and elastic waves. Morgan Claypool, San Francisco

Chew WC, Tong MS, Hu B (2009) Integral equations methods for electromagnetic and elastic waves. Morgan & Claypool, San Rafael

Chou HT, Ho HK, Chung TY (2005) A discrete-time uniform geometrical theory of diffraction for the fast transient analysis of scattering from curved wedges. IEEE Trans Antennas Propag 53(11):3633–3643

Clark B (1985) Well logging apparatus and method using transverse magnetic mode. U.S. Patent 4,553,097

DOI 10.1007/978-981-4560-75-7_6-2

Coifman R, Rokhlin V, Wandzura S (1993) The fast multipole method for the wave equation: a pedestrian prescription. IEEE Antennas Propag Mag 35(3):7–12

Collin RE (1991) Field theory of guided waves. IEEE Press, Piscataway

Collins RE (1966) Foundations for microwave engineering. McGraw-Hill, New York

Conde OM, Pérez J, Cátedra MF (2001) Stationary phase method application for the analysis of radiation of complex 3-D conducting structures. IEEE Trans Antennas Propag 49(5):724–731

Cools K, Andriulli FP, Olyslager F, Michielssen E (2009) Improving the MFIE's accuracy by using a mixed discretization. In: IEEE antennas and propagation society international symposium, pp 1–4, Charleston, SC

Corvi M, Schaich WL (1986) Hydrodynamic-model calculation of second-harmonic generation at a metal surface. Phys Rev B 33:3688

Cui YX, Xu J, Fung KH et al (2011) A thin film broadband absorber based on multi-sized nanoantennas. Appl Phys Lett 99:253101

Dai QI, Chew WC, Lo YH, Liu YG, Jiang LJ (2012) Generalized modal expansion of electromagnetic field in 2-D bounded and unbounded media. IEEE Antennas Wirel Propag Lett 11:1052–1055

Dai QI, Lo YH, Chew WC, Jiang LJ (2013) An efficiently preconditioned eigenanalysis of inhomogeneously loaded rectangular cavities. IEEE Antennas Wirel Propag Lett 12:58–61

Dai QI, Lo YH, Chew WC, Liu YG, Jiang LJ (2014) Generalized modal expansion and reduced modal representation of 3-D electromagnetic fields. IEEE Trans Antennas Propag 62:783–793

Davis TA, Duff IS (1997) An unsymmetric-pattern multifrontal method for sparse LU factorization. SIAM J Matrix Anal Appl 18:140–158

Devilez A, Stout B, Bonod N (2010) Compact metallo-dielectric optical antenna for ultra directional and enhanced radiative emission. ACS Nano 4:3390–3396

Dong YD, Itoh T (2010) Miniaturized substrate integrated waveguide slot antennas based on negative order resonance. IEEE Trans Antennas Propag 58:3856–3864

Fallahpour M, Ma Z, Li MK, Chew WC (2014) Using equivalence principle algorithm to analyze and design reconfigurable pixelled antennas. In: Allerton antenna applications symposium, IL, USA

Fock VA (1946) The distributions of currents induced by a plane wave on the surface of a conductor. J Phys 10:130–136

Garbacz RJ, Turpin RH (1971) A generalized expansion for radiated and scattered fields. IEEE Trans Antennas Propag 19:348–358

Gerard D, Devilez A, Aouani H et al (2009) Efficient excitation and collection of single-molecule fluorescence close to a dielectric microsphere. J Opt Soc Am B 26:1473–1478

Giannini V, Fernandez-Dominguez AI, Heck SC, Maier SA (2011) Plasmonic nanoantennas: fundamentals and their use in controlling the radiative properties of nanoemitters. Chem Rev 111:3888–3912

Graglia RD (1993) On the numerical integration of the linear shape functions times the 3D Green's function or its gradient on a plane triangle. IEEE Trans Antennas Propag 41(10):1448–1456

Greengard L, Rokhlin V (1987) A fast algorithm for particle simulations. J Comput Phys 73:325–348

Hansen RC (ed) (1981) Geometric theory of diffraction. IEEE Press, Piscataway

Hansen TB, Shore RA (1998) Incremental length diffraction coefficients for the shadow boundary of a convex cylinder. IEEE Trans Antennas Propag 46(10):1458–1466

Harrington RF, Mautz JR (1971) Theory of characteristic modes for conducting bodies. IEEE Trans Antennas Propag 19:622–628

He SQ, Sha WEI, Jiang LJ, Choy WCH, Chew WC, Nie ZP (2012) Finite-element-based generalized impedance boundary condition for modeling plasmonic nanostructures. IEEE Trans Nanotechnol 11:336–345

Heaviside O (1888) On electromagnetic waves, especially in relation to the vorticity of the impressed forces, and the forced vibration of electromagnetic systems. Philos Mag 25:130–156. Also see Nahin PJ (1990) Oliver Heaviside. Scientific American 122–129

Ho C, Ruehli A, Brennan P (1975) The modified nodal approach to network analysis. IEEE Trans Circ Syst CAS-32:504–509

Hsiao GC, Kleinman RE (1997) Mathematical foundations for error estimation in numerical solutions of integral equations in electromagnetics. IEEE Trans Antennas Propag 45(3):316–328

Jin JM (2002) The finite element method in electromagnetics, 2nd edn. Wiley-IEEE Press, New York

Jin JM (2014) The finite element method in electromagnetics, 3rd edn. Wiley-IEEE Press, Hoboken

Joannopoulos JD, Johnson SG, Winn JN, Meade RD (2008) Photonic crystals: molding the flow of light. Princeton University Press, Princeton

Johnson SG, Joannopoulos JD (2001) Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis. Opt Express 8:173–190

Kalkbrenner T, Hakanson U, Schadle A et al (2005) Optical microscopy via spectral modifications of a nanoantenna. Phys Rev Lett 95:200801

Kauranen M, Zayats AV (2012) Nonlinear plasmonics. Nat Photonics 6:737–748

Keller JB (1962) Geometrical theory of diffraction. JOSA 52(2):116–130

Kelley DF, Luebbers RJ (1996) Piecewise linear recursive convolution for dispersive media using FDTD. IEEE Trans Antennas Propag 44:792–797

King RWP (1956) The theory of linear antennas. Harvard University Press, Cambridge, MA

Kinkhabwala A, Yu ZF, Fan SH et al (2009) Large single-molecule fluorescence enhancements produced by a bowtie nanoantenna. Nat Photonics 3:654–657

Kouyoumjian RG, Pathak PH (1974) A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface. Proc IEEE 62(11):1448–1461

Krasnok AE, Maksymov IS, Denisyuk AI et al (2013) Optical nanoantennas. Phys Usp 56:539–564

Kwon DH, Werner PL, Werner DH (2008) Optical planar chiral metamaterial designs for strong circular dichroism and polarization rotation. Opt Express 16:11802–11807

Lee R, Cangellaris AC (1992) A study of discretization error in the finite element approximation of wave solution. IEEE Trans Antennas Propag 40(5):542–549

Lee SW, Deschamps GA (1976) A uniform asymptotic theory of electromagnetic diffraction by a curved wedge. IEEE Trans Antennas Propag 24(1):25–34

Lee SW, Mittra R (1983) Fourier transform of a polygonal shape function and its application in electromagnetics. IEEE Trans Antennas Propag 31(1):99–103

Lee JF, Lee R, Burkholder RJ (2003) Loop star basis functions and a robust preconditioner for EFIE scattering problems. Trans Antennas Propag 51(8):1855–1863

Lehoucq R, Sorensen D (1996) Deflation techniques for an implicitly restarted Arnoldi method. SIAM J Matrix Anal Appl 17:789–821

Li M-K, Chew WC (2007) Wave-field interaction with complex structures using equivalence principle algorithm. IEEE Trans Antennas Propag 55(1):130–138

Li M-K, Chew WC (2008) Multiscale simulation of complex structures using equivalence principle algorithm with high-order field point sampling scheme. IEEE Trans Antennas Propag 56(8):2389–2397

Ling H, Chou RC, Lee SW (1989) Shooting and bouncing rays: calculating the RCS of an arbitrarily shaped cavity. IEEE Trans Antennas Propag 37(2):194–205

Liu YG, Li Y, Sha WEI (2011) Directional far-field response of a spherical nanoantenna. Opt Lett 36:2146–2148

DOI 10.1007/978-981-4560-75-7_6-2

Liu YG, Choy WCH, Sha WEI, Chew WC (2012) Unidirectional and wavelength-selective photonic sphere-array nanoantennas. Opt Lett 37:2112–2114

Liu QS, Sun S, Chew WC (2014) Convergence of low-frequency EFIE-based systems with weighted right-hand-side effect. IEEE Trans Antennas Propag 62(10):5108–5116

Lo YT, Solomon D, Richards W (1979) Theory and experiment on microstrip antennas. Antennas Propag IEEE Trans 27(2):137–145

Lo YH, Jiang LJ, Chew WC (2013) Finite-width feed and load models. IEEE Trans Antennas Propag 61(1):281–289

Lu CC, Chew WC (1995) The use of Huygens equivalence principle for solving 3-D volume integral equation of scattering. IEEE Trans Antennas Propag 43(5):500–507

Lu Z-Q, An X, Hong W (2008) A fast domain decomposition method for solving three dimensional large-scale electromagnetic problems. IEEE Trans Antennas Propag 56(8):22002210

Luebbers R, Hunsberger FP, Kunz KS et al (1990) A frequency-dependent finite-difference time-domain formulation for dispersive materials. IEEE Trans Electromagn Compat 32:222–227

Macdonald HM (1913) The effect produced by an obstacle on a train of electric waves. Philos Trans R Soc Lond A Math Phys Sci 212:299–337

Maier SA (2007) Plasmonics: fundamentals and applications. Springer, New York.

Makitalo J, Suuriniemi S, Kauranen M (2011) Boundary element method for surface nonlinear optics of nanoparticles. Opt Express 19:23386–23399

Maksymov IS, Davoyan AR, Kivshar YS (2011) Enhanced emission and light control with tapered plasmonic nanoantennas. Appl Phys Lett 99:083304

Manges JB, Cendes ZJ (1995) A generalized tree-cotree gauge for magnetic field computation. IEEE Trans Magn 31(3):1342–1347

Martin-Moreno L, Garcia-Vidal FJ, Lezec HJ, Degiron A, Ebbesen TW (2003) Theory of highly directional emission from a single subwavelength aperture surrounded by surface corrugations. Phys Rev Lett 90:167401

Maxwell JC (1865) A dynamical theory of the electromagnetic field. Phil Trans R Soc Lond 155:459–512

Merlein J, Kahl M, Zuschlag A et al (2008) Nanomechanical control of an optical antenna. Nat Photonics 2:230–233

Michiels B, Bogaert I, Fostier J, Zutter DD (2014) A well-scaling parallel algorithm for the computation of the translation operator in the MLFMA. IEEE Trans Antennas Propag 62(5):2679–2687

Mishchenko MI, Travis LD, Mackowski DW (2010) T-Matrix computations of light scattering by nonspherical particles: a review. J Quant Spectrosc Radiat Transf 111:1704–1744

Moharam MG, Grann EB, Pommet DA, Gaylord TK (1995) Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings. J Opt Soc Am A 12:1068–1076

Muhlschlegel P, Eisler HJ, Martin OJF et al (2005) Resonant optical antennas. Science 308:1607–1609

Nagel LW, Pederson DO (1973) SPICE (Simulation Program with Integrated Circuit Emphasis). Memorandum No. ERL-M382, University of California, Berkeley

Novotny L, Hecht B (2006) Principles of nano-optics. Cambridge University Press, New York

Novotny L, Hulst NV (2011) Antennas for light. Nat Photonics 5:83–90

Pellegrini G, Mattei G, Mazzoldi P (2009) Light extraction with dielectric nanoantenna arrays. ACS Nano 3:2715–2721

Peng Z, Wang XC, Lee JF (2011) Integral equation based domain decomposition method for solving electromagnetic wave scattering from non-penetrable objects. IEEE Trans Antennas Propag 59(9):3328–3338

Pillegi L, Rohrer R, Visweswariah C (1995) Electronic circuits and system simulation methods. McGraw-Hill, New York

Pozar DM (2011) Microwave engineering, 4th edn. Wiley, New York

Pu Y, Grange R, Hsieh CL, Psaltis D (2010) Nonlinear optical properties of core-shell nanocavities for enhanced second-harmonic generation. Phys Rev Lett 104:207402

Qian ZG, Chew WC (2008) An augmented electric field integral equation for low frequency electromagnetics analysis. In: IEEE international symposium on antennas and propagation, San Diego

Qian Z-G, Chew WC (2009) Fast full-wave surface integral equation solver for multiscale structure modeling. IEEE Trans Antennas Propag 57(11):3594–3601

Qian ZG, Chew WC (2010) Enhanced A-EFIE with perturbation method. IEEE Trans Antennas Propag 58(10):3256–3264

Qiao PF, Sha WEI, Choy WCH, Chew WC (2011) Systematic study of spontaneous emission in a two-dimensional arbitrary inhomogeneous environment. Phys Rev A 83:043824

Rahmat-Samii Y, Galindo-Israel V (1980) Shaped reflector antenna analysis using the Jacobi–Bessei series. IEEE Trans Antennas Propag 28(4):427–432

Rao SM, Wilton GR, Glisson AW (1982) Electromagnetic scattering by surfaces of arbitrary shape. IEEE Trans Antennas Propag 30(3):409–418

Ren XG, Sha WEI, Choy WCH (2013) Tuning optical responses of metallic dipole nanoantenna using graphene. Opt Express 21:31824–31829

Rokhlin V (1990) Rapid solution of integral equations of scattering theory in two dimensions. J Comput Phys 86(2):414–439

Rong A, Cangellaris AC (2001) Generalized PEEC models for three-dimensional interconnect structures and integrated passives of arbitrary shapes. Digest Electr Perf Electron Packag, Boston, MA 10:225–228

Ruehli AE (1972) Inductance calculations in a complex integrated circuit environment. IBM J Res Dev 16(5):470–481

Ruehli AE (1974) Equivalent circuit models for three dimensional multiconductor systems. IEEE Trans Microw Theory Tech MTT-22(3):216–221

Ruehli AE, Brennan PA (1973) Efficient capacitance calculations for three-dimensional multiconductor systems. IEEE Trans Microw Theory Tech 21(2):76–82

Ruehli AE, Antonini G, Esch J, Ekman J, Mayo A, Orlandi A (2003) Non-orthogonal PEEC formulation for time and frequency domain EM and circuit modeling. IEEE Trans Electromagn Compat 45(2):167–176

Ruehli A, Antonio G, Jiang LJ (2013) Skin-effect loss models for time- and frequency-domain PEEC solver. Proc IEEE 101(2):451–472

Sacks ZS, Kingsland DM, Lee R, Lee JF (1995) A perfectly matched anisotropic absorber for use as an absorbing boundary condition. IEEE Trans Antennas Propag 43(12):1460–1463

Sha WEI, Meng LL, Choy WCH, Chew WC (2014) Observing abnormally large group velocity at the plasmonic band edge via a universal eigenvalue analysis. Opt Lett 39:158–161

Shao H, Hu J, Nie ZP, Han G, He S (2011) Hybrid tangential equivalence principle algorithm with MLFMA for analysis of array structures. Prog Electromagn Res 113:127–141

Shao H, Hu J, Guo H, Ye F, Lu W, Nie Z (2012) Fast simulation of array structures using T-EPA with hierarchical LU decomposition. IEEE Antennas Wirel Propag Lett 11:1556–1559

Shao H, Hu J, Chew WC (2013) Solving array structures using single-source equivalence principle algorithm. In: International Conference on Electromagnetics in Advanced Applications (ICEAA), Torino, pp 774–777

DOI 10.1007/978-981-4560-75-7_6-2

Shao H, Hu J, Nie Z, Jiang L (2014) Simulation of multiscale structures using equivalence principle algorithm with grid-robust higher order vector basis. J Electromagn Waves Appl 28(11):1333–1346. doi:10.1080/09205071.2014.921123

Sommerfeld A (1896) Mathematische Theorie der Diffraction. Math Ann 47(s319):317–374

Song JM, Lu CC, Chew WC (1997) Multilevel fast multipole algorithm for electromagnetic scattering by large complex objects. IEEE Trans Antennas Propag 45:1488–1493

Staude I, Miroshnichenko AE, Decker M et al (2013) Tailoring directional scattering through magnetic and electric resonances in subwavelength silicon nanodisks. ACS Nano 7:7824–7832

Sun S, Jiang LJ, Chew WC (2013a) Enhanced A-EFIE with Calderón multiplicative preconditioner. In: Symposium on antennas and propagation (IEEE APS13), Orlando, FL

Sun S, Liu YG, Chew WC, Ma ZH (2013b) Calderón multiplicative preconditioned EFIE with perturbation method. IEEE Trans Antennas Propag 61(1):247–255

Taflove A (1995) Computational electrodynamics: the finite difference time domain method. Artech House, Boston

Taflove A, Hagness SC (2005) Computational electrodynamics: the finite-difference time-domain method, 3rd edn. Artech House, Boston

Taminiau TH, Stefani FD, Hulst NFV (2008) Enhanced directional excitation and emission of single emitters by a nano-optical Yagi-Uda antenna. Opt Express 16:10858–10866

Tong MS, Chew WC, Rubin BJ, Morsey JD, Jiang L (2009) On the dual basis for solving electromagnetic surface integral equations. IEEE Trans Antennas Propag 57(10):3136–3146

Trefethen LN, Bau D (1997) Numerical Linear Algebra, No. 50, SIAM

Tsang L, Kong JA, Ding KH (2000) Scattering of electromagnetic waves: theories and applications. Wiley, New York

Ufimtsev PY (2005) Backscatter. Wiley, New York

Ufimtsev PY (2008) New insight into the classical macdonald physical optics approximation. IEEE Antennas Propag Mag 50:11–20

Valdés F, Andriulli FP, Bağci H, Michielssen E (2011) A Calderón-preconditioned single source combined field integral equation for analyzing scattering from homogeneous penetrable objects. IEEE Trans Antennas Propag 59(6):2315–2328

Van Bladel J (1985) Electromagnetic fields. Hemisphere, New York

Vandervorst HA (1992) Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems. SIAM J Sci Stat Comput 13:631–644

Vecchi G (1999) Loop-star decomposition of basis functions in the discretization of EFIE. IEEE Trans Antennas Propag 47(2):339–346

Veysoglu ME, Shin RT, Kong JA (1993) A finite-difference time-domain analysis of wave scattering from periodic surfaces-oblique-incidence case. J Electromagn Waves Appl 7:1595–1607

Vico-Bondia F, Ferrando-Bataller M, Valero-Nogueira A (2010) A new fast physical optics for smooth surfaces by means of a numerical theory of diffraction. IEEE Trans Antennas Propag 58(3):773–789

Volakis JL, Chaterjee A, Kempel LC (1998) Finite element method electromagnetics: antennas, microwave circuits, and scattering applications. Wiley-IEEE Press, New York

Wagner RL, Otto GP, Chew WC (1993) Fast waveguide mode computation using wavelet-like basis functions. IEEE Microw Guid Wave Lett 3(7):208–210

Warwick C (2009) Everything you always wanted to know about SPICE* (*But were afraid to ask). EMC J (Nutwood UK Limited) 82:27–29

Weinan E (2012) Principles of multiscale modeling. Cambridge University Press, Cambridge, UK

DOI 10.1007/978-981-4560-75-7_6-2

Wu YM, Jiang LJ, Chew WC (2012a) An efficient method for computing highly oscillatory physical optics integral. In: Symposium on antennas and propagation (IEEE APS12), Chicago, IL

Wu YM, Jiang LJ, Chew WC (2012b) An efficient method for computing highly oscillatory physical optics integral. Prog Electromagn Res PIER 127:211–257

Wu YM, Jiang LJ, Chew WC (2013a) Computing highly oscillatory physical optics integral on the polygonal domain by an efficient numerical steepest descent path method. J Comput Phys 236:408–425

Wu YM, Jiang LJ, Chew WC (2013b) Reducing computational workload of electromagnetic scattered fields from electrically large quadratic surface at high frequency. In: Symposium on antennas and propagation (IEEE APS13), Orlando, FL

Wu YM, Jiang LJ, Sha WEI, Chew WC (2013c) The numerical steepest descent path method for calculating physical optic integrals on smooth conducting surfaces. IEEE Trans Antennas Propag 61(8):4183–4193

Wu YM, Jiang LJ, Chew WC (2014) The contour deformation method for calculating the high frequency scattered fields by the Fock current on the surface of the 3-D convex cylinder. In: Symposium on antennas and propagation (IEEE APS14)

Xiong XYZ, Jiang LJ, Lo YH, Chew WC (2014) Second-harmonic generation in metal nanoparticles modeling by surface integral equation. In: 2014 I.E. international symposium on antennas and propagation and USNC-URSI radio science meeting, Memphis

Yaghjian AD (1984) Equivalence of surface current and aperture field integrations for reflector antennas. IEEE Trans Antennas Propag 32(12):1355–1358

Yaghjian AD, Shore RA, Woodworth MB (1996) Shadow boundary incremental length diffraction coefficients for perfectly conducting smooth, convex surfaces. Radio Sci 31(12):1681–1695

Yagi H, Uda S (2014) Projector of the sharpest beam of electric waves. Proc Imp Acad Jpn (Imp Acad) 2(2):49–52. Retrieved Sept 11, 2014

Yao Y, Kats MA, Genevet P et al (2013) Broad electrical tuning of graphene-loaded plasmonic antennas. Nano Lett 13:1257–1264

Ye YQ, He SL (2010) 90 degrees polarization rotator using a bilayered chiral metamaterial with giant optical activity. Appl Phys Lett 96:203501

Ye YQ, Jin Y, He SL (2010) Omnidirectional, polarization-insensitive and broadband thin absorber in the terahertz regime. J Opt Soc Am B 27:498–504

Yee KS (1966) Numerical solution of initial boundary value problems involving Maxwell's equation in isotropic media. IEEE Trans Antennas Propag 14:302–307

Yonekura J, Ikeda M, Baba T (1999) Analysis of finite 2-D photonic crystals of columns and lightwave devices using the scattering matrix method. J Light Technol 17:1500–1508

Yu ZF, Raman A, Fan SH (2010) Fundamental limit of nanophotonic light trapping in solar cells. Proc Natl Acad Sci 107:17491–17496

Zhang Y, Cui TJ, Chew WC, Zhao JS (2003) Magnetic field integral equation at very low frequencies. IEEE Trans Antennas Propag 51(8):1864–1871

Zhang J, Xu B, Cui TJ (2014) An alternative treatment of saddle stationary phase points in physical optics for smooth surfaces. IEEE Trans Antennas Propag 62(2):986–991

Zhao JS, Chew WC (2000) Integral equation solution of Maxwell's equations from zero frequency to microwave frequency. IEEE Trans Antennas Propag 48(10):1635–1645

Zhao K, Vouvakis M, Lee JF (2004) Application of the multilevel adaptive cross-approximation on ground plane designs. IEEE Int Symp Electromagn Compat 1:124–127

Zhu Y, Cangellaris AC (2006) Multigrid finite element methods for electromagnetic field modeling. Wiley-IEEE Press, Hoboken

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_8-1

Localized Waves: Theory, Techniques and Applications

Mohamed A. Salem* and Christophe Caloz
Polytechnique Montreal, Montreal, QC, Canada

Abstract

In the first part of the chapter, an introduction to localized waves (LWs) is presented as polychromatic superposition of propagation-invariant beams (PIBs) with specific spatiotemporally coupled spectra. In the second part of the chapter, the focus is shifted towards some of the peculiar characteristics of electromagnetic LWs that distinguish them from other types of electromagnetic waves. In the last part, a presentation of the state-of-the-art techniques and experiments to generate electromagnetic PIBs is illustrated. Since PIBs are near-field phenomena, the electromagnetic structures that generate them differ significantly from conventional radiating antennas.

Keywords

Localized waves; X-waves; Propagation-invariant beams; Bessel beams; Mathieu beams; Weber beams; Modal synthesis; Antenna arrays; Metasurfaces

Introduction

Dispersion is the phenomenon where the group velocity of the wave depends on its frequency (Hecht 1998). In three dimensions, the group velocity is defined as the gradient of the angular frequency ω with respect to the wave vector $\mathbf{k}$, $\mathbf{v}_g = \nabla_{\mathbf{k}}\omega$. It is also equivalently valid to define dispersion as the dependence of phase velocity on frequency, yet such dependence may not be immediately obvious from the expression of the phase velocity, especially in monochromatic (single angular frequency) waves. Dispersion affects the propagation of waves in space and time and often introduces distortions that limit the applicability range where the wave characteristics need to be maintained. Wave dispersion is essentially an intrinsic propagation characteristic of the wave and is present independent of the propagation medium. Dispersive media and guiding structures introduce additional dispersion to waves propagating in them and thus further limit the propagation distance where wave distortion is within application tolerance.

Spatial wave dispersion is the dependence of the wave propagation direction on its wave vector components. This effect is always present even in monochromatic waves, with the exception of ideal plane waves. Spatial dispersion gradually changes the transverse profile of the wave along its propagation direction, which limits the propagation range in applications where the transverse profile of the wave needs to be maintained, such as free-space communication, imaging, remote sensing, lithography, etc.

Temporal wave dispersion affects polychromatic waves (pulses) and is the dependence of the velocity of each spectral component on the wavelength. Temporal dispersion gradually changes the extent of the wave (longitudinal profile) in time, which limits the propagation range in applications where the longitudinal profile needs to be maintained, such as optical and wideband communication, chromatic imaging, etc.

*Email: mohamed.salem@polymtl.ca

DOI 10.1007/978-981-4560-75-7_8-1

Localized waves (LWs), therefore, emerge as a special class of waves that ideally propagate in homogeneous media free from intrinsic spatial and temporal dispersion. In practice, ideal LWs cannot be generated for reasons that will be shortly explained; nevertheless, it is possible to mold nonideal LWs with dispersion-resistant characteristics for arbitrary propagation distances. Currently, LWs are well established theoretically, verified experimentally, and have plethora of engineering applications.

Historical Overview

The existence of waves that propagate without dispersion was predicted in the works of Bateman (1915) and Courant and Hilbert (1966), yet waves that propagate rigidly (without distortion) attracted considerable attention in 1983 after the publication of the work of Brittingham (1983). In his seminal work, Brittingham set forth a free-space solution to Maxwell's equations that is packet-like in nature and propagates luminally (at the speed of light). This solution, named focus wave mode (FWM), possessed infinite energy and thus was deemed physically impossible.

In 1985, Sezginer (1985) demonstrated that finite-energy luminal pulses can be constructed to propagate nearly without distortion for a certain "large" depth of field, much larger than ordinary pulses, such as Gaussian ones. In a series of subsequent papers, Ziolkowski et al. (Ziolkowski 1989; Besiries et al. 1989; Shaarawi et al. 1990; Ziolkowski 1991; Ziolkowski et al. 1993; Donnelly and Ziolkowski 1993) developed a framework for constructing and launching finite-energy FWM variants.

About the same time, Lu and Greenleaf (1992a, b) mathematically constructed and experimentally verified a new class of waves that can propagate without distortion. These solutions had a characteristic X-shaped pulse profile (see Fig. 1) and hence were given the name X-waves. In three-dimensional space, X-waves look like two cones joined at the tip and propagate along the axis of the cones. X-waves also possessed the peculiar property that their pulse peak (often referred to as the centroid) propagates with superluminal velocity, i.e., propagates faster than the speed of light. Lu et al. initially studied acoustic X-waves and their experimental work showed that the centroid propagates with supersonic velocity. The centroid of electromagnetic X-waves propagates with superluminal velocities, and since the main interest is in electromagnetic waves here, the term superluminal is used and the reader should understand that it generally refers to velocities larger than the phase velocity in a homogeneous medium. Since the centroid velocity was considered equivalent to the group velocity of the pulse, the nature of the superluminal propagation of X-waves generated controversy to whether it contradicts the special theory of relativity or not. The superluminal propagation of X-waves was experimentally demonstrated in the optical regime by Saari and Reivelt in 1997 (Saari and Reivelt 1997) and in the microwave regime by Mugnai et al. in 2000 (Mugnai et al. 2000). It is now well understood that X-waves do not violate the special theory of relativity

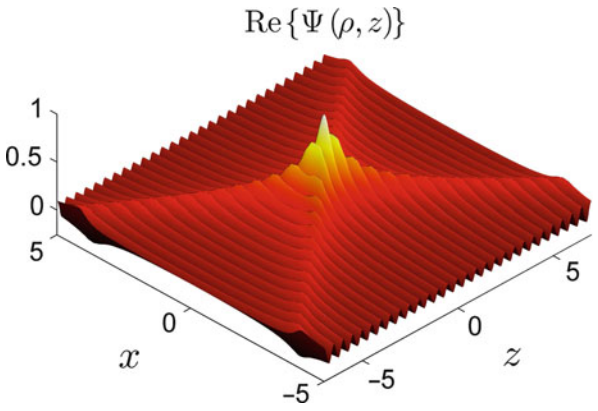


Fig. 1 Plot of the real part of the scalar X-wave introduced in Lu and Greenleaf (1992b) showing its characteristic X-shaped arms

DOI 10.1007/978-981-4560-75-7_8-1

since their superluminal centroids do not carry information. A detailed description of this phenomenon is given in section "Spectral Structure," but for now, the reader can think of this paradox as an analogy to the "superluminal scissors paradox," with the centroid analogous to the vertex and the advanced cone analogous to the blades.

The forementioned historical overview is by no means comprehensive and is meant to familiarize the reader with this topic. A comprehensive account on the historical development of LWs can be found in the introductory chapters of Hernández-Figueroa et al. (2008, 2013).

On the Name "Localized Waves": In his work, Brittingham described the localized wave solutions he found as "focus wave modes" (Brittingham 1983) pertaining to their self-reconstructing nature. Brittingham described these solutions as modes because they were propagating without distortion similar to guided wave modes. The term "localized wave" first appeared in the title of Ziolkowski's work on localized transmission of energy (Ziolkowski 1989). Ziolkowski went on and emphasized this name in his subsequent work. Later on, several terms describing LW solutions showed up in the literature and were mostly related to the Strategic Defense Initiative applications, such as electromagnetic missiles (Wu 1985; Wu and Lehmann 1985), electromagnetic bullets (Moses and Prosser 1986, 1990), and directed energy pulse trains (Ziolkowski 1989, 1991). When Lu and Greenleaf introduced the term "X-waves," it was regarded a separate class of waves as the common fundamental spectral structure of both wave families was not established yet. Recently, the term "nondiffracting waves" has been used to described the same waves (Hernández-Figueroa et al. 2013), probably pertaining to the work of Durnin on Bessel beams (Durnin 1987; Durnin et al. 1987). The term "nondiffracting" may be misleading in the sense that LWs are initially conceived as propagating solutions in a homogeneous media, whereas diffraction is conventionally associated with parts of the field scattered by an obstacle. Moreover, diffraction phenomena are also equivalently described as interference phenomena (bending of waves due to the interference of the spherical waves emitted by all unobstructed points on the wavefront). According to this description, LWs are in fact diffractive waves, since their undistorted propagation essentially depends on the interference between their different spectral components as will be shown in section "Spectral Structure." Hence, in what follows, the term "nondiffracting waves" will not be used, and the term "localized waves" (as a proper noun) will be used instead, to describe the unique spatiotemporally coupled solutions to the wave equation.

Applications

Localized waves are potentially superior to conventional waves and pulses in various applications. Areas of interest are not limited to acoustics, microwaves, and optics, but also extend to mechanics, geophysics, and even gravitational waves and elementary particle physics. The applications of LWs are numerous and span a broad range of scientific and industrial interests, such as communication, wireless energy, nondestructive testing, remote sensing, imaging, and biomedical radiography. For example, since their inception in acoustics, X-waves have been successfully employed in high-resolution ultrasound scanning of moving organs in the human body (Lu et al. 1994; Lu 1997; Cheng and Lu 2006). In electromagnetism, PIBs are extensively used in electromagnetic tweezers (Arlt et al. 2001; MacDonald et al. 2002, 2003), optical scalpels, particle manipulation and transportation (Arlt et al. 2000; Fan et al. 2000; Rhodes et al. 2002), optical lithography (Erdélyi et al. 1997; Garcés-Chávez et al. 2002; Yu et al. 2009), imaging (Herman and Wiggins 1991; Palma et al. 1996; Zhang et al. 2014), and free-space communication (Ziolkowski 1989, 1991; Salem and Bağci 2012a).

Applications and application areas where LWs can have an impact are plentiful. Past works have only revealed a small part of what LWs can offer in different areas. This limited exploitation of LWs was mainly due to two factors, the first being the incomplete understanding of the physical nature of LWs and how they interact and the second being the difficulty of generating them. In the following two sections, the

DOI 10.1007/978-981-4560-75-7_8-1

physical nature of LWs will be explored, in particular their spectral structure, and then some of the state-of-the-art approaches to generate them will be introduced.

Spectral Structure

In this section, the fundamental physics of LWs are explored in terms of their spectral structure. The strategy adopted here is to first explore the spectral structure of PIBs, then generalize from PIBs to LWs. It will be demonstrated that all the physical properties of LWs follow directly from the corresponding properties of their PIB constituents under a specific spatiotemporal coupling condition. In section "Historical Overview," PIB solutions of the source-free scalar wave equation are derived and their unique and characteristic spectral structure is determined. In section "Spatio-Temporal Coupling in Localized Wave Spectra," the spatiotemporal coupling condition that is necessary to form LWs from PIBs is established. In section "Electromagnetic Bessel-Type Localized Waves," vectorial electromagnetic LW expressions are derived from the scalar solutions, thus completing the spectral structure description of LWs.

It is interesting to note that the analysis of LW characteristics did not historically follow from PIBs as described here, but directly from Bateman constraint (Bateman 1915; Brittingham 1983; Ziolkowski 1989). Formulation in terms of polychromatic superposition of PIBs, specifically Bessel beams, appeared in the literature much later based on the work of Lu and Greenleaf (1992a, b).

Propagation-Invariant Beam Families

Propagation-invariant beams are defined through four essential characteristics: the first characteristic is being monochromatic waves, the second is that PIBs propagate along an axis, the third is that they propagate without distortion in their transverse profiles and with harmonic phase variation along their propagation axis, and the fourth is that their energy is concentrated in the transverse plane, i.e., they have a beam "spot." Propagation-invariant beams are thus similar to guided modes in waveguiding structures, except that, since they are propagating in a homogeneous medium, their spectra are not discretized. They may thus be considered as eigensolutions of the source-free scalar wave equation in homogeneous media.

Analyzing PIBs starts by writing the source-free wave equation in free space as

$$\left[\nabla^2 - \frac{\partial^2}{c^2 \partial t^2}\right] \Psi(\mathbf{r}, t) = 0, \tag{1}$$

where ∇^2 is the Laplacian, Ψ is the wave function, $\mathbf{r}$ is the position vector, and c is the speed of light in free space. To find the eigensolutions, which are of interest, the separation of variables is used to find solutions to Eq. 1, $\Psi(\mathbf{r}, t) = R(\mathbf{r})T(t)$. From the first characteristic of PIBs, only monochromatic waves are of interest; thus, the temporal dependence is fixed as $T(t) = \exp(j\omega t)$, where ω is the angular frequency of the beam. Alternatively, for the $\exp(-i\omega t)$ time dependence, replace j by $-i$ throughout. Substituting $R(\mathbf{r})\exp(j\omega t)$ into Eq. 1 yields Helmholtz equation

$$\left[\nabla^2 + k^2\right] R(\mathbf{r}) = 0, \tag{2}$$

where $k = \omega/c$ is the wave number. Equation 2 is not written in any specific coordinate system. However, to proceed, the set of coordinate systems need to be specified, where PIB solutions can be obtained. The second and third PIB characteristics suggest the need to use separation of variables again for the spatial dependence. From Morse and Feshbach (1953, pp. 509–510), it is known that there are only eleven

DOI 10.1007/978-981-4560-75-7_8-1

coordinate systems where Helmholtz equation is separable. From these coordinate systems, those that have at least one coordinate that spans $-\infty$ to $+\infty$ are chosen. This particular coordinate will correspond to the propagation axis of the PIB. This choice reduces the number of matching coordinate systems to four, namely, the Cartesian, the circular cylindrical, the parabolic cylindrical, and the elliptic cylindrical coordinate systems. In the case of the Cartesian coordinate system, all coordinates span $-\infty$ to $+\infty$, whereas in the other three cylindrical coordinate systems, only the z coordinate spans the same range. The last PIB characteristic, which requires them to have a "spot" in the transverse plane, further eliminates the Cartesian coordinates, since the eigensolutions in the Cartesian coordinates are plane waves with uniform transverse profiles.

The Laplacian in Eq. 2 is separated in a transverse term and a longitudinal term as $\nabla^2 = \nabla_\perp^2 + \partial_z^2$. Accordingly, the PIB spatial function can be also separated in transverse and longitudinal functions as $R(\mathbf{r}) = R_\perp(\rho)\exp(-j\beta z)$, where ρ is the transverse position vector and β is the wave vector component along the propagation axis, z. Substituting into Eq. 2 yields

$$\left[\nabla_\perp^2 + \chi^2\right] R_\perp(\rho) = 0, \tag{3}$$

with $\chi^2 = k^2 - \beta^2$. The transverse Laplacian $\nabla_\perp^2$ depends on the choice of the coordinate system. Table 1 summarizes the description of the three coordinate systems of interest and gives expressions of the transverse Laplacian in each system.

Eigensolutions of Eq. 2 in the circular, parabolic, and elliptic cylindrical coordinate systems are, respectively, known as Bessel (Durnin 1987), Weber (or parabolic) (Bandres et al. 2004), and Mathieu (or elliptic) beams (Gutiérrez-Vega et al. 2000). Any PIB is thus either one of these eigensolutions or a superposition thereof.

A pictorial representation of the PIB spectral structure is shown in Fig. 2. In the figure, k_x and k_y are the Cartesian transverse components of the wave vector. Waves with their spectrum confined within the red dashed circle at $k_x^2 + k_y^2 = \chi^2 = k^2$ are propagating waves. Waves with spectral components outside this circle are evanescent since their longitudinal wave vector component is imaginary according to $\beta^2 = k^2 - \chi^2$ and $\chi^2 > k^2$. Any single point within the circle represents a propagating plane wave in the direction defined by the direction cosines (k_x/k, k_y/k, β/k), e.g., the black star at $k_x = k_y = 0$ represents the spectrum of a plane wave propagating in the z-direction. The blue solid circle with radius χ/k is the locus of any propagating PIB. A PIB is thus understood as a continuum of an infinite number of plane waves propagating along a conical surface with a conical angle given by arcsin (χ/k). The distribution of the

Table 1 Transverse Laplacian in the coordinate systems supporting PIB eigensolutions Morse and Feshbach (1953, chapter 6)

Coordinate system	Relation to Cartesian	Transverse Laplacian
Circular (ρ, ϕ, z) $\rho \in [0, \infty)$ $\phi \in [0, 2\phi)$ $z \in (-\infty, \infty)$	$x = \rho \cos(\theta)$ $y = \rho \sin(\theta)$ $z = z$	$\nabla_\perp^2 R_\perp = \frac{\partial}{\rho\partial\rho}\left[\rho \frac{\partial R_\perp}{\partial \rho}\right] + \frac{\partial^2 R_\perp}{\rho^2 \partial \phi^2}$
Parabolic (σ, τ, z) $\sigma \in [0, \infty)$ $\tau \in [0, \infty)$ $z \in (-\infty, \infty)$	$x = \sigma\tau$ $y = \frac{1}{2}[\tau^2 - \sigma^2]$ $z = z$	$\nabla_\perp^2 R_\perp = \frac{1}{\sigma^2+\tau^2}\left[\frac{\partial^2 R_\perp}{\partial \sigma^2} + \frac{\partial^2 R_\perp}{\partial \tau^2}\right]$
Elliptic (μ, v, z) $\mu \in [0, \infty)$ $v \in [0, 2\pi)$ $z \in (-\infty, \infty)$	$x = a \cosh(\mu) \cos(v)$ $y = a \sinh(\mu) \sin(v)$ $z = z$	$\nabla_\perp^2 R_\perp = \frac{1}{a^2\left[\sinh^2(\mu) + \sin^2(v)\right]}\left[\frac{\partial^2 R_\perp}{\partial \mu^2} + \frac{\partial^2 R_\perp}{\partial v^2}\right]$

DOI 10.1007/978-981-4560-75-7_8-1

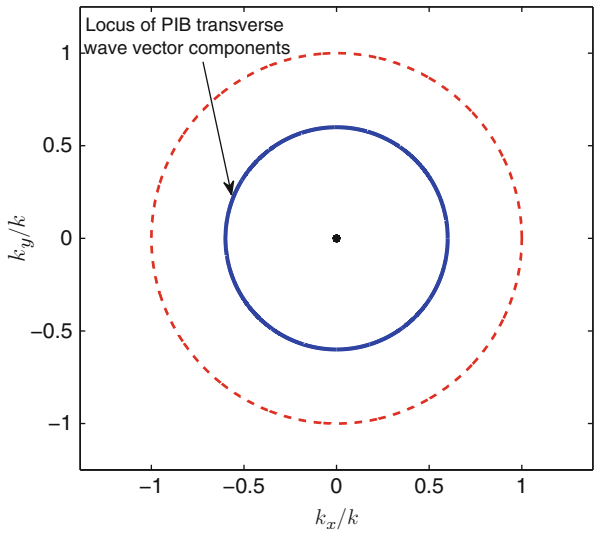


Fig. 2 Pictorial representation of the spectral structure of PIBs. Any point in the plane $k_x - k_y$ represents a plane wave of the form $\exp\left(-j\left[k_x x + k_y x + \sqrt{k^2 - k_x^2 - k_y^2}z\right]\right)$. All plane waves within the *red dashed* circle are propagating waves, while those outside the circle are evanescent. A PIB is a superposition of an infinite number of propagating plane waves having the same transverse wave number $\chi^2 = k_x^2 + k_y^2$. The distribution of the plane-wave magnitudes along the locus χ/k determines the family of the PIB, while the distribution of the phases of the plane waves determines the order of the PIB

magnitudes of these plane waves along the circle is what determines the family of the beam, and the distribution of the phases is what determines the order of the beam, e.g., all the plane waves in a Bessel beam have the same magnitude, while Weber beams are characterized by an increased relative magnitude in the vicinity of the two opposite points.

Spatiotemporal Coupling in Localized Wave Spectra

In this section, LW solutions will be constructed based on Bessel PIBs, noting from the discussion in the previous section that this process can be generalized to any type of PIBs. The choice of Bessel PIBs is due to their familiarity and the availability of mathematical framework to manipulate them.

Using separation of variables, the transverse function is expressed as $R_\perp = P(\rho)\,\Phi(\phi)$, where $P(\rho)$ is the radial dependence function and $\Phi(\phi)$ is the azimuthal dependence function. Substituting into Eq. 2,

$$P''(\rho)\Phi(\phi) + \frac{1}{\rho}P'(\rho)\Phi(\phi) + \frac{1}{\rho^2}P(\rho)\Phi''(\phi) + \chi^2 P(\rho)\Phi(\phi) = 0, \tag{4}$$

where the dash designates derivation with respect to the argument. Dividing by $P(\rho)\Phi(\phi)/\rho^2$ and rearranging the terms yields

$$\frac{\rho^2 P''(\rho)}{P(\rho)} + \frac{\rho P'(\rho)}{P(\rho)} + \chi^2\rho^2 = -\frac{\Phi''(\phi)}{\Phi(\phi)}, \tag{5}$$

where the left-hand side of (5) is only a function in ρ and the right-hand side is only a function in ϕ. The equation is now separated, and since solutions that are azimuthally periodic are sought after, the separation constant must be negative.

The azimuthal equation is written as

DOI 10.1007/978-981-4560-75-7_8-1

$$\Phi''(\phi) + m^2\Phi(\phi) = 0, \tag{6}$$

with $-m^2$ the separation constant. Equation 6 has the solution

$$\Phi(\phi) = Ce^{jm\phi} + De^{-jm\phi}, \tag{7}$$

where C and D are constants. Note that Eq. 7 may be alternatively written as $\Phi(\phi) = C\cos(m\phi) + D\sin(m\phi)$.

The radial equation is multiplied by ρ^2 and rearranged to read

$$P''(\rho) + \frac{1}{\rho}P'(\rho) + \left(\chi^2 + \frac{m^2}{\rho^2}\right)P(\rho) = 0. \tag{8}$$

Equation 8 is a modified version of Bessel differential equation and has the solution

$$P(\rho) = AJ_m(\chi\rho) + BY_m(\chi\rho), \tag{9}$$

where A and B are constants, $J_m(z)$ and $Y_m(z)$ are, respectively, the ordinary Bessel functions of first and second kind, and order m.

From Eqs. 7 and 9, the expression of the monochromatic Bessel beam of order m is written as

$$\psi(\rho, \phi, z, t; \chi, m, \beta, \omega) = J_m(\chi\rho)e^{jm\phi}e^{-j\beta z}e^{j\omega t}. \tag{10}$$

Note that $B = 0$ is set in the Bessel beam expression, since $Y_m(z) \to -\infty$ as $z \to 0$ and is thus unphysical. Also $\exp(-jm\phi)$ is omitted from the expression, since it is directly recovered from $\exp(jm\phi)$ by changing the sign of m.

Since Bessel beams are eigensolutions in the circular cylindrical coordinate system and thus form a complete and orthogonal set, they may be used as an expansion set for any arbitrary wave. The expansion is called the Fourier-Bessel expansion and takes the form

$$\Psi(\rho, \phi, z, t) = \sum_{m=-\infty}^{\infty}\int_0^{\infty} d\chi \int_{-\infty}^{\infty} d\beta \int_{-\infty}^{\infty} d\omega \chi \tilde{\Psi}_m(\chi, \beta, \omega) J_m(\chi\rho)e^{-j\beta z}e^{j\omega t}e^{jm\phi}, \tag{11}$$

where $\tilde{\Psi}_m(\chi, \beta, \omega) = \tilde{\Psi}_m(\beta, \omega)\delta\left(\chi^2 - \left[(\omega/c)^2 - \beta^2\right]\right)$ is the spectrum of $\Psi(\rho, \phi, z, t)$ and the Dirac delta is the spectral representation of Eq. 1.

If Ψ is an ideal LW, one may expect it to maintain its form along its axis, with the exception local variations, as it propagates. This propagation property is expressed as (Zamboni-Rached et al. 2002)

$$\Psi(\rho, \phi, z, t) = \Psi\left(\rho, \phi, z + \Delta z_0, t + \frac{\Delta z_0}{V}\right), \tag{12}$$

where Δz_0 is an arbitrary distance and V is the centroid speed.

Using Eq. 11, the translated wave in Eq. 12 is expressed as

DOI 10.1007/978-981-4560-75-7_8-1

$$\Psi\left(\rho,\phi,z+\Delta z_0,t+\frac{\Delta z_0}{V}\right)= \sum_{m=-\infty}^{\infty}\int_0^{\infty}d\chi\int_{-\infty}^{\infty}d\beta\int_{-\infty}^{\infty}d\omega\chi\widetilde{\Psi}_m(\chi,\beta,\omega)J_m(\chi\rho)e^{-j\beta[z+\Delta z_0]}e^{j\omega[t+(\Delta z_0/v)]}e^{jm\phi}. \tag{13}$$

One can immediately recognize that Eq. 12 is satisfied when $\exp(-j\beta[z+\Delta z_0])\exp(j\omega[t+(\Delta z_0/V)]) = \exp(-j\beta z)\exp(j\omega t)$, which yields the periodic reconstruction constraint

$$\omega = V\beta + \alpha, \tag{14}$$

where $\alpha = 2n\pi\frac{V}{\Delta z_0}$ and n an integer. This constraint does not contradict the spectral representation of the wave, which is taken into account by means of the Dirac delta in $\widetilde{\Psi}$ and the fact that the spectrum of the wave is indeed a function of β and ω as observed in $\hat{\Psi}$. Spectra of an LW solution may thus be written as

$$\tilde{\Psi}_m(\chi,\beta,\omega) = \tilde{\Psi}_m(\omega)\delta\left(\chi^2-\left[\left(\frac{\omega}{c}\right)^2-\beta^2\right]\right)\delta\left(\beta-\left[\frac{\omega-\alpha}{V}\right]\right). \tag{15}$$

This spectral representation effectively reduces the free spectral parameters to a single parameter, ω in this case. Note that the LW spectrum may be equally formulated for any of the other spectral components, except for m. The reduction in the degrees of freedom is the characteristic spatiotemporal coupling property of LWs, which essentially dictates the nature of the spatial spectrum upon choosing the temporal one and vice versa.

Electromagnetic Bessel-Type Localized Waves

So far, only the scalar LW solutions are discussed, and it was shown that they are naturally represented by superposition of PIBs with a specific spatiotemporal coupling constraint. In this section, the electromagnetic vectorial LWs are derived from the scalar solutions.

The strategy here is to make use of the Hertz vector potentials to derive the electromagnetic PIB fields from scalar PIB solutions and then construct LW superpositions from the electromagnetic PIBs under the spatiotemporal coupling constraint Eq. 14. The electric and magnetic fields are derived from the Hertz vector potentials as (Stratton 1941)

$$\mathbf{E} = \nabla(\nabla\cdot\Pi_e) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\Pi_e - \mu_0\nabla\times\left(\frac{\partial}{\partial t}\Pi_h\right), \tag{16}$$

$$\mathbf{H} = \varepsilon_0\nabla\times\left(\frac{\partial}{\partial t}\Pi_e\right) + \nabla(\nabla\cdot\Pi_h) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\Pi_h, \tag{17}$$

where ε_0 and μ_0 are the free-space permittivity and permeability, respectively, and Π_e and Π_h are the electric and magnetic Hertz vector potentials, which satisfy the source-free vector wave equation

$$\nabla\times\nabla\times\Pi(\mathbf{r},t) - \varepsilon_0\mu_0\frac{\partial^2}{\partial t^2}\Pi(\mathbf{r},t) = 0. \tag{18}$$

Assuming the same harmonic time dependence $\exp(j\omega t)$, Eq. 18 reduces to the vector Helmholtz equation

DOI 10.1007/978-981-4560-75-7_8-1

$$\nabla \times \nabla \times \Pi(\mathbf{r}, \omega) + k^2 \Pi(\mathbf{r}, \omega) = 0, \tag{19}$$

where $\varepsilon_0 \mu_0 = 1/c^2$.

Equation 19 is separable in transverse electric (TE) and transverse magnetic (TM) fields in six coordinate systems, among them the three cylindrical coordinate systems (Morse and Feshbach 1953, pp. 494–523) where PIBs are eigensolutions. Moreover, TE and TM separation in these three coordinate systems requires that the Hertz vector potentials have a single component, namely, in the z-direction. This reduces Eqs. 19 to 2, for which PIB solutions are already established. In such case, Π_e yields the TE field while Π_h yields the TM field. Hence, the Hertz vector potentials is written as

$$\Pi_{e/h}(\mathbf{r}, t) = A_{e/h} \Psi(\mathbf{r}, t) \hat{\mathbf{z}}, \tag{20}$$

where Ψ is scalar PIB function, $A_{e/h}$ is an arbitrary amplitude for the electric/magnetic Hertz vector potential component, and $\hat{\mathbf{z}}$ is the unit vector in the z-direction.

Without loss of generality, the TE and TM fields are derived in the circular cylindrical coordinate system from scalar Bessel PIBs. Substituting Eqs. 10 into 20 and then into Eqs. 16 and 17 yields the TE PIB electric and magnetic fields as

$$\begin{aligned} E_\rho(\rho, \phi, z, t) &= -A_e \frac{m\mu_0 k}{\chi^2 \rho} e^{jm\phi} e^{-j[\beta z - \omega t]} J_m(\chi\rho), \\ E_\phi(\rho, \phi, z, t) &= A_e \frac{j\mu_0 k}{\chi^2} e^{jm\phi} e^{-j[\beta z - \omega t]} \frac{\partial}{\partial \rho} J_m(\chi\rho), \\ E_z(\rho, \phi, z, t) &= 0, \end{aligned} \tag{21}$$

$$\begin{aligned} H_\rho(\rho, \phi, z, t) &= -A_e \frac{j\beta}{\chi^2} e^{jm\phi} e^{-j[\beta z - \omega t]} \frac{\partial}{\partial \rho} J_m(\chi\rho), \\ H_\phi(\rho, \phi, z, t) &= -A_e \frac{m\beta}{\chi^2 \rho} e^{jm\phi} e^{-j[\beta z - \omega t]} J_m(\chi\rho), \\ H_z(\rho, \phi, z, t) &= A_e e^{jm\phi} e^{-j[\beta z - \omega t]} J_m(\chi\rho), \end{aligned} \tag{22}$$

and the TM PIB electric and magnetic fields as

$$\begin{aligned} E_\rho(\rho, \phi, z, t) &= -A_h \frac{j\beta}{\chi^2} e^{jm\phi} e^{-j[\beta z - \omega t]} \frac{\partial}{\partial \rho} J_m(\chi\rho), \\ E_\phi(\rho, \phi, z, t) &= -A_h \frac{m\beta}{\chi^2 \rho} e^{jm\phi} e^{-j[\beta z - \omega t]} J_m(\chi\rho), \\ E_z(\rho, \phi, z, t) &= A_h e^{jm\phi} e^{-j[\beta z - \omega t]} J_m(\chi\rho), \end{aligned} \tag{23}$$

$$\begin{aligned} H_\rho(\rho, \phi, z, t) &= -A_h \frac{m\varepsilon_0 k}{\chi^2 \rho} e^{jm\phi} e^{-j[\beta z - \omega t]} J_m(\chi\rho), \\ H_\phi(\rho, \phi, z, t) &= -A_h \frac{j\varepsilon_0 k}{\chi^2} e^{jm\phi} e^{-j[\beta z - \omega t]} \frac{\partial}{\partial \rho} J_m(\chi\rho), \\ H_z(\rho, \phi, z, t) &= 0. \end{aligned} \tag{24}$$

One notes here that in the case of zeroth-order ($m = 0$) PIBs, the TE field is purely azimuthally polarized, while the TM field is radially polarized. The polarization of higher-order PIBs is a combination of radial and azimuthal polarizations, while linear and elliptic polarizations cannot be achieved for these PIBs.

DOI 10.1007/978-981-4560-75-7_8-1

One also notes that no transverse electromagnetic (TEM) PIB solutions exist. This could be directly shown from the spectral structure of the beams, since they are superpositions of plane waves propagating oblique to the beam axis.

Electromagnetic LW solutions are obtained by superposition of electromagnetic PIBs under the constraint Eq. 14. This is carried out by replacing $A_{e/h}$ with the LW spectrum Eq. 15 and performing the Fourier-Bessel transform.

Moreover, the superposition of TE and TM LW fields with the same spectral function but different amplitudes can lead to new interesting wave phenomena. Some of these interesting phenomena resulting from this superposition were explored in the literature, such as truncated propagation (Salem and Bağci 2010), reflection and transmission at dielectric interfaces and slabs (Salem and Bağci 2012b), and reverse power flow (Salem and Bağci 2011).

Generation Techniques

In the previous section, the unique coupling between the spatial and temporal frequency components of LWs is explored. In this section, some techniques to generate PIBs are explored as the primary step to launch LWs. The first reported experimental generation of a Bessel beam was by Durnin et al. in 1987 (Durnin et al. 1987) for scalar optical beams using an annular slit at the focal plane of a collimating lens. A series of experiments followed, where scalar optical Bessel beams are generated using various techniques, such as axicons (Indebetouw 1989), computer generated holograms (CGHs) (Vasara et al. 1989), a combination of an annulus and a Fabry-Pérot resonator (Cox and Dibble 1992), spatial light modulators (SLMs) (Chattrapiban et al. 2003), and Mach-Zehnder interferometers (López-Mariscal et al. 2004). Generating vectorial electromagnetic Bessel beams has not been extensively reported in the literature. The lack of vectorial Bessel beam experiments is due to two main difficulties associated with generating and observing vectorial Bessel beams. The first difficulty is related to the polarization nature of the beam. As was shown in the previous section, Bessel beams posses a combination of radial and azimuthal polarizations, which require a new paradigm in designing the launching structures or antennas, since conventional structures are generally designed for linear or elliptic polarizations. The second difficulty is related to near-field nature of the beam. Since Bessel beams are near-field interferometric phenomena, their propagation distance is limited by the effective aperture of the launching structure, which effectively prevents any observation of Bessel beams in the far zone.

In this section, three different techniques to launch vectorial electromagnetic Bessel beams are detailed. The first approach is based on waveguide modal synthesis, where the propagating modes inside a waveguide section are superimposed to form a Bessel beam at the launching open end of the waveguide section. The second approach utilizes an antenna array with subsampled distribution to form an effective Bessel beam aperture. The third approach employs a metasurface, which is effectively thin metamaterial sheet, to transform an incident electromagnetic field into a Bessel beam field.

Waveguide Modal Synthesis

A truncated Bessel beam is set to be the aperture field at the open end of a flanged metallic circular waveguide section as shown in Fig. 3. As this aperture field would propagate outside the waveguide, it is taken as the transmitted field. Reflection from the aperture is ignored due to the presence of the metallic flange. The coefficients of the propagating modes inside the waveguide section are determined from the expansion of the aperture field in terms of the modes using the modes completeness and orthogonality properties. This further allows the matching of the tangential fields at the open end of the waveguide section. In order to excite the modes inside the waveguide with the required coefficients, a set of loop

DOI 10.1007/978-981-4560-75-7_8-1

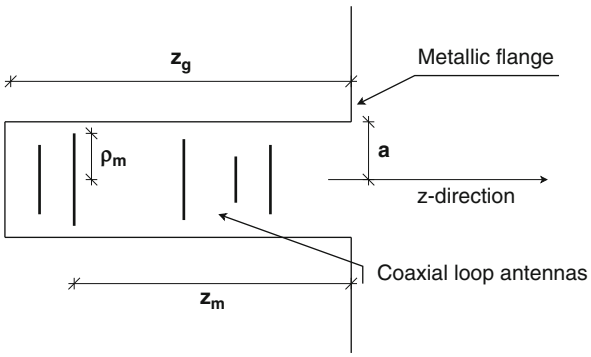


Fig. 3 Schematic showing the Bessel beam launcher setup (Reprinted from Salem et al. (2011) with permission)

antennas are placed coaxially inside the waveguide section. For power efficiency, a finite length waveguide section in the *z*-direction terminated by a perfect electric conductor is considered. The antennas are placed well away from each other to avoid cross-coupling. For simplicity, only the TE field is considered. This does not limit the applicability of the method, but simplifies the analysis and elucidates the steps of the scheme. Hence, from the axial symmetry of this configuration, only uniform currents are required to drive the antennas and excite TE modes. A relation between the current on any antenna and the excitation coefficient of each TE mode is thus established. A system of linear equations in the unknown antenna currents is constructed by choosing a number of antennas equal to the number of excited TE modes. Solving the linear system determines the antenna currents, and hence, the forward problem is solved to reconstruct the Bessel beam at the aperture. Finally, the generated beam is propagated away from the aperture and compared against the ideal truncated beam.

This approach to Bessel beam generation offers a lot of flexibility. As the generated beam is controlled directly by the antenna currents, it could be relatively easy to manipulate the beam intensity, spot size, and modulation, by controlling the excitation currents. Moreover, it allows for superposition of several Bessel beams by exciting the loop antennas with pulsed currents to launch LWs using the same setup.

Axisymmetric Transverse Electric Bessel Beams

Equation 10 gives the scalar Bessel beam solution of order *m*. The axisymmetric Bessel beam is the zero-order solution, which has no azimuthal variation. This Bessel beam has the form

$$\Psi(\rho, z, t) = AJ_0(\chi\rho)e^{j[\omega t - \beta z]}. \tag{25}$$

The time-average intensity of the beam $I(\rho, z)$ is independent of z, as

$$I(\rho, z) \propto |\langle \Psi(\rho, z)\rangle|^2 = |A|^2 J_0^2(\chi\rho), \tag{26}$$

where $\langle \Psi(\rho, z)\rangle$is the time-average function of $\Psi(\rho, z, t)$.

One also notes that the energy contained in each annular section (i.e., the area between two successive zeros of the Bessel function) is approximately equal to the energy contained in the central spot. Bessel and Gaussian beams have comparable power transport efficiencies in the Fraunhofer limit, yet the depth of field of a Bessel beam can be made arbitrary larger than that of a Gaussian beam with the same spot size, at the expense of higher power (Durnin et al. 1987).

While an ideal Bessel beam cannot be practically realized due to its infinite transverse extension and consequently infinite energy content, a truncated version could be realized. However, using an aperture of a finite radius in the transverse plane to limit the radial extension of the beam introduces edge diffraction

DOI 10.1007/978-981-4560-75-7_8-1

effects. Such effects cause the beam to spread as it propagates, hence limiting the invariant propagation range of the beam. It was shown in Durnin (1987) that the maximum propagation range $z_{\max}$ of a truncated version of the $J_0(\chi\rho)$ Bessel is

$$z_{\max} = a\sqrt{\left(\frac{k}{\chi}\right)^2 - 1}, \tag{27}$$

where a is the radius of the aperture.

The electromagnetic axisymmetric TE Bessel electric field is derived from Eqs. 25 and 21 as

$$\begin{aligned} E_\rho(\rho) &= 0, \\ E_\phi(\rho) &= AJ_1(\chi\rho), \\ E_z(\rho) &= 0, \end{aligned} \tag{28}$$

and the magnetic field is derived from Eqs. 25 and 22 as

$$\begin{aligned} H_\rho(\rho) &= -A\frac{\beta}{\omega\mu}J_1(\chi\rho), \\ H_\phi(\rho) &= 0, \\ H_z(\rho) &= -jA\frac{\chi}{\omega\mu}J_0(\chi\rho), \end{aligned} \tag{29}$$

where a (z, t) dependence of exp(j[$\omega t - \beta z$]) is assumed and omitted.

Field Modal Expansion

The TE Bessel beam field expressions Eqs. 28 and 29 can be expanded in terms of the propagating TE modes of a circular waveguide. Due to the axisymmetric nature of the Bessel beam, only axisymmetric TE modes, designated as TE_{0n} modes, would yield nonzero expansion coefficients. For a circular waveguide with radius a, the field components of the set of TE_{0n} (Collin 1990) read

$$\begin{aligned} e_\rho(\rho,z) &= 0, \\ e_\phi(\rho,z) &= jA_n\frac{\omega\mu}{\chi_{cn}}J_1(\chi_{cn}\rho)e^{\mp j\beta_n z}, \\ e_z(\rho,z) &= 0, \end{aligned} \tag{30}$$

and

$$\begin{aligned} h_\rho(\rho,z) &= \pm\frac{\beta_n}{\omega\mu}e_\phi(\rho,z), \\ h_\phi(\rho,z) &= 0, \\ h_z(\rho,z) &= A_n J_0(\chi_{cn}\rho)e^{\mp j\beta_n z}, \end{aligned} \tag{31}$$

where χ_{cn} satisfies $J_1(\chi_{cn}a) = 0$, $\chi_{cn}^2 = k^2 - \beta_n^2$, β_n is the propagation constant of the nth mode, A_n is a constant amplitude to be determined by the excitation, and the $\mp$; sign designates forward and backward propagation, respectively.

If the beam's transverse wave number χ is chosen such that $\chi = \chi_{cn}$, the Bessel beam is represented by a single TE mode. The excitation coefficient for that mode is thus $A_n = -j\chi/(\omega\mu)$. Other choices for the

DOI 10.1007/978-981-4560-75-7_8-1

beam's transverse wave number will yield some errors in the modal representation as all TE modes have to terminate at $\rho = a$. If the choice of χ results in a beam with transverse profile that has nonzero values at $\rho = a$, the modal representation will not be able to exactly replicate such profile.

The modal expansion of the beam in terms of the TE_{0n} modes utilizes the power orthogonality relation of modes (Collin 1990) such as

$$E_{\phi}^{\mathrm{BB}} = \sum_{n} c_n^{(e)} e_{\phi n}, \tag{32}$$

where E_{ϕ}^{BB} is the Bessel beam electric field component and $c_n^{(e)}$ are the expansion coefficients of the electric field, which are given by

$$c_n^{(e)} = \frac{1}{P_n} \int_0^{2\pi} \int_0^a E_{\phi}^{\mathrm{BB}} e_{\phi n}^* \rho d\rho d\phi, \tag{33}$$

where the symbol * stands for the complex conjugate and P_n is the normalization integral given by

$$P_n = \int_0^{2\pi} \int_0^a \left| e_{\phi n} \right|^2 \rho d\rho d\phi. \tag{34}$$

The calculated coefficients in Eq. 33 represent the expansion of the TE Bessel beam at $z = 0$, the open-end face of the waveguide. To avoid dealing with the diffraction effects at the edge of the waveguide, a metallic flange is introduced around the waveguide aperture (see Fig. 3) to force the electric field to be zero in the radial direction where $\rho > a$ at $z = 0^+$. The numerical results of calculating the magnetic field across the discontinuity show that such maneuver is an acceptable approximation as the value of the magnetic field along the discontinuity surface is several orders of magnitude less than its value on the aperture. Thus, the computed expansion coefficients in Eq. 33 are also the transmission coefficients of the field.

Excitation

The field inside the waveguide is excited by means of a series of loop antennas placed coaxially inside the waveguide (see Fig. 3 for a schematic). As previously stated, the field components are derived from the longitudinal magnetic field, which is now the solution to the inhomogeneous Helmholtz equation

$$\left(\nabla^2 + k^2\right) h_z(\rho) = -\sum_{m} I_m \delta(z - z_m) \frac{\delta(\rho - \rho_m)}{\rho}, \tag{35}$$

where I_m is the excitation current density on the mth antenna, z_m is the position of the mth antenna inside the guide, and ρ_m is its radius. The loop antennas are thin; thus, two Dirac delta functions $\delta(x)$ are sufficient to fully describe the geometry and position of each one. The antennas are assumed to be far enough from each other to avoid cross-coupling. The antennas are also placed adequately far enough from the ends of the waveguide to prevent interaction between the edges and the evanescent fields generated in the vicinity of each antenna.

Equation 35 is tackled by first solving the corresponding Green's function equation

$$\left(\frac{d^2}{dz^2} + \beta_n^2\right) g_n(z) = -I_m J_0(\chi_{cn} \rho_m) \delta(z - z_m), \tag{36}$$

DOI 10.1007/978-981-4560-75-7_8-1

subject to the boundary condition $g_n(z) = 0$ at $z = z_g$, where z_g is the length of the waveguide section. Here it is assumed that no field is reflected back from the open end of the waveguide section at $z = 0$. The solution of Eq. 36 is (Felsen and Marcuvitz 1994)

$$g_n(z) = \begin{cases} \tau_n e^{-j\beta_n z}, & z > z_m \\ \gamma_n \sin\left(\beta_n \left[z_g - z\right]\right), & z < z_m \end{cases} \tag{37}$$

where τ_n and γ_n are unknown coefficients and $\mathrm{Im}\{\beta_n\} > 0$.

The unknown coefficients are determined by applying the source condition which implies that at $z = z_m$, $g_n(z)$ is continuous, and its first derivative with respect to z is discontinuous by $-I_m J_0(\chi_{cn}\rho_m)$. Hence, for the field emerging at $z > z_m$

$$\tau_n = -\frac{j}{2\beta_n} I_m J_0(\chi_{cn}\rho_m) e^{j\beta_n z_m} \xi_{nm}, \tag{38}$$

where $\xi_{nm} = 1 - \exp(-2j\beta_n[z_g - z_m])$. The magnetic field is readily available as

$$h_z(\rho, z) = \frac{\sqrt{2}}{a} \sum_n \sum_m \tau_n \frac{J_0(\chi_{cn}\rho)}{J_0(\chi_{cn}a)} e^{-j\beta_n z} \xi_{nm}. \tag{39}$$

Using Maxwell's equations to determine the electric field

$$e_{\phi n}(\rho, z) = \frac{J\omega\mu\beta_n}{\chi_{cn}^2} \frac{\partial}{\partial\rho} h_{zn}(\rho, z), \tag{40}$$

and hence, the expansion coefficients in the azimuthal direction, c_n, are

$$c_n = \frac{1}{\sqrt{2}\chi_{cn}a} \sum_m I_m \frac{J_0(\chi_{cn}\rho_m)}{J_0(\chi_{cn}a)} e^{j\beta_n z_m} \xi_{nm}. \tag{41}$$

The set of coefficients c_n relate the current density on the mth antenna to the excitation coefficient of the nth mode. Thus, by choosing a number of antennas m equal to the number of modes n, one can construct a linear system that can be inverted to compute the unknown excitation currents from the known excitation coefficients. The linear system is given by

$$\mathbf{LJ} = \mathbf{C}, \tag{42}$$

where $\mathbf{J}$ is the vector of unknown current densities I_m, $\mathbf{C}$ is the vector of expansion coefficients $c_n^{(e)}$, and the elements of the matrix $\mathbf{L}$ are

$$L_{nm} = \frac{1}{\sqrt{2}\chi_{cn}a} \frac{J_0(\chi_{cn}\rho_m)}{J_0(\chi_{cn}a)} e^{j\beta_n z_m} \xi_{nm}. \tag{43}$$

The excitation currents are obtained by multiplying Eq. 42 by the inverse if the matrix $\mathbf{L}$.

In the case of $\chi = \chi_{cp}$, where the truncated Bessel beam could be fully represented by the mode number p, one should still use $m = n > 1$ number of antennas and modes, with the expansion coefficients $c_n = 0$ for $n \neq p$. This is essential to obtain a good approximation of the ideal truncated Bessel beam.

DOI 10.1007/978-981-4560-75-7_8-1

Table 2 Excitation currents $I_m(\times 10^{-12}\lambda)$ $\angle 14.2835°$ for configurations C_1 and C_2

m	C_1	C_2
1	−2.232	1.300
2	1.541	−3.614
3	1.400	2.551
4	1.156	−7.540
5	1.236	–
6	−0.169	–
7	−1.104	–
8	−0.262	–

Numerical Example

To illustrate the technique, a numerical example is given, where a Bessel beam is generated inside the waveguide then propagated in free space. A beam with $\chi = \chi_{c4} = 13.3237/a$ and a propagation constant $\beta_4 = \sqrt{k^2 - \chi^2} \approx 5.5418/\lambda$ are chosen, which has four distinctive annular sections in the waveguide cross section. The launcher structure is a truncated circular waveguide section of length $z_g = 30\lambda$, where λ is the wavelength of the beam. The walls of the waveguide are perfect electric conductors and the waveguide has a radius $a = 4.5\lambda$. The choice of this value for the radius ensures that all TE_{0n} modes below TE_{09} are above cutoff.

Eight loop antennas are chosen to excite the beam. Here, the antennas are chosen to have uniformly decreasing radii as they are placed farther away from the open end of the waveguide starting with the largest radius of $\rho_m = 4.25\lambda$ and decreasing by half wavelength in each successive loop. The antennas are separated by 3λ from each other, and the first antenna is placed 4.5λ away from the open end of the waveguide.

The generation technique is demonstrated using two configurations: in the first configuration C_1, eight loop antennas are used for the analysis to match all propagating TE_{0n} in this waveguide section, while in the second configuration C_2, only the first four antennas are used, and consequently only the first four modes excitation coefficients are matched to the beam's coefficients. In both configurations, all expansion coefficients are identically zero except for $c_4^{(e)} = -j\chi/\omega\mu$.

The matrix elements relating the expansion coefficients to the excitation current densities are constructed using Eq. 43. The condition numbers of the resulting matrices $\mathbf{L}$ in ℓ^2 norm are 29.6325 and 8.2867 for C_1 and C_2, respectively, which indicate that the matrices are not ill-conditioned. The excitation current densities obtained after inverting $\mathbf{L}$ and multiplying with $\mathbf{C}$ are given in Table 2.

The relative error in the mode excitation coefficients obtained by solving the forward excitation problem is calculated as $\Delta_n = \left|\tilde{c}_n^{(e)} - c_n^{(e)}\right|$, where $\tilde{c}_n^{(e)}$ are the reconstruction coefficients. Results show that in C_1 and C_2, the generated matched modes have relative errors in the order of the machine precision, while in C_2 the unmatched modes have relative errors in the order of $\mathcal{O}(-14)$. The effect of launching the unwanted modes is apparent in Fig. 4, where the transverse intensity profiles of the ideal truncated and the reconstructed beams of C_1 and C_2 are plotted. The figure shows an excellent matching between the ideal truncated beam and that generated in C_1, while the beam generated in C_2 suffers from deformations due to the interference effects with the launched higher-order modes.

To further investigate the behavior of the generated beams, a finite-difference frequency-domain (FDFD) scheme is employed to simulate the propagation of the beam in free space. The developed FDFD scheme uses a 2D staggered grid in the $\rho - z$ plane following the body of revolution cell grid scheme in Taflove and Hagness (2005). The cell dimensions are $\Delta z = \Delta\rho = \lambda/20$ and the computation domain is 20λ in the z-direction and 15λ in the ρ-direction. An absorbing boundary of thickness 5λ is

DOI 10.1007/978-981-4560-75-7_8-1

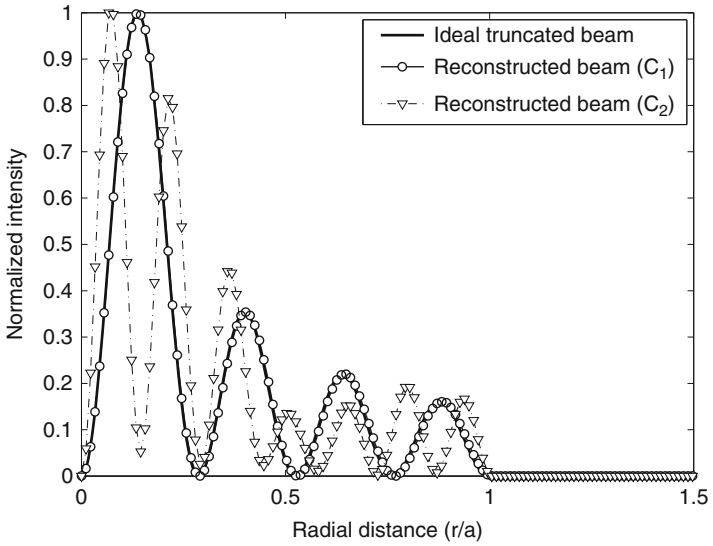


Fig. 4 Normalized intensity plot of the ideal truncated and the reconstructed Bessel beams at the open end of the circular waveguide section (Reprinted from Salem et al. (2011) with permission)

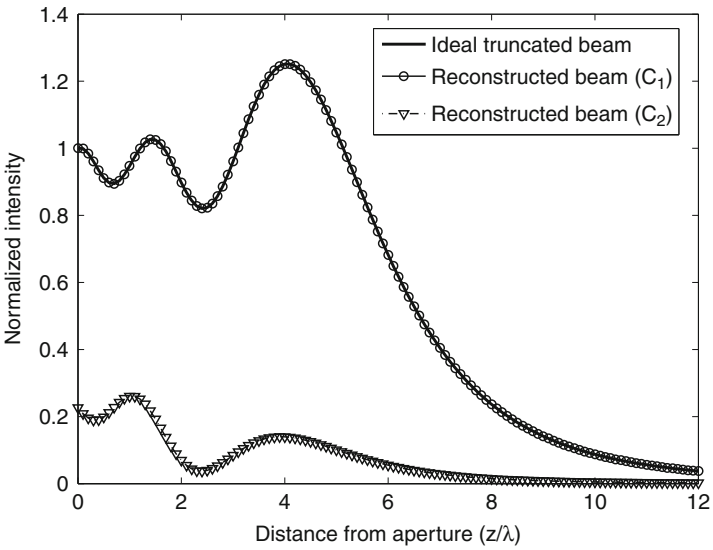


Fig. 5 Comparison between the intensities of the ideal truncated and the reconstructed Bessel beams along the propagation direction (Reprinted from Salem et al. (2011) with permission)

added before and after the computation domain in the z-direction, and an absorbing boundary of thickness 10λ is added after the computation domain in the ρ-direction. The absorbing layer is implemented by adding $i\varepsilon|x - x_{CD}|^{\delta}$ to the permittivity, where ε and δ are constants x and x_{CD} are cell number and the boundary cell number of the computation domain, respectively. For the z-direction, $\varepsilon_z = 0.075\Delta z$ and $\delta_z = 2.65$; and for the ρ-direction, $\varepsilon_\rho = 0.025\Delta\rho$ and $\delta_\rho = 0.65$. The FDFD scheme is tested with a unit strength impressed magnetic line source excitation at $\rho = 0$ and is verified to produce the magnetic field $H_z(\rho) = -k^2/(4\omega\mu_0)H_0^{(1)}(k\rho)$, with $H_0^{(1)}(k\rho)$ the Hankel function of first kind and zero order.

Figure 5 shows the normalized intensity plot of the reconstructed beams in comparison to an ideal truncated beam along the z-direction. The normalized intensity plot along the z-axis (at $\rho_a = 1.8412/\chi$, the maximum of the first annular section) in Fig. 5 shows that for the ideal beam and in the C_1

DOI 10.1007/978-981-4560-75-7_8-1

configuration, the beam intensity oscillates before it exhibits some focusing then sharply drops off. The simulation shows that the beam intensity drops below half of its initial value after propagating for a distance $z_{\max} \approx 6.8\lambda$, which is in the order of the geometrical optics approximation Eq. 27. For the beam in C_2, the intensity oscillates but does not exhibit the same focusing before decaying. The intensity drops below half of its initial value at $z_{\max} \approx 4.9\lambda$, which is less than the expected distance.

Figure 6 depicts the intensity of the generated Bessel beams in C_1 and C_2, respectively, in the transverse and propagation directions. The figure shows that the generated beams maintain their distinctive dark spot (due to the annulation of the intensity) at its center, while the beam reconstructed in C_1 maintains its shape in the transverse plane for longer propagation distance compared to that in C_2. The faster spread in the C_2 beam is due to the presence of higher-order modes with larger χ values.

Summary of the Technique

In summary, this Bessel beam generation technique takes advantage of completeness and orthogonality of the circular waveguide's modes. The transverse profile of the beam is decomposed in a finite number of

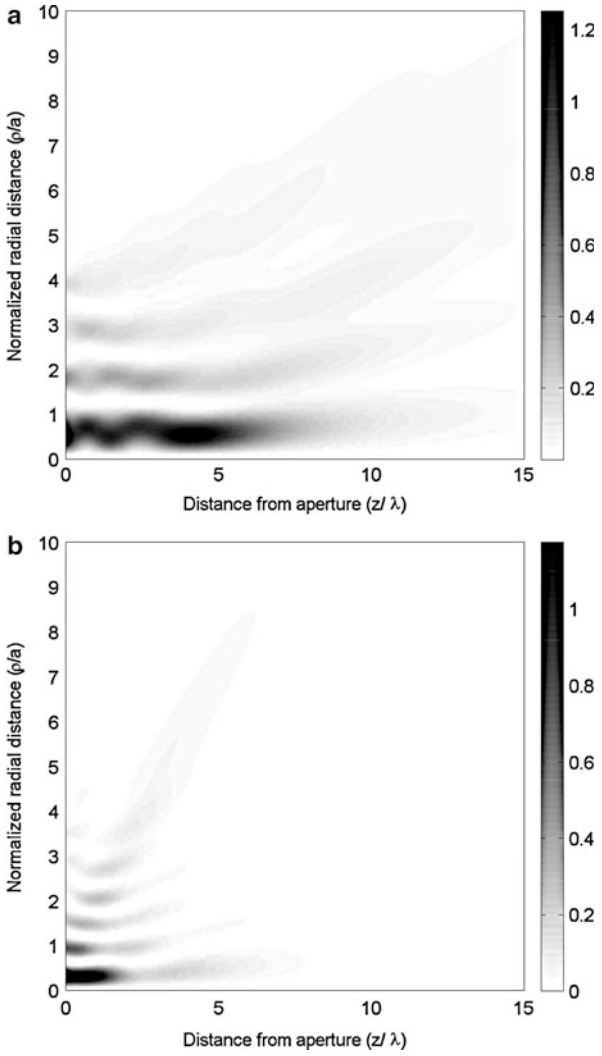


Fig. 6 Intensity plots of the reconstructed Bessel beams in the transverse plane and along the propagation direction for configurations (**a**) C_1 and (**b**) C_2 (Reprinted from Salem et al. (2011) with permission)

DOI 10.1007/978-981-4560-75-7_8-1

propagating TE_{0n} modes. Except for the case where the beam's transverse wave number χ matches one of the mode's cutoff wave numbers χ_{cn}, then the beam could be fully represented by a single mode.

Thin circular loop antennas are used to excite the modes. The antennas are placed coaxially inside the waveguide with proper separations from each other and the waveguide's ends. A relation between the antenna currents and the excitation coefficients of the modes is established. Using this relation, a system of linear equations is constructed with the antenna currents as the unknowns. It was shown that even in the case where the beam is fully represented by a single mode, it is recommended to include all propagating modes in the linear system for better matching. Solving the system yields the necessary antennas' currents. The obtained currents are used in the reverse process, and the generated beam is constructed at the open end of the waveguide. To simplify the calculations, a metallic flange is placed around the open end of the waveguide; thus, the reconstructed beam was taken as the aperture field and propagated in free space using an FDFD algorithm. Obtained numerical results for the propagation agree with the expected behavior of TE Bessel beams.

While the given numerical example serves the purpose of demonstrating the proposed scheme, several additional optimizations could be incorporated into it. An example could be optimizing the locations and radii of the antennas in order to minimize the effect of launched unwanted modes, in the case of using a number of antennas less than the total number of propagating modes. Another example of optimization is to solve for more than one parameter (antennas current) at the same time. Such optimizations could be achieved using numeric multi-objective optimization techniques, where the numeric technique simultaneously searches for the optimal parameter values that fulfill certain design constraints, such as error in excitation coefficients, antennas' current range, desired separation between antennas, or range of permissible antenna radii values.

Antenna Arrays

In this technique, the proposed concept is using a two-dimensional antenna array with a subsampled distribution for launching a zero-order pseudo-Bessel beam as an alternative and practical solution to the lack of simple, stable, and compact PIB sources. The antenna array is subsampled because, otherwise, the number of antennas in the array would be so large that the subsequent complexity would render any practical realization difficult. The term subsampled here refers to a spatial sampling of the Bessel distribution that does not meet the Nyquist criterion. It is shown that the generated beam may be engineered to propagate to several hundreds of wavelengths while maintaining its transverse localization.

The concept of an array to generate localized waves was initially proposed for launching acoustic LWs by Ziolkowski in 1992 (Ziolkowski 1992). The proposed array consists of patch or dipole antennas arranged within a circle of radius R_{array}. The most natural distribution in the case of circular symmetry is the hexagonal lattice, but a square lattice or a polar lattice could also be adequate, as shown in the Fig. 7. An almost paraxial approximation is assumed, where $\beta/\chi \gg 1$, and thus, the polarization of each array element is linear in a first approach. It is important to note that the use of the array (and thus its design) is quite different from classical arrays employed in radar, sensor, or communication applications. Ideally, the beam produced is intended to be used at distances ranging from a few tens to several hundreds of wavelengths from the launcher. To reach such distances using a pseudo-Bessel beam with a main lobe diameter of a few wavelengths only, the diameter of the launcher (D) is in the order of 50 to 100 wavelengths. From the classical far-zone limit, $2D^2/\lambda$ (Balanis 2005), it is found that a pseudo-Bessel beam extending to distances in the order of a few hundreds of wavelengths is in the near zone of the array. This means that classical array synthesis techniques and the concept of far-field radiation patterns are not pertinent to the proposed antenna array. Conversely, each antenna element has the dimension D of $\lambda/2$, so a distance of 10λ or more can be considered to be in its far field. Classical formulas for the far-field

DOI 10.1007/978-981-4560-75-7_8-1

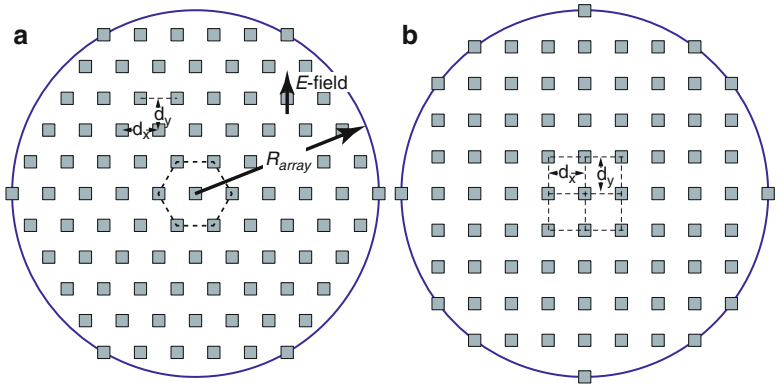


Fig. 7 Schematic representation of the pseudo-Bessel beam array antenna launcher using patch antenna elements. The antenna elements are arranged within a circle of radius R_{array} according to (**a**) a hexagonal lattice or (**b**) a square lattice (Reprinted from Lemaître-Auger et al. (2013) with permission)

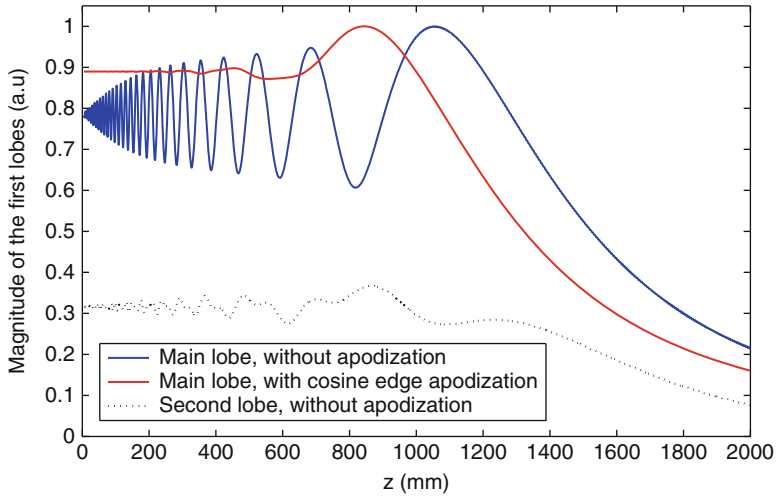


Fig. 8 Magnitude of the first two lobes along z for tapered and non-tapered truncated Bessel excitations (Reprinted from Lemaître-Auger et al. (2013) with permission)

radiation of dipoles or patch antennas can thus be used in the region of existence of the pseudo-Bessel beam (Balanis 2005).

The most intuitive excitation approach would be to sample the Bessel source function with a spacing that is less than the wavelength both to meet the Nyquist criterion and to avoid grating lobes, with each sampling point being the center of an antenna. However, this would lead to a prohibitively large number of required antenna elements, implying a complexity that would make any realization unpractical.

To relax this constraint, the possibility of subsampling the excitation function is investigated, i.e., using a spacing between antennas which is greater than the wavelength following an empirical perspective.

With an array of finite size, the tapering required to flatten the oscillations of the beam (see Fig. 8) forces the excitation function to depart from the Bessel function. Furthermore, by employing subsampling, one can expect that the excitation function will depart even more, up to a point where it becomes completely different from the Bessel function. The different excitation functions that lead to a Bessel or pseudo-Bessel beam are listed in Table 3.

The challenge is thus to find the optimal excitation function that would ideally produce a non-oscillating pseudo-Bessel beam of high quality.

DOI 10.1007/978-981-4560-75-7_8-1

Table 3 Excitation for different generator sources

Generator	Excitation function	Comment
Infinite, continuous	$J_0(\chi\rho)$	Ideal Bessel beam
Finite, continuous	$J_0(\chi\rho)$, $0 < \rho < R_{\max}$	Oscillating beam
Finite, continuous	$J_0(\chi\rho)T(\rho)$, $0 < \rho < R_{\max}$	Non-oscillating beam, $T(\rho)$: tapering
Finite, sampled	Unknown function	

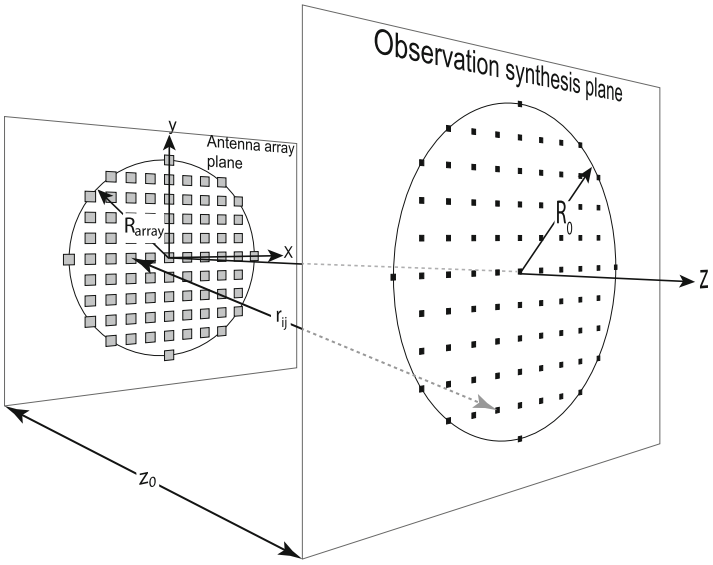


Fig. 9 Schematic representation of the antenna array and the observation synthesis plane with the coordinate systems (Reprinted from Lemaître-Auger et al. (2013) with permission)

Subsampling Synthesis

Least Mean Square Synthesis Technique: A numerical method was used for the synthesis of the antenna element excitations. The method was first proposed by Hernandez et al. (1992) for the synthesis of acoustic arrays launching LWs. The original method was intended for the generation of short pulsed signals, but it is easily adaptable to monochromatic beams. It works for scalar fields and is based on the minimization of the mean square error between the desired beam's transverse distribution, ψ, and the real field, $\hat{\psi}$. The error is computed at several points i contained inside a comparison region. Because the Bessel beam is a natural solution to the wave equation, it is assumed that if the beam is properly formed at a single transverse plane in a homogeneous medium, then it will also exist before and after that plane. In the present case, the comparison region is a circle of radius R_0 located at a transverse plane at a distance z_0 from the antenna array at an arbitrary position inside the Bessel beam region, the observation synthesis plane (OSP). This is illustrated in Fig. 9.

A second assumption is that in the region of existence of the Bessel beam, the distance between a point of interest and the antenna elements is large enough so that the fields produced by the latter are spherical. Therefore, all the antenna elements can be considered as point sources.

The scalar field at a point i of the OSP is

$$\hat{\psi}_i(x,y,z) = \sum_{l=1}^{L} \frac{e^{-jkr_{li}}}{r_{li}} f_l, \tag{44}$$

DOI 10.1007/978-981-4560-75-7_8-1

where L is the number of point sources, f_l is the complex amplitude excitation of the source l, and r_{li} is the distance between the source l and the observation point i. The mean square error (MSE) is

$$\mathrm{MSE} = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{\psi}_i - \psi_i\right)^2, \tag{45}$$

where M is the total number of observation points. The minimization of the MSE with respect to each complex amplitude f_l results in the following linear system:

$$\sum_{i=1}^{M} h_{pi} g_i = \sum_{i=1}^{M} h_{pi} \sum_{l=1}^{L} h_{li} f_l, \tag{46}$$

where the terms h_{pi} are defined as follows:

$$h_{pi} = \frac{e^{-jkr_{pi}}}{r_{pi}}. \tag{47}$$

Equation 46 can take the matrix form

$$\mathbf{H}\psi = \mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{f}, \tag{48}$$

where the elements of the matrix $\mathbf{H}$ are h_{li} and $\mathbf{f}$ is a vector with elements f_i. If the matrix $\mathbf{A} = \mathbf{H}\mathbf{H}^{\mathrm{T}}$ can be inverted, then the solution to Eq. 46 is immediately given as

$$\mathbf{f} = \mathbf{A}^{-1}\psi. \tag{49}$$

In practice, $\mathbf{A}$ is ill-conditioned and classical techniques like Gauss or LU method cannot be used. For example, for an array having 367 antennas, $\mathbf{A}$ is 367×367 and its condition number is 3.13×10^{17}.

Nevertheless, it was found that for the present problem, it is still possible to solve it directly without the use of any iterative numerical minimization techniques, contrarily as in Hernandez et al. (1992). The numerical technique is based on the singular value decomposition (SVD) of and on the neutralization effect of ill-conditioning by replacing all the small terms of the $\mathbf{A}$ matrix by infinity (Press et al. 2007).

Parametric Study: Four parameters have to be chosen in the numerical synthesis method, namely:

1. Sampling lattice of the array
2. Position of the OSP (z_0)
3. Radius of the OSP (R_0)
4. Sampling lattice of OSP

Confusion should be avoided between the sampling lattice of the array, which corresponds to the locations of the physical antenna elements of the array launcher, and the sampling lattice of the OSP, which corresponds to the points in the observation plane where the difference between the cross section of the ideal beam and that of the synthesized beam is realized according to Eq. 44.

DOI 10.1007/978-981-4560-75-7_8-1

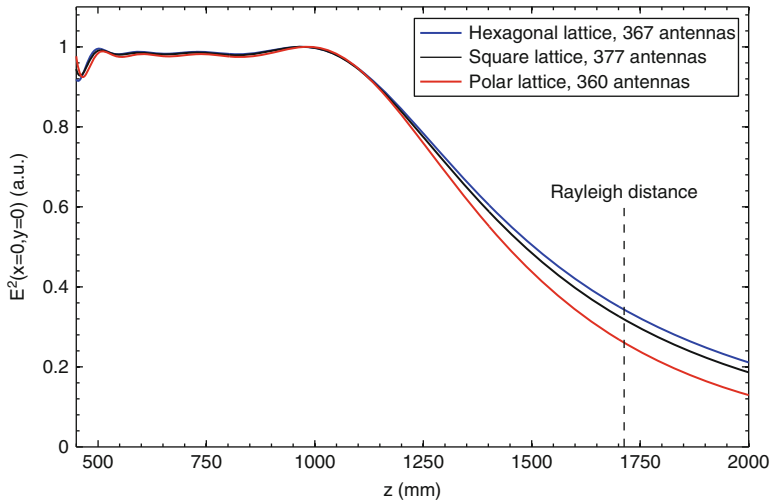


Fig. 10 Axial intensity for three different lattices having a comparable number of antennas. Array diameter is 360 mm, $\beta/\chi = 9.5, f = 50$ GHz, $R_0 = 90$ mm, and $z_0 = 650$ mm (Reprinted from Lemaître-Auger et al. (2013) with permission)

It was found that as long as the spatial sampling rate of the OSP satisfies the Nyquist criterion, its influence on the synthesized beam is almost unnoticeable. OSP sampling with a hexagonal and a square lattice are thus both possible choices.

1. *Sampling lattice of the array*: Three different lattices were tested: hexagonal, square, and polar. Using Eqs. 44 and 49, the complex amplitudes of the point source antennas were obtained.

It was found that, for a comparable number of antennas under the subsampling condition, the nature of the lattice has a minor impact on the pseudo-Bessel beam quality and on its maximum propagation length, as seen in Fig. 10. These results and all subsequent results are computed at a frequency of 50 GHz, chosen as an acceptable compromise between the size of a realizable proof-of-concept prototype and adequate propagation lengths that can be measured in a lab environment. The value of β/χ was chosen arbitrarily. The amplitudes of the obtained beams exhibit almost no oscillation and are similar up to a distance of 1,200 mm (200λ) from the array. Beyond that point, the beam obtained with the hexagonal lattice exhibits the slowest decrease in amplitude while the beam obtained with the polar grid exhibits the highest one. Before the distance of 500 mm($\sim 80\lambda$), the beam is not yet formed and the field distribution is different from the Bessel function, as it is shown in Fig. 11. Naturally, the transition is smooth and the limit of 500 mm is arbitrary. This is an effect of the array subsampling. However, for $z > 500$ mm, the Bessel beam is clearly formed and behaves as expected. An example of the field obtained for the hexagonal lattice is shown in Fig. 12, which almost exhibits no oscillation and thus no deformation along the propagation direction.

2. *Position of the observation synthesis plane (z_0)*: Of the four synthesis parameters, the distance of the OSP from the array has the greatest impact on the synthesized beam. If the OSP is too close to the array, the beam will rapidly form but it will not propagate over great distances. For larger values of z_0, the beam will propagate over greater distances, but will also be formed at a greater distance from the array. Example of simulation results is shown in Fig. 13 for a hexagonal lattice, while examples of beam evolution are shown in Fig. 14. It is a representative of similar behavior using other lattices.

DOI 10.1007/978-981-4560-75-7_8-1

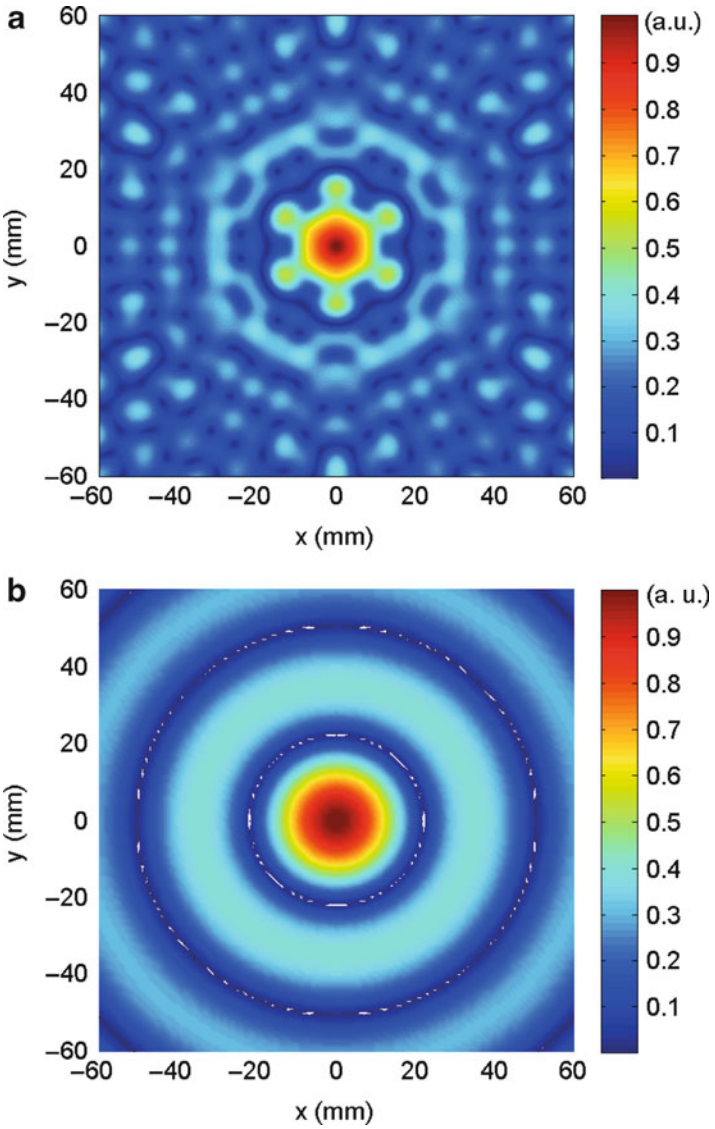


Fig. 11 Magnitude of the field at the position $z = 300$ mm. (**a**) Before the pseudo-Bessel launched with the array beam is formed. (**b**) For comparison, pseudo-Bessel beam launched with the tapered continuous excitation (Reprinted from Lemaître-Auger et al. (2013) with permission)

Figure 13 may however be misleading because, whatever the OSP position is, after the Rayleigh distance (approximated by Eq. 27) is reached, the main lobe starts to spread. The zone where the pseudo-Bessel beam exists is thus limited. Other figures of merit are thus helpful as guidelines for the synthesis: first, the maximum distance from the array reached by the main lobe before the intensity spot size is doubled; second, the useful propagation length, i.e., the difference between parameter one and the starting position of the pseudo-Bessel beam; and third, the fraction of the total power that reaches an orthogonal target of diameter equals half of the smallest diameter of the intensity spot size. The three figures of merit are shown in Fig. 15. The first parameter is increasing with the increase of the OSP position, and it is the most pertinent one for applications like industrial surface inspections or secure high-speed point-to-point communication. The second parameter varies much less except for close OSP positions. Nevertheless, this physical parameter could be important to consider for quasi-optical millimeter-wave laboratory experimental benches, high-resolution hyperthermia, and confined-beam THz spectroscopy. Finally, the

DOI 10.1007/978-981-4560-75-7_8-1

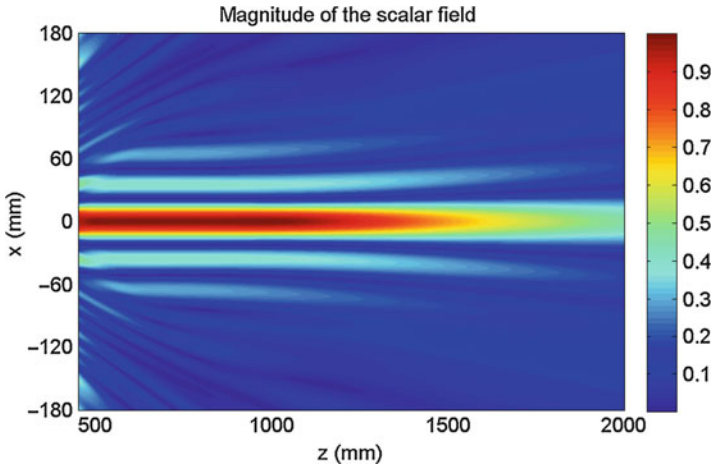


Fig. 12 Intensity of the scalar field along the propagation direction in the plane $x - z$. The field is produced by 367 antennas placed on a hexagonal lattice (Reprinted from Lemaître-Auger et al. (2013) with permission)

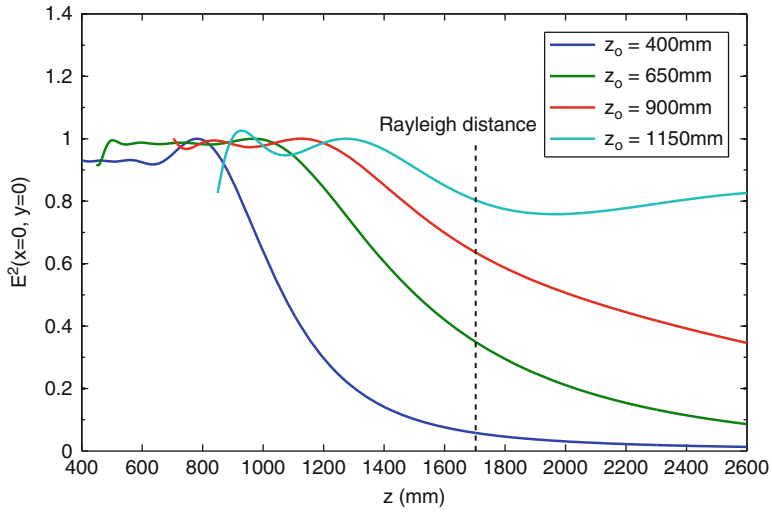


Fig. 13 Axial intensity produced with a hexagonal lattice array with 367 antennas synthesized at different OSP positions. Array diameter is 360 mm, $\beta/\chi = 9.5$, $f = 50$ GHz, and $R_0 = 90$ mm (Reprinted from Lemaître-Auger et al. (2013) with permission)

last parameter is pertinent to all applications. It globally decreases with increasing OSP position. A trade-off must be made between the power delivered by the main lobe and either the distance at which power can be delivered or on the delivery range.

3. *Radius of the observation synthesis plane (R_0)*: The last parameter to consider is the radius of the OSP. Once again, there is an optimal value. As can be seen in Fig. 16, if R_0 is too small, the beam will have a maximum propagation length shorter than the Rayleigh distance, but will present very low oscillations. This is easily explained by the fact that the total width is so small that the beam has less secondary lobes which prevents the main beam's diffraction.

For R_0 values close to the optimal choice, the beam's propagation length is the longest possible while displaying very low amplitude oscillations. For increasingly larger values of R_0, the beam suffers from increasingly larger amplitude oscillations. This can be explained by the fact that R_0 is larger than what the

DOI 10.1007/978-981-4560-75-7_8-1

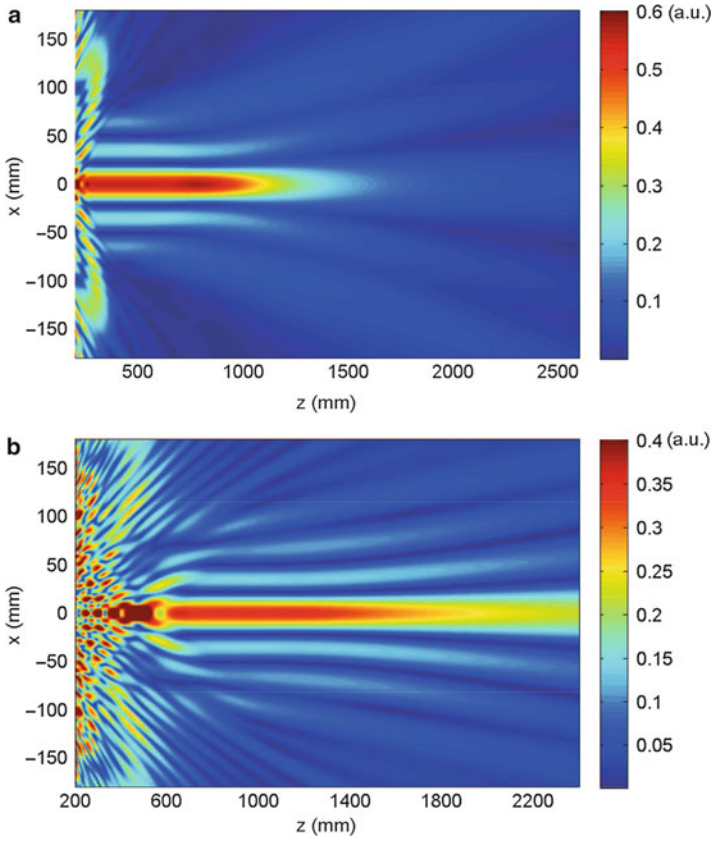


Fig. 14 Intensity of the scalar field along the propagation direction in the plane $x - z$ for two different positions of the OSP: (**a**) $z_0 = 400$ mm and (**b**) $z_0 = 900$ mm (Reprinted from Lemaître-Auger et al. (2013) with permission)

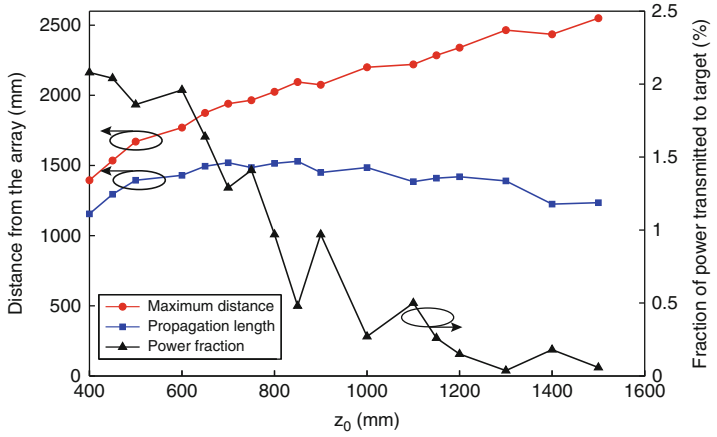


Fig. 15 Evolution of three figures of merit used versus the position of the OSP (Reprinted from Lemaître-Auger et al. (2013) with permission)

synthesis method for an array of fixed radius can naturally produce. This results in point sources close to the array edge having more power than the ones used in an array with optimal R_0 radii values. The efficiency of the natural tapering produced by the synthesis method thus decreases. The optimal value of R0 is smaller than the one indicated by geometrical optics, R_0', which is easily found from Eq. 27 as $R_0' = R_{\text{array}} - z_0\chi/\beta$.

DOI 10.1007/978-981-4560-75-7_8-1

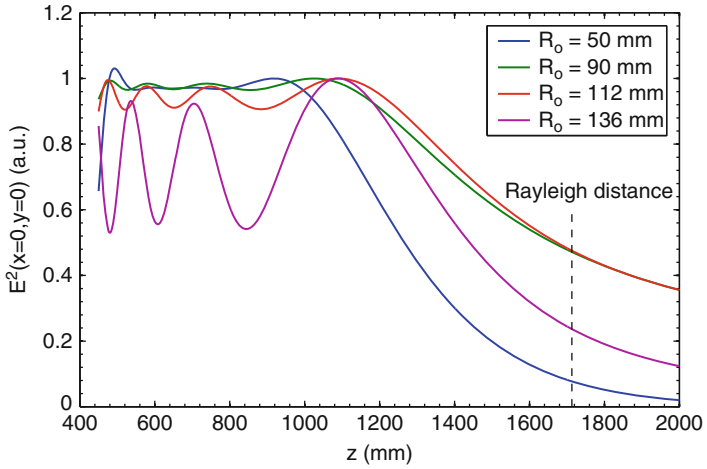


Fig. 16 Axial intensity produced with a hexagonal lattice array with 367 antennas synthesized with different OSP radii. Array diameter is 360 mm, $\beta/\chi = 9.5$, $f = 50$ GHz, and $z_0 = 650$ mm (Reprinted from Lemaître-Auger et al. (2013) with permission)

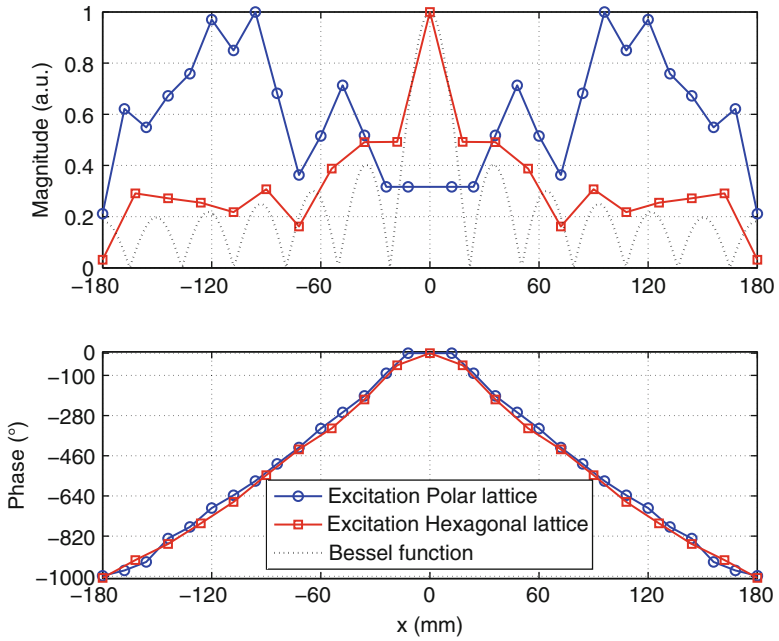


Fig. 17 Excitation along the x-axis for the hexagonal and polar lattices. (**a**) Excitation magnitude. (**b**) Excitation phase (Reprinted from Lemaître-Auger et al. (2013) with permission)

In the present case, $R_0' = 112$ mm, while the optimal value found is close to 90 mm. The axial intensity of the beams obtained with these two values of R_0 is shown in Fig. 16.

Array Implementation

Synthesis Function: As previously mentioned, the excitation function is expected to greatly depart from the Bessel function using a subsampled distribution of the radiation aperture. This was confirmed for all the tested configurations. An illustration of this is presented in Fig. 17 for the hexagonal and polar lattices. To simplify the graphs, only the excitation values along the x-axis are presented. The Bessel function is

DOI 10.1007/978-981-4560-75-7_8-1

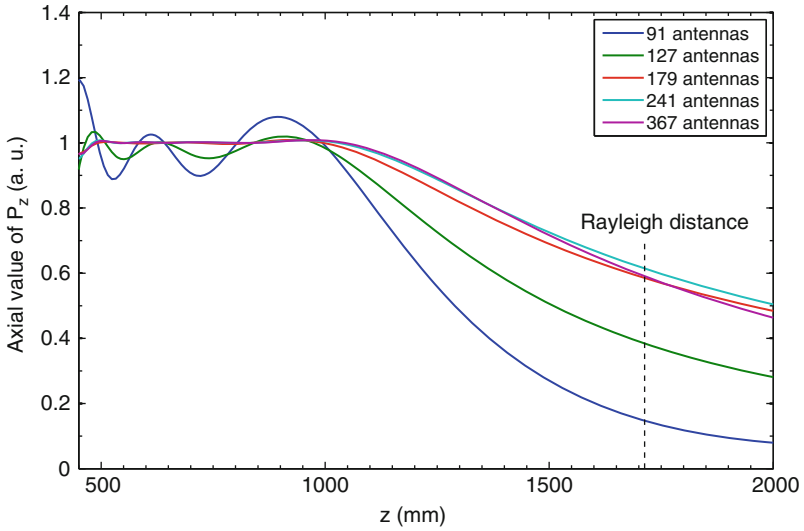


Fig. 18 Evolution of the axial *z*-component of the Poynting vector for hexagonal patch antenna arrays having different number of antennas (Reprinted from Lemaître-Auger et al. (2013) with permission)

also plotted for comparison. The differences between the excitation functions obtained with a hexagonal lattice and a polar one (see Fig. 19b) are only in the magnitude, the phase variations being almost identical for both lattices. This fact is intuitively explained by the antenna densities over the aperture surface which greatly differs between the two lattices. For the hexagonal lattice, the antenna density is uniform and the excitation amplitude has a main lobe in the center and decreases with some oscillations from the center to the outside. For the polar lattice, the antenna density decreases with increasing radius (see Fig. 19b). To compensate for this effect, the excitation amplitude first increases (still with oscillations though) and reaches its maximum at approximately 2/3 of the radius before decreasing. In both cases, the outer amplitude decrease is due to the natural tapering effect of the synthesis method.

As also noted, the magnitude and phase distributions for the two lattices are dramatically different from the Bessel function. Because the differences are seen all along the *x*-axis and not only at the edges, the underlying conclusion is that the subsampling distribution is the main cause of these differences.

Effect of Antenna Reduction in the Antenna Array: The antenna arrays discussed up to now all have 360 antennas or more. Such a large number of antennas render practical implementation very challenging. Therefore, the possibility of diminishing the number of antennas while conserving high-quality beams is investigated now. An example of results is shown in Fig. 18 for the hexagonal lattice. It is seen that a reduction of antennas from 367 to 179 antennas has almost no impact on the evolution of the axial intensity along the propagation distance. A greater decrease, however, has impact on the maximum propagation length and on the oscillation level. A good choice is thus 179 antennas. Other lattices present a similar behavior.

The drawback of decreasing the number of antennas in the array is the decrease of the power contained in the main beam with respect to the overall power emitted by the array. For example, with an 823-antenna array, 14 % of the power is contained inside the pseudo-Bessel beam. This power decreases to 8 % for a 179-antenna array and to 4 % for a 127-antenna array.

The relatively small fraction of useful power in the main beam is explained by the presence of secondary beams radiated by the array. This is a direct and unavoidable consequence of subsampling. A uniform array designed to radiate at broadside having an antenna spacing greater than λ radiates secondary beams, which are the near-field counterparts of the grating lobes in the far field. The only

DOI 10.1007/978-981-4560-75-7_8-1

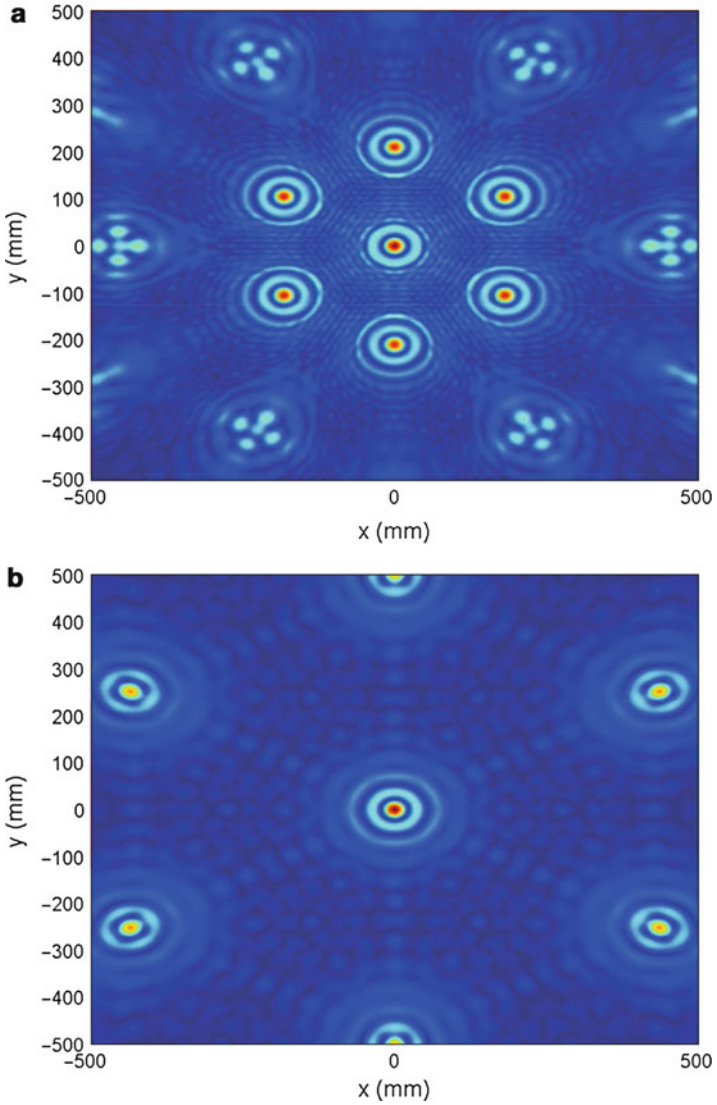


Fig. 19 Magnitude of the electric field in transverse planes located at (**a**) $z = 500$ mm; (**b**) $z = 1200$ mm (Reprinted from Lemaître-Auger et al. (2013) with permission)

difference is that the secondary lobes are also Bessel-like beams. They propagate with the same collimated behavior along directions diverging away of the z-axis. These "grating beams" are shown in Fig. 19.

Validation with Full-Wave Electromagnetic Simulations: Theoretical results for the hexagonal and polar lattice arrays were compared with electromagnetic full-wave simulations using the commercial software Ansoft Designer based on method-of-moments (MoM). The hexagonal and polar lattice arrays, shown in Fig. 20, employ 179 and 112 resonant patch antennas, respectively, operating at $f = 50$ GHz ($\lambda = 6$ mm) placed on a low-loss substrate ($\varepsilon_r = 2.2$, $h = 0.254$ mm, $\tan\delta = 0.0009$). To simplify the validation, the patch antennas were directly excited by a complex distribution (similar to Fig. 17) using excitation ports at the end of the feeding transmission line. The MoM results are from the near field of the antenna array.

Comparisons between the theoretical and MoM results for both lattices are presented in Fig. 21 which show good agreement. The small discrepancies between the results always appear in the vicinity of the maximum field, as it can be seen in Fig. 22. For all other transverse positions, the agreement is very good.

DOI 10.1007/978-981-4560-75-7_8-1

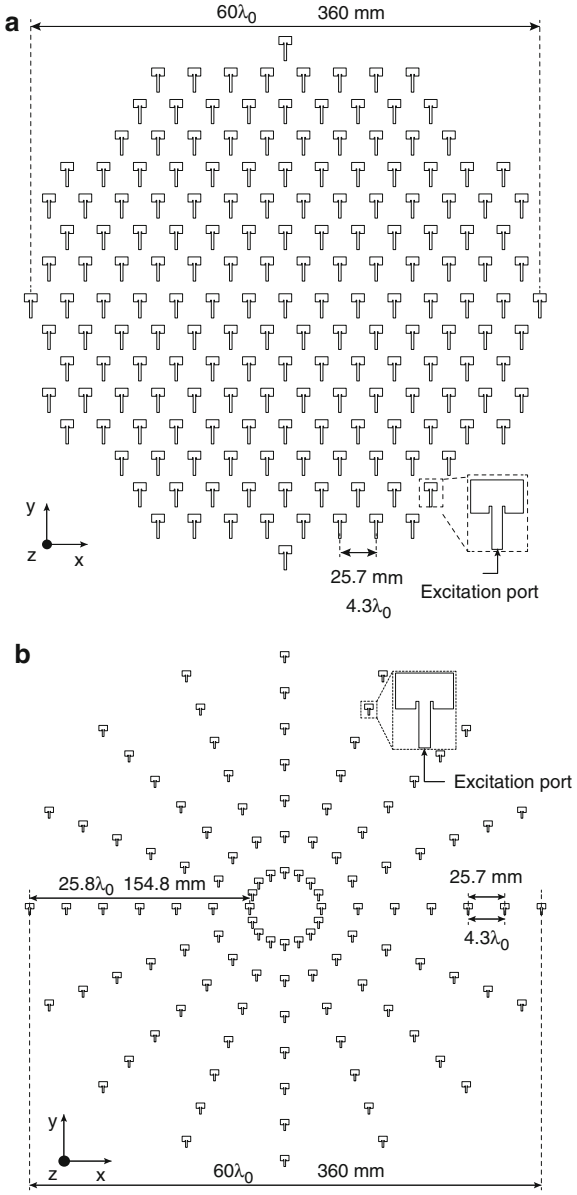


Fig. 20 Lattice configurations used in the method-of-moment full-wave simulator Ansoft Designer. The substrate is RO5880 ($\varepsilon_r = 2.2$, $h = 0.254$ mm, $\tan\,\delta = 0.0009$), $f = 50$ GHz. (**a**) Hexagonal lattice with 179 antennas. (**b**) Polar lattice with 112 antennas (Reprinted from Lemaître-Auger et al. (2013) with permission)

Figure 22 shows a comparison of the transverse value at different z-position of the z-directed Poynting vector along the x-axis. Good agreement is also observed between the theoretical and MoM results.

Summary of the Technique

Using a two-dimensional antenna array launcher for pseudo-Bessel beam generation with a minimal number of required antennas is discussed for practical implementation through the use of a subsampled distribution. It was shown that the subsampling distribution has a direct impact on the excitation function of the array greatly departing from a Bessel function. In order to obtain this excitation function, an analytical method was used based on the minimization of the least mean square of error between the desired field in a transverse plane and the real one. It was found that imposing the Bessel-like shape to the

DOI 10.1007/978-981-4560-75-7_8-1

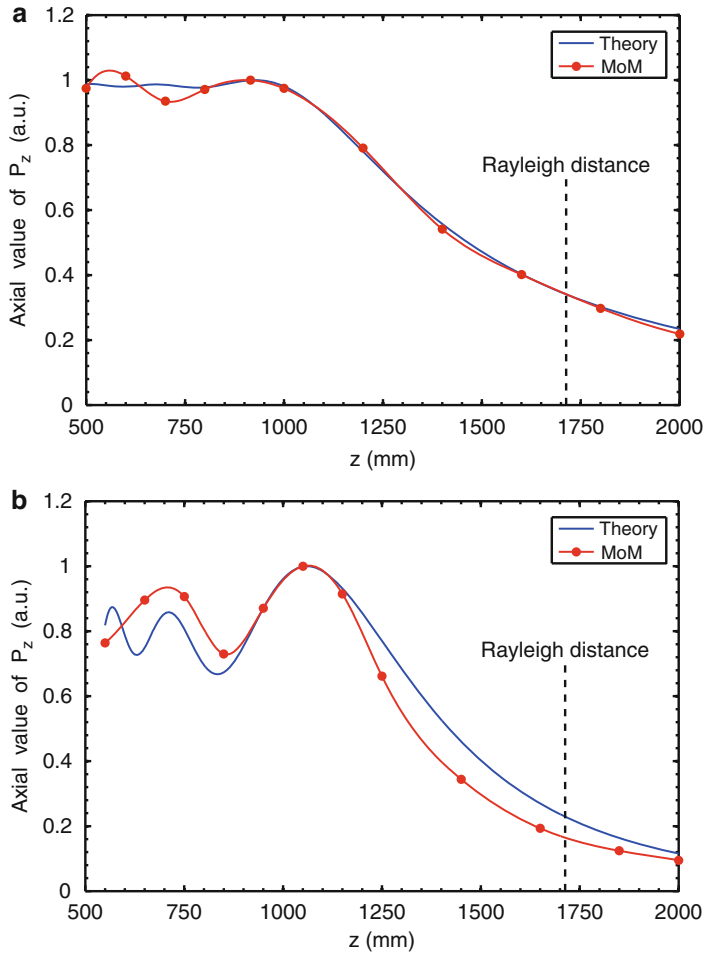


Fig. 21 Comparison between theoretical and method-of-moment (MoM) results for two different array configurations. (**a**) Hexagonal lattice with 179 antennas. (**b**) Polar lattice with 112 antennas. The commercial software Ansoft Designer was used (Reprinted from Lemaître-Auger et al. (2013) with permission)

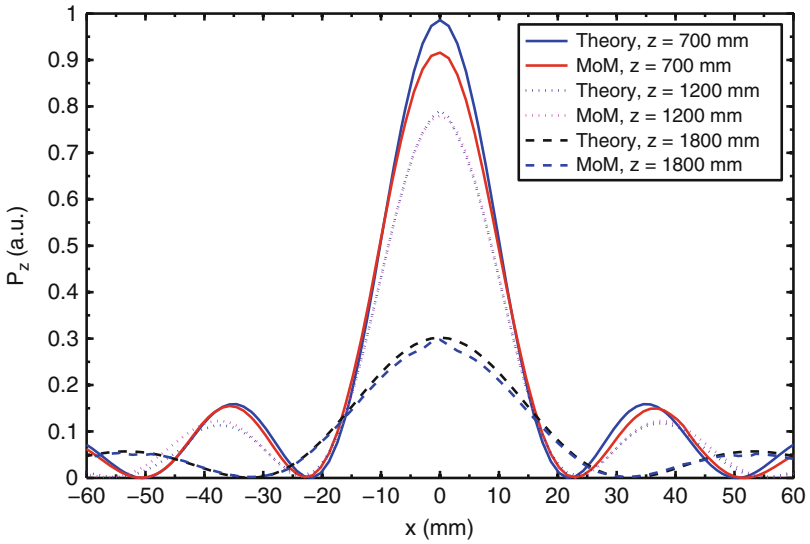


Fig. 22 Transverse value of P_z along the x-axis: comparison between theoretical and MoM simulation results at three distances along the z-direction (Reprinted from Lemaître-Auger et al. (2013) with permission)

DOI 10.1007/978-981-4560-75-7_8-1

desired field in a single plane is sufficient to obtain a pseudo-Bessel beam that propagates over several hundreds of wavelength. The synthesis method also naturally does the tapering of the excitation function to obtain very low oscillations. Full-wave simulation results are in good agreement with the theoretical predictions.

Variations to this technique, where leaky radiation is the main launching technique, were reported in Ettorre et al. (2012) and Ettorre and Grbic (2012), and another variation, where radially slotted-line antennas (RLSAs) are used, was reported by Mazzinghi et al. (2014).

Metasurfaces

Metasurfaces (Kuester et al. 2003; Holloway et al. 2009, 2012) may be considered as dimensional reductions of volumetric metamaterials (Caloz and Itoh 2005; Engheta and Ziolkowski 2006; Capolino 2009) and functional extensions of frequency selective surfaces (FSSs) (Munk 2000). They are composed of two-dimensional arrays of sub-wavelength scattering elements engineered to transform incident waves into desired reflected and transmitted waves. Compared to volumetric metamaterials, metasurfaces offer the advantage of being lighter in weight, easier to fabricate, and less lossy due to their reduced dimensionality, while compared to FSSs, they provide greater flexibility in design and much diversified functionalities.

A myriad of metasurfaces have been reported in the literature. Such metasurfaces are reported to perform electromagnetic transformation that was conventionally not possible using planar surfaces or FSSs, such as providing tunable reflection and transmission coefficients (Holloway et al. 2005), reflectionless plane-wave refraction (Pfeiffer and Grbic 2013), single-layer perfect absorption (Ra'di et al. 2013), polarization twisting (Shi et al. 2014), and vortex wave generation (Yu et al. 2011).

Due to their recent inception, metasurface synthesis techniques are in their developing stage. So far, only two universal approaches to synthesize metasurfaces have been introduced: one technique, based on transforming the wave momenta, the spatial spectral components of the wave and is suitable for scalar and paraxial wave transformation (Salem and Caloz 2014) and the second is based on the susceptibilities of the surface and is suitable for vectorial wave transformation (Achouri et al. 2014).

The technique discussed here is based on the surface susceptibility method as introduced in Achouri et al. (2014), which allows for closed-form solutions, and is hence extremely fast to compute. The technique describes the metasurface in terms of surface susceptibility tensors, where the susceptibility tensors are related through generalized sheet transition conditions (GSTCs) (Kuester et al. 2003) to the incident, reflected, and transmitted fields in the vicinity of the surface.

First, the metasurface synthesis problem is described and its general solution is provided, and then this synthesis technique is employed to transform a plane-wave incident on the metasurface into a transmitted Bessel beam.

Metasurface Synthesis Problem

A metasurface is an electromagnetic structure with sub-wavelength thickness ($\delta \ll \lambda$). The metasurface may be finite, with dimensions $L_x \times L_y$, or theoretically infinite. It is typically composed of a nonuniform arrangement of planar scattering elements that transforms incident waves into specified reflected and transmitted waves.

Figure 23 shows the sketch of the synthesis problem to solve, where it is required to find the metasurface composition that transforms an arbitrary specified incident wave, $\psi^{\mathrm{i}}(r)$, into an arbitrary specified reflected wave, $\psi^{\mathrm{r}}(r)$, and an arbitrary specified transmitted wave, $\psi^{\mathrm{t}}(r)$, assuming monochromatic waves. The solution to this synthesis problem will be expressed in terms of the surface transverse susceptibility tensor functions of $\rho = x\hat{x} + y\hat{y}$, $\overline{\overline{\chi}}_{\mathrm{ee}}(\rho), \overline{\overline{\chi}}_{\mathrm{mm}}(\rho), \overline{\overline{\chi}}_{\mathrm{em}}(\rho)$ and $\overline{\overline{\chi}}_{\mathrm{me}}(\rho)$, which represent the

DOI 10.1007/978-981-4560-75-7_8-1

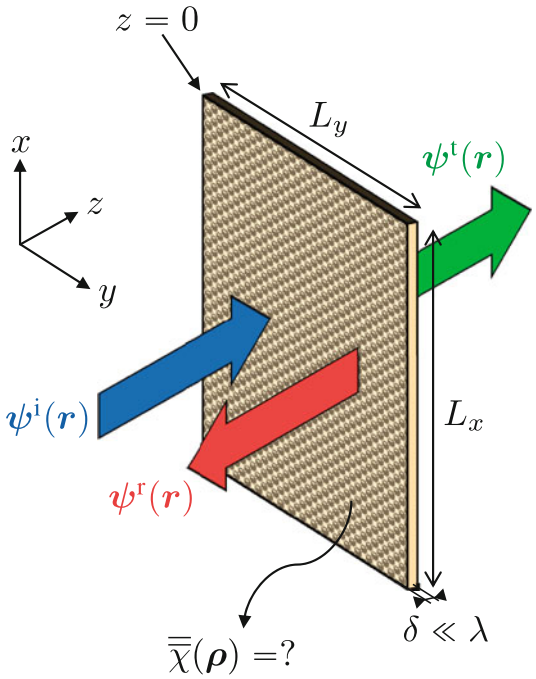


Fig. 23 Metasurface synthesis (inverse) problem to solve. A metasurface, generally defined as an electromagnetic two-dimensional nonuniform structure of extent $L_x \times L_y$ with sub-wavelength thickness ($\delta \ll \lambda$), is placed at $z = 0$. Determine the surface susceptibility tensors $\chi(\rho)$ of the metasurface transforming an arbitrary specified incident wave $\psi^{\mathrm{i}}(r)$ into an arbitrary specified reflected wave $\psi^{\mathrm{r}}(r)$ and an arbitrary specified transmitted wave $\psi^{\mathrm{t}}(r)$ (Reprinted from Achouri et al. (2014) with permission)

electric/magnetic transverse polarization responses (first subscript) to transverse electric/magnetic field excitations (second subscript).

The synthesis procedure will always yield $\overline{\overline{\chi}}_{\mathrm{ee}}(\rho), \overline{\overline{\chi}}_{\mathrm{mm}}(\rho), \overline{\overline{\chi}}_{\mathrm{em}}(\rho)$ and $\overline{\overline{\chi}}_{\mathrm{me}}(\rho)$ results, but will not guarantee that these results can be practically implemented using planar scattering elements. For instance, if the susceptibilities exhibit multiple spatial variations per wavelength, it may be difficult or impossible to realize. In such cases, one has to determine whether some features may be neglected or one may have to relax the design constraints (e.g., allow higher reflection or increase the metasurface dimensions).

The complete synthesis of a metasurface typically requires the additional step of determining the geometry of the scattering elements. This step is usually done using full-wave parametric analysis of judiciously selected scattering elements (Yu et al. 2011) or, under certain special conditions such as normal plane-wave incidence, using scattering elements with well-known electric and magnetic dipolar responses (Niemi et al. 2013). Since this second step involves scattering parameters, transformation formulas between susceptibilities and scattering parameters are provided to enable the complete synthesis of the metasurface.

Metasurface Boundary Conditions in Terms of Surface Susceptibility Tensors

A metasurface may be considered as an electromagnetic discontinuity in space. Conventional textbook electromagnetic boundary conditions do not apply to such a discontinuity, as was pointed out by Schelkunoff (1972), although they yield satisfactory results away from the discontinuities. Assuming an interface at $z = 0$, the conventional boundary conditions relate the fields at $z = 0^{\pm}$, but fail to describe the field behavior at the discontinuity itself ($z = 0$). This discrepancy is due to the fact that Stokes and Gauss theorems used to derive them assume field continuity in all the regions they apply to, including the

interface, whereas the fields may be discontinuous. For example, consider the conventional boundary condition for the normal component of the displacement vector D in the presence of surface charges ρ_s,

$$\hat{z} \cdot D|_{z=0^-}^{0^+} = \rho_s. \tag{50}$$

This relation is derived by applying Gauss theorem, $\int\int\int_V \nabla \cdot \mathbf{D} dV = \oiint_S \mathbf{D} \cdot \hat{\mathbf{n}} dS$, to a volume V enclosed by the surface S including the interface discontinuity with $\hat{\mathbf{n}}$ the normal unit vector to S. This theorem rigorously applies only if $\mathbf{D}$ is continuous inside the entire volume V, whereas in the case of a discontinuous $\mathbf{D}$, its projection onto S is not defined at the interface, and application of this theorem is not rigorously correct. Thus, since a metasurface may be modeled by Huygens sources (Pfeiffer and Grbic 2013), the correct field behavior on the metasurface cannot be determined using the conventional boundary conditions and rigorous boundary conditions, namely, GSTCs, must be applied. It should be noted that, from a physical perspective, a metasurface structure is not a single interface but rather a thin inhomogeneous slab and may be naturally treated as such. However, it is much simpler to treat the metasurface as a single interface using rigorous GSTCs, which is allowed by the fact that it is electromagnetically thin.

Rigorous GSTCs, treating discontinuities in the sense of distributions, were derived by Idemen (2011). The corresponding relations pertaining to this work were first applied by Kuester et al. to metasurfaces (Kuester et al. 2003). A detailed derivation is found in the appendix of Achouri et al. (2014), but the final form of the GSTC is expressed as

$$\hat{z} \times \Delta H = j\omega P_{\parallel} - \hat{z} \times \nabla_{\parallel} M_z, \tag{51a}$$

$$\Delta E \times \hat{z} = j\omega\mu M_{\parallel} - \nabla_{\parallel} \left(\frac{P_z}{\varepsilon} \right) \times \hat{z}, \tag{51b}$$

$$\hat{z} \cdot \Delta D = -\nabla \cdot P_{\parallel}, \tag{51c}$$

$$\hat{z} \cdot \Delta B = -\mu \nabla \cdot M_{\parallel}. \tag{51d}$$

In these relations, the terms in the left-hand sides represent the differences between the fields on the two sides of the metasurface, whose Cartesian components are defined as

$$\Delta \Psi_u = \hat{u} \cdot \Delta \Psi(\rho)|_{z=0^-}^{0^+} = \Psi_u^{\mathrm{t}} - \left(\Psi_u^{\mathrm{i}} + \Psi_u^{\mathrm{r}} \right),\ u = x, y, z, \tag{52}$$

where $\Psi(\rho)$ represents any of the fields H, E, D, or B; where the superscripts i, r, and t denote incident, reflected, and transmitted fields; and where P and M are the electric and magnetic surface polarization densities, respectively. In the most general case of a bi-anisotropic medium, these densities are related to the acting (or local) fields, E_{act} and H_{act}, by (Kong 1986; Lindell 1994)

$$P = \varepsilon N \langle \overline{\overline{\alpha}}_{\mathrm{ee}} \rangle E_{\mathrm{act}} + \varepsilon N \eta \langle \overline{\overline{\alpha}}_{\mathrm{em}} \rangle H_{\mathrm{act}}, \tag{53a}$$

$$M = N \langle \overline{\overline{\alpha}}_{\mathrm{mm}} \rangle H_{\mathrm{act}} + \frac{N}{\eta} \langle \overline{\overline{\alpha}}_{\mathrm{mm}} \rangle E_{\mathrm{act}}, \tag{53b}$$

DOI 10.1007/978-981-4560-75-7_8-1

where the $\langle \overline{\overline{\alpha}}_{ab} \rangle$ terms represent the averaged polarizabilities of a given scattering element, N is the number of scattering elements per unit area, and $\eta = \sqrt{\mu/\varepsilon}$. The acting fields are, by definition, the average fields on both sides of the surface taking into account the contributions of all the scattering elements (coupling effects) except that of the element being considered. The contribution of this element may be modeled by replacing it with a disk of radius R encompassing its electric and magnetic current dipoles. Kuester et al. express the fields of this disk as functions of the polarization densities P and M (Kuester et al. 2003), from which relations Eq. 53 can be rewritten as functions of the average fields. Their relations, with averaged polarizabilities replaced by surface susceptibilities for macroscopic description, read

$$P = \varepsilon \overline{\overline{\chi}}_{ee} \langle E \rangle + \varepsilon \overline{\overline{\chi}}_{em} \eta \langle H \rangle, \tag{54a}$$

$$M = \overline{\overline{\chi}}_{mm} \langle H \rangle \frac{+\overline{\overline{\chi}}_{me}}{\eta} \langle E \rangle, \tag{54b}$$

where the average fields are defined as

$$\langle \Psi_u \rangle = \hat{u} \cdot \langle \Psi \rangle(\rho) = \frac{\Psi_u^t + \left(\Psi_u^i + \Psi_u^r\right)}{2}, \; u = x, y, z, \tag{55}$$

where $\Psi(\rho)$ represents either H or E. The utilization of surface susceptibilities, which represent the actual macroscopic quantities of interest, allows for an easier description of the metasurface than element averaged polarizabilities and densities in Eq. 53.

The surface may be infinite or finite with dimensions $L_x \times L_y$. The two problems are automatically solved by specifying the fields Ψ_u^i, Ψ_u^r and Ψ_u^t in Eqs. 52 and 55 to be of infinite or finite $L_x \times L_y$ extent in the former and latter cases, respectively. In the finite case, truncation practically corresponds to placing a sheet of absorbing material around the metasurface. This operation neglects diffraction at the edges of the metasurface, as is safely allowed by the fact that a metasurface is generally electrically very large, but properly accounts for the finiteness of the aperture via the GSTCs (Eqs. 51a and 51b).

Synthesis Method

The proposed synthesis method solves the inverse problem depicted in Fig. 23, where the electromagnetic fields are specified everywhere (for all ρ) in the $z = 0$ plane on both sides of the metasurface and the properties of the metasurface are the unknowns to be determined. It is required to specifically find the surface susceptibilities that transform specified incident waves into specified transmitted and reflected waves. The method essentially consists in solving Eq. 51 for the components of the susceptibility tensors in Eq. 54.

The last terms in Eqs. 51a and in 51b involve the transverse derivatives of the normal components of the polarization densities, namely, $\nabla_{||} M_z$ and $\nabla_{||} P_z$. Solving the inverse problem for nonzero M_z and/or P_z would be quite involved since this would require solving the set of coupled non-homogenous partial differential equations formed by Eqs. 51a and 51b with nonzero $\nabla_{||} M_z$ and $\nabla_{||} P_z$. Although such a problem could be generally addressed by means of numerical analysis, the condition $P_z = M_z = 0$ is enforced, which will lead to convenient closed-form solutions for the susceptibilities. This restriction may limit the physical realizability of the metasurface in some cases, in the sense that the corresponding synthesized susceptibilities might be excessively difficult to realize with practical scattering elements. In such cases, the restriction might be removed without changing the main spirit of the method but at the cost of losing the closed-form nature of the solution. As shall be seen next, this restriction still allows the metasurface to

DOI 10.1007/978-981-4560-75-7_8-1

realize a large number of operations, given the large number of degrees of freedom provided by combinations of its bi-anisotropic susceptibility tensor components.

The method needs considering only Eqs. 51a and 51b as these two equations involve all the transverse field components, which is sufficient to completely describe the fields at each side of the metasurface according to the uniqueness theorem. These two equations, with $P_z = M_z = 0$, represent a set a four linear equations relating the transverse electric and magnetic fields to the effective surface susceptibilities. Thus, the solution of the inverse problem will consist in determining four transverse effective susceptibility tensors in Eq. 54.

Using the notation in Eqs. 52, 55, 51b, and 51a are rearranged to read

$$\begin{pmatrix} -\Delta H_y \\ \Delta H_x \end{pmatrix} = j\omega\varepsilon \begin{pmatrix} \chi_{\mathrm{ee}}^{xx} & \chi_{\mathrm{ee}}^{xy} \\ \chi_{\mathrm{ee}}^{yx} & \chi_{\mathrm{ee}}^{yy} \end{pmatrix} \begin{pmatrix} \langle E_{\mathrm{x}} \rangle \\ \langle E_{\mathrm{y}} \rangle \end{pmatrix} + j\omega\varepsilon\eta \begin{pmatrix} \chi_{\mathrm{em}}^{xx} & \chi_{\mathrm{em}}^{xy} \\ \chi_{\mathrm{em}}^{yx} & \chi_{\mathrm{em}}^{yy} \end{pmatrix} \begin{pmatrix} \langle H_{\mathrm{x}} \rangle \\ \langle H_{\mathrm{y}} \rangle \end{pmatrix}, \tag{56a}$$

$$\begin{pmatrix} \Delta E_y \\ -\Delta E_x \end{pmatrix} = j\omega\mu \begin{pmatrix} \chi_{\mathrm{mm}}^{xx} & \chi_{\mathrm{mm}}^{xy} \\ \chi_{\mathrm{mm}}^{yx} & \chi_{\mathrm{mm}}^{yy} \end{pmatrix} \begin{pmatrix} \langle H_x \rangle \\ \langle H_y \rangle \end{pmatrix} + \frac{j\omega\mu}{\eta} \begin{pmatrix} \chi_{\mathrm{me}}^{xx} & \chi_{\mathrm{me}}^{xy} \\ \chi_{\mathrm{me}}^{yx} & \chi_{\mathrm{me}}^{yy} \end{pmatrix} \begin{pmatrix} \langle E_x \rangle \\ \langle E_y \rangle \end{pmatrix}. \tag{56b}$$

Assuming single incident, reflected, and transmitted waves (only one wave of each of the three types), the system (Eq. 56) contains four equations for a total number of sixteen unknown susceptibility components. It is thus underdetermined as such, and it can be solved only by restricting the number of independent susceptibilities to four. This single-transformation underdetermination of Eq. 56 reveals two important facts: the first is that there exist many different combinations of susceptibilities that produce the same fields and second that a single metasurface has the fundamental capability to simultaneously manipulate several linearly independent incident, reflected, and transmitted waves. Specifically, a metasurface, as defined by Eq. 56b, can in principle manipulate up to four sets of incident, reflected, and transmitted waves (Achouri et al. 2014).

Two approaches may be considered to reduce the number of independent unknown susceptibilities. A first approach could consist in using more than four susceptibilities but enforcing relationships between some of them to ensure a maximum of four independent unknowns. For example, the conditions of reciprocity and losslessness would be a possible way to link some susceptibilities together, if this is compatible with design specifications. According to Kong (1986) and Lindell (1994), the conditions for reciprocity are

$$\overline{\overline{\chi}}_{\mathrm{ee}}^{\mathrm{T}} = \overline{\overline{\chi}}_{\mathrm{ee}}, \overline{\overline{\chi}}_{\mathrm{mm}}^{\mathrm{T}} = \overline{\overline{\chi}}_{\mathrm{mm}}, \overline{\overline{\chi}}_{\mathrm{me}}^{\mathrm{T}} = -\overline{\overline{\chi}}_{\mathrm{em}}, \tag{57}$$

and the conditions for losslessness are

$$\overline{\overline{\chi}}_{\mathrm{ee}}^{\mathrm{T}} = \overline{\overline{\chi}}_{\mathrm{ee}}^{*}, \overline{\overline{\chi}}_{\mathrm{mm}}^{\mathrm{T}} = \overline{\overline{\chi}}_{\mathrm{mm}}^{*}, \overline{\overline{\chi}}_{\mathrm{me}}^{\mathrm{T}} = \overline{\overline{\chi}}_{\mathrm{em}}^{*}, \tag{58}$$

where the superscripts T and * denote the matrix transpose and complex conjugate operations, respectively. Enforcing conditions between susceptibilities also enforces conditions on the fields on both sides of the metasurface. Therefore, this approach restricts the diversity of electromagnetic transformations achievable with the metasurface.

A second approach provides a more general synthesis method for quasi-arbitrary electromagnetic transformations. This approach consists in selecting only one susceptibility tensor component in each of the four equations included in Eq. 56. The number of possible combinations is huge; thus, the analysis here is limited to solutions for a single transformation by eliminating a number of susceptibility components from Eq. 56.

DOI 10.1007/978-981-4560-75-7_8-1

Single-Transformation Metasurface

The problem of single transformation with only one specified wave triplet is considered: (Ψ^{i}, Ψ^{r}, Ψ^{t}) for a mono-anisotropic $\left(\overline{\overline{\chi}}_{\mathrm{em}} \equiv \overline{\overline{\chi}}_{\mathrm{me}} = 0\right)$ and uniaxial $\left(\chi_{\mathrm{ee}}^{xy} \equiv \chi_{\mathrm{ee}}^{yx} \equiv \chi_{\mathrm{mm}}^{xy} \equiv \chi_{\mathrm{mm}}^{yx} = 0\right)$ and, hence, non-gyrotropic and reciprocal, metasurface. Solving Eq. 56 under these conditions yields the following simple relations for the remaining four susceptibilities:

$$\begin{aligned}
\chi_{\mathrm{ee}}^{xx} &= \frac{-\Delta H_y}{j\omega\varepsilon\langle E_x\rangle}, \\
\chi_{\mathrm{ee}}^{yy} &= \frac{\Delta H_x}{j\omega\varepsilon\langle E_y\rangle}, \\
\chi_{\mathrm{mm}}^{xx} &= \frac{\Delta E_y}{j\omega\mu\langle H_x\rangle}, \\
\chi_{\mathrm{mm}}^{yy} &= \frac{-\Delta E_x}{j\omega\mu\langle H_y\rangle},
\end{aligned} \tag{59}$$

where, according to Eqs. 52 and 55, $\Delta H_y = H_y^t - \left(H_y^i + H_y^{\mathrm{r}}\right)$, $\langle E_x\rangle = \frac{E_x^{\mathrm{t}} + \left(E_x^{\mathrm{i}} + E_x^{\mathrm{r}}\right)}{2}$ and so on.

By synthesis, a metasurface with the susceptibilities given by Eq. 59 will produce exactly the specified reflected and transmitted transverse components of the fields when the metasurface is illuminated by the specified incident field. Since the longitudinal fields are completely determined from the transverse components, according to the uniqueness theorem, the complete specified electromagnetic fields are exactly generated by the metasurface.

Consistency with Maxwell equations can be easily verified. Consider, for instance, Eq. 51c along with the relation $D = \varepsilon E + P$,

$$\begin{aligned}
D_z\big|_{z=0^-}^{0^+} &= \varepsilon E_z\big|_{z=0^-}^{0^+} + P_z = \varepsilon\Delta E_z + P_z \\
&= -\nabla\cdot P_\perp .
\end{aligned} \tag{60}$$

Substituting in this relation the relations Eq. 54 for $P_\perp$ and remembering the assumption that $P_z = 0$,

$$\Delta E_z = -\frac{\partial}{\partial x}\left(\chi_{\mathrm{ee}}^{xx}\langle E_x\rangle\right) - \frac{\partial}{\partial y}\left(\chi_{\mathrm{ee}}^{yy}\langle E_y\rangle\right), \tag{61}$$

which upon substitution of Eq. 59 reads

$$\begin{aligned}
\Delta E_z &= E_z^{\mathrm{t}} - E_z^{\mathrm{i}} - E_z^{\mathrm{r}} \\
&= \frac{j}{\omega\varepsilon}\left[\frac{\partial}{\partial y}\left(H_x^{\mathrm{t}} - H_x^{\mathrm{i}} - H_x^{\mathrm{r}}\right) - \frac{\partial}{\partial x}\left(H_y^{\mathrm{t}} - H_y^{\mathrm{i}} - H_y^{\mathrm{r}}\right)\right].
\end{aligned} \tag{62}$$

This equation represents a relation between linear combinations of the longitudinal electric fields and derivatives of the transverse magnetic fields of the incident (k = i), reflected (k = r), and transmitted (k = t) waves. From linearity, and subsequent superposition, these equations may be decomposed as

$$E_z^{\mathrm{k}} = \frac{j}{\omega\varepsilon}\left(\frac{\partial H_x^{\mathrm{k}}}{\partial y} - \frac{\partial H_y^{\mathrm{k}}}{\partial x}\right) = \frac{j}{\omega\varepsilon}\left(\nabla_t \times H^{\mathrm{k}}\right)_z, \tag{63}$$

which is nothing but the projection of Maxwell-Ampère equation upon the z-direction. This shows that the longitudinal fields are well defined with the relations Eq. 59 in accordance with the uniqueness theorem.

To establish the relationship linking the transmitted field to the incident field and the susceptibilities, the case of a reflectionless metasurface is considered. Inserting Eqs. 52 and 55 with $\Psi_u^{\mathrm{r}} = 0$ ($u = x, y$) into Eq. 59 and solving for the transmitted components of the fields yields

$$E_x^{\mathrm{t}} = -E_x^{\mathrm{i}} + \frac{8E_x^{\mathrm{i}} - j4\chi_{\mathrm{mm}}^{yy}\mu\omega H_y^{\mathrm{i}}}{4 + \chi_{\mathrm{ee}}^{xx}\chi_{\mathrm{mm}}^{yy}\varepsilon\mu\omega^2}, \tag{64a}$$

$$E_y^{\mathrm{t}} = -E_y^{\mathrm{i}} + \frac{8E_y^{\mathrm{i}} - j4\chi_{\mathrm{mm}}^{xx}\mu\omega H_x^{\mathrm{i}}}{4 + \chi_{\mathrm{mm}}^{xx}\chi_{\mathrm{ee}}^{yy}\varepsilon\mu\omega^2}, \tag{64b}$$

$$H_x^{\mathrm{t}} = -H_x^{\mathrm{i}} + \frac{8H_x^{\mathrm{i}} + j4\chi_{\mathrm{ee}}^{yy}\varepsilon\omega E_y^{\mathrm{i}}}{4 + \chi_{\mathrm{mm}}^{xx}\chi_{\mathrm{ee}}^{yy}\varepsilon\mu\omega^2}, \tag{64c}$$

$$H_y^{\mathrm{t}} = -H_y^{\mathrm{i}} + \frac{8H_y^{\mathrm{i}} - j4\chi_{\mathrm{ee}}^{yy}\varepsilon\omega E_x^{\mathrm{i}}}{4 + \chi_{\mathrm{ee}}^{xx}\chi_{\mathrm{mm}}^{yy}\varepsilon\mu\omega^2}. \tag{64d}$$

These relations show how each of the transmitted field components depends on their incident field counterparts and orthogonal duals, e.g., $E_x^{\mathrm{t}} = E_x^{\mathrm{t}}\left(E_x^{\mathrm{i}}, H_y^{\mathrm{i}}\right)$, etc. They have to be considered after the susceptibilities Eq. 59 have been synthesized for given specifications to determine whether they may be realized by a passive metasurface ($|E^{\mathrm{t}}| \leq |E^{\mathrm{i}}|$ and $|H^{\mathrm{t}}| \leq |H^{\mathrm{i}}|$) or require active elements.

The susceptibilities in Eq. 59 represent the synthesis (inverse problem) results of the proposed method, while Eq. 64 expresses the transmitted field components in terms of these susceptibilities (direct problem). Next, the relationships existing between the susceptibilities and the scattering parameters are established in order to enable the second step of the synthesis.

In the plane-wave approximation, which is naturally valid when the source of the incident wave is far enough from the metasurface, the response of each scattering element may be expressed in terms of its reflection and transmission coefficients (Pfeiffer and Grbic 2013; Asadchy and Fanyaev 2011; Asadchy et al. 2014). Since according to Eq. 64, the pairs (E_x^{t}, H_y^{t}) and (E_y^{t}, H_x^{t}) are proportional to their incident counterparts and orthogonal duals only, the problem splits into an x-polarized incident plane-wave problem and a y-polarized incident plane-wave problem, whose fields at normal incidence are, respectively, given by

$$E^{\mathrm{i}} = \hat{x}, \quad E^{\mathrm{r}} = R_x\hat{x}, \quad E^{\mathrm{t}} = T_x\hat{x}, \tag{65a}$$

$$H^{\mathrm{i}} = \frac{1}{\eta}\hat{y}, \quad H^{\mathrm{r}} = -\frac{R_x}{\eta}\hat{y}, \quad H^{\mathrm{t}} = \frac{T_x}{\eta}\hat{y}, \tag{65b}$$

and

$$E^{\mathrm{i}} = \hat{y}, \quad E^{\mathrm{r}} = R_y\hat{y}, \quad E^{\mathrm{t}} = T_y\hat{y}, \tag{66a}$$

DOI 10.1007/978-981-4560-75-7_8-1

$$H^{\mathrm{i}} = -\frac{1}{\eta}\hat{x}, \quad H^{\mathrm{r}} = \frac{R_y}{\eta}\hat{x}, \quad H^{\mathrm{t}} = \frac{T_y}{\eta}\hat{x}, \tag{66b}$$

where R_u and $T_u (u = x, y)$ represent reflection and transmission coefficients, respectively. The waves in Eqs. 65 and 66 are defined as rectilinear (i.e., they do not change direction at the metasurface) for consistency with periodic boundary conditions to be used in full-wave simulations for the second step of the synthesis. Inserting Eqs. 65 and 66 into Eq. 56 with the four nonzero susceptibilities given in Eq. 59 yields the transmission and reflection coefficients

$$T_x = \frac{4 + \chi_{\mathrm{ee}}^{xx}\chi_{\mathrm{mm}}^{yy}k^2}{\left(2 + jk\chi_{\mathrm{ee}}^{xx}\right)\left(2 + jk\chi_{\mathrm{mm}}^{yy}\right)}, \tag{67a}$$

$$R_x = \frac{2jk\left(\chi_{\mathrm{mm}}^{yy} - \chi_{\mathrm{ee}}^{xx}\right)}{\left(2 + jk\chi_{\mathrm{ee}}^{xx}\right)\left(2 + jk\chi_{\mathrm{mm}}^{yy}\right)} \tag{67b}$$

and

$$T_y = \frac{4 + \chi_{\mathrm{ee}}^{yy}\chi_{\mathrm{mm}}^{xx}k^2}{\left(2 + jk\chi_{\mathrm{ee}}^{yy}\right)\left(2 + jk\chi_{\mathrm{mm}}^{xx}\right)}, \tag{68a}$$

$$R_y = \frac{2jk\left(\chi_{\mathrm{mm}}^{xx} - \chi_{\mathrm{ee}}^{yy}\right)}{\left(2 + jk\chi_{\mathrm{ee}}^{yy}\right)\left(2 + jk\chi_{\mathrm{mm}}^{xx}\right)}, \tag{68b}$$

These relations may be used in the second step of the synthesis to determine the scattering parameters corresponding to the synthesized susceptibilities. Solving Eqs. 67 and 68 for the susceptibilities yields

$$\chi_{\mathrm{ee}}^{xx} = \frac{2j(T_x + R_x - 1)}{k(T_x + R_x + 1)}, \tag{69a}$$

$$\chi_{\mathrm{ee}}^{yy} = \frac{2j\left(T_y + R_y - 1\right)}{k\left(T_y + R_y + 1\right)}, \tag{69b}$$

$$\chi_{\mathrm{mm}}^{xx} = \frac{2j\left(T_y - R_y - 1\right)}{k\left(T_y - R_y + 1\right)}, \tag{69c}$$

$$\chi_{\mathrm{mm}}^{yy} = \frac{2j(T_x - R_x - 1)}{k(T_x - R_x + 1)}. \tag{69d}$$

In Eq. 68 and 69, the reflection and transmission coefficients are associated with scattering parameters S_{pq} with $p, q = 1, \ldots, 4$ accounting for the two ports (incident and transmitted waves) and two polarization (x and y). Specifically, assigning ports 1, 2, 3, and 4 to x-polarized input, y-polarized input, x-polarized output, and y-polarized output, respectively, one has $R_x = S_{11}$, $T_x = S_{31}$, $R_y = S_{22}$, and $T_y = S_{42}$, while the other 12 scattering parameters are not required since the chosen tensors are uniaxial so that the metasurface is not gyrotropic (i.e., does not involve transformations between x-polarized and y-polarized waves).

DOI 10.1007/978-981-4560-75-7_8-1

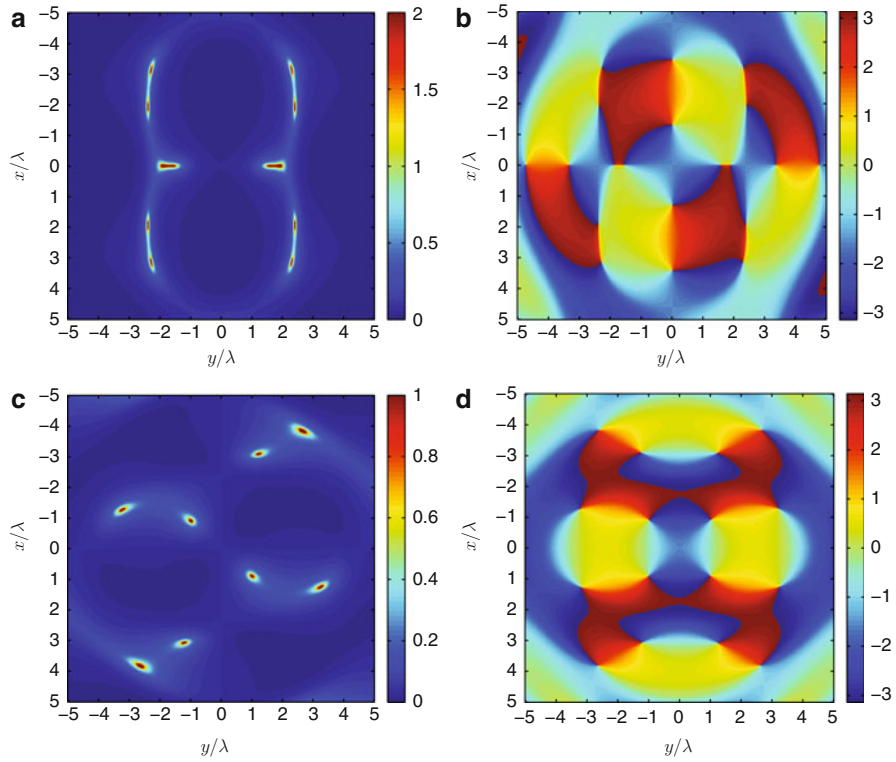


Fig. 24 Plot of the (**a**) magnitude and (**b**) phase of χ_{ee}^{xx} and (**c**) magnitude and (**d**) phase of χ_{ee}^{yy} for transformation of a normally incident plane wave into a normally transmitted Bessel of order $m = 3$ (Reprinted from Achouri et al. (2014) with permission)

Bessel Beam Transformation

The fundamental metasurface transformation relations are established, and one may thus use this synthesis technique to transform a normally incident plane wave into a normally transmitted Bessel beam. For diversity purpose, it is required to transmit a Bessel beam of order $m = 3$; such higher-order beams carry orbital angular momentum and may thus be used for various application, such as communication and particle manipulation.

Assuming the metasurface is located at $z = 0$, the incident plane-wave field reads

$$\begin{aligned} E^{\mathrm{i}}(x,y) &= (\hat{x}+\hat{y})\frac{\sqrt{2}}{2}, \\ H^{\mathrm{i}}(x,y) &= (\hat{y}-\hat{x})\frac{\sqrt{2}}{2\eta}, \end{aligned} \tag{70}$$

and the transmitted Bessel beam is assumed to be TE and is directly found from Eqs. 21 to 22 with $m = 3$. The amplitude of the Bessel beam is chosen such that the Poynting vectors $\left(S = \frac{1}{2}\mathrm{Re}\{E \times H^{*}\}\right)$ of the incident and transmitted fields are equal over the area of the metasurface (here, the surface is assumed to be $10\lambda \times 10\lambda$ in size), namely, $\int\int_S \mathrm{Re}\{E^{\mathrm{i}} \times H^{\mathrm{i}*}\}dS = \int\int_S \mathrm{Re}\{E^{\mathrm{t}} \times H^{\mathrm{t}*}\}dS$. This ensures that the power is conserved at the interface and, hence, that the metasurface is passive. The susceptibilities are found in a straightforward fashion by substituting the expressions of the incident and transmitted fields into Eq. 59 and χ_{ee}^{xx} and χ_{ee}^{yy} are plotted in Fig. 24. One notes here that the x- and y-components of the Bessel beam electromagnetic field are nonzero. Therefore, the incident fields must have electric and magnetic

DOI 10.1007/978-981-4560-75-7_8-1

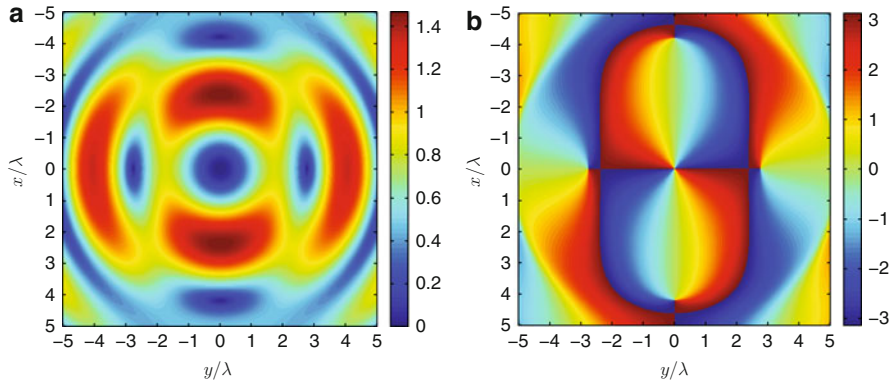


Fig. 25 Plot of the (**a**) magnitude and (**b**) phase of T_x for the Bessel beam transformation synthesis as depicted in Fig. 24 (Reprinted from Achouri et al. (2014) with permission)

components along both the x- and y-directions, as chosen in Eq. 70. In other words, the incident field must be properly polarized with respect to the prescribed transmitted wave.

Figure 25 plots the x-components of the transmission parameter. The y-components (not shown) are similar to the x-components and rotated by 90° with respect to them.

Summary of the Technique

A metasurface synthesis method based on transverse susceptibility tensors has been employed to transform an incident plane wave into a Bessel beam. The technique provides closed-form expressions for selected electric and magnetic susceptibility components required for the transformation. This synthesis method is essentially theoretical at this stage, and the physical scattering elements that would correspond to the synthesized ideal susceptibilities might be, in some cases, practically difficult or even impossible to realize. However, even in relatively extreme cases, typically corresponding to fast susceptibility variations in comparison with the wavelength, the proposed synthesis might be used as an initial and insightful step of the complete synthesis. Further discussion on this synthesis method is found in Achouri et al. (2014).

Conclusion

The early work of Brittingham on "focus wave modes" has stimulated much interest since it was first published in 1983. The possibility of generating beams and pulses with unprecedented properties and propagation characteristics has inspired many fundamental and applied researches and resulted in a deeper understanding on the mathematical, physical, and engineering aspects of LWs.

In this chapter, light has been shed on part of the historical development of LWs. A unified framework has been established to understand the physics of LWs via the spectral structure of PIBs. Finally, some of the state-of-the-art engineering techniques that could be applied to generate LWs have been unveiled.

Since the fundamental properties of LWs are now well understood, research has moved from the mathematical concepts into real applications, especially in the areas where LWs have an advantage over other waveforms, such as communication, manipulation of microparticles, remote sensing, and biophotonics. The future use of such wave fields seems very promising though more research is needed to establish innovative and practical techniques to generate and manipulate them.

DOI 10.1007/978-981-4560-75-7_8-1

Cross-References

- ▶ Antennas in Microwave Wireless Power Transmission
- ▶ Frequency Selective Surfaces
- ▶ Loop Antennas
- ▶ Optimization Methods in Antenna Engineering

References

Achouri K, Salem MA, Caloz C (2014) General metasurface synthesis based on susceptibility tensors. arXiv:1408.0273 [physics.optics]

Arlt J, Hitomi T, Dholakia K (2000) Atom guiding along Laguerre-Gaussian and Bessel light beams. Appl Phys B 71(4):549–554

Arlt J, Garces-Chavez V, Sibbett W, Dholakia K (2001) Optical micromanipulation using a Bessel light beam. Opt Commun 197(46):239–245

Asadchy VS, Fanyaev IA (2011) Simulation of the electromagnetic properties of helices with optimal shape, which provides radiation of a circularly polarized wave. J Adv Res Phys 2(1):011107

Asadchy VS, Faniayeu IA, Ra'di Y, Tretyakov SA (2014) Determining polarizability tensors for an arbitrary small electromagnetic scatterer. arXiv:1401.4930 [physics.optics]

Balanis CA (2005) Antenna theory: analysis and design, 3rd edn. Wiley, Hoboken

Bandres MA, Gutiérrez-Vega JC, Chávez-Cerda S (2004) Parabolic nondiffracting optical wave fields. Opt Lett 29(1):44–46

Bateman H (1915) Electrical and optical wave motion on the basis of Maxwell's equations. Cambridge University, Cambridge, UK. Reprinted (Dover, New York, 1955)

Besiries IM, Shaarawi AM, Ziolkowski RW (1989) A bidirectional traveling plane wave representation of exact solutions of the scalar wave equation. J Math Phys 30(6):1254–1269

Brittingham JN (1983) Focus waves modes in homogeneous Maxwell's equations: transverse electric mode. J Appl Phys 54(3):1179–1189

Caloz C, Itoh T (2005) Electromagnetic metamaterials: transmission line theory and microwave applications. Wiley, Hoboken

Capolino F (2009) Theory and phenomena of metamaterials. CRC Press, Boca Raton

Chattrapiban N, Rogers EA, Cofield D, Hill WD III, Roy R (2003) Generation of nondiffracting Bessel beams by use of a spatial light modulator. Opt Lett 28(22):2183–2185

Cheng J, Lu J-Y (2006) Extended high-frame rate imaging method with limited-diffraction beams. IEEE Trans Ultrason Ferroelectr Freq Control 53(5):880–899

Collin RE (1990) Field theory of guided waves, 2nd edn. Wiley-IEEE Press, New York

Courant R, Hilbert D (1966) Methods of mathematical physics, vol 2. Wiley, New York, p 760

Cox AJ, Dibble DC (1992) Nondiffracting beam from a spatially filtered Fabry-Perot resonator. J Opt Soc Am A 9(2):282–286

Donnelly R, Ziolkowski RW (1993) Designing localized waves. Proc Royal Soc London A 440(1910):541–565

Durnin J (1987) Exact solutions for nondiffracting beams. I. the scalar theory. J Opt Soc Am A 4(4):651–654

Durnin J, Miceli JJ, Eberly JH (1987) Diffraction-free beams. Phys Rev Lett 58:1499–1501

Engheta N, Ziolkowski RW (2006) Metamaterials: physics and engineering explorations. Wiley, Hoboken

Erdélyi M, Horvàth ZL, Szabó G, Bor Z, Tittel FK, Cavallaro JR, Smayling MC (1997) Generation of diffraction-free beams for applications in optical microlithography. J Vac Sci Technol B 15(2):287–292

Ettorre M, Grbic A (2012) Generation of propagating Bessel beams using leaky-wave modes. IEEE Trans Antennas Propag 60(8):3605–3613

Ettorre M, Rudolph S, Grbic A (2012) Generation of propagating Bessel beams using leaky-wave modes: experimental validation. IEEE Trans Antennas Propag 60(6):2645–2653

Fan J, Parra E, Milchberg HM (2000) Resonant self-trapping and absorption of intense Bessel beams. Phys Rev Lett 84:3085–3088

Felsen LB, Marcuvitz N (1994) Radiation and scattering of waves. IEEE Press, Piscatawy

Garcés-Chávez V, McGloin D, Melville H, Sibbett W, Dholakia K (2002) Simultaneous micromanipulation in multiple planes using a self-reconstructing light beam. Nature 419:145–147

Gutiérrez-Vega JC, Iturbe-Castillo MD, Chávez-Cerda S (2000) Alternative formulation for invariant optical fields: Mathieu beams. Opt Lett 25(20):1493–1495

Hecht E (1998) Optics, 4th edn. Addison-Wesley, Reading

Herman RM, Wiggins TA (1991) Production and uses of diffraction less beams. J Opt Soc Am A 8(6):932–942

Hernandez JE, Ziolkowski RW, Parker SR (1992) Synthesis of the driving functions of an array for propagating localized wave energy. J Acoust Soc Am 92(1):550–562

Hernández-Figueroa HE, Zamboni-Rached M, Recami E (eds) (2008) Localized waves. Wiley, Hoboken

Hernández-Figueroa HE, Recami E, Zamboni-Rached M (eds) (2013) Non-diffracting waves. Wiley-VCH, Weinheim

Holloway C, Mohamed M, Kuester EF, Dienstfrey A (2005) Reflection and transmission properties of a metafilm: with an application to a controllable surface composed of resonant particles. IEEE Trans Electromagn Compat 47(4):853–865

Holloway C, Dienstfrey A, Kuester EF, O'Hara JF, Azad AK, Taylor AJ (2009) A discussion on the interpretation and characterization of metafilms/metasurfaces: the two-dimensional equivalent of metamaterials. Metamaterials 3(2):100–112

Holloway C, Kuester EF, Gordon J, O'Hara J, Booth J, Smith D (2012) An overview of the theory and applications of metasurfaces: the two-dimensional equivalents of metamaterials. IEEE Antennas Propag Mag 54(2):10–35

Idemen MM (2011) Discontinuities in the electromagnetic field. Wiley, Hoboken

Indebetouw G (1989) Nondiffracting optical fields: some remarks on their analysis and synthesis. J Opt Soc Am A 6(1):150–152

Kong JA (1986) Electromagnetic wave theory. Wiley, New York

Kuester EF, Mohamed M, Piket-May M, Holloway C (2003) Averaged transition conditions for electromagnetic fields at a metafilm. IEEE Trans Antennas Propag 51(10):2641–2651

Lemaître-Auger P, Abielmona S, Caloz C (2013) Generation of Bessel beams by two-dimensional antenna arrays using sub-sampled distributions. IEEE Trans Antennas Propag 61(4):1838–1849

Lindell IV (1994) Electromagnetic waves in chiral and bi-isotropic media. The Artech House Antenna Library. Artech House, Boston

López-Mariscal C, Gutiérrez-Vega JC, Chávez-Cerda S (2004) Production of high-order Bessel beams with a Mach-Zehnder interferometer. Appl Opt 43(26):5060–5063

Lu J-Y (1997) 2D and 3D high frame rate imaging with limited diffraction beams. IEEE Trans Ultrason Ferroelectr Freq Control 44(4):839–856

Lu J-Y, Greenleaf JF (1992a) Experimental verification of nondiffracting X waves. IEEE Trans Ultrason Ferroelectr Freq Control 39(3):441–446

Lu J-Y, Greenleaf JF (1992b) Nondiffracting X waves-exact solutions to free-space scalar wave equation and their finite aperture realizations. IEEE Trans Ultrason Ferroelectr Freq Control 39(1):19–31

Lu J-Y, Zou H, Greenleaf JF (1994) Biomedical ultrasound beam forming. Ultrasound Med Biol 20(5):403–428

MacDonald MP, Paterson L, Volke-Sepulveda K, Arlt J, Sibbett W, Dholakia K (2002) Creation and manipulation of three-dimensional optically trapped structures. Science 296(5570):1101–1103

Mazzinghi A, Balma M, Devona D, Guarnieri G, Mauriello G, Albani M, Freni A (2014) Large depth of field pseudo-Bessel beam generation with a RLSA antenna. IEEE Trans Antennas Propag 62(8):3911–3919

McGloin D, Garcés-Chávez V, Dholakia K (2003) Interfering Bessel beams for optical micromanipulation. Opt Lett 28(8):657–659

Morse PM, Feshbach H (1953) Methods of theoretical physics, vol 1. McGraw-Hill, New York

Moses HE, Prosser R (1986) Initial conditions, sources, and currents for prescribed time-dependent acoustic and electromagnetic fields in three dimensions, part I: the inverse initial value problem. Acoustic and electromagnetic "bullets," expanding waves, and imploding waves. IEEE Trans Antennas Propag 34(2):188–196

Moses HE, Prosser RT (1990) Acoustic and electromagnetic bullets: derivation of new exact solution of the acoustic and Maxwell's equations. J Appl Math 50(5):1325–1340

Mugnai D, Ranfagni A, Ruggeri R (2000) Observation of superluminal behaviors in wave propagation. Phys Rev Lett 84:4830–4833

Munk BA (2000) Frequency selective surfaces: theory and design. Wiley, New York

Niemi T, Karilainen A, Tretyakov SA (2013) Synthesis of polarization transformers. IEEE Trans Antennas Propag 61(6):3102–3111

Palma C, Cincotti G, Guattari G, Santarsiero M (1996) Imaging of generalized Bessel-gauss beams. J Mod Opt 43(11):2269–2277

Pfeiffer C, Grbic A (2013) Metamaterial Huygens' surfaces: tailoring wave fronts with reflectionless sheets. Phys Rev Lett 110:197401

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical recipes: the art of scientific computing. Cambridge University Press, Cambridge, UK

Ra'di Y, Asadchy V, Tretyakov S (2013) Total absorption of electromagnetic waves in ultimately thin layers. IEEE Trans Antennas Propag 61(9):4606–4614

Rhodes DP, Lancaster GPT, Livesey J, McGloin D, Arlt J, Dholakia K (2002) Guiding a cold atomic beam along a co-propagating and oblique hollow light guide. Opt Commun 214(16):247–254

Saari P, Reivelt K (1997) Evidence of *X*-shaped propagation-invariant localized light waves. Phys Rev Lett 79:4135–4138

Salem MA, Bağci H (2010) On the propagation of truncated localized waves in dispersive silica. Opt Express 18(25):25482–25493

Salem MA, Bağci H (2011) Energy flow characteristics of vector X-waves. Opt Express 19(9):8526–8532

Salem MA, Bağci H (2012a) Modulation of propagation-invariant localized waves for FSO communication systems. Opt Express 20(14):15126–15138

Salem MA, Bağci H (2012b) Reflection and transmission of normally incident full-vector X waves on planar interfaces. J Opt Soc Am A 29(1):139–152

Salem MA, Caloz C (2014) Manipulating light at distance by a metasurface using momentum transformation. Opt Express 22(12):14530–14543

Salem MA, Kamel AH, Niver E (2011) Microwave Bessel beams generation using guided modes. IEEE Trans Antennas Propag 59(6):2241–2247

Schelkunoff SA (1972) On teaching the undergraduate electromagnetic theory. IEEE Trans Educ 15(1):15–25

Sezginer A (1985) A general formulation of focus wave modes. J Appl Phys 57(3):678–683

Shaarawi AM, Besieris IM, Ziolkowski RW (1990) A novel approach to the synthesis of nondispersive wave packet solutions to the Klein-Gordon and Dirac equations. J Math Phys 31(10):2511–2519

Shi H, Zhang A, Zheng S, Li J, Jiang Y (2014) Dual-band polarization angle independent 90° polarization rotator using twisted electric-field-coupled resonators. Appl Phys Lett 104(3):034102

Stratton J (1941) Electromagnetic theory. McGraw-Hill, New York

Taflove A, Hagness SC (2005) Computational electrodynamics: the finite-difference time-domain method, 3rd edn. Artech House, Norwood

Vasara A, Turunen J, Friberg AT (1989) Realization of general non-diffracting beams with computer-generated holograms. J Opt Soc Am A 6(11):1748–1754

Wu TT (1985) Electromagnetic missiles. J Appl Phys 57(7):2370–2373

Wu TT, Lehmann H (1985) Spreading of electromagnetic pulses. J Appl Phys 58(5):2064–2065

Yu Y-Y, Lin D-Z, Huang L-S, Lee C-K (2009) Effect of subwavelength annular aperture diameter on the nondiffracting region of generated Bessel beams. Opt Express 17(4):2707–2713

Yu N, Genevet P, Kats MA, Aieta F, Tetienne J-P, Capasso F, Gaburro Z (2011) Light propagation with phase discontinuities: generalized laws of reflection and refraction. Science 334(6054):333–337

Zamboni-Rached M, Recami E, Hernndez-Figueroa H (2002) New localized superluminal solutions to the wave equations with finite total energies and arbitrary frequencies. Eur Phys J D 21(2):217–228

Zhang P, Phipps ME, Goodwin PM, Werner JH (2014) Confocal line scanning of a Bessel beam for fast 3D imaging. Opt Lett 39(12):3682–3685

Ziolkowski RW (1989) Localized transmission of electromagnetic energy. Phys Rev A 39:2005–2033

Ziolkowski RW (1991) Localized wave physics and engineering. Phys Rev A 44:3960–3984

Ziolkowski R (1992) Properties of electromagnetic beams generated by ultra-wide bandwidth pulse-driven arrays. IEEE Trans Antennas Propag 40(8):888–905

Ziolkowski RW, Besieris IM, Shaarawi AM (1993) Aperture realizations of exact solutions to homogeneous-wave equations. J Opt Soc Am A 10(1):75–87

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_12-1

Maxwell's Original Presentation of Electromagnetic Theory and Its Evolution

Tapan K. Sarkar[a]* and Magdalena Salazar-Palma[b]
[a]Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, USA
[b]Department of Signal Theory & Communications, Universidad Carlos III de Madrid, Leganés, Madrid, Spain

Abstract

One of the objectives of this presentation is to illustrate how Maxwell came to his mathematical constructs of the work done before him by Oersted, Ampère, Faraday, Gauss, and so on, into a concise and precise mathematical form. In addition, the chapter addresses two specific topics which are fundamental in engineering electromagnetic education: how did Maxwell reached the conclusion that light was electromagnetic in nature and thereby revolutionized the last-century physics and the concept of displacement current? Maxwell first published his famous equations 20 in number in the early 1860s, and yet they were not accepted by the scientific community and were not put in the proper form till the early 1880s. The question is why it took over 20 years for the scientific community to grasp Maxwell's ideas. One of the reasons why Maxwell's theory was so difficult to follow was due to the development of Maxwell's thought process through different times. This made Maxwell not to identify his physical pictures with reality. Maxwell felt free to discard one picture and adopt another as often as expediency or convenience demanded. Maxwell's theory over the years evolved in two different steps. That is the other subject matter of this chapter.

Maxwell never believed that light could be generated electromagnetically. In fact, he was always silent about electromagnetic waves and their generation and detection. It took almost 25 years before Hertz discovered electromagnetic waves, and his ingenious experiments confirmed Maxwell's theory. Maxwell's ideas and equations were expanded, modified, and made understandable by the efforts of Hertz, FitzGerald, Lodge, and Heaviside. The last three are being referred to as the "Maxwellians." The early history of electromagnetic waves, up to the death of Hertz in 1894, is briefly discussed. The work of Hertz and the Maxwellians is briefly reviewed in the context of electromagnetic waves. It is found that historical facts do not support the views proposed by some, in the past, that Hertz's epoch-making findings and contributions were "significantly influenced by the Maxwellians." Finally, Maxwell's macroscopic theory was transformed to the microscopic theory based on electrons as its discovery came 18 years after Maxwell's death.

Keywords

James Clerk Maxwell; Heinrich Hertz; Oliver Heaviside; George FitzGerald; Hans Christian Oersted; Oliver Lodge; André-Marie Ampère; Joseph Larmor; Henry Augustus Rowland; Michael Faraday; Hermann von Helmholtz; Carl Friedrich Gauss; History; Electromagnetism

*Email: tksarkar@syr.edu

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_12-1

Introduction

The name of James Clerk Maxwell is synonymous with electromagnetics and electromagnetic waves. In this chapter, an attempt is made to describe what actually Maxwell did to come to the conclusion that light was electromagnetic in nature which revolutionized last-century physics. The bottom line is how Maxwell did actually make the connection between electricity and optics and magnetics and optics, as the connection between electricity and magnetism was already known at that time due to the works of Michael Faraday. However, it is also imperative to point out that what Maxwell presented as his theory in the 1800s and what is presented today are totally different even though some of the equations have remained the same over this long period. Maxwell did not reach the conclusion that light was electromagnetic in nature from the four equations that is presented in the various books today. To start with, Maxwell did not even write the four equations that bear his name in current time, but rather 20 in number starting with the potentials and not from the sources of currents and charges that are used today. It is important to remember the discovery of electron was made much later. Secondly, Maxwell's conclusion that revolutionized modern physics that light was electromagnetic in nature was based primarily on dimensional analysis, and he arrived to that conclusion long before he wrote his 20 equations. In addition, Maxwell could not have solved those 20 equations as he did not have the proper boundary conditions. Also, even though Maxwell introduced the concept of displacement current, he gave it up as quickly and yet in modern days textbooks have retained that terminology even after 150 years! Hence, it is difficult to present the historical development in a way that flows smoothly and easy to understand. Another factor that makes this presentation on this historical aspect difficult to present is that the genius of Maxwell was that he would start on a subject, work for a few years, give a gap of few years, and then restart it from a different viewpoint than when he left off as his thinking in the meantime has progressed along a different line of thought and he had reached a different conclusion than what he had started with. He would take the development of the subject to a new height in that new phase restarting after a gap in time with the old theory. He did that thrice with electromagnetic theory, and at each phase his way of thinking was quite different. This makes it difficult to follow what Maxwell had in his mind unless the context and the time are specified. It is to be recognized that there was a difference of about 20 years between Maxwell first writing the 20 equations and their modifications by Heinrich Hertz and Oliver Heaviside, reducing them to the scalar and vector versions, respectively, of the four equations that we use today. Since Maxwell never identified his physical pictures with reality, his theory was very difficult and sometimes even impossible to follow. This is because Maxwell left himself free to discard one picture and adopt another as often as expediency or convenience demanded. He described his method in the following words as written in the preface of his *Treatise* (Maxwell 1881):

> In the application of mathematics to the calculation of electrical quantities, I shall endeavour in the first place to deduce the most general conclusions from the data at our disposal, and in the next place to apply the results to the simplest cases that can be chosen. I shall avoid, as much as I can, those questions which, though they have elicited the skill of mathematicians, have not enlarged our knowledge of science. If we adopt a physical hypothesis, we see the phenomenon only through a medium, and are liable to that blindness to facts and rashness in assumption which a partial explanation encourages. We must therefore discover some method of investigation which allows the mind at every step to lay hold of a clear physical conception, without being committed to any theory founded on physical science for which that conception is borrowed.

Hence, unless the discussion of a particular topic is linked with Maxwell's thought process in that particular phase of the development of his theory, it is difficult to comprehend and even follow Maxwell's line of thinking (Sarkar et al. 2006, 2009; Sengupta and Sarkar 2003). So, with this introduction, an effort is made to recall the history.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_12-1

Prologue

As the English physicist and mathematician Sir James Hopwood Jeans (1877–1946) Glazebrook (1896) pointed out:

> In his hands electricity first became a mathematically exact science and the same might be said of other larger parts of Physics. In whatever area he worked, he brought new innovation.

For example, as Sir John Ambrose Fleming wrote (Commemorative Volume 1931, p. 118):

> In electricity course, he gave us a new and powerful method of dealing with problems in networks and linear conductors. Kirchhoff's corollaries of Ohm's Law had provided a means only applicable in the case of simple problems in which one could foresee the direction of flow of current in each conductor. But that was not possible in complicated networks. Maxwell initiated a new method by considering the actual current in each wire to be the difference of two imaginary currents circulating in the same direction round each mesh of the network. In this way, the difficulty of foreseeing the direction of the real current was eliminated. The solution of the problem was then reduced to the solution of a set of linear equations and the current in any wire could be expressed as the quotient of two determinants. After Maxwell's death, in 1885 I communicated a paper to the Physical Society of London in which the method was extended so as to give an expression for the electrical resistance of any network between any two points.

One would immediately recognize this as a method for writing the loop equations that is currently available in all undergraduate electrical engineering textbooks dealing with electrical circuits, and yet no mention is made of Maxwell, the inventor of this technique! The point here is that Kirchhoff no doubt wrote the theorems on the loop and the node equations but did not provide a methodology on how to solve for them in an arbitrary circuit. That job was completed by Maxwell.

In 1871, Maxwell showed how a circuit containing both capacitance and inductance would respond when connected to generators containing alternating currents of different frequencies. Maxwell thus developed the phenomenon of electrical resonance in parallel to acoustic resonance first presented by Sir John William Strutt, Lord Rayleigh, and Maxwell also suggested naming Rayleigh's book on acoustics *The Theory of Sound*. Maxwell developed the electrical analog for the phenomenon of resonance when reviewing Rayleigh's paper *On Progressive Waves* (Harman 2002c, pp. 547–549) and wrote about it to Rayleigh (Harman 2002b, pp. 598–608). Maxwell provided a simpler mathematical expression for the wave velocity and group velocity than what was described by Rayleigh.

Maxwell developed a coherent set of units of measurement of electricity and magnetism in 1863. They were later adopted almost unchanged as the first internationally accepted system of units, which became known misleadingly as the Gaussian system, which is a combination of the electrostatic units and the electromagnetic units. Maxwell also introduced the M, L, and T notation in dimensional analysis in physics which is used currently, and yet nobody wonders who first thought about it. He also produced the first standard of electrical resistance in 1868. When creating his standard for electrical resistance, he wanted to design a governor to keep a coil spinning at a constant rate. He made the system stable by using the idea of negative feedback. He worked out the conditions of stability under various feedback arrangements in terms of the poles and zeros of the system. This was the first mathematical analysis of control systems (Maxwell 1881). This work did not get any attention till 1940, when gun control radars were in demand during the Second World War. After the war, the American mathematician Norbert Wiener (1894–1964) took things further and developed the science of cybernetics, based on his paper *On Governors* in 1868 (Sarkar et al. 2006, 2009).

With this background, it is now possible to discuss Maxwell's work on electromagnetic theory and outline his innovations at each phase of his work.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_12-1

First Phase of Maxwell's Work on Electromagnetics

During the first phase of his work on electromagnetic theory, Maxwell thought about the following framework while he was in between jobs and before accepting a position at King's College. During this initial phase, Maxwell wrote down the following aspects of his development of electromagnetic theory. He tried to translate the experiments carried out by Faraday to suitable mathematical forms. As Maxwell wrote (Maxwell 1881):

> Before I began the study of electricity I resolved to read no mathematics on the subject till I had first read through Faraday's Experimental Researches on Electricity. I was aware that there was supposed to be a difference between Faraday's way of conceiving phenomena and that of the mathematicians, so that neither he nor they were satisfied with each other's language. I had also the conviction that this discrepancy did not arise from either party being wrong. As I proceeded with the study of Faraday, I perceived that his method of conceiving the phenomena was also a mathematical one, though not exhibited in the conventional form of mathematical symbols. I also found that these methods were capable of being expressed in the ordinary mathematical forms, and thus compared with those of the professed mathematicians. For instance, Faraday, in his mind's eye, saw lines of force traversing all space where the mathematicians saw centres of force attracting at a distance; Faraday saw a medium where they saw nothing but distance. Faraday sought the seat of the phenomena in real actions going on in the medium, they were satisfied that they had found it in a power of action at a distance impressed on the electric fluids. When I had translated what I considered to be Faraday's ideas into a mathematical form, I found that in general the results of the two methods coincided, so that the same phenomena were accounted for, and the same laws of action deduced by both methods, but that Faraday's methods resembled those in which we begin with the whole and arrive at the parts by analysis, while the ordinary mathematical methods were founded on the principle of beginning with the parts and building up the whole by synthesis. I also found that several of the most fertile methods of research discovered by the mathematicians could be expressed much better in terms of ideas derived from Faraday than in their original form. The whole theory, for instance, of the potential, considered as a quantity which satisfies a certain partial differential equation, belongs essentially to the method which I have called that of Faraday. According to the other method, the potential, if it is to be considered at all, must be regarded as the result of a summation of the electrified particles divided each by its distance from a given point. Hence many of the mathematical discoveries of Laplace, Poisson, Green and Gauss find their proper place in this treatise, and their appropriate expression in terms of conceptions mainly derived from Faraday.
>
> I shall endeavour to place in as clear a light as I can the relations between the mathematical form of this theory and that of the fundamental science of Dynamics, in order that we may be in some degree prepared to determine the kind of dynamical phenomena among which we are to look for illustrations or explanations of the electromagnetic phenomena. There is also a considerable mass of mathematical memoirs which are of great importance in electrical science, but they lie concealed in the bulky Transactions of learned societies; they do not form a connected system; they are of very unequal merit, and they are for the most part beyond the comprehension of any but professed mathematicians. I have therefore thought that a treatise would be useful which should have for its principal object to take up the whole subject in a methodical manner, and which should also indicate how each part of the subject is brought within the reach of methods of verification by actual measurement.

At this point, it is interesting to note that neither Michael Faraday nor Oliver Heaviside went to college for a formal education! This implies that the major contributions in the development of various concepts in electromagnetics were rather intuitive and were not carried out by an individual engrained in the contemporary mathematical sciences, and perhaps that is why new revolutionary developments were possible by Maxwell as he tried to put all these intuitive concepts into a consistent mathematical framework.

In 1855, Maxwell published the first part of his paper *On Faraday's Lines of Force* (Harman 2002a, pp. 337–375, Niven 1890a, pp. 155–229), where he used fluid flow as an analogy for lines of force shaped in his exposition based on physical geometry. His objective was to find a physical analogy which would help the mind to grasp the results of previous investigations without being committed to any theory founded on the physical science from which that conception is borrowed so that it is neither drawn aside from the subject in the pursuit of analytical subtleties nor carried beyond the truth by a favorite hypothesis.

DOI 10.1007/978-981-4560-75-7_12-1

The laws of electricity were compared with the properties of an incompressible fluid, the motion of which was retarded by a force proportional to the velocity, and the fluid was supposed to possess no inertia. The geometrical relations between lines and surfaces were crucial elements in Maxwell's field theory of electricity and magnetism, where the use of vectors, integral theorems, and topology became central to his mathematical method (Harman 2002a, pp. 14). Here, Maxwell introduced the concept of image theory for efficient calculation of the fields. Maxwell provided the mathematical basis of the theory of electric images which was earlier developed by Sir William Thomson (better known as Lord Kelvin) to illustrate that equivalent fields can be obtained from them (Maxwell 1881, pp. 226–227). He applied his analogy to the theory of electrostatics, dielectrics, paramagnetism, and diamagnetism, to Michael Faraday's ideas of magnetic properties of crystalline materials, and to electric currents. The second part of his paper proposed a theory based on his distinction between electric and magnetic quantities (acting through surfaces) and intensities acting along lines (Harman 2002a, pp. 15 & 371). Maxwell, thus, proceeded to consider the phenomenon of electromagnetism and showed how the laws discovered by André-Marie Ampère lead to conclusions identical to those of Faraday.

In 1861 Maxwell published the first two parts (Niven 1890a, pp. 451–488) of his paper *On Physical Lines of Force* and parts 3 and 4 (Niven 1890b, pp. 489–513) the following year. In the first part, he addressed that in all phenomena involving attractions or repulsions, or forces depending on the relative position of the bodies, he determined the magnitude and direction of the force which would act on a given body, if placed in a given position. The second part dealt with the question: Is there any mechanical hypothesis as to the condition of the medium indicated by lines of force, by which the observed resultant forces may be accounted for? What is the mechanical cause of this difference of pressures produced by the lines of force?

In the third part of the paper, Maxwell himself says:

> I showed how the forces acting between magnets, electric currents, and matter capable of magnetic induction may be accounted for on the hypothesis of the magnetic field being occupied with innumerable vortices of revolving matter, their areas coinciding with the direction of the magnetic force at every point of the field. The centrifugal force of these vortices produces pressures distributed in such a way that the final effect is a force identical in direction and magnitude with that which we observe. In addition he described the mechanism by which these rotations may be made to coexist, and to be distributed according to the known laws of magnetic lines of force.
>
> I conceived the rotating matter to be substance of certain cells, divided from each other by cell-wells composed of particles which are very small compared with the cells, and that it is by the motions of these particles, and their tangential action on the substance in the cells, that the rotation is communicated from one cell to another. [This is illustrated in Fig. 1] I have not attempted to explain this tangential action, but it is necessary to suppose in order to account for the transmission of rotation from the exterior to the interior parts of each cell, that the substances in the cells possess electricity of figure, similar in kind, though different in degree, that observed in solid bodies. The undulatory theory of light requires us to admit this kind of electricity in a luminiferous medium in order to account for transverse vibrations. We need not then be surprised if the magneto-electric medium possess the same property.

In addition, he introduced

> the concept of displacement current in dielectrics force acting between two electric field bodies and the rate of propagation of transverse vibrations through the elastic medium of which the cells are composed, on the supposition that its elasticity is due entirely to forces acting between pairs of particles.

Maxwell extends his model by providing the medium with elastic properties. Because the vortices' rotation is no longer of interest, Maxwell speaks of "elastic cells" surrounded by the layer of particles which play the role of electricity. This is illustrated in Fig. 1 as drawn by Maxwell. When these particles are displaced from their equilibrium positions, they distort the cells and call into play a force arising from their elasticity, equal and opposite to that which urges the particles away from their equilibrium position. This is illustrated in Fig. 2. The state of the medium, in which the particles are displaced (and the cells

DOI 10.1007/978-981-4560-75-7_12-1

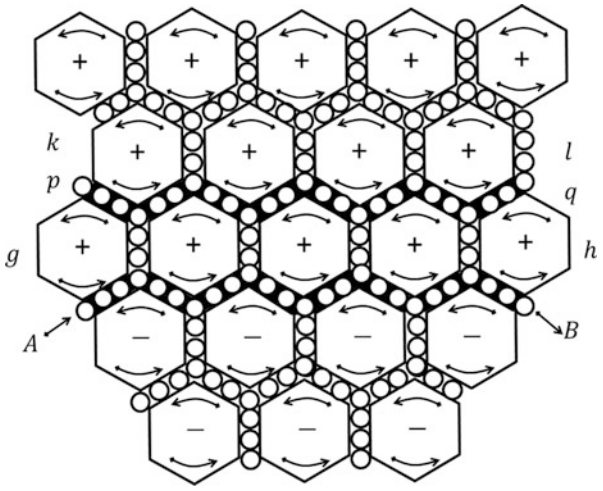


Fig. 1 Maxwell's drawing depicting his mechanical model of electromagnetic fields. Direction of rotation of the vortex tubes with hexagonal cross section is shown by *arrows*. The line of ball *A–B* moving to the right represents an electric current. Because of the rigidity of the structure, the motion of all vortices is synchronous

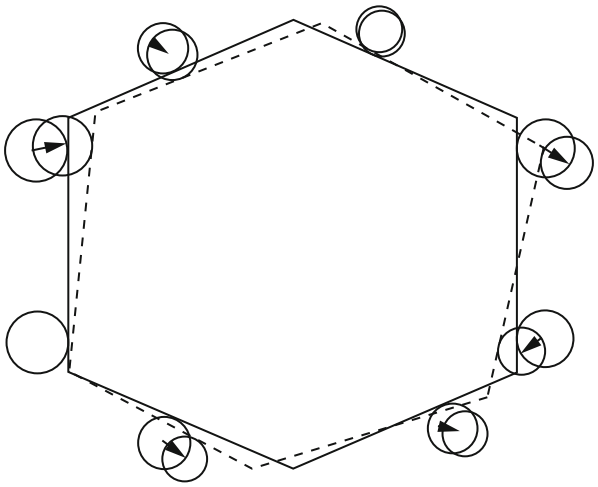


Fig. 2 Assuming vortex tubes elastic so that fixed balls could be displaced by electric force made Maxwell's model elastic and capable of carrying mechanical waves. Displacement of electricity corresponds to current creating a magnetic field

distorted) from their equilibrium position, is assumed to represent an electrostatic field. To Maxwell, this displacement current is like the real current as the vortices of Fig. 1 get distorted/displaced due to an externally applied field as shown in Fig. 2. Maxwell called the additional current generated due to the displacement of the fixed balls. But what happens *during* the displacement, the distortion of the magnetic vortices due to an applied field? In the light of his model, the answer is apparently obvious, and Maxwell makes a statement which will be of paramount relevance:

> This displacement does not amount to a current because when it has attained a certain value it remains constant, but it is a commencement of a current, and its variations constitute currents, in the positive or negative direction, according as the displacement is increasing or diminishing. The amount of the displacement depends on the nature of the body, and on the electromotive force.

Part 3 of his classic paper closes with the derivation of the relationship between the index of refraction of a substance and its relative electric and magnetic constants, $n^2 = \varepsilon_r \mu_r$ which helped him link light to

DOI 10.1007/978-981-4560-75-7_12-1

electricity. He made use of a bee's wax as a dielectric and measured its permittivity in a capacitor configuration. He then measured the refractive index by making a slab and light rays diffract through them. Thus, he established this equality both from a theoretical and experimental point of view. He mentioned in a letter to Faraday that the relationship between the refractive index and the dielectric constant could obviously provide a further check of the close connection between optical and electromagnetic phenomena.

In the fourth and last part of his classic paper (*The Theory of Molecular Vortices Applied to the Action of Magnetism on Polarized Light*), Maxwell applies his model to the analysis of the Faraday effect, and he succeeds in determining a quantitative law for the rotation of the polarization plane, which was in agreement with all known experimental results. The law involves also the mean radius of the molecular vortices, which motivates the abovementioned efforts to detect their existence in an independent, nonoptical way.

As Sir James Jeans points out:

> Maxwell pictured electromagnetic theory of light in terms of a medium whose properties could be specified completely in terms of a single mathematical constant. He saw that if the value of this constant could once be discovered, it ought to become possible to predict all phenomena of optical theory with complete mathematical precision. Maxwell showed that the constant in question ought to be merely the ratio of electromagnetic and electrostatic units of electricity and his first calculations suggested that this was in actual fact equal to the constant of the medium which measured the velocity of light.

The first mention of the great discovery comes in a letter, which he wrote to Michael Faraday on 19th October 1861:

> I suppose the elasticity of a sphere to react on the electrical matter surrounding it, and press it downwards. From the determination by Kohlrausch and Weber of the numerical relation between static and the magnetic effects of electricity, I have determined the elasticity of the medium in air, and assuming that it is the same with the luminiferous aether, I have determined the velocity of propagation of transverse vibrations. The result is 193,088 miles per second. Fizeau has determined the velocity of light as 193,118 miles per second by direct experiment.

Maxwell obtains the velocity for the wave propagation as 310,740,000 m/s. Furthermore, he states that:

> The velocity of transverse undulation in our hypothetical medium, calculated from the electromagnetic experiments of M. M. Kohlrausch and Weber, agrees so exactly with the velocity of light calculated from the optical instruments of M. Fizeau that we can scarcely avoid the inference that light consist in the transverse undulation of the same medium which is the cause of electric and magnetic phenomenon.

Even though the two numbers quoted above agreed to within 30 miles per second, oddly enough both are in error by more than 6,000 miles a second. When Maxwell came to publish his paper, *A Dynamical Theory of the Electromagnetic Field*, probably the most far-reaching paper he ever wrote, he gave the two velocities in terms of kilometers per second, Kohlrausch and Weber (310,740,000 m/s) and Fizeau (314,858,000 m/s), and nowhere near each other. Happily he seems to have realized that the velocity of light was not at all accurately known, and so he did not allow himself to be deterred, as Newton had been, by a substantial numerical disagreement in his law of universal gravitation.

Since the beginning, Maxwell was fully persuaded of the correctness of his hypothesis on the nature of light. This is clearly shown by the above reported letters, as well as by his interest in a precise determination of the "electromagnetic" value of the light velocity, as witnessed in the written letters to Henry Charles Fleeming Jenkin (27 August 1863) and Thomson (27 September and 15 October 1864), as well as the four parts of his classic paper published from September 1864 to February 1865.

Also in 1861, in setting up the standard for electrical resistance, he established the electrostatic units and the electromagnetic units to set up a coherent system of units and presented a thorough dimensional analysis of the electrical quantities. Thus, he put dimensional analysis on a scientific footing which was

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_12-1

626.] *Table of Dimensions.*

	Symbol	Dimensions in Electrostatic System	Dimensions in Electromagnetic System
Quantity of electricity	e	$[L^{\frac{3}{2}} M^{\frac{1}{2}} T^{-1}]$	$[L^{\frac{1}{2}} M^{\frac{1}{2}}]$.
Line-integral of electro-motive force	E	$[L^{\frac{1}{2}} M^{\frac{1}{2}} T^{-1}]$	$[L^{\frac{3}{2}} M^{\frac{1}{2}} T^{-2}]$.
Quantity of magnetism / Electrokinetic momentum of a circuit	m / p	$[L^{\frac{1}{2}} M^{\frac{1}{2}}]$	$[L^{\frac{3}{2}} M^{\frac{1}{2}} T^{-1}]$.
Electric current / Magnetic potential	C / Ω	$[L^{\frac{3}{2}} M^{\frac{1}{2}} T^{-2}]$	$[L^{\frac{1}{2}} M^{\frac{1}{2}} T^{-1}]$.
Electric displacement / Surface-density	$\mathfrak{D}$	$[L^{-\frac{1}{2}} M^{\frac{1}{2}} T^{-1}]$	$[L^{-\frac{3}{2}} M^{\frac{1}{2}}]$.
Electromotive force at a point	$\mathfrak{E}$	$[L^{-\frac{1}{2}} M^{\frac{1}{2}} T^{-1}]$	$[L^{\frac{1}{2}} M^{\frac{1}{2}} T^{-2}]$.
Magnetic induction	$\mathfrak{B}$	$[L^{-\frac{3}{2}} M^{\frac{1}{2}}]$	$[L^{-\frac{1}{2}} M^{\frac{1}{2}} T^{-1}]$.
Magnetic force	$\mathfrak{H}$	$[L^{\frac{1}{2}} M^{\frac{1}{2}} T^{-2}]$	$[L^{-\frac{1}{2}} M^{\frac{1}{2}} T^{-1}]$.
Strength of current at a point	$\mathfrak{C}$	$[L^{-\frac{1}{2}} M^{\frac{1}{2}} T^{-2}]$	$[L^{-\frac{3}{2}} M^{\frac{1}{2}} T^{-1}]$.
Vector potential	$\mathfrak{A}$	$[L^{-\frac{1}{2}} M^{\frac{1}{2}}]$	$[L^{\frac{1}{2}} M^{\frac{1}{2}} T^{-1}]$.

Fig. 3 List of dimensions for the various electrical quantities in both *ESU* and the *EMU* units. They all differ by the quantity LT^{-1}

discovered much earlier by Jean-Baptiste Fourier and others. The ESU and the EMU systems [Mahon 2003; Tolstoy 1981] of units were later mislabeled as the Gaussian system of units. In an 1863 memorandum, he introduced the dimensional notation which was to become the standard of using the powers of mass, length, and time. He also showed that the relation between the two electromagnetic units, ESU and EMU, has a dimension LT^{-1}, which has a value very close to that of the velocity of light (Harman 2002b, pp. 8). This is seen in Fig. 3. Later on, he also made an experiment to evaluate this number (Maxwell 1881) using purely electromagnetic principles.

This dimensional analysis prompted him to write the paper *A Dynamical Theory of the Electromagnetic Field* (Niven 1890a, pp. 526–597) in 1864 where he set out

> to establish the theory of the electromagnetic field because it has to do with the space in the neighborhood of the electric and magnetic bodies, and it may be called a Dynamical Theory, because it assumes that in that space there is matter in motion by which the electromagnetic phenomena are produced. In order to bring these results within the power of symbolical calculation, I then express them in the form of general equations of the electromagnetic fields (Maxwell 1881). The equations express –

A. The relation between the electric displacement, true conduction and the total current compounded of both. This equation expresses the total current. The variation of the electrical displacement must be added to the currents p, q, r to get the total motion of electricity which we may call p', q', r' to yield $p' = p + \frac{df}{dt}$; $q' = q + \frac{dg}{dt}$ and $r' = r + \frac{dh}{dt}$ where f, g, h denote the electric displacement along x, y, z.
B. The relation between the lines of magnetic force and the inductive coefficients of a circuit, as already deduced from the laws of induction. This equation represents the equation of magnetic force and is given by: $\mu\alpha = \frac{dH}{dy} - \frac{dG}{dz}$; $\mu\beta = \frac{dF}{dz} - \frac{dH}{dx}$; $\mu\gamma = \frac{dG}{dx} - \frac{dF}{dy}$; where α, β, γ are the magnetic force acting on a unit magnetic pole and F, G, H are the electromagnetic momentum.
C. The relation between the strength of a current and its magnetic effects, according to the electromagnetic system of measurements. This provides an equation for electric currents and expressed by $\frac{d\gamma}{dy} - \frac{d\beta}{dz} = 4\pi p'$; $\frac{d\alpha}{dz} - \frac{d\gamma}{dx} = 4\pi q'$; $\frac{d\beta}{dx} - \frac{d\alpha}{dy} = 4\pi r'$.

DOI 10.1007/978-981-4560-75-7_12-1

Table 1 Twenty variables originally introduced by Maxwell

Variable name used by Maxwell (*equivalent modern name*)	Symbol used by Maxwell	Modern equivalent Vector/scalar
Electromagnetic momentum (*magnetic vector potential*)	*F, G, H*	$\boldsymbol{A}$
Magnetic force (*magnetic field intensity*)	α, β, γ	$\boldsymbol{H}$
Electromotive force (*Electric field intensity*)	*P, Q, R*	$\boldsymbol{E}$
Current due to true conduction (*conduction current density*)	*p, q, r*	$\boldsymbol{J}$
Electric displacement (*electric flux density*)	*f, g, h*	$\boldsymbol{D}$
Total current (p^1, q^1, r^1) including variation of displacement (*conduction plus displacement current density*)	$\left\{ \begin{array}{l} p^1 = p + \frac{\mathrm{d}f}{\mathrm{dt}} \\ q^1 = q + \frac{\mathrm{d}g}{\mathrm{dt}} \\ r^1 = r + \frac{\mathrm{d}h}{\mathrm{dt}} \end{array} \right\}$	$\boldsymbol{J}_T$
Quantity of free electricity (*volume density of electric charge*)	*e*	ρ
Electric potential (*electric scalar potential*)	ψ	ψ

D. The value of the electromotive force in a body, as arising from the motion of the body in the field, the alteration of the field itself, and the variation of electric potential from one part of the field to another. This results in the equation of the electromotive force given by *P, Q, R* on a moving conductor as: $P = \mu\left(\gamma\frac{dy}{dt} - \beta\frac{dz}{dt}\right) - \frac{dF}{dt} - \frac{d\psi}{dx}$; $Q = \mu\left(\alpha\frac{dz}{dt} - \gamma\frac{dx}{dt}\right) - \frac{dG}{dt} - \frac{d\psi}{dy}$; $R = \mu\left(\beta\frac{dx}{dt} - \alpha\frac{dy}{dt}\right) - \frac{dH}{dt} - \frac{d\psi}{dz}$.

E. The relation between electric displacement, and the electromotive force which produces it. So when an electromotive force acts on a dielectric, it puts every part of the dielectric in a polarized condition, in which its opposite sides are oppositely electrified. If k is the ratio of the electromotive force to the electric displacement, then the equation for electric elasticity is given by $P = kf$; $Q = kg$; $R = kh$.

F. The relation between an electric current, and the electromotive force which produces it. In isotropic substances if ρ is the specific resistance referred to unit of volume, then the equation for electric resistance is given by $P = -\rho p$; $Q = -\rho q$; $R = -\rho\gamma$.

G. The relation between the amount of free electricity at any point, and the electric displacements in the neighborhood. If e is the quantity of free positive electricity contained in unit of volume at any point of the field, then the equation for free electricity is given by $e + \frac{df}{dx} + \frac{dg}{dy} + \frac{dh}{dz} = 0$.

H. The relation between the increase or diminution of free electricity and the electric currents in the neighborhood. This gives rise to the equation of continuity $\frac{de}{dt} + \frac{dp}{dx} + \frac{dq}{dy} + \frac{dr}{dz} = 0$.

Hence, there are 20 equations in all, involving 20 variable quantities summarized in Table 1 including the symbols used by Maxwell and the modern-day notations.

Maxwell also derived the coefficients of induction between two circuits in this paper. As he stated in the abstract of his paper:

> The proposed theory seeks for the origin of electromagnetic effects in the medium surrounding the electric or magnetic bodies, and assumes that they act on each other not immediately at a distance, but through the intervention of a medium which is progressive in nature.

It is also important to note that Maxwell believed in the ether theory as he states the following [Niven 1890a, b]:

DOI 10.1007/978-981-4560-75-7_12-1

The existence of the medium is assumed as probable, since the investigations of Optics have led philosophers to believe that in such a medium the propagation of light takes place.

He stated that the following quantities:

Electric currents by conduction, electric displacements, and Total currents; Magnetic forces, Electromotive forces, and Electromagnetic Momenta are involved in the mathematical expression for the electromagnetic fields.
Each of these quantities being directed quantity has three components; and beside these we have two others, the Free electricity and the Electric potential, making twenty quantity in all. There are twenty equations between these quantities, namely Equations of Total Currents, of Magnetic Force, of Electric Currents, of Electromotive Force, of Electric Elasticity, and of Electric Resistance, making six sets of three equations, together with one equation of Free Electricity, and another of Electric Continuity. The equations show that the transverse disturbances, and transverse disturbances only, will be propagated through the field, and that the number which expresses the number of electrostatic units of electricity in one electromagnetic unit, the standards of space and time being the same. The first of these results agrees with the undulatory theory of light as deduced from optical experiments. The second may be judged of by a comparison of the electromagnetical experiments of Wilhelm Eduard Weber and Rudolf Hermann Arndt Kohlrausch with the velocity of light as determined by astronomers in heavenly spaces, and by M. Foucault in the air of his laboratory.

Electrostatic units in an Electromagnetic Unit: 310,740,000 m/s
Velocity of light as found by Armand Hippolyte Louis Fizeau : 314,858,000 m/s
Velocity of light as found by Jean Bernard Leon Foucault : 298,000,000 m/s
Velocity of light deduced from aberration : 308,000,000 m/s

At the outset of the paper, the dynamical theory of the electromagnetic field borrowed from the undulatory theory of light the use of its luminiferous medium. It now restores the medium, after having tested its powers of transmitting undulations, and the character of those undulations, and certifies that the vibrations are transverse, and that the velocity is that of light. With regard to normal vibrations, the electromagnetic theory does not allow of their transmission. What then is light according to electromagnetic theory? It consists of alternate and opposite rapidly recurring transverse magnetic disturbances, accompanied with electric displacements, the direction of the electric displacement being at right angles to the magnetic disturbance, and both at right angles to the direction of the ray. The theory does not attempt to give a mechanical explanation of the nature of magnetic disturbance or of electric displacement, it only asserts the identity of these phenomena. . ... It discloses a relation between the inductive capacity of a dielectric and its index of refraction. .. The propagation of vibrations in a conducting medium is then considered and it is shown that the light is absorbed at a rate depending on the conducting power of the medium.

To answer now the specific question in more detail as to how did Maxwell reach the definite conclusion that light was electromagnetic in nature and at which point did it cease to be speculated, it is necessary to look at his letters. The first mention of the great discovery comes in a letter which he wrote to Faraday on the date 19 October 1861 [7, pp. 683–688, 5, p. 102]. For this part, it is relevant to refer to his book which was first published in 1873 and in a revised form in the second edition after his death in 1881 (Maxwell 1881). In (Maxwell 1881), Maxwell first presents the 20 equations as mentioned earlier. Even though he talks about the usage of the quaternion convention, the final expressions are in scalar form (Maxwell 1881, Ch. IX). In chapter X, he introduces the two systems of units, ESU and EMU, and illustrates that if the units of length, mass, and time are the same in the two systems, the number of electrostatic units of electricity contained in one electromagnetic unit is numerically equal to a certain velocity, the absolute value of which does not depend on the magnitude of the fundamental units employed [Maxwell 1881, Part II, p. 245]. In chapter XX, he shows that the disturbances in a media are propagated through a transverse electromagnetic wave which propagates at a velocity close to that of light. He furthermore compares the velocity of light measured by Fizeau (314,000,000 m/s) that measured through aberrations and the sun's parallax (308,000,000 m/s) and by Foucault (298,360,000 m/s) and that measured by Weber and Kohlrausch from the ratio between the ESU and the EMU units (310,740,000 m/s) and by Maxwell (288,000,000 m/s) and Lord Kelvin (282,000,000 m/s). From this comparison, Maxwell asserts that the velocity of light and the ratio of the two units are quantities of the same order of magnitude as also demonstrated from the previous quote.

DOI 10.1007/978-981-4560-75-7_12-1

Next, Maxwell observes in his theory that the refractive index of the media is related to the square root of the dielectric constant of a transparent media. He considers melted paraffin. The dielectric constant of paraffin had been measured by J. C. Gibson and T. Barclay to be 1.975 (Maxwell 1881, p. 398), and the refractive index of the same material had also been measured by Dr. J. H. Gladstone to be 1.43 [Maxwell 1881, p. 398]. But Maxwell confirms this by making a capacitor with a bee's wax as the dielectric to measure its dielectric constant and also fabricates a lens to measure the refractive index. In this way, he relates the electrical properties of a medium to that of optical properties of the medium. Finally, in chapter XXII, he explains the properties of ferromagnetism and diamagnetism by molecular currents and tries to explain the *Faraday effect* of the rotation of the polarization of the light passing through a magnetic medium, thus relating the magnetic properties to the optical properties. His Treatise (Maxwell 1881) thus illustrates that light is electromagnetic in nature, even though he lacked the information about the boundary conditions to solve the wave equation completely!

In 1868, he published the paper *On a Method of Making a Direct Comparison of Electrostatic with Electromagnetic Force; with a Note on the Electromagnetic Theory of Light* (Niven 1890b, pp. 125–143), where he described how he measured the speed of light by using 2,600 batteries to produce 3,000 V (Mahon 2003; Tolstoy 1981). The goal was to balance the electrostatic attraction between two charged metal plates against the magnetic repulsion between two current-carrying coils and build a balance arm to do this. He got a result of 288,000 km/s as compared to the current accepted value of 299,792.5 km/s.

He liked to use the quaternion convention. However, he wrote the final equations in the scalar form even though he used the terms "curl," "convergence," and "gradient." Nowadays, convergence is replaced by its negative, which is called divergence, and the other two are still in the standard mathematical literature. These are available in his paper *On the Mathematical Classification of Physical Quantities* (Niven 1890b, pp. 257–266). Maxwell identified light with electromagnetic waves and introduced the concept of ether as the basic medium of the electromagnetic field to retain the possibility of a mechanical interpretation. He presented his first paper on the new theory before the Royal Society in 1864 and published the comprehensive *A Treatise on Electricity and Magnetism* in 1873.

Maxwell considered magnetism as a phenomenon of rotation and electric currents as a phenomenon of translation. So in any magnetic field the medium is in rotation about the lines of magnetic force. Maxwell's theory sought unity through a highly plastic set of field equations coupled to Hamilton's principle, named after the Irish mathematician William Rowan Hamilton. In current bearing linear circuits, Maxwell thought that the currents were linked by rigid constraints to an intervening medium called the ether. Thus, the most difficult concepts for the modern reader to grasp in the original Maxwell's theory are the concepts of "charge" and "current." In modern theory, charge is the source of the electric field, and current is the source of the magnetic field. In his theory, charge is produced by the electric field; current, in the usual sense, is the rate of change of charge over time and is only indirectly related to the magnetic field. Therefore, in the Maxwellian theory, charge is a discontinuity of the displacement $\boldsymbol{D}$ and not in $\boldsymbol{E}$. Maxwell's goal was to create a theory of electromagnetism which made no use whatsoever of the microstructure of the matter. To Maxwell, the conduction current was effectively a continuous series of charging and discharging. The conduction current then is the process and growth of displacement. Maxwell proposed all currents are closed. In this fashion, he introduced the displacement current. Maxwell quite explicitly limited electric polarization to the boundary conditions on the flux characteristics of electric displacement and magnetic induction. Maxwell did not apply boundary conditions to explain the phenomenon of reflection and refraction. This was achieved by Hertz and Heaviside.

Maxwell introduced the term displacement current as he visualized that when an external field is applied to these molecular vortices, they are deformed or displaced and that also produces a current which he termed as the displacement current. To Maxwell, displacement current was the same as the real current and also flowed in the same direction (but in the modern theory used today, one uses either a conduction

DOI 10.1007/978-981-4560-75-7_12-1

current or a displacement current, and the flow of the displacement current is opposite to that of the conduction current).

In summary, in the first phase, Maxwell reached the conclusion that light was electromagnetic in nature from dimensional analysis and even constructed an experiment to measure the velocity of light by establishing how many ESU units are there in one EMU unit (Maxwell 1881).

Second Phase of Maxwell's Work on Electromagnetics

At this point, he resigned from King's College and went back to his home in the countryside. While he was there and before he accepted the Cavendish chair at Cambridge University, his thought process for the second phase began as during this stage he started writing his two-volume treatise covering about 1000 pages (Maxwell 1881). It is important to observe that Maxwell's thought process has radically changed during this phase when he started to write his treatise on electromagnetics. In this phase, Maxwell has abandoned his mechanical model of electromagnetic theory and starts the subject from an abstract point of view mainly using his equations. As he describes it:

> **I have confined myself almost entirely to the mathematical treatment of the subject, but I would recommend the student, after he has learned, experimentally if possible, what are the phenomena to be observed, to read carefully Faraday's Experimental Researches in Electricity**. He will there find a strictly contemporary historical account of some of the greatest electrical discoveries and investigations, carried on in an order and succession which could hardly have been improved if the results had been known from the first, and expressed in the language of a man who devoted much of his attention to the methods of accurately describing scientific operations and their results. It is of great advantage to the student of any subject to read the original memoirs on that subject, for science is always most completely assimilated when it is in the nascent state, and in the case of Faraday's Researches this is comparatively easy, as they are published in a separate form, and may be read consecutively. If by anything I have here written I may assist any student in understanding Faraday's modes of thought and expression, I shall regard it as the accomplishment of one of my principal aims to communicate to others the same delight which I have found myself in reading Faraday's Researches.

It is therefore quite relevant to point out that Maxwell by this time has discarded the mechanical model, along with it he also discarded the concept of displacement of current. In his thousand-page treatise, he never once mentions the term *displacement current*, whereas he confines himself to terms like currents in the dielectric. Perhaps it is high time that in the current textbooks the same thought process of Maxwell is followed with the concept of displacement current as it is not a current in a true sense as it does not produce a magnetic field.

In his treatise, Maxwell was able to show (Maxwell 1881, Art. 783 & 784; Equations 8 and 9) that in free space $\left(d^2\boldsymbol{J}\right)/(d\,t^2) + \left\{d\left(\nabla^2\Psi\right)\right\}/(d\,t) = 0$, where ψ is the scalar electric potential. He then made an important assertion, for which he provided no real justification: "$\nabla^2\Psi$," he said, *which is proportional to the volume density of the free electricity, is independent of t*, i.e., he claimed that the electric potential is determined solely by the spatial distribution of charge which in a nonconductor does not change. This is the assumption usually made in electrostatics, and Maxwell simply extended it to general electromagnetic theory without alteration or explanation. Time independence implied that the electric potential adjusted instantaneously across all space to any changes in the positions or magnitudes of the charges. It also implied $(d^2\boldsymbol{J})/(d\,t^2) = 0$, so that as Maxwell wrote:

> $\boldsymbol{J}$ must be a linear function of t, or a constant, or zero and we may therefore leave $\boldsymbol{J}$ and ψ out of account in considering wave disturbances.

In practice, Maxwell generally took $\boldsymbol{J} = 0$ and so worked out what we now call the *Coulomb gauge*, a gauge well suited to electrostatic problems but with the serious drawback in treating changing fields that it

DOI 10.1007/978-981-4560-75-7_12-1

requires the electric potential to be propagated instantaneously (Maxwell 1881). FitzGerald differed from Maxwell on this point, and instead of assuming the two potentials to be independent as Maxwell did, FitzGerald put $\boldsymbol{J} = -(d\Psi)/(dt)$ or equivalently $\nabla \bullet \boldsymbol{A} + (d\Psi)/(dt) = 0$. This *Lorenz gauge*, as it was later called, is much better suited to treating propagation phenomenon than was Maxwell's *Coulomb gauge* with $\boldsymbol{J} = 0$. This new gauge thus eliminated the question of the instantaneous propagation of the electric potential. However, FitzGerald found that Heaviside had independently done it already (Mahon 2003; Tolstoy 1981; Hunt 1991; Buchwald 1985)! As FitzGerald and Rowland put it in 1888, *That ψ should be murdered from treating propagation problems* (Heaviside 1893). Soon after, the same fate happened to $\boldsymbol{A}$, as Heaviside puts it:

> not merely the murder of Maxwell's ψ, but of that wonderful three legged monster with a scalar parasite on its back, the so called electrokinetic momentum at a point – that is the vector potential itself. (Heaviside 1893)

That is the first complete modification of Maxwell's theory done by Heaviside (1893) and Hertz [1962] to get rid of the potentials and to start the problem with the sources, i.e., currents and charges.

Even though the final results of the original Maxwell's theory are valid even today, the intermediary steps used to arrive at the conclusion were in question (Mahon 2003; Buchwald 1985). The first problem was associated with the definition of the charge.

Third Phase of Maxwell's Work on Electromagnetics

Maxwell, along with the English biologist Thomas Henry Huxley, was the joint scientific editor of the ninth edition of *Encyclopedia Britannica* (1879). There he provided an account of the motion of earth through ether. He had done experiments in 1860 which failed to detect any effect; however, he strongly believed in the existence of ether. At that time, his paper was rejected by Lord Kelvin, who was the reviewer, and so Maxwell put it in the *Encyclopedia Britannica*! Maxwell suggested that ether could perhaps be detected by measuring the velocity of light when light was propagated in opposite directions. He had further discussions in a letter to David Peck Todd, an astronomer at Yale. Maxwell's suggestion of a double-tracked arrangement led A. A. Mickelson, when he was working under Helmholtz as a student, to read Maxwell's work and then later to undertake his famous experiments on ether drag in the 1880s, and the rest is history.

The Evolution of the Original Maxwell's Theory to the Modern Times

After Maxwell's books on the Treatise were published in the European Continent, Helmholtz had been trying to understand Maxwell's theory of electromagnetism and to compare it with a theory based mostly on Newtonian mechanics (Bryant 1988). In 1879 Helmholtz called for an experimental validation of Maxwell's theory and had it published as a prize problem of the Prussian Academy of Science, often referred to as the *Berlin Prize*. For the 1882 prize, the problem stated (Hertz 1962; Bryant 1988):

> The theory of electrodynamics which was brought forth by Faraday and was mathematically executed by Mr. Cl. Maxwell presupposed that the formation and disappearance of the dielectric polarization in insulating media – as well as in space – is a process that has the same electrodynamic effects as an electrical current and that this process, just like a current, can be excited by electrodynamically induced forces. According to that theory, the intensity of the mentioned currents would have to be assumed equal to the intensity of the current that charges the contact surfaces of the conductor. The Academy demands that decisive experimental proof be supplied either
>
> for or against the existence of electrodynamic effects of forming or disappearing dielectric polarization in the intensity as assumed by Maxwell or

for or against the excitation of dielectric polarization in insulating media by magnetically or electrodynamically induced electromotive forces.

Answers to these questions had to be submitted by 1 March 1882, and the prize of 955 marks will be awarded. At that time, Hertz was working with Helmholtz at the Physical Institute in Berlin. Helmholtz thought that one of his students, Heinrich Hertz, would be most likely to succeed in this experimentation. He suggested to Hertz that, should he address this problem, the resources of the institute will be available to him. However, at that time, Hertz gave up the idea because he thought a solution was not possible, as he found no adequate sources for generation of high frequencies. However, he continued thinking about this problem. Hertz did an analytical thesis on the induced currents in a rotating metal sphere in a magnetic field (Bryant 1988; Hunt 1991). After graduating in 1880, he stayed on as an assistant to Helmholtz and then went to the University of Kiel as an instructor of theoretical physics. At Kiel, Hertz had no laboratory and was very impatient working only in theoretical physics. In 1884 at Kiel, he published the paper *On the relation between Maxwell's fundamental electromagnetic equations and the fundamental equations of the opposing electromagnetics* (Hertz 1884). He concluded that if he had to make a choice, he would choose Maxwell's theory. This work convinced him to carry out experimental work to verify this theory. In his book, he himself addresses the question as to what exactly is Maxwell's theory. In his words (Hertz 1962; Buchwald 1985, p.191):

Maxwell left us as the result of his mature thought a great treatise on Electricity and Magnetism; it might therefore be said that Maxwell's theory is the one propounded in that work. But such an answer will scarcely be regarded as satisfactory by all scientific men who have considered the question closely. Many a man has thrown himself with zeal into the study of Maxwell's work, and even when he has not stumbled upon unwanted mathematical difficulties, has never the less been compelled to abandon the hope of forming for himself an altogether consistent view of Maxwell's ideas. I have fared no better myself. Notwithstanding the greatest admiration for Maxwell's mathematical conceptions, I have not always felt quite certain of having grasped the physical significance of the statements. Hence, it was not possible for me to be guided directly by Maxwell's book. I have rather been guided by Helmholtz's work, as indeed may plainly be seen from the manner in which the experiments are set forth. But unfortunately, in the special limiting case of Helmholtz's theory which lead to Maxwell's equations, and to which the experiments pointed, the physical basis of Helmholtz's theory disappears, and indeed it does, as soon as action-at-a-distance is disregarded. **I therefore endeavored to form for myself in a consistent manner the necessary physical conceptions, starting from Maxwell's equations, but otherwise simplifying Maxwell's theory as far as possible by eliminating or simply leaving out of consideration those portions which could be dispensed within as much as they could not affect any possible phenomena**... To the question, "What is Maxwell's Theory?" I know of no shorter or more definite answer than the following: – Maxwell's theory is Maxwell's system of equations. Every theory which leads to the same system of equations, and therefore comprises the same possible phenomena, I would consider as being a form or special case of Maxwell's theory; every theory which leads to different equations, and therefore to different possible phenomena, is a different theory. Hence, in this sense, and in this sense only, may the two theoretical dissertations in the present volume be regarded as representations of Maxwell's theory? **In no sense can they claim to be precise rendering of Maxwell's ideas. On the contrary, it is doubtful whether Maxwell, were he alive, would acknowledge them as representing his own views in all respects**.

The 1884 paper (Hertz 1884), in which Hertz developed the coordinate form of Maxwell's equations discarding the potentials and getting the four equations that is used today in scalar form, 12 in number, helped Hertz to get his next appointment at the Technical High School at Karlsruhe in 1885, where he started his experimental work. There, he discovered how to generate and detect electromagnetic energy, in trying to address the problem of spurious electromagnetic radiation from high-power transformers during closing and opening of circuits (Tai and Bryant 1994; D'Agostino 1975). Finally, in 1887, he provided experimental results for the Berlin Prize, and it was published in his 1888 paper. It is not known whether Hertz collected the Berlin Prize money since the due date was over! He also made experiments to determine the velocity of electromagnetic wave propagation using a wire over a ground plane. With the transmission line open-ended at one end, he moved a detector along the lines and observed the distance between nulls is one half-wavelength and had a value of 2.8 m (Bryant 1988; Tai and Bryant 1994;

DOI 10.1007/978-981-4560-75-7_12-1

D'Agostino 1975). Hertz calculated the frequency of the source from the estimate of the capacitance and the inductance of the dipole resonator. For the velocity, the product of *wavelength* × *frequency*, he arrived at a value of 2×10^8 m/s not knowing of a computational error of $1/\sqrt{2}$ for the frequency. Without this error, his result would have been 2.8×10^8 m/s. After he published his results, this error was pointed out to him in a letter from H. Poincaré in 1890 (Bryant 1988; Tai and Bryant 1994; D'Agostino 1975).

However, what Hertz missed was the core idea of the discontinuity in the displacement (Buchwald 1985, p.193). The quandary of Hertz forced him into an uneasy compromise with the traditional Helmholtzian view on charge. Hertz distinguished the free electricity from which one calculates forces, and which is alterable by nonconducting means, from the true electricity, which is alterable only by conduction. So, though Hertz referred the measure of true charge to the divergence of the displacement, he preserved Helmholtzian wording because he had not seen how to avoid it. Whereas a Maxwellian would write apparent charge as $\nabla \bullet \boldsymbol{E}$, Hertz wrote of free charge and felt it necessary to retain the idea of bound charge to grant free charge physical significance, though he refused to consider why such a thing as bound charge exists. The significance of Hertz's famed rejection of the Maxwellian distinction between electric intensity and displacement in the free ether is now examined. Without this distinction, it is impossible, in Maxwellian theory, to understand the existence of a charged surface in vacuum because charge is due to the discontinuity in the displacement. The fact that in free ether $\boldsymbol{D}$ reduces to $\boldsymbol{E}$ is merely a mathematical artifact. This is due to the definition of capacity of the ether being unity. The conceptual and physical distinction between displacement and intensity is still essential. Not knowing or understanding this distinction goes to the heart of Maxwell's theory. Hertz felt free to ignore it where it seemed mathematically to make no difference. In addition, Hertz started from the sources of the fields which were charges in electrostatics and currents in magnetostatics and not treat the potentials as fundamental quantities in analogy to a mechanical model as Maxwell did. Hertz showed that at a dielectric boundary the tangential components of the electric fields are continuous and the normal component is discontinuous. He obtained similar boundary conditions for a magnetic media.

In short (Buchwald 1985), Hertz defined the electric and magnetic constants as unity in free space, thus eliminating displacement as a primary quantity. While this somewhat simplified the mathematical structure of Maxwell's theory, it was fatal to its dynamical basis. Heaviside raised this point with Hertz in 1890 after reading his first paper:

> On the Fundamental Equations of Electrodynamics. He asked, Can you conceive of a medium for electromagnetic disturbances which has not at least two physical constants, analogous to density and elasticity? If not, is it not well to explicitly symbolize them, leaving to the future their true interpretation?

Heaviside was calling on Hertz not to sacrifice for the sake of a small and perhaps illusory mathematical simplification, the dynamical ether, filled with stresses, strains, and stored energy on which Maxwell had built his theory. The British Maxwellians took this dynamical ether much more seriously than did their Continental counterparts, insisting that even on those points where Maxwell's theory required clarification and correction, this could and should be done without reducing the theory to a mere set of equations. The important modifications Heaviside had made to Maxwell's theory, including his abandonment of the potentials in favor of his four equations in the vector form that is termed Maxwell's equations today, he said, meant as sustentative changes or new departures, but were directed solely at bringing out the leading points of the theory more clearly than had Maxwell himself. Both Heaviside and FitzGerald drew a careful distinction between *Maxwell's treatise* and *Maxwell's theory* and said that Maxwell's book gave only an imperfect account of the real nature of the theory (Buchwald 1985).

In 1886, Hertz observed while examining some of the apparatus used in lecture demonstrations that the oscillatory discharge of a Leyden jar or induction coil through a wire loop caused sparks to jump a gap in a similar loop a short distance away. He recognized this as a resonance phenomenon and saw that such

DOI 10.1007/978-981-4560-75-7_12-1

sparking loops could serve as very sensitive detectors of oscillating currents and, thus, of electromagnetic waves. This provided him with the proper experimental tool which had eluded Lodge, FitzGerald, and others. His two papers:

> On the finite velocity of propagation of electromagnetic actions and The forces of electric oscillations, treated according to Maxwell's theory

both published in 1888 vindicated Maxwell's theory, and from then on it was accepted by the electromagnetic community (Tai and Bryant 1994; D'Agostino 1975).

Heaviside, on the other hand, differentiated between the absolute and the relative permeability and permittivity, defining the relative quantities as the ratio of the absolute value for a medium and that for free space. Heaviside rewrote Maxwell's equations in the vector form that is used today using the above modifications. There are conductors and nonconductors or insulators, and since the finite speed of propagation in the nonconducting space outside conductors was unknown, attention was almost entirely concentrated upon the conductors and an assumed field which was supposed to reside upon or in them and to move about, upon, or through them. And the influence on distant conductors was attributed to instantaneous action at a distance, ignoring an intermediate agent. Maxwell explained these actions of the intermediate agent of an intervening medium being transmitted at finite speed. In Heaviside's words (1893):

> What is Maxwell's Theory or what should we agree to understand by Maxwell's theory? The first approximation to the answer is to say there is Maxwell's book as he wrote it; there is his text and there are his equations, together they make his theory. But when we come to examine it closely, we find that this answer is unsatisfactory. To begin with, it is sufficient to refer to papers by physicists, written say during the twelve years following the first publication of Maxwell's Treatise, to see that there may be much difference of opinion as what his theory is. It may be, and has been, differently interpreted by different men, which is a sign that it is not set forth in a perfectly clear and unmistakable form. There are many obscurities and inconsistencies. Speaking for myself, it was only by changing its form of presentation that I was able to see it clearly, and so as to avoid inconsistencies. **It is therefore impossible to adhere strictly to Maxwell's theory as he gave it to the world, if only on account of its inconvenient form**. . . **But it is clearly not admissible to make arbitrary changes in it and still call it his. He might have repudiated them utterly. But if we have good reason to believe that the theory as stated in his treatise does require modification to make it self-consistent and to believe that he would have admitted the necessity of the change when pointed out to him, then I think the resulting modified theory may well be called Maxwell's**.

To understand the problem, one must introduce the Maxwellian concept of a current. It was apparently not possible to incorporate conductivity into the dynamical structure of the Maxwellian theory (Buchwald 1985). Maxwell did not himself incorporate conductivity directly into the dynamical structure of the theory. Rather, he treated Ohm's law as an empirical, independent fact and subtracted the electromagnetic intensity it requires from the induced intensity.

It is important to point out that Heaviside's duplex equations, as he called them, can be found in Maxwell's own writing in 1868, *Note on the Electromagnetic Theory of Light*. Maxwell gave a clear statement of the second circuital law relating the magnetic current to the curl of the electric force, crediting it to Faraday and asserting that it afforded "the simplest and most comprehensive" expression of the facts. But despite the advantages of this simple equation, particularly in treating electromagnetic waves, Maxwell did not use it at all in his *Treatise* in 1873, and it remained unknown even to his close students till 1890 when Maxwell's scientific papers (Niven 1890a, b) were published. It is rather amusing to note that even Lodge (one of the Maxwellians) actually credited this equation to Heaviside in his Presidential Address to the Physical Society in February 1899, when the Irish physicist and mathematician Sir Joseph Larmor pointed out to him that it was in the writing of Maxwell himself (Mahon 2003; Tolstoy 1981; Tai and Bryant 1994; D'Agostino 1975)!

DOI 10.1007/978-981-4560-75-7_12-1

In summary, why Maxwell's theory was not accepted for a long time by contemporary physicists was that there were some fundamental problems with his theory. Also, it is difficult to explain using modern terminology as to what those problems were. As stated in Buchwald (1985):

> Maxwellian theory cannot be translated into familiar to the modern understanding because the very act of translation necessarily deprives it of its deepest significance, and it was this significance which guided British research. It assumed that field and matter can always be treated as a single dynamical system, subject to modification according to the circumstances.

In modern days, we reject this very basis of the Maxwellian theory! Maxwell's paper in 1862 *On Physical Lines of Force* described an elaborate mechanism for ether. He was deeply attached to the mechanism despite certain problems with it and remained throughout his life till 1879 strongly committed to the principle to model building. Yet, in 1864, his paper on *A Dynamical Theory of the Electromagnetic Field* avoided specifying the structure of the ether, but, nevertheless, presumed the field to be governed by what he called dynamical laws. This was done by treating $\boldsymbol{H}$ as a velocity and $\boldsymbol{D}$ as the curl of the corresponding mechanical displacement. Through such assumption, it does not mean that the true structure of the ether is fully understood. Maxwell's theory was based on the assumption that all electromagnetic phenomena, including boundary conditions, can be obtained by applying Hamilton's principle to suitably chosen field energy densities which contain the appropriate electrical parameters like permittivity and permeability for the material medium. Modern theory implies that this can at best work on certain occasions as the macroscopic fields ($\boldsymbol{D}$, $\boldsymbol{H}$) are not a simple dynamical system but a construct obtained by averaging over the true state and the combined field vectors ($\boldsymbol{E}$, $\boldsymbol{B}$) with material vectors ($\boldsymbol{P}$, $\boldsymbol{M}$). Hence, where modern theory introduces the electron, the Maxwellian theory invented new forms of energy. This was possible because the Maxwellians were quite willing to invent modifications to the basic equations governing the electromagnetic field – as long as the results held up experimentally. Modern theory seeks unified explanations in an unmodifiable set of field equations coupled through the electron motions to intricate microphysical models. Maxwellian theory sought unity through a set of field equations coupled to Hamilton's principle. In Maxwell's theory, the goal was to create a theory of electromagnetism which made no use whatsoever of the microstructure of matter. Hence, the basic problem in understanding Maxwell's theory by a modern reader is to decipher what exactly did Maxwell mean by the words *charge* and *current*! Some of this confusion depends on Maxwell's having altered at least once his choice for the sign for the charge density in the equation which links it to the divergence of the electric flux $\boldsymbol{D}$. The conduction current in Maxwell's theory cannot be simply explained! The details of these subtleties may be found in Buchwald (1985). Finally, as pointed out by Bruce J. Hunt (1991):

> When the 1880s began, Maxwell's theory was virtually a trackless jungle. By the second half of the decade, guided by the principle of energy flow, Poynting, FitzGerald and above all Heaviside have succeeded in taming and pruning that jungle and in rendering it almost civilized.

In modern electrodynamics, we do not regard the field itself as a material structure, so we do not consider the stresses that may act upon it. Electromagnetic radiation, for example, transports energy and momentum, but stresses arise only when the radiation impinges on material structures. The Maxwellians did not think in this way. For them energy inhomogeneity, whether matter is present or not, implies stress. Indeed after the discovery of Poynting's theorem, they realized that the free ether must be stressed when transmitting radiation and so must move. In Maxwellian theory, the electromagnetic field transmits stress and is itself acted upon by stress. In modern theory, the field only acts; it is not acted upon (Buchwald 1985). There is an alternate point of view to this, however (Sarkar et al. 2006). For example, one side of Poynting's theorem has $[\boldsymbol{J} \bullet \boldsymbol{E}]$, while the other side has only fields. So one can say that the fields act on $\boldsymbol{J}$ or that the fields act on themselves, especially in cases like perfect electric conductors where no total work is done on the current and yet energy ends up in the fields.

DOI 10.1007/978-981-4560-75-7_12-1

In summary, the great difficulty with Maxwellian theory is that one cannot set up correspondences between mechanical and field variables which lead to consistent results unless one ignores conductivity, as illustrated by Heaviside. Heaviside's demonstration of this fact was based on Green's potential function. He suggested using $\boldsymbol{E}$ as the velocity instead of $\boldsymbol{H}$ in the mechanical model, and this requires a complete reconstruction of Maxwellian theory, as Maxwell did it the other way around (Buchwald 1985).

The mathematical formulation of *macroscopic* electromagnetic phenomena was accomplished by Maxwell (Larmor 1937). According to Maxwell's theory, electromagnetic radiation is produced by the oscillation of electric charges, the charges that produce light were unknown. Since it was generally believed that an electric current was made up of charged particles, Hendrik Antoon Lorentz in 1892 theorized that the atoms of matter might also consist of charged particles and suggested that the oscillations of these charged particles (electrons) inside the atom were the source of light. It should be noted that this occurred prior to the experimental clarification of the basic properties of the electron; those were not measured until 1897 and 1898 when Joseph John Thomson succeeded in a determination of the charge to mass ratio and in a first measurement of the elementary charge for an isolated electron. Lorentz thus constructed a *microscopic* theory by using Maxwell's equations and adding to it an expression for the force which a charged particle experiences in the presence of electric and magnetic fields. This microscopic theory is a description of matter in terms of its charged atomic fragments, ions, and electrons. The success of this microscopic theory lay in the proof first provided by Lorentz that the macroscopic Maxwell theory can be deduced from this microscopic theory by a suitable averaging process over the motion of the individual ions and electrons. The Maxwellian and Continental ideas were so profound that only Lorentz was able to retain the substantial character of charge while incorporating certain Maxwellian elements that did not violate basic preconceptions. Lorentz's theory computes interparticle actions directly by using retarded forces, and it employs careful microphysical averaging. It is interesting to note that the Lorentz force equation is very similar to Maxwell's equation (D) as described in section First Phase of Maxwell's Work on Electromagnetics. However, Lorentz's theory very much obscures the difference between ionic motions and the basic field equations (Buchwald 1985).

From 1894 to 1897, the British electromagnetic theory abandoned the basic principles of Maxwellian theory, and the entire subject was reconstructed on a new foundation – the electron, developed by Joseph Larmor in consultation with George FitzGerald (Buchwald 1985). Larmor thus effected a second revolution in the Maxwellian theory, one in which the electron had become the fundamental for generation of the field. Heaviside felt that Larmor's idea of the electron was not sufficiently fundamental. The major impact of Larmor's theory was the destruction of the idea that continuum theory can serve as a sufficient basis for electromagnetism. Maxwell's theory with its fundamental assumption that the electromagnetic field can be subjected to precisely the same type of analysis as the material continuum was an artifact after 1898. This was the second conceptual modification of Maxwell's theory after the first modifications carried out by Hertz and Heaviside, even though the basic equations still remained the same. Larmor's student, John Gaston Leathem, connected mathematically displacement with polarization by writing $\boldsymbol{D} = \boldsymbol{E} + \boldsymbol{P}$ and thus effectively destroying the Maxwell-FitzGerald theory of using Hamilton's principle as a starting point based on a continuous energy function (Buchwald 1985).

From this time onward, future work on electromagnetism, in Britain and in the Continent, depended directly on microphysics based on the electron rather on the macrophysics. Hence, there was a complete divorce between the matter and the field (Buchwald 1985). Thus, the nature of conduction, the stumbling block in Maxwell's theory, led to the final modification by Larmor and Lorentz by incorporating the microscopic view of matter through the introduction of the electron as the source of charge, the flow of which results in a current. These sources now produce the fields.

DOI 10.1007/978-981-4560-75-7_12-1

In short, the current perception of Maxwell's theory consists of Maxwell's system of equations supplemented by his initial radical concept of displacement current. A relevant question here is: Is this displacement current a physical reality? If so, then how is it that it produces radiation similar to that by conduction current? Currently accepted forms of Maxwell's equations as obtained through the work of Hertz, Heaviside, Lodge, FitzGerald, Lorentz, and Larmor are given as follows by these equations which are available in any modern textbook on electromagnetic theory: $\boldsymbol{J} + \frac{\partial \boldsymbol{D}}{\partial t} = \nabla \times \boldsymbol{H}$, $-\frac{\partial \boldsymbol{B}}{\partial t} = \nabla \times \boldsymbol{E}$, $\nabla \bullet \boldsymbol{D} = q_v$, $\nabla \bullet \boldsymbol{B} = 0$, and $\nabla \bullet \boldsymbol{J} = \frac{-\partial q_v}{\partial t}$, along with $\boldsymbol{B} = \mu \boldsymbol{H}$ and $\boldsymbol{D} = \varepsilon \boldsymbol{E}$. Equation of continuity here is included for completeness.

Epilogue

Finally, it would be interesting to draw the curtain on James Clerk Maxwell. In his hands electricity became an exact science and this can be said of many other branches of physics. In fact, it is not well appreciated that Maxwell can be called the greatest scientist of the last century even if he did nothing on electromagnetics. His contributions are everywhere and often few researchers in electromagnetics are aware of them. For completeness, Maxwell's contribution to electrical engineering is listed numerically in this section.

1. It is true that G. R. Kirchhoff wrote the laws related to the voltages and currents in a circuit, but as Sir Ambrose Fleming pointed out, the method to solve for the loop (and the node) equations for a general circuit by assuming flow of current in each wire to be the difference between two imaginary currents circulating in opposite directions in the network and then writing the solution as a ratio of two determinants was invented by him.
2. Maxwell showed how a circuit containing both capacitance and inductance would respond when connected to AC generators of different frequencies. He developed the phenomenon of electrical resonance in parallel to acoustic resonance developed by Lord Rayleigh. Maxwell developed the electrical analog when reviewing Rayleigh's paper and wrote about it to Rayleigh. He introduced a simpler concept for the group and the phase velocities.
3. Maxwell developed a coherent set of units of measurement of electricity and magnetism. They were later adopted almost unchanged as the first internationally accepted system of units, which became known misleadingly as the Gaussian system, which is a combination of the electrostatic units and the electromagnetic units. He introduced the *M*, *L*, and *T* notation in dimensional analysis.
4. He developed an experiment to measure the velocity of light from purely electrical means. He measured the speed of light by using 2,600 batteries to produce 3,000 V. The goal was to balance the electrostatic attraction between two charged metal plates against the magnetic repulsion between two current-carrying coils and build a balance arm to do this. He got a result of 288,000 km/s as compared to the current accepted value of 299,792.5 km/s. This was a direct application of his dimensional analysis, namely, to find how many *ESU* units are there in one *EMU* unit.
5. He wrote on the various ways to assess the color of the macula which in the eye is responsible for the quality of the vision. Nowadays when an ophthalmologist is checking the quality of the macula in a person's eyes, he/she is performing what is known as *Maxwell's yellow spot test*.
6. He developed the fish-eye lens to look into the retina of the eye.
7. A normal eye has three sorts of receptors for color. Maxwell demonstrated that in color-blind people their eyes are sensitive to only two rather than to three.

DOI 10.1007/978-981-4560-75-7_12-1

8. Maxwell chose the three primary colors to generate white light as red, green, and blue. Today, color television works on this principle, but Maxwell's name is rarely mentioned. However, other choices for the primary colors are equally viable. Maxwell demonstrated that other colors can be generated by using the barycentric coordinate system by placing the three primary colors red, green, and blue at the apex of a triangle, while white is achieved at the centroid when using light. If instead of light one uses color paints, then the centroid will be black instead of white.
9. He took the first color photograph using the same principles as it is done today, by superposing the three images taken in different colors!
10. He also showed that radiation pressure from the sun exists, which has a force of 4 pounds per square mile, which was later used to explain why the tails of the comets move away from the sun as they approach it.
11. He introduced and first applied a statistical law to a physical process. His discovery opened up an entirely new approach to physics, which led to statistical mechanics and to the use of probability distributions in quantum mechanics.
12. He introduced the concept of ensemble averaging. This enabled people to explain the properties of matter in terms of the behavior, en masse, of its molecules. Thus, using the method of ensemble averaging, the whole system is much easier to analyze, rather than dealing with individual components. He thus introduced the Maxwellian distribution.
13. He first asserted that the second law of thermodynamics is a statistical law. So if one throws a bucket of water into the sea, one cannot get the same bucket of water out again as the law applies to molecules in masses and not to individuals.
14. He developed the thought process related to Maxwell's demon which could violate the second law of thermodynamics and in this way create a perpetual motion. Through the application of Maxwell's demon, Leo Szilard developed the concept of quantization of information in terms of entropy which was later used in the development of information theory and was redeveloped by Claude Shannon.
15. He was the first to analyze the motion of Saturn's rings using mathematical analysis.
16. He illustrated how to generate objects having multiple foci using graphical means. Today, the same principles are used to generate optical lenses with multiple foci.
17. He developed a platometer on how to measure area of a planar arbitrary shaped object.
18. He produced a standard for electrical resistance.
19. When creating his standard for electrical resistance, he designed a governor to keep a coil spinning at a constant rate. He made the system stable by using the idea of negative feedback. He worked out the conditions of stability under various feedback arrangements. This was the first mathematical analysis of control systems. This work did not get any attention till 1940, when gun control radars were in demand during the Second World War. After the war, the American mathematician Norbert Wiener took things further and developed the science of cybernetics.
20. He introduced the idea of relaxation time into physics, engineering, and glaciology, by drawing on a suggestion of the Scottish physicist J. D. Forbes that a glacier behaves as a solid at times shorter than the relaxation time but like a liquid at longer times. This is related to the collision between gas molecules and replacing the concept of mean free path between collisions as a function of temperature.
21. In the 1870s during two discussions, he first emphasized what is now called the *butterfly effect – the fact that tiny differences in initial conditions can produce huge final effects* – the starting point of chaos theory.
22. Maxwell also wrote a paper on the "Protection of Building from Lightning." In that paper, he pointed out that most of the published writing dealt with were on the necessity of obtaining what telegraph engineers call a good earth connection. The telegraphist uses the earth to complete his circuit;

DOI 10.1007/978-981-4560-75-7_12-1

therefore, it is of great importance to him; but the protection of buildings from electric discharges has a different aim and a different method.

23. He also extended the work of the English mathematician A. Cayley published in 1859. The surface of the earth has high areas or hills and low areas with a bottom point. There are also ridges, valleys or dales, and passes. He thought that the numbers of each of these features must be somehow related by mathematical rules. His original ideas about the earth's surface have now evolved into a branch of topology called global analysis.
24. Maxwell introduced Hamilton's word *RELATIVITY*, in the way that physicists now understand it, in his small book *Matter and Motion* of 1876 (Maxwell 1876). He had noticed that the interpretation of electromagnetic induction was different depending on whether one considers a magnet approaching a wire loop or a loop approaching a magnet. The French mathematician, physicist, and engineer J. H. Poincaré read the work; A. Einstein learned of it from Poincaré; and the rest is history.
25. Last but not the least, he developed the fundamental principles of electromagnetic theory including the assertion that light is electromagnetic in nature.

Conclusion

This chapter has provided an introduction of the developments made up to the time of Maxwell in the creation of the theory of electromagnetism based on merging the different theories related to magnetism, electricity, and light. This is a unique theory where the fundamental equations did not change even though the philosophy went through two complete modifications: the first one by Hertz and Heaviside and the second by Larmor by introducing the microscopic structure of the matter into the macroscopic equations developed by Maxwell. The four of his equations that are used today were simultaneously developed by Hertz who wrote in the scalar form before he carried out his experiments, and what is available in modern textbooks under the heading of Maxwell's equations was written by Oliver Heaviside who first put them in the vector form. Also, a brief outline of Maxwell's other works is provided, as his name is not only synonymous with electromagnetic waves, but also Maxwell single-handedly laid the basic foundations of electrical engineering. In his hands electricity first became an exact science. The bottom line is that Maxwell can be called the greatest scientist of the last century even if he did absolutely nothing on electromagnetics as illustrated by the list of his accomplishments provided in the last section.

Finally, it is important to also note that the first patent on wireless was filed by an American dentist by the name of Mahlon Loomis, exactly the same year in 1864 when Maxwell first wrote his equations (Sarkar et al. 2006). This illustrates that important discoveries in the area of electromagnetics and wireless were done by researchers who had little or no formal education like Faraday and Heaviside or even education in physics like Loomis. Also Maxwell, who was a mathematician by training, wrote that he did not want to employ any mathematical principles until he went through all of Faraday's experiments in order to understand first the physical phenomena involved. Thus, one other conclusion of this presentation could be that in our teaching of electromagnetics, we should first provide the context and highlight the physical phenomena before going through all the mathematical details. Perhaps it is time for a change on how we teach electromagnetics to the students!

DOI 10.1007/978-981-4560-75-7_12-1

References

Bryant JH (1988) Heinrich Hertz, the beginning of microwaves: discovery of electromagnetic waves and opening of the electromagnetic spectrum by Heinrich Hertz in the years 1886–1892. IEEE Service Center, Piscataway

Buchwald JC (1985) From Maxwell to microphysics: aspects of electromagnetic theory in the last quarter of the nineteenth century. University of Chicago Press, Chicago

Commemoration Volume – James Clerk Maxwell (1931) Cambridge University Press, London

D'Agostino S (1975) Hertz's researches on electromagnetic waves. Hist Stud Phys Sei 6:261–323

Glazebrook RT (1896) James Clerk Maxwell and modern physics. Macmillan, London

Harman PH (2002a) The Scientific Letters and Papers of James Clerk Maxwell, vol I. Cambridge University Press (1846–1862), London

Harman PH (2002b) The Scientific Letters and Papers of James Clerk Maxwell, vol II. Cambridge University Press (1862–1873), London

Harman PH (2002c) The Scientific Letters and Papers of James Clerk Maxwell, vol III. Cambridge University Press (1874–1879), London

Heaviside O (1893) Electromagnetic theory, vol I and II. The Electrician Printing and Publishing Company, London

Hertz H (1962) Electric waves. Dover Publications, New York

Hertz H (1884) On the relations between Maxwell's fundamental equations of the opposing electromagnetics. Wiedemann's Annalen 23. pp 84–103. English translation is given in Hertz H. pp 127–145

http://www.britannica.com/. Accessed 25 Mar 2015

Hunt BJ (1991) The Maxwellians. Cornell University Press, Ithaca

Larmor J (1937) Origins of James Clerk Maxwell's electric ideas. Cambridge University Press, Cambridge

Mahon B (2003) The man who changed everything – the life of James Clerk Maxwell. Wiley, New York

Maxwell JC (1881) A treatise on electricity and magnetism, vol I and II. Oxford University Press, London

Maxwell JC (1876) Matter and motion. Cambridge University Press, London

Niven WD (1890a) The scientific papers of James Clerk Maxwell, vol I. Cambridge University Press, Cambridge

Niven WD (1890b) The scientific papers of James Clerk Maxwell, vol II. Cambridge University Press, Cambridge

Sarkar TK, Mailloux RJ, Oliner AA, Salazar-Palma M, Sengupta DL (2006) History of wireless. IEEE Press & Wiley, New Jersey

Sarkar TK, Salazar Palma M, Sengupta DL (2009) Who was James Clerk Maxwell and what was and is his electromagnetic theory? IEEE Antenn Propag Lett Mag 51:97–116

Sengupta DL, Sarkar TK (2003) Maxwell, Hertz, the Maxwellians and the early history of electromagnetic waves. IEEE Antenn Propag Lett Mag 45:13–19

Tai CT, Bryant JH (1994) New insights into Hertz's theory of electromagnetism. Radio Sci 29:685–690

Tolstoy I (1981) James Clerk Maxwell. University of Chicago Press, Chicago

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_13-2

Concept and Applications of Receiving Mutual Impedance

Hon Tat Hui
University of Macau, Macau, China

Abstract

The concept of receiving mutual impedance is introduced through derivation from a method of moments (MoM) analysis. The theoretical and experimental methods for the determination of the receiving mutual impedance are given and illustrated with typical examples of dipole and monopole antenna arrays. The fundamental difference, namely, the truly isolated state between the receiving mutual impedance and the conventional mutual impedance, is explained. Typical examples for the application of the receiving mutual impedance are given to demonstrate the validity and accuracy of using this concept. These examples include applications in direction of arrival (DOA) estimation, in interferences suppression, in magnetic resonance imaging (MRI) phased-array design, and in multiple-input and multiple-output (MIMO) communication systems.

Keywords

Receiving mutual impedance; Mutual coupling; Mutual impedance; Antenna arrays; Phased arrays; Method of moments; DOA estimation; Interference suppression; Magnetic resonance imaging; MIMO antenna arrays

Introduction

Mutual coupling is a common problem in the application of antenna arrays. It significantly affects the operation of all types of antenna arrays. The study of mutual coupling problem started several decades ago and attracted the attention of not just antenna engineers and researchers but also many researchers from other disciplines such as communications and biomedical imaging where antenna arrays are frequently used. Compared to a single antenna, an antenna array is able to provide the spatial information of a signal distribution. However, this function critically relies on the independence or distinctiveness of the signals received or transmitted from different antenna elements in the array. In reality, this is simply impossible because antenna elements interact with each other, i.e., they mutually couple with each other. In order to restore the independent signals received or transmitted from the antenna elements, the mutual coupling effect has to be removed or mitigated. Hence, it is very important to find ways to decouple the array signals received or transmitted from antenna arrays. In Gupta and Ksienski (1983), the effect of mutual coupling on the performance of adaptive arrays was studied by deriving the open-circuit voltages from the measured voltages across the antenna terminal loads. This method was later used in Yeh et al. (1989) to compensate for the mutual coupling effect of a dipole array used in direction findings. In Pasala and Friel (1994) and Yuan et al. (1998), the mutual coupling effect was investigated by solving the entire electromagnetic boundary value problem of the antenna array, but this requires the knowledge of the current distributions on the antenna elements which are not available in real situations. In Adve and Sarkar

Hon Tat Hui: deceased.

DOI 10.1007/978-981-4560-75-7_13-2

(2000), the principle to remove the mutual coupling effect is basically the same as in Pasala and Friel (1994), but a known incoming elevation angle of the signal and the interferences was assumed. This enables the reduction of the number of equations to the number of antenna elements in the array. This assumption makes it possible to eliminate the mutual coupling effect in an exact manner rather than approximately as by the method in Gupta and Ksienski (1983).

Notwithstanding these previous efforts, what makes this mutual coupling problem more complicated was that some recent findings (Hui 2003, 2004a, b; Lui et al. 2009; see also Wu and Nie 2008; Daniel 1974) pointed out that it is inappropriate to characterize the mutual coupling effect in transmitting and receiving antenna arrays in the same way. More specifically, they pointed out that a different characterization method from those used for transmitting antenna arrays is needed for receiving antenna arrays. They showed that the conventional mutual impedance concept is only suitable for characterizing the mutual coupling effect in transmitting antenna arrays and that a new concept, called the "receiving mutual impedance", has to be used for characterizing the mutual coupling effect in receiving antenna arrays. As receiving antenna arrays have applications in many different areas such as radar detection, remote sensing, interference cancellation, MIMO communications, magnetic resonance imaging (MRI), etc., it becomes very necessary to explore this concept in a greater depth and to explain the salient differences in using this concept from the conventional one.

The Development of the Concept of Receiving Mutual Impedance from the Method of Moments (MoM)

This section will introduce the development of the concept of receiving mutual impedance from the method of moments (Balanis 2005). Consider a linear dipole antenna array with n dipole antennas as shown in Fig. 1. The dipole antenna elements are assumed to be thin wires of equal length $\ell = 0.5\lambda$ and equal thickness with a radius $a << \lambda$, where λ is the operation wavelength. The dipole elements are placed along the x-axis with an equal separation $d = 0.5\lambda$ between them. The dipole axes are parallel to the z-axis. All the dipole elements are connected at their feeding points to a terminal load Z_L.

Assume that the dipole array is to receive an incoming signal which is a plane wave of electromagnetic field propagating in a three-dimensional space. By using the standard moment method analysis (Harrington 1993) and discretizing each antenna into m segments, a system of $m \times n$ linear equations can be obtained in the following matrix form:

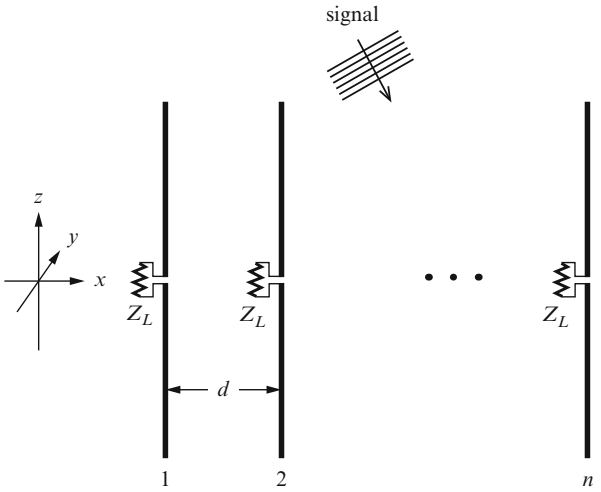


Fig. 1 The linear dipole antenna array for receiving a signal

DOI 10.1007/978-981-4560-75-7_13-2

$$\begin{bmatrix} [Z^{11}] & [Z^{12}] & \cdots & [Z^{1n}] \\ [Z^{21}] & [Z^{22}] & \cdots & [Z^{2n}] \\ \vdots & \vdots & \ddots & \vdots \\ [Z^{n1}] & [Z^{n2}] & \cdots & [Z^{nn}] \end{bmatrix} \begin{bmatrix} [I^1] \\ [I^2] \\ \vdots \\ [I^n] \end{bmatrix} = \begin{bmatrix} [V^1] \\ [V^2] \\ \vdots \\ [V^n] \end{bmatrix} \tag{1}$$

where the matrix blocks are respectively

$$[Z^{\alpha\beta}] = \begin{bmatrix} z_{11}^{\alpha\beta} - Z_L & z_{12}^{\alpha\beta} & \cdots & z_{1m}^{\alpha\beta} \\ z_{21}^{\alpha\beta} & z_{22}^{\alpha\beta} & \cdots & z_{2m}^{\alpha\beta} \\ \vdots & \vdots & \ddots & \vdots \\ z_{m1}^{\alpha\beta} & z_{m2}^{\alpha\beta} & \cdots & z_{mm}^{\alpha\beta} \end{bmatrix} \tag{2}$$

$$[I^{\alpha}] = [\, I_1^{\alpha} \quad I_2^{\alpha} \quad \cdots \quad I_m^{\alpha} \,]^T \tag{3}$$

$$[V^{\alpha}] = [\, V_1^{\alpha} \quad V_2^{\alpha} \quad \cdots \quad V_m^{\alpha} \,]^T. \tag{4}$$

For the matrix in Eq. 2, the element $z_{\mu\nu}^{\alpha\beta}$ (with $\alpha, \beta = 1, 2, \cdots, n$ and $\mu, \nu = 1, 2, \cdots, m$) denotes the impedance between the μth segment of the αth antenna element and the νth segment of the βth antenna element. The first segment is assumed to be connected to the terminal load Z_L. The vectors in Eqs. 3 and 4 are, respectively, the current distributions and the voltage distributions on the αth dipole element. The superscript "T" in Eqs. 3 and 4 denotes the matrix transpose operation. The voltage vector in Eq. 4 represents the excitation for the array due to the incoming signal. For the matrix elements $z_{\mu\nu}^{\alpha\beta}$ in Eq. 2, they can be expressed explicitly in terms of the dyadic Green's functions as follows:

$$z_{\mu\nu}^{\alpha\beta} = -j\omega\mu_0 \int_{\ell} I_{\mu}^{\alpha}(z) G_{zz}(z, z') I_{\nu}^{\beta}(z') dz' dz \tag{5}$$

where

$$G_{zz}(z, z') = \left[3 + j3k_0R - (k_0R)^2\right] \frac{(z - z')^2}{R^2} + \left[-1 - jk_0R + (k_0R)^2\right] \tag{6}$$

and

$$R = \left| \vec{R} - \vec{R} \right| = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}. \tag{7}$$

From Eq. 1, it can be seen that the existence of the mutual coupling effect is due to the off-diagonal matrix blocks, i.e., $[Z^{\alpha\beta}]$ with $\alpha \neq \beta$. If the current distributions on the dipole elements, $[I^{\alpha}]\, \alpha = 1, 2, \cdots, n$, are exactly known, the mutual coupling effect can be exactly quantified. For example, if the mutual coupling effect due to all other antenna elements on the pth antenna element is to be eliminated, then the pth row of Eq. 1 can be first rewritten as

$$[Z^{pp}][I^p] = -\left[Z^{p1}\right]\left[I^1\right] - \cdots - \left[Z^{p(p-1)}\right]\left[I^{p-1}\right] - \left[Z^{p(p+1)}\right]\left[I^{p+1}\right] - \cdots - [Z^{pn}][I^n] + [V^p]. \tag{8}$$

DOI 10.1007/978-981-4560-75-7_13-2

It is seen that the left-hand side of Eq. 8 is exactly the system of equations resulted from the moment method analysis for the pth antenna element as if it were standing alone without the presence of all other antenna elements. However, the right-hand side of Eq. 8 has two excitation sources now:

(i) The excitation voltage induced by the current distributions on the other antenna elements =

$$-\left[Z^{p1}\right]\left[I^{1}\right]-\cdots-\left[Z^{p(p-1)}\right]\left[I^{p-1}\right]-\left[Z^{p(p+1)}\right]\left[I^{p+1}\right]-\cdots-\left[Z^{pn}\right]\left[I^{n}\right] \tag{9}$$

(ii) The excitation voltage due to the incoming signal on the pth antenna element =

$$\left[V^{p}\right] \tag{10}$$

By the principle of superposition, the effects on the pth antenna element due to these two excitation sources are independent, meaning that the field generated by any one of these two excitation sources satisfies the boundary conditions on the pth antenna alone and hence can exist independent of the field generated by the other source. Thus, if the current distributions on the antenna elements are completely and exactly known, the quantity $-\left[Z^{p1}\right]\left[I^{1}\right]-\cdots-\left[Z^{p(p-1)}\right]\left[I^{p-1}\right]-\left[Z^{p(p+1)}\right]\left[I^{p+1}\right]-\cdots-\left[Z^{pn}\right]\left[I^{n}\right]$ can be readily calculated, and the exact mutual coupling voltages from all other antenna elements can be determined.

The problem now is that only the voltages or currents on the antenna terminal loads can be measured (known). The knowledge of the *distributions* of the currents or voltages on every point of the antenna elements is not available. That is, at most only one point (the feeding point) of the current or voltage distribution in Eq. 3 or Eq. 4 can be measured. To solve this problem, a "single-mode" approximation can be made. That is, the current distribution on each antenna element in the array is treated as consisting of a single-mode current. Furthermore, the terminal currents and voltages on the antenna elements are used to describe the antenna array as an n-port network. By doing so, Eq. 8 is reduced to

$$\left(Z_t'^{pp}-Z_L\right)I_t'^{p}=-Z_t'^{p1}I_t'^{1}-\cdots-Z_t'^{p(p-1)}I_t'^{p-1}-Z_t'^{p(p+1)}I_t'^{p+1}-\cdots-Z_t'^{pn}I_t'^{n}+V_t'^{p} \tag{11}$$

in which a subscript "t" is added to signify that these variable/quantities are referred to the antenna *terminals*. Note also that the variables in Eq. 3 have been added with a prime to indicate that they are obtained under the single-mode assumption. The measurable quantities in Eq. 3 are the terminal currents $I_t'^{1},\ I_t'^{2},\ \cdots, I_t'^{p}, \cdots,\ I_t'^{n}$. On the left-hand side of Eq. 3, the term $\left(Z_t'^{pp}-Z_L\right)I_t'^{p}$ is the sum of the voltage dropped across the self-impedance of the pth antenna, $Z_t'^{pp}I_t'^{p}$, and the voltage dropped across the terminal load of the pth antenna, $-Z_L I_t'^{p}$. The voltage dropped across the self-impedance of the pth antenna $Z_t'^{pp}I_t'^{p}$ is not practically measurable. So this can further be removed from the left-hand side of Eq. 3 so that only the voltage dropped across the terminal load Z_L of the pth dipole element remains on the left-hand side of Eq. 3. However, to do so, it is also required to remove correspondingly that part of coupled voltages and that part of the incident voltage from the right-hand side of Eq. 3 which excite the voltage drop across the self-impedance. Finally, after the removal of these voltages from both sides of Eq. 3, the following new equation is obtained

$$-Z_L I_t^{p}=V_p=-Z_t^{p1}I_t^{1}-\cdots-Z_t^{p(p-1)}I_t^{p-1}-Z_t^{p(p+1)}I_t^{p+1}-\cdots-Z_t^{pn}I_t^{n}+U_t^{p} \tag{12}$$

DOI 10.1007/978-981-4560-75-7_13-2

where $-Z_t^{p1}I_t^1 - \cdots - Z_t^{p(p-1)}I_t^{p-1} - Z_t^{p(p+1)}I_t^{p+1} - \cdots - Z_t^{pn}I_t^n$ is the total coupled voltage dropped across the terminal load Z_L of the pth antenna due to the current distributions (single mode) on the other antenna elements in the array, U_t^p is the excitation voltage across the terminal load Z_L of the pth antenna due to the incoming signal alone, and V_p is the voltage across the terminal load Z_L of the pth antenna. Note that the prime notation on the term $I_t'^p$ on the left-hand side of Eq. 4 has been dropped to simplify the expression.

Thus from Eq. 4, it can be seen that the total terminal voltages V_p across the terminal load of the pth antenna can be expressed as a sum of products, each of which (e.g., $-Z_t^{pk}I_t^k$) is a multiplication of the terminal current (I_t^k) on an antenna element (other than the pth antenna) with a quantity known as the "modified mutual impedance" Z_t^{pk}. It is called "modified" because the meaning of this mutual impedance (Z_t^{pk}) is different from the conventional mutual impedance (Balanis 2005; Jordan 1968) (commonly denoted by Z_{pk}) in that this modified mutual impedance Z_t^{pk} measures only the *coupled voltage* across the terminal load of the pth antenna while the conventional mutual impedance Z_{pk} measures the *open-circuit voltage* on the pth antenna. The open-circuit voltage is the voltage dropped across the total impedance (i.e., the terminal load plus the self-impedance) of the pth antenna. Another difference is that the modified mutual impedance Z_t^{pk} refers to the situation where both antenna elements (the pth and kth) are in the passive mode (receiving mode) so that the current I_t^k is induced by an external signal whereas the conventional mutual impedance Z_{pk} refers to the situation where the pth antenna is in the passive mode while the kth antenna is in an active transmitting mode. Although this latter difference is not of a fundamental one, it does result in two different current distributions on the kth antenna under these two different situations. The two different current distributions on the kth antenna, in turn, induce different coupled voltages on the pth antenna. Thus, the modified mutual impedance Z_t^{pk} is different from the conventional mutual impedance Z_{pk}, and so a new name is given for this modified mutual impedance – the receiving mutual impedance. **Its definition is the ratio of the coupled voltage dropped across the terminal load of the *p*th antenna (due to the reradiation of the current distribution on the *k*th antenna) to the terminal current on the *k*th antenna, while the antenna array is in the passive receiving mode and excited by an external electromagnetic plane wave.** As the definition of the receiving mutual impedance requires an external plane wave source (which is to simulate the signal the antenna array is to receive), the result is that the value of the receiving mutual impedance is in general dependent on the direction (incident angle) of the plane wave. The difference between the receiving mutual impedance Z_t^{pk} and the conventional mutual impedance Z_{pk} will be demonstrated in the next section.

Another observation from Eq. 4 can be obtained by rewriting it as

$$\begin{aligned} V_p &= -Z_L I_t^p = -Z_L\left(I_c^p + I_s^p\right) \\ &= W + U_t^p \\ &= -Z_t^{p1}I_t^1 - \cdots - Z_t^{p(p-1)}I_t^{p-1} - Z_t^{p(p+1)}I_t^{p+1} - \cdots - Z_t^{pn}I_t^n + U_t^p \end{aligned} \tag{13}$$

where I_t^p has been split into two parts: I_c^p and I_s^p so that I_c^p is that part of the current on the pth antenna that is induced by the scattered fields from the other antenna elements and I_s^p is that part of the current on the pth antenna that is induced by the incoming signal U_t^p along. I_c^p can be termed as the *coupled current*, and I_s^p can be termed as the *isolated current* on the pth antenna. In Eq. 5, W is used to denote the total coupled voltage, i.e.,

$$W = -Z_t^{p1}I_t^1 - \cdots - Z_t^{p(p-1)}I_t^{p-1} - Z_t^{p(p+1)}I_t^{p+1} - \cdots - Z_t^{pn}I_t^n \tag{14}$$

DOI 10.1007/978-981-4560-75-7_13-2

and hence

$$\begin{cases} W = -Z_L I_c^p = \text{total coupled voltage} \\ U_t^p = -Z_L I_s^p = \text{signal voltage} \end{cases}. \tag{15}$$

From Eqs. 5 and 7, it can be seen that if all the receiving mutual impedances Z_t^{pk} and the terminal currents $I_t^k \ (k = 1, 2, \cdots, (p-1), (p+1), \cdots, n)$ on the antenna elements are known, it is able to quantify and remove the total coupled voltage W from the terminal voltage across the terminal load of the pth antenna, leaving behind the isolated signal voltage U_t^p. Note that this result cannot be obtained by using the conventional mutual impedances Z_{pk} (Balanis 2005; Jordan 1968; Gupta and Ksienski 1983). The set of isolated signal voltages $U_t^p \ (p = 1, 2, \cdots, n)$ is known as the ideal array manifold and is one of the important fundamental concepts in array signal processing.

Theoretical and Experimental Methods for the Determination of the Receiving Mutual Impedance

As explained in the last section, for characterizing the mutual coupling effect in a receiving antenna array, it is more accurate to use the receiving mutual impedance. The next question is how to determine it. As the receiving mutual impedance is different from the conventional mutual impedance, different methods are needed. In this section, theoretical and measurement methods will be introduced to determine the receiving mutual impedance.

Theoretical Methods for the Determination of the Receiving Mutual Impedance

Theoretical Method Based on the Calculation of Current Distributions

This section will introduce the development of the concept of receiving mutual impedance from the method of moments (Hui 2004a). Consider a two-antenna array as shown in Fig. 2. For an antenna array with more than two elements, it can be considered as consisted of different pairs of two-element arrays with different separations. By the definition of receiving mutual impedance, it is the ratio of the coupled voltage dropped across the terminal load of the first antenna to the source terminal current on the second antenna, whereas both antennas are in the passive receiving mode and excited by an external electromagnetic plane wave. Hence the simplest theoretical method for determining the receiving mutual impedance is to excite the two antennas by a plane wave and obtain the current distribution on antenna 2 (denote its

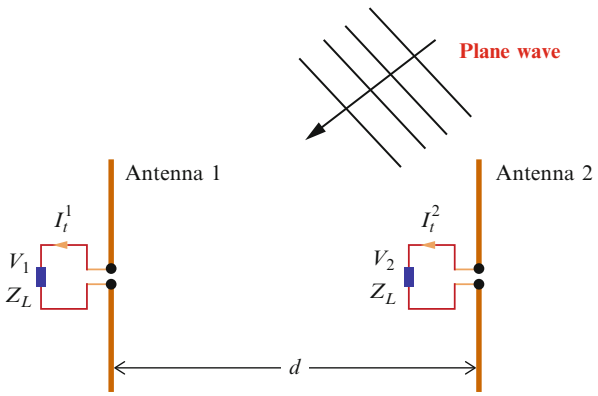


Fig. 2 The scenario for the theoretical calculation of the receiving mutual impedance

DOI 10.1007/978-981-4560-75-7_13-2

value at the terminal load as I_t^2). Then remove the plane wave source and excite antenna 1 by the current distribution on antenna 2 (reradiation). Calculate the voltage dropped across the terminal load of antenna 1 (denote it as W_1). Then the receiving mutual impedance Z_t^{12} can be determined by (see Eq. 21)

$$Z_t^{12} = -\frac{W_1}{I_t^2}. \tag{16}$$

The receiving mutual impedance Z_t^{21} can also be determined similarly. By this method, it is required to first calculate the current distribution on antenna 1 or antenna 2, which later served as the reradiation source. Some software may not be able to assign and use a current distribution as the excitation source. In this case, an alternative method as introduced below can be used.

Theoretical Method Based on the Calculation of Antenna Terminal Voltages

This section will introduce the development of the concept of receiving mutual impedance from the method of moments (Hui 2003). For the two receiving antennas shown in Fig. 2, V_1 and V_2 are the total received voltages across the terminal loads of antennas 1 and 2, respectively, when the array is excited by a plane wave. The corresponding terminal currents on the two antennas are I_t^1 and I_t^2. Using these terminal quantities (which can be directly measured), the receiving mutual impedance Z_t^{12} can be calculated as follows:

$$\begin{aligned} Z_t^{12} &= \frac{\text{Coupled voltage across antenna 1}'\text{s terminal load}}{\text{Received current on antenna 2}'\text{s terminal load}} \\ &= -\frac{V_1 - U_t^1}{I_2} \end{aligned} \tag{17}$$

where U_t^1 is the received voltage across the terminal load of antenna 1 when antenna 1 is excited by the plane wave alone (with antenna 2 removed from the array). U_t^1 can be termed as the isolated voltage on antenna 1, and it can be calculated as shown in Fig. 3. Using this method, there is no need to calculate the current distribution on the antennas. Because of this advantage, this method is also suitable for experimental determination of the receiving mutual impedance, as will be shown later. In a similar way, Z_t^{21} can also be determined.

A typical calculation of the receiving mutual impedance is shown in Fig. 4. This is the receiving mutual impedance Z_t^{12} between two monopole antennas at a frequency of 2.4 GHz and is shown against the separation between the two antennas. The dimensions of the monopole antennas are length = 3 cm (0.24λ at 2.4 GHz) and radius = 0.3 mm. Each monopole antenna is connected to a terminal load of $Z_L = 50\Omega$. The plane wave used to excite the array comes from the horizontal direction with $\theta = 90°$ and

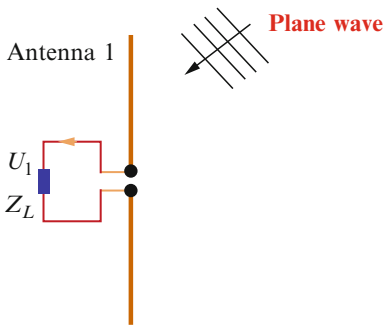


Fig. 3 The situation for the calculation of the isolated voltage U_t^1

DOI 10.1007/978-981-4560-75-7_13-2

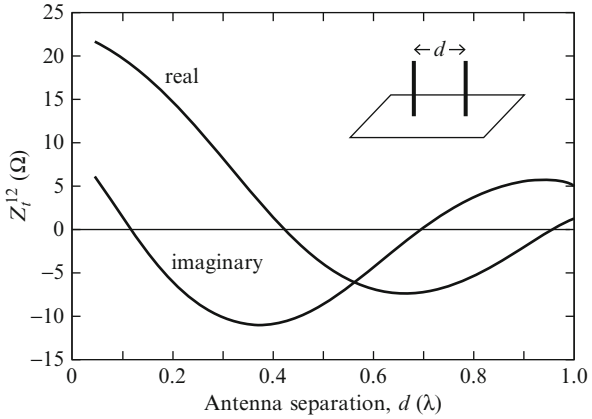


Fig. 4 The receiving mutual impedance Z_t^{12} between two monopole antennas with their separation at a frequency of 2.4 GHz

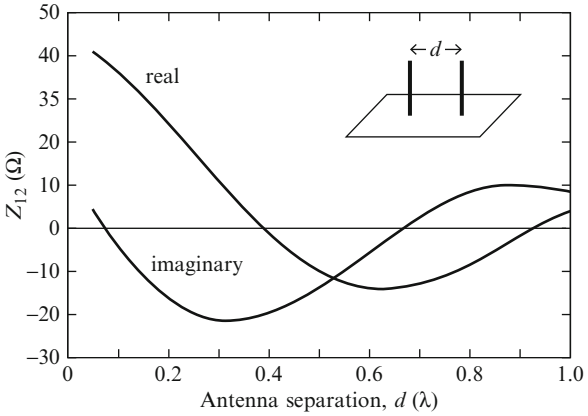


Fig. 5 The corresponding conventional mutual impedance Z_{12} between the two monopole antennas as shown in Fig. 4

from an azimuth angle $\phi = 0$. Note that due to the circular symmetry of the monopole antennas, the received currents on the two monopoles are independent of ϕ, and hence the value of ϕ can be arbitrary. For comparison, the corresponding conventional mutual impedance Z_{12} between these two monopole antennas is shown in Fig. 5.

Experimental Methods for the Determination of the Receiving Mutual Impedance

Experimental Method Considering Two Antenna Elements at a Time

This subsection will describe the experimental method considering two antenna elements at a time (Hui 2004b). To determine the receiving mutual impedance experimentally, a vector network analyzer (VNA) is required to take measurements inside an anechoic chamber, as shown in Fig. 6. The same two monopoles as in Fig. 4 are considered again here for the experimental measurement. The operating frequency is 2.4 GHz, and the two monopoles are kept at a fixed separation of 6.25 cm (half wavelength at 2.4 GHz). The monopoles are labeled as monopole 1 and monopole 2. By using the toggle switch as shown in Fig. 6, the scattering parameters (S_{12} or S_{21}) of the two monopoles are measured in turn. The measured scattering parameter, for example, S_{21}, is converted into a voltage according to Pozar (1990)

$$S_{21} = \frac{\beta}{\alpha} \tag{18}$$

DOI 10.1007/978-981-4560-75-7_13-2

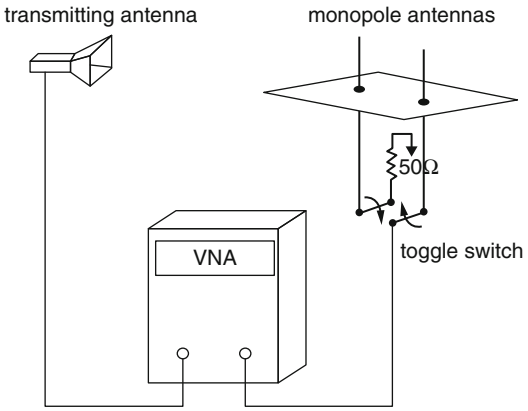


Fig. 6 The measurement of the receiving mutual impedances inside an anechoic chamber using a vector network analyzer

where α is the square root of the power emitted by the transmitting horn inside the anechoic chamber and $\beta = V/\sqrt{Z_0}$ is the square root of the power received by a monopole element with V being the terminal voltage of the monopole element and Z_0 being the system impedance (50Ω). Both α and β are complex values with magnitudes and phases. The detailed measurement procedures are as follows (Hui 2004b):

1. Measure S_{21} at monopole 1's terminal with monopole 2's terminal connected to a load Z_0. Denote it as S_{21_1}.
2. Measure S_{21} at monopole 2's terminal with monopole 1's terminal connected to a load Z_0. Denote it as S_{21_2}.
3. Measure S_{21} at monopole 1's terminal with monopole 2 removed (taken away from the array). Denote it as S'_{21_1}.
4. Measure S_{21} at monopole 2's terminal with monopole 1 removed (taken away from the array). Denote it as S'_{21_2}.

Throughout the whole measurement process, the positions of the two monopoles must be kept fixed since any change in their positions will alter the phase and magnitude of S_{21}. This is especially true when one of the antennas is being taken away from or put back into its original position. Now convert the measured scattering parameters S_{21} into its respective terminal voltages using Eq. 10. That is,

$$V_t^1 = S_{21\_1}\alpha\sqrt{Z_0} \tag{19}$$

$$V_t^2 = S_{21\_2}\alpha\sqrt{Z_0} \tag{20}$$

$$V_t'^1 = S'_{21\_1}\alpha\sqrt{Z_0} \tag{21}$$

$$V_t'^2 = S'_{21\_2}\alpha\sqrt{Z_0} \tag{22}$$

where

V_t^1 is the measured voltage at monopole 1's terminal with monopole 2's terminal connected to a load Z_0.
V_t^2 is the measured voltage at monopole 2's terminal with monopole 1's terminal connected to a load Z_0.
$V_t'^1$ is the measured voltage at monopole 1's terminal with monopole 2 removed.
$V_t'^2$ is the measured voltage at monopole 2's terminal with monopole 1 removed.

DOI 10.1007/978-981-4560-75-7_13-2

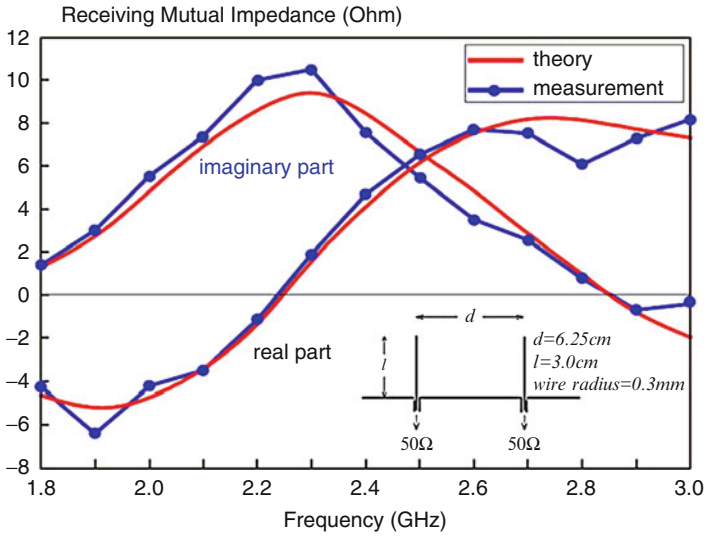


Fig. 7 The receiving mutual impedance Z_t^{12} between two monopole antennas with frequency at a fixed separation of 6.25 cm (0.5λ at 2.4 GHz). The two antennas are the same as those in Fig. 4

Then from Eq. 9,

$$V_t^1 = U_t^1 - Z_t^{12} I_t^2, \quad V_t^2 = U_t^2 - Z_t^{21} I_t^1 \tag{23}$$

$$V'^1_t = U_t^1, \quad V'^2_t = U_t^2 \tag{24}$$

$$I_t^1 = -\frac{V_t^1}{Z_0}, \quad I_t^2 = -\frac{V_t^2}{Z_0}, \quad I'^1_t = -\frac{V'^1_t}{Z_0}, \quad I'^2_t = -\frac{V'^2_t}{Z_0}. \tag{25}$$

From Eqs. 15 and 16, the following two relations can be obtained:

$$-\left(V_t^1 - V'^1_t\right) = Z_t^{12} I_t^2, \quad -\left(V_t^2 - V'^2_t\right) = Z_t^{21} I_t^1. \tag{26}$$

Finally, using Eq. 17, the measured values of Z_t^{12} and Z_t^{21} can be determined as

$$\begin{aligned} Z_t^{12} &= \frac{\left(V_t^1 - V'^1_t\right)}{V_t^2} Z_0 & Z_t^{21} &= \frac{\left(V_t^2 - V'^2_t\right)}{V_t^1} Z_0 \\ &= \frac{\left(S_{21\_1} - S'_{21\_1}\right)}{S_{21\_2}} Z_0, & &= \frac{\left(S_{21\_2} - S'_{21\_2}\right)}{S_{21\_1}} Z_0. \end{aligned} \tag{27}$$

Thus, the receiving mutual impedances can be obtained in terms of the measured scattering parameters. An example measurement result of Z_t^{12} with frequency is shown in Fig. 7 for two monopole antennas (same as the two antennas in Fig. 4) at a fixed separation of 6.25 cm. The result is compared with the theoretically calculated one.

DOI 10.1007/978-981-4560-75-7_13-2

Experimental Method Considering All Antenna Elements Together

Strictly speaking, the previous experimental method is an approximation method only this section will introduce the development of the concept of receiving mutual impedance from the method of moments (Lui and Hui 2010a, b, c). This is because it only considers two antenna elements at a time. But doing this would ignore the scattering effect from other antenna elements in the array, if the array has more than two antenna elements. This shortcoming can be overcome in some types of antenna arrays. Two typical such antenna arrays are the monopole and dipole arrays because the element pattern of a monopole or a dipole is omnidirectional in the azimuth plane. For this type of antenna arrays, the receiving mutual impedances of all the antenna elements can be determined simultaneously (not two at a time) and accurately by measurement or calculation.

Consider an n-element dipole antenna array as shown in Fig. 1. When an incoming plane wave signal is incident on the array from a horizontal direction ϕ (with $\theta = 90$), the terminal voltage on the antenna elements can be represented as (see Eq. 19)

$$\begin{cases} V_1 = -Z_L I_t^1 = -Z_t^{12} I_t^2 - \cdots - Z_t^{1n} I_t^n + U_t^1 \\ V_2 = -Z_L I_t^2 = -Z_t^{21} I_t^1 - \cdots - Z_t^{2n} I_t^n + U_t^2 \\ \vdots \\ V_p = -Z_L I_t^p = -Z_t^{p1} I_t^1 - \cdots - Z_t^{p(p-1)} I_t^{p-1} - Z_t^{p(p+1)} I_t^{p+1} - \cdots - Z_t^{pn} I_t^n + U_t^p \\ \vdots \\ V_n = -Z_L I_t^n = -Z_t^{n1} I_t^1 - \cdots - Z_t^{n(n-1)} I_t^{n-1} + U_t^n \end{cases}. \tag{28}$$

The terminal currents I_t^p can be expressed using the terminal voltages V_p as

$$I_t^1 = -\frac{V_1}{Z_L},\ I_t^2 = -\frac{V_2}{Z_L},\ \cdots,\ I_t^n = -\frac{V_n}{Z_L}. \tag{29}$$

Putting Eq. 21 into Eq. 20, the following matrix equation can be obtained:

$$\begin{bmatrix} 1 & -\frac{Z_t^{12}}{Z_L} & \cdots & -\frac{Z_t^{1(n-1)}}{Z_L} & -\frac{Z_t^{1n}}{Z_L} \\ -\frac{Z_t^{21}}{Z_L} & 1 & \cdots & -\frac{Z_t^{2(n-1)}}{Z_L} & -\frac{Z_t^{2n}}{Z_L} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{Z_t^{(n-1)1}}{Z_L} & -\frac{Z_t^{(n-1)2}}{Z_L} & \cdots & 1 & -\frac{Z_t^{(n-1)n}}{Z_L} \\ -\frac{Z_t^{n1}}{Z_L} & -\frac{Z_t^{n2}}{Z_L} & \cdots & -\frac{Z_t^{n(n-1)}}{Z_L} & 1 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_{n-1} \\ V_n \end{bmatrix} = \begin{bmatrix} U_t^1 \\ U_t^2 \\ \vdots \\ U_t^{n-1} \\ U_t^n \end{bmatrix}. \tag{30}$$

Consider the first row of Eq. 22. It is written below:

$$V_1 - Z_t^{12}\frac{V_2}{Z_L} - \cdots - Z_t^{1k}\frac{V_k}{Z_L} \cdots - Z_t^{1n}\frac{V_n}{Z_L} = U_t^1. \tag{31}$$

The terminal voltages in Eq. 23, $V_1, V_2, \cdots, V_n$, and U_t^1 are measured when the incoming signal comes from an azimuth angle ϕ (with $\theta = 90$). It is not difficult to understand that due to the different positions

DOI 10.1007/978-981-4560-75-7_13-2

of the antenna elements and the unequal mutual coupling effect, these terminal voltages (especially their phases) are dependent on the incoming direction ϕ of the signal. But on the other hand, the receiving mutual impedances in Eq. 23 are independent of the incoming direction of the signal due to the circular symmetry and the omnidirectional radiation patterns of the antenna elements. These receiving mutual impedances depend only on the relative position of the antenna elements. Because of these two observations, the measurement of the terminal voltages can be carried out at different azimuth angles (but same θ). If this measurement is repeated $(n-1)$ times at $(n-1)$ different azimuth angles, the following system of equations can be obtained:

$$\begin{cases} V_1^{\phi_1} - Z_t^{12}\dfrac{V_2^{\phi_1}}{Z_L} - \cdots - Z_t^{1k}\dfrac{V_k^{\phi_1}}{Z_L}\cdots - Z_t^{1n}\dfrac{V_n^{\phi_1}}{Z_L} = U_t^{1,\phi_1} \\ V_1^{\phi_2} - Z_t^{12}\dfrac{V_2^{\phi_2}}{Z_L} - \cdots - Z_t^{1k}\dfrac{V_k^{\phi_2}}{Z_L}\cdots - Z_t^{1n}\dfrac{V_n^{\phi_2}}{Z_L} = U_t^{1,\phi_2} \\ \vdots \\ V_1^{\phi_{n-1}} - Z_t^{12}\dfrac{V_2^{\phi_{n-1}}}{Z_L} - \cdots - Z_t^{1k}\dfrac{V_k^{\phi_{n-1}}}{Z_L}\cdots - Z_t^{1n}\dfrac{V_n^{\phi_{n-1}}}{Z_L} = U_t^{1,\phi_{n-1}} \end{cases} . \tag{32}$$

The system in Eq. 24 contains $(n-1)$ unknowns, i.e., the receiving mutual impedances $Z_t^{12}, Z_t^{13}, \cdots, Z_t^{1n}$. But there are also $(n-1)$ equations in this system. Hence Eq. 24 can be uniquely solved, and the $(n-1)$ receiving mutual impedances can be determined accurately. The $(n-1)$ azimuth angles, $\phi_1, \phi_2, \cdots, \phi_{n-1}$, for the incoming signal directions can be chosen to be equally divided in the angular range $0 \sim \pi$. When the second row of Eq. 22 is considered and the above procedure is repeated, the receiving mutual impedances, $Z_t^{21}, Z_t^{23}, \cdots, Z_t^{2n}$, can be determined in the same way. This procedure can be repeated $(n-1)$ times, and all the receiving mutual impedances in Eq. 22 can then be determined. Obviously, this method can also be used to theoretically calculate the receiving mutual impedances of an antenna array. As this method accounts for the scattering effect from all the antenna elements in the array, it is accurate, especially when the element separation is very small. A typical determination of the receiving mutual impedances of an eight-element dipole array by using this method and the method described in section "Experimental Method Considering Two Antenna Elements at a Time" is shown in Table 1 (Lui

Table 1 Comparison of receiving mutual impedances determined using the two measurement methods in sections "Experimental Method Considering Two Antenna Elements at a Time" and "Experimental Method Considering All Antenna Elements Together." The antenna array is an eight-element uniform linear array with dipole antenna elements. The length of the monopoles is 0.48λ and diameter of the dipoles is 0.6 mm. The operation frequency is 2.4 GHz. The element separation is denoted by d

	$d = 0.30\lambda$		$d = 0.15\lambda$	
Receiving mutual impedances	By Eq. 19 section "Experimental Method Considering Two Antenna Elements at a Time" (Ω)	By Eq. 24 section "Experimental Method Considering All Antenna Elements Together" (Ω)	By Eq. 19 section "Experimental Method Considering Two Antenna Elements at a Time" (Ω)	By Eq. 24 section "Experimental Method Considering All Antenna Elements Together" (Ω)
Z_t^{12}	$-10.38 + j15.07$	$-10.40 + j15.06$	$-24.11 + j5.47$	$-24.76 + j5.35$
Z_t^{13}	$10.02 + j5.42$	$10.07 + j5.44$	$-10.38 + j15.07$	$-8.90 + j15.78$
Z_t^{14}	$2.24 - j7.75$	$2.14 - j7.74$	$3.02 + j13.83$	$0.67 + j12.69$
Z_t^{15}	$-6.20 - j0.14$	$-6.09 - j0.12$	$10.02 + j5.42$	$12.29 + j6.65$
Z_t^{16}	$1.28 + j4.86$	$1.21 + j4.83$	$8.76 - j3.59$	$7.28 - j4.18$
Z_t^{17}	$3.59 - j2.19$	$3.61 - j2.20$	$2.24 - j7.75$	$2.89 - j7.61$
Z_t^{18}	$-2.72 + j2.40$	$-2.71 + j2.38$	$-4.13 - j5.67$	$-4.20 - j5.66$

DOI 10.1007/978-981-4560-75-7_13-2

and Hui 2010a). It can be seen that the differences between these two methods are significant only when the element separation is very small, i.e., $d = 0.15\lambda$.

Comparison of the Receiving Mutual Impedance and the Conventional Mutual Impedance

In order to compare the receiving mutual impedance and the conventional mutual impedance, a brief revision of the conventional mutual impedance is required. Consider a two-element transmitting dipole antenna array with the dipole elements parallel to the z-axis as shown in Fig. 8. When the dipole antennas are excited by excitation sources V_{s1} and V_{s2} at their feeding terminals, the tangential electric field $E_1(z)$ on the surface of the antenna 1 can be written as follows (Ishimaru 1991):

$$E_1(z) = E_{11}(z) + E_{12}(z) \tag{33}$$

where $E_{11}(z)$ and $E_{12}(z)$ are the tangential electric fields on the surface of antenna 1 caused by currents on antenna 1 and antenna 2, respectively. The conventional mutual impedances Z_{12} can be derived through the following procedure. Multiply Eq. 25 with the current on antenna 1, $I_1(z)$, on both sides and integrate throughout the length of antenna 1, the following expression is obtained:

$$-I_t^1\left(V_{s1} - V_{g1}\right) = -I_t^1 V_1 = \int_\ell I_1(z)E_{11}(z)d\ell + \int_\ell I_1(z)E_{12}(z)d\ell \tag{34}$$

where I_t^1 is the terminal current on antenna 1 ($I_1(z)$ is the current distribution on antenna 1), V_{g1} is the voltage dropped across the source impedance, and $V_1 = V_{s1} - V_{g1}$ is the voltage across the terminal of antenna 1. From Eq. 26, the following expression can be obtained:

$$\begin{aligned} V_1 &= I_t^1\left[\frac{-1}{\left(I_t^1\right)^2}\int_\ell I_1(z)E_{11}(z)d\ell\right] + I_t^2\left[\frac{-1}{I_t^1 I_t^2}\int_\ell I_1(z)E_{12}(z)d\ell\right] \\ &= I_t^1 Z_{11} + I_t^2 Z_{12} \end{aligned} \tag{35}$$

where Z_{11} is the self-impedance of antenna 1, I_t^2 is terminal current on antenna 2, and

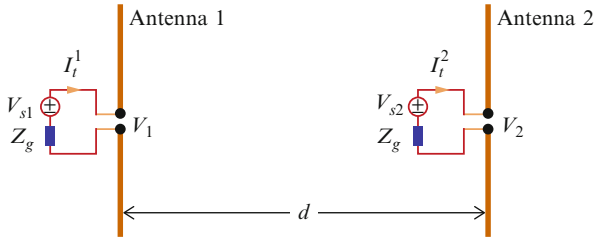


Fig. 8 A two-element transmitting dipole antenna array

DOI 10.1007/978-981-4560-75-7_13-2

$$Z_{12} = \left[\frac{-1}{I_t^1 I_t^2}\int_{\ell} I_1(z)E_{12}(z)d\ell\right] = \text{Conventional mutual impedance.} \tag{36}$$

Equations 28 and 27 show that the conventional mutual impedance Z_{12} measures the "*open-circuit voltage*" dropped across the terminal of antenna 1 due to the *scattered* field $E_{12}(z)$ produced by the current on antenna 2. On the other hand, the receiving mutual impedance Z_t^{12} defined in section "Theoretical and Experimental Methods for the Determination of the Receiving Mutual Impedance" for two receiving antennas measures the voltage dropped across the *terminal load* of antenna 1 due to the *total* field produced by the current on antenna 2. Thus, the terminal voltage across the terminal load of antenna 1 can be written as

$$V_1 = Z_t^{12} I_t^2 + U_t^1. \tag{37}$$

Except this difference, another more fundamental difference can be seen from Eqs. 27 and 29. For the receiving mutual impedance in Eq. 29, the coupled voltage term $Z_t^{12} I_t^2$ truly characterizes the *complete* mutual coupling effect on antenna 1 due to the current on antenna 2. Hence when this coupled voltage $Z_t^{12} I_t^2$ is taken away from the terminal voltage V_1, the remaining voltage U_t^1 is the truly isolated voltage across the terminal load of antenna 1. That is, U_t^1 is the voltage on antenna 1 when antenna 2 is taken away (completely absent) from the array. On the other hand, this is not the case with the conventional mutual impedance in Eq. 27 for the transmitting dipole antenna array. The coupled voltage term $I_t^2 Z_{12}$ in Eq. 27 is *not* the *complete* mutual coupling effect on antenna 1 due to the current on antenna 2. This is because this term only accounts for the coupled voltage on antenna 1 due to the scattered field from antenna 2, $E_{12}(z)$. The total coupled voltage should also include the additional coupled voltage on antenna 1 produced by the additional current on antenna 1 (on top of the original isolated current on antenna 1 when it stands alone) that is induced by the current on antenna 2. Because of this, when the coupled voltage $I_t^2 Z_{12}$ in Eq. 27 is taken away from the total terminal voltage V_1, the remaining voltage $I_t^1 Z_{11}$ does not represent the truly isolated voltage across the terminal of antenna 1 as it were alone. That means the uncoupled voltage $I_t^1 Z_{11}$ is different from the voltage on antenna 1 when antenna 2 is taken away from the array. This difference is not trivial as in many antenna array applications, it is required to have the ideal undistorted array signals (ideal array manifold), i.e., the isolated voltages, $U_t^1, U_t^2, \cdots, U_t^n$, and these can only be obtained by using the receiving mutual impedances. This difference is a fundamental one.

The Application of the Receiving Mutual Impedance in the Direction-of-Arrival (DOA) Estimation

The application of the concept of receiving mutual impedance in the direction-of-arrival (DOA) estimation using an antenna array is to be demonstrated in this section. Typical DOA estimation algorithms, such as MUSIC (Schmidt 1986; Schmidt and Franks 1986) and ESPRIT (Roy and Kailath 1989), are known to be highly sensitive to the errors in the array manifold. In practical antenna arrays, one common error in the array manifold is the mutual coupling effect among the antenna elements. When there is mutual coupling, compensation methods are usually required in order to restore the correct array manifold. There have been many compensation methods proposed before (e.g., Hui 2003; Gupta and Ksienski 1983; Dandekar et al. 2002; Lau et al. 2004; Yang and Nie 2005; Goossens and Rogier 2007; Ye and Liu 2008; Wu and Nie 2009; Hui 2007; Yu et al. 2011; Yu and Hui 2011; Niow and Hui 2012; Yeh et al. 1989). Among these

DOI 10.1007/978-981-4560-75-7_13-2

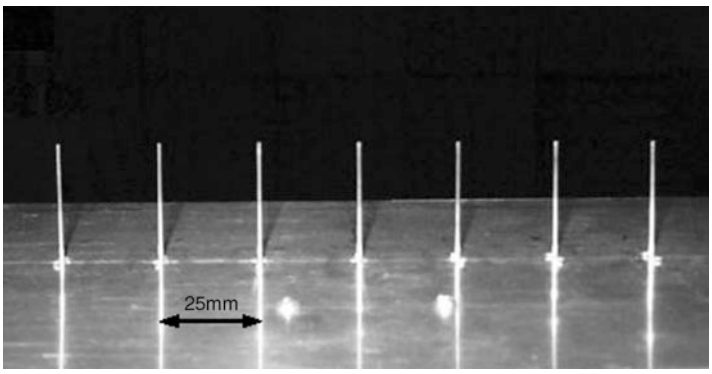


Fig. 9 The seven-element monopole antenna array above a metallic ground plane. The monopole antennas have a length of 30 mm, a wire diameter of 1 mm, and a separation between adjacent elements of 25 mm and work at 2.4 GHz

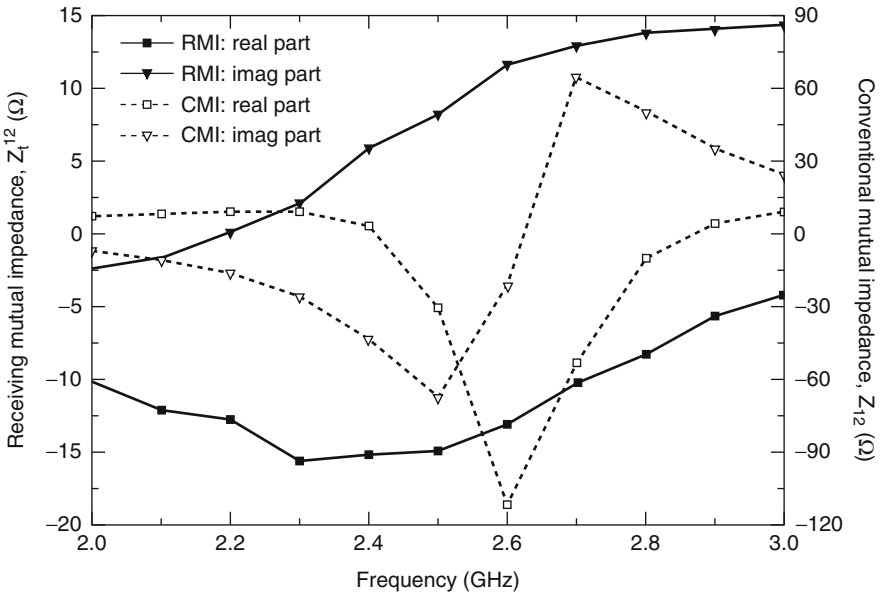


Fig. 10 The measured receiving mutual impedance (RMI) Z_t^{12} and the conventional mutual impedance (CMI) Z_{12} of the monopole array shown in Fig. 9

compensation methods, the one proposed originally in Gupta and Ksienski (1983) which uses the conventional mutual impedances to characterize the mutual coupling effect is probably the most popular one and can be termed as the conventional mutual impedance method (CMIM). However, CMIM uses the conventional mutual impedances to characterize the mutual coupling effect in receiving antenna arrays (direction finding arrays) and hence suffers from the drawback of inaccuracy as it fails to recover the isolated voltages (the ideal array manifold) on the antenna elements as explained in section "Comparison of the Receiving Mutual Impedance and the Conventional Mutual Impedance." In this section, the use of receiving mutual impedances to improve the accuracy of DOA estimation is illustrated through several examples.

In the Presence of Uncorrelated Array Noise

As shown in Fig. 9, a seven-monopole array is constructed to operate at 2.4 GHz to detect two coherent signals. The dimensions of the individual monopole antennas are optimized to be 30 mm (0.24λ), and the

DOI 10.1007/978-981-4560-75-7_13-2

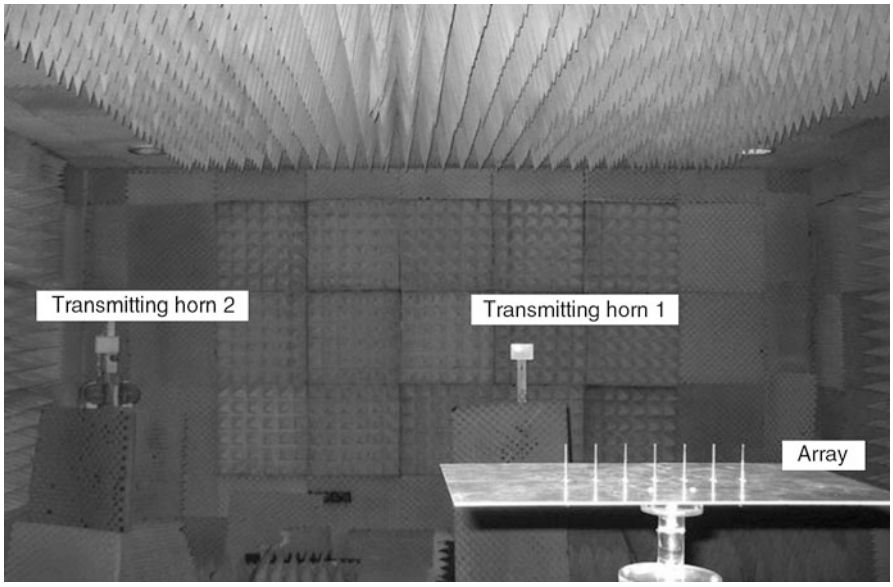


Fig. 11 The experimental setup for the two-source DOA estimation inside an anechoic chamber

diameter of the monopoles is 1 mm. The separation between adjacent monopole elements is 25 mm (0.2λ at 2.4 GHz). An isolated monopole antenna with these dimensions has a return loss of about 16 dB at 2.4 GHz. The measured receiving mutual impedances Z_t^{12} (the receiving mutual impedance between the first and the second elements) are shown in Fig. 10. The corresponding measured conventional mutual impedances are also shown for comparison. The conventional mutual impedances are obtained from the measurement of the two-port S parameters of the pair of antenna elements. The experimental setup is shown in Fig. 11. Two coherent sources are simulated by two transmitting horn antennas inside the anechoic chamber. The two horn antennas are fed from a VNA (vector network analyzer) output port through a power divider such that the powers' output to the two horn antennas is equal. The measured S_{21} parameter, which indicates the relative power received by the monopoles, is converted into voltages (see Eq. 27). In this way, all the received voltages V_i $(i = 1,\ 2, \cdots, 7)$ on the antenna elements can be measured one by one. The angular separation between the two horn signals is varied from about 50–30°. Using Yeh's method (Yeh et al. 1989) for DOA estimation with the MUSIC algorithm, a covariance matrix $\mathbf{R}$ can be formed as follows:

$$\mathbf{R} = E\{\mathbf{v}(t)\mathbf{v}^T(t)\} = \frac{1}{2}\mathbf{V}\mathbf{V}^H \tag{38}$$

where $E\{\}$ is the expectation operator, $\mathbf{v}(t)$ is the time-domain vector of the antenna terminal voltages (including noise), $\mathbf{V}$ is the corresponding phasor representation of $\mathbf{v}(t)$, and "^{H}" is the conjugate transpose operator.

Note that in Eq. 30, the time-domain expectation values of the covariance matrix ($E\{\mathbf{v}(t)\mathbf{v}^T(t)\}$) elements are replaced by their equivalent phasor-domain values (the elements of $\mathbf{V}\mathbf{V}^H$) which are exactly the output from the VNA machine. The equivalence between the time-domain expectation operation and the multiplication operation in the phasor-domain can be obtained from many textbooks, for example, Cheng's book (Cheng 1983). Once $\mathbf{R}$ is obtained, a transformed covariance matrix $\mathbf{R}'$ is formed according to Yeh's method (Yeh et al. 1989):

DOI 10.1007/978-981-4560-75-7_13-2

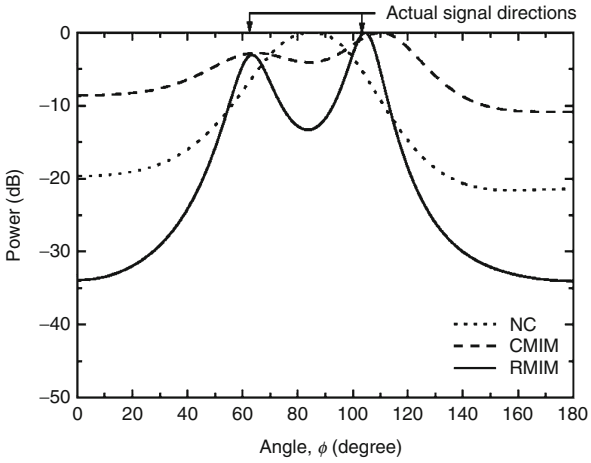


Fig. 12 The MUSIC spectra for the estimation of two sources at $\phi_1 = 62.4°$ and $\phi_2 = 103.2°$ with an angular separation of 40.8°. The spectra were obtained with the mutual coupling effect being compensated by different methods: (i) with no compensation (NC), (ii) using the conventional mutual impedance method (CMIM), and (iii) using the receiving mutual impedance method (RMIM). The averaged SNR of the received voltages is 34.6 dB (Taken from (Yu et al. 2011)

$$\mathbf{R}' = \mathbf{ZRZ}^H = \frac{1}{2}\mathbf{ZVV}^H\mathbf{Z}^H = \frac{1}{2}\mathbf{UU}^T + \frac{1}{2}\sigma^2\mathbf{ZZ}^T \tag{39}$$

where $\mathbf{U}$ is the vector for the compensated signal voltages, σ^2 is the noise power received at the antenna terminals, and $\mathbf{Z}$ is the mutual impedance matrix on the left-hand side of Eq. 22, with $n = 7$. That is,

$$\mathbf{Z} = \begin{bmatrix} 1 & -\frac{Z_t^{12}}{Z_L} & \cdots & -\frac{Z_t^{16}}{Z_L} & -\frac{Z_t^{17}}{Z_L} \\ -\frac{Z_t^{21}}{Z_L} & 1 & \cdots & -\frac{Z_t^{26}}{Z_L} & -\frac{Z_t^{17}}{Z_L} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{Z_t^{61}}{Z_L} & -\frac{Z_t^{62}}{Z_L} & \cdots & 1 & -\frac{Z_t^{67}}{Z_L} \\ -\frac{Z_t^{71}}{Z_L} & -\frac{Z_t^{72}}{Z_L} & \cdots & -\frac{Z_t^{76}}{Z_L} & 1 \end{bmatrix}. \tag{40}$$

The reason to form the transformed covariance matrix $\mathbf{R}'$ is that it can be used in the spatial smoothing method for detecting two coherent sources. Note that the noise associated with $\mathbf{R}'$ in Eq. 31 is correlated now ($\sigma^2\mathbf{ZZ}^T$) although the noise itself is assumed to be uncorrelated among the antenna elements. The correlation in noise is due to antenna mutual coupling, and this will affect the MUSIC operation to some extent. (In the next section, the case with spatial correlated noise will be investigated, i.e., the case in which the array noise is itself correlated even before received by the array.) With the transformed covariance matrix $\mathbf{R}'$, a generalized eigenvalue decomposition was performed to enable the construction of a MUSIC spatial spectrum function (Yeh et al. 1989).

Figure 12 shows the MUSIC estimation results when the two sources are coming from angles of $\phi_1 = 62.4°$ and $\phi_2 = 103.2°$ with an angular separation of 40.8°. It can be seen that even at this relatively large angular separation, both the NC (no compensation) and CMIM cases fail to correctly detect the two source directions. On the other hand, when the RMIM (receiving mutual impedance method) was used, the MUSIC spectrum correctly indicates the source directions with relatively small errors. The results in

DOI 10.1007/978-981-4560-75-7_13-2

Fig. 12 are obtained at an SNR of 39.1 dB and averaged over the antenna elements with a standard deviation of 7.0 dB. More details can be obtained from Yu's paper (Yu et al. 2011).

In the Presence of Correlated Array Noise: An Intensive Treatment of Array Noise

Recent studies (Torrieri and Bakhru 1997; Zhao and Zhang 2000; Craeye et al. 2004; Kisliansky et al. 2007) have shown that under practical situations, array noise tends to be correlated, even at a moderate antenna element separation such as half wavelength. Because of this, many modified DOA estimation methods have been proposed to tackle the problem of correlated noise (e.g., Zhao and Zhang 2000; Prasad et al. 1988; Rajagopal and Rao 1991; Qi et al. 2005; Werner and Jansson 2005; Qi et al. 2007). In Prasad et al. (1988) and Zhao and Zhang (2000), a covariance differencing method was developed to remove the correlated noise covariance matrix which was characterized as a symmetric Toeplitz matrix. In Qi et al. (2005) and Qi et al. (2007), the spatial difference smoothing (SDS) method was developed, relying similarly on the Toeplitz nature of the correlated noise covariance matrix to handle the correlated noise. Most of the previous correlated noise handling methods did not consider the effect of antenna mutual coupling.

In the following, a new method is introduced to tackle correlated noise in DOA estimations in the presence of antenna mutual coupling. Unlike the previous methods in Craeye et al. (2004) and Kisliansky et al. (2007) which considered noise as an internal and an external parts, in this method, the array noise is divided into a coupled and an uncoupled components whose origins can be traced back into the array environment and the internal circuitry of the antenna elements. Through this division of noise, the coupled and uncoupled noise components can be handled separately. More importantly, it is shown that antenna mutual coupling in the coupled noise component can be compensated in the same way as the signals, leading to a simple but very effective MUSIC DOA estimation algorithm.

Coupled Array Noise Modeling Using Receiving Mutual Impedances

Consider an antenna array with n elements employed in DOA estimation in the presence of both external and internal noise. The external noise is the noise that comes from the environment of the array, while the internal noise is the noise generated inside the antenna elements. The received array signal vector $\mathbf{x}(t)$ at time t is given by

$$\mathbf{x}(t) = \mathbf{v}(t) + \mathbf{n}_c(t) + \mathbf{n}_u(t) \tag{41}$$

where $\mathbf{v}(t)$ is an $n \times 1$ column vector for the signal voltages due to the incoming signals, $\mathbf{n}_c(t)$ is an $n \times 1$ column vector for the coupled noise voltages, and $\mathbf{n}_u(t)$ is an $n \times 1$ column vector for the uncoupled noise voltages. $\mathbf{v}(t)$, $\mathbf{n}_c(t)$, and $\mathbf{n}_u(t)$ are assumed to be mutually uncorrelated. Note that all the signals and noise are expressed in their analytic forms unless stated otherwise. Here the division of noise into the coupled and uncoupled components has an advantage for the MUSIC DOA estimation algorithm, especially when compensation for antenna mutual coupling effect is required. This will be explained in the next section (section "MUSIC DOA Estimation Algorithm with Coupled Array Noise").

For an antenna array, noise mainly comes from two physical sources: external environment and the internal circuitries (Craeye et al. 2004). The external environmental noise source induces coupled noise currents on the antenna elements and generates the coupled noise voltages $\mathbf{n}_c(t)$ in Eq. 33. The internal circuit noise sources (mainly the input stages of the low-noise amplifiers (LNAs) or the resistive terminal loads of the antennas) produce noise currents on the antenna elements which consist of two parts: the uncoupled noise current and the coupled noise current. For example, in the kth antenna, its internal circuit noise source excites an uncoupled noise current $i_{u,kk}(t)$ on itself as when it is in the isolated mode. The fields from this current in turn induce coupled noise currents $i_{c,pk}(t)$ ($p \neq k,$ and $p = 1, 2, \cdots, n$) on all the

DOI 10.1007/978-981-4560-75-7_13-2

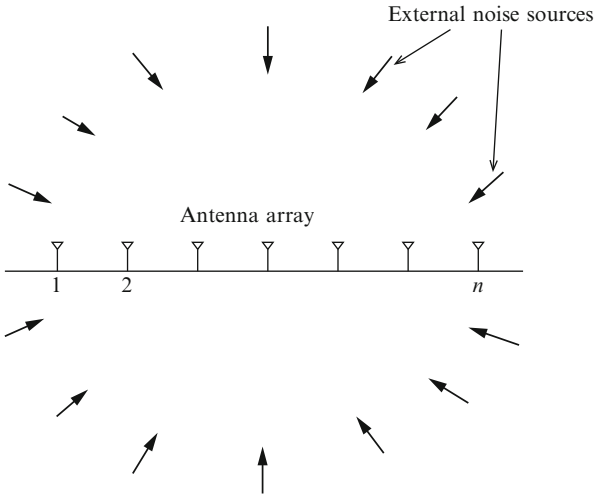


Fig. 13 The scattering field model for the generation of coupled noise with s plane EM waves coming from random directions to an antenna array

other antenna elements. It is assumed that the coupled and uncoupled currents excited by the internal noise source of a particular antenna are weakly correlated as they appear on different antennas. Furthermore, the internal noise currents of different antennas are uncorrelated. When all the internal circuit noise sources of the antenna array are considered in this way, the uncoupled noise currents $i_{u,kk}(t)$ $(k = 1, 2, \cdots, n)$ give rise to the uncoupled noise voltages $\mathbf{n}_u(t)$ in Eq. 33, while the coupled noise currents $i_{c,pk}(t)$ give rise to coupled noise which is modeled in the same way as (see below) and as part of the environmental noise and contribute to total coupled noise voltages $\mathbf{n}_c(t)$ in Eq. 33.

To model the coupled noise coming from the outside environment of the antenna elements, a scattering field model as shown in Fig. 13 can be used. The coupled noise voltages $\mathbf{n}_c(t)$ are considered to be produced by a large number of external random noise sources in the form of scattered electromagnetic (EM) plane waves with random amplitudes and phases. These random EM plane waves give rise to $\mathbf{n}_c(t)$ through the antenna mutual impedance matrix $\mathbf{Z}$ as,

$$\mathbf{n}_c(t) = \mathbf{Z}^{-1}\mathbf{e}(t) \tag{42}$$

where $\mathbf{e}(t)$ is an $n \times 1$ column vector of the random noise voltages at the antenna elements when the antenna mutual coupling is not considered. The reason to express the coupled noise voltages in the form of Eq. 34 is that both $\mathbf{Z}$ and $\mathbf{e}(t)$ can be calculated or measured directly. As the antenna array is in the receiving mode, the elements of the antenna mutual impedance matrix $\mathbf{Z}$ in Eq. 34 are the receiving mutual impedances of the array which can be obtained by calculation or measurement (Hui 2004b). The form of $\mathbf{Z}$ is given by Eq. 32 for an n element case. In Eq. 34, the noise voltage vector $\mathbf{e}(t)$ without mutual coupling can be calculated using a standard numerical method such as MoM (Harrington 1993). Based on the couple noise model in Fig. 13, consider s narrow-band EM plane waves impinging on the n-element antenna array from directions ϕ_k $(k = 1, 2, \ldots s)$ at time t_i $(i = 1, 2, \ldots)$. The amplitudes and phases of these plane waves are random Gaussian variables, which are both independently generated for each direction and each time instance (Kisliansky et al. 2007). Assuming each antenna (considering dipole antennas here for ease of illustration) is divided into m segments in the MoM procedure, then the element $e_j(t_i)$ of the noise voltage vector $\mathbf{e}(t_i)$ at time instant $t = t_i$ can be calculated as follows:

DOI 10.1007/978-981-4560-75-7_13-2

$$\begin{bmatrix} z_{j,11} & z_{j,12} & \cdots & z_{j,1m} \\ z_{j,21} & z_{j,22} & \cdots & z_{j,2m} \\ \vdots & \vdots & \ddots & \vdots \\ z_{j,m1} & z_{j,m2} & \cdots & z_{j,mm} \end{bmatrix} \begin{bmatrix} I_{j,1}(t_i) \\ I_{j,2}(t_i) \\ \vdots \\ I_{j,k}(t_i) \\ \vdots \\ I_{j,m}(t_i) \end{bmatrix} = \begin{bmatrix} n_{j,11}(t_i) + n_{j,12}(t_i) + \cdots + n_{j,1s}(t_i) \\ n_{j,21}(t_i) + n_{j,22}(t_i) + \cdots + n_{j,2s}(t_i) \\ \vdots \\ n_{j,m1}(t_i) + n_{j,m2}(t_i) + \cdots + n_{j,ms}(t_i) \end{bmatrix} \tag{43}$$

$$e_j(t_i) = I_{j,1}(t_i) Z_L \tag{44}$$

where the first matrix on the left-hand side of Eq. 35 is the MoM impedance matrix for the *j*th antenna and the second column vector on the left-hand side of Eq. 35 is the current distribution for the *j*th antenna. The column vector on the right-hand side of Eq. 35 is the voltage vector due to the *s* noise source plane waves impinging on each segment of the *j*th antenna. Equation 36 simply calculates the noise voltage of the *j*th antenna by multiplying the current on the terminal segment (first segment) with the antenna terminal load Z_L. The noise voltages of all other antenna elements can be calculated in the same way. Once $\mathbf{e}(t_i)$ is known, the covariance matrix $\mathbf{N}_c$ of the coupled noise voltage vector can be expressed as

$$\mathbf{N}_c = E\{\mathbf{n}_c(t)\mathbf{n}_c^H(t)\} = \mathbf{Z}^{-1} E\{\mathbf{e}(t)\mathbf{e}^H(t)\}\mathbf{Z}^{-H} \tag{45}$$

where the noise correlation matrix $E\{\mathbf{e}(t)\mathbf{e}^H(t)\}$ represents only the spatial correlation and its elements are known for many typical wave distributions, such as the Bessel function for a uniform incident wave distribution (Clarke 1968) received by a uniform linear array (ULA).

To model the uncoupled noise voltages $\mathbf{n}_u(t)$, the equivalent circuit of a receiving antenna element as shown in Fig. 14 is considered. In this equivalent circuit, Z_{ant} is the input impedance of the antenna, and Z_{in} is the impedance looking into the input terminal of the LNA, which is connected to the antenna terminal (Craeye et al. 2004). The two sources, V_n and I_n, are the equivalent noise voltage and current generated by the noise resistance of the LNA. The uncoupled noise voltage generated by V_n and I_n is the noise voltage dropped across the input impedance Z_{in} of the LNA when this antenna element is receiving in an isolated mode. For example, for the *p*th antenna element, uncoupled noise voltage is

$$n_{u,p} = \frac{V_n Z_{in}}{Z_{ant} + Z_{in}} + \frac{I_n Z_{ant} Z_{in}}{Z_{ant} + Z_{in}}. \tag{46}$$

Since $n_{u,p}$ represents the *uncoupled* noise components, the covariance matrix of $\mathbf{n}_u(t)$ is a diagonal matrix

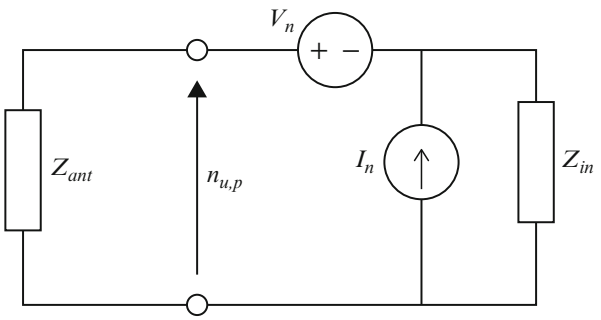


Fig. 14 The circuit noise model for the *p*th antenna element connected to an LNA with an input impedance of Z_{in} and noise sources V_n and I_n

whose diagonal elements are all the same and can be calculated from the following equation (cf. eq. (19) in Craeye et al. 2004):

$$\begin{aligned}\mathbf{N}_u &= E\{\mathbf{n}_u(t)\mathbf{n}_u^H(t)\} \\ &= \frac{4k_BT_0B|Z_{in}|^2}{|Z_{ant}+Z_{in}|^2}\left(R_n + |Z_{ant}|^2G_n + 2\mathrm{Re}\left\{Z_{ant}^*Y_\gamma^*\right\}R_n\right)\mathbf{I}_{n\times n}\end{aligned} \tag{47}$$

where k_B is the Boltzmann's constant, T_0 is the 290 K standard noise temperature, B is the instantaneous bandwidth of observation, and $\mathbf{I}_{n\times n}$ is the $n \times n$ identity matrix. The other symbols R_n, G_n, and Y_γ in Eq. 39 are the noise resistance, conductance, and admittance, respectively, of the LNA whose values can be obtained from the data sheet of the LNA. The uncoupled noise power σ_u^2 dissipated in the respective LNAs can be calculated by dividing the magnitude squares of the diagonal elements of Eq. 39 by Z_{in}^* and taking the real part. The advantage of formulating the uncoupled noise covariance matrix in Eq. 39 is that it can be determined beforehand by calculation using known parameters in Eq. 39 or by a direct measurement on an LNA. This is an important observation which enables us to subtract the uncoupled noise covariance matrix $\mathbf{N}_u$ from the received signal-plus-noise covariance matrix, as shown in the next section.

MUSIC DOA Estimation Algorithm with Coupled Array Noise

Using the noise covariance matrices in Eqs. 37 and 39, the covariance matrix of the received signal with noise is

$$\mathbf{R} = E\{\mathbf{x}(t)\mathbf{x}^H(t)\} = E\{\mathbf{v}(t)\mathbf{v}^H(t)\} + \mathbf{N}_c + \mathbf{N}_u. \tag{48}$$

As explained in the last section, the uncoupled noise covariance matrix $\mathbf{N}_u$ can be determined beforehand. Hence after the signal-plus-noise covariance matrix is obtained, the following difference covariance matrix can be calculated immediately:

$$\begin{aligned}\mathbf{R}_d = \mathbf{R} - \mathbf{N}_u &= E\{\mathbf{v}(t)\mathbf{v}^H(t)\} + \mathbf{N}_c \\ &= E\{\mathbf{v}(t)\mathbf{v}^H(t)\} + \mathbf{Z}^{-1}E\{\mathbf{e}(t)\mathbf{e}^H(t)\}\mathbf{Z}^{-H}\end{aligned} \tag{49}$$

Now the noise covariance matrix in Eq. 41 contains only the coupled noise. This is a great advantage as the signal covariance matrix and noise covariance matrix can be decoupled at the same time. Furthermore, as the signals $\mathbf{v}(t)$ are subjected to the same coupling process as the coupled noise $\mathbf{n}_c(t)$, the coupled signal voltage vector can also be expressed in terms of the mutual impedance matrix $\mathbf{Z}$ as (cf. Eq. 34)

$$\mathbf{v}(t) = \mathbf{Z}^{-1}\mathbf{u}(t) \tag{50}$$

where $\mathbf{u}(t)$ is the uncoupled signal voltage vector (cf. Eq. 22). By applying Eq. 42 into Eq. 41 and multiplying Eq. 41 on the left and right-hand sides with $\mathbf{Z}$ and $\mathbf{Z}^H$, respectively, the following decoupled signal-plus-noise covariance matrix is obtained:

$$\begin{aligned}\mathbf{R}_d' &= \mathbf{Z}\mathbf{R}_d\mathbf{Z}^H \\ &= E\{\mathbf{u}(t)\mathbf{u}^H(t)\} + \left(\sigma_c^2/\mathrm{Re}\{1/Z_{in}\}\right)E\{\mathbf{s}(t)\mathbf{s}^H(t)\}\end{aligned} \tag{51}$$

where

DOI 10.1007/978-981-4560-75-7_13-2

$$\mathbf{s}(t)\mathbf{s}^H(t) = \mathbf{e}(t)\mathbf{e}^H(t)/\left(\sigma_c^2/\mathrm{Re}\{1/Z_{in}\}\right) \tag{52}$$

is the normalized spatial noise correlation matrix and σ_c^2 is the power of the coupled noise at each element when there is no spatial correlation. Note that Eq. 43 is an important new result in array signal theory for DOA estimation. In Eq. 43, the correct uncoupled array manifold has been restored in $\mathbf{u}(t)$. The decoupled noise covariance matrix $E\{\mathbf{e}(t_i)\mathbf{e}^H(t_i)\}$ is not uncorrelated as previous studies assumed. Instead, noise is still spatially correlated though the antenna mutual coupling effect has been removed. This is due to the random distribution of the incoming noise directions. The exact noise correlation structure of $E\{\mathbf{e}(t_i)\mathbf{e}^H(t_i)\}$ depends on the signal environment. Furthermore, for the MUSIC DOA estimation algorithm, the method used in Yeh et al. (1989; see eq. (13) in Yeh et al. 1989) can still be directly used for either noncoherent or coherent signals by the following changes:

$$\begin{aligned} SAS^+ &\rightarrow E\{\mathbf{u}(t)\mathbf{u}^H(t)\}, \\ \sigma^2 &\rightarrow \frac{\sigma_c^2}{\mathrm{Re}\{1/Z_{in}^H\}}, \\ \mathbf{Z}_0\mathbf{Z}_o^+ &\rightarrow E\{\mathbf{s}(t)\mathbf{s}^H(t)\}. \end{aligned} \tag{53}$$

In the next section, it will demonstrate the improved performance of using Eq. 43 in MUSIC DOA estimation.

Simulation Examples

Coupled Array Noise To investigate the correlation of the coupled noise, the magnitudes of the noise current correlation coefficients (Kisliansky et al. 2007) between two center-fed half-wavelength dipole antennas at different antenna separations, d, are recalculated using a numerical method. The dipole wire radius is $\lambda/500$ (where λ is the wavelength in free space), and the two dipoles are shorted through their terminals, i.e., $Z_L = 0$. In the numerical method, an MoM computer simulation tool, FEKO (http://www.feko.info/), is used. Each dipole is discretized into 63 segments, each of which has a length of $\lambda/125$. To simulate the coupled noise, 50 plane waves are used to excite the array. These plane waves have random amplitudes and phases and are vertically polarized. They come uniformly from the horizontal directions along the plane $\theta = \pi/2$. This number of plane waves is obtained after several convergence tests. The results are shown in Fig. 15 and compared with the results obtained in Kisliansky et al. (2007). It can be seen that these results are almost the same as those in Kisliansky et al. (2007).

For the normalized spatial noise correlation matrix in Eq. 44, it is well known that its elements follow the variation of the Bessel function $J_0(kd)$ across the array if the noise comes uniformly on the horizontal plane. This result can be demonstrated using a modeling method (as mentioned below). Shown in Fig. 16 is the calculated spatial correlation coefficients of the coupled noise for different elements with respect to the first element in a ten-element dipole antenna array with an element separation d equal to 0.5λ. The variation of the Bessel function is also shown for comparison. The dipole antennas are same as the two dipole antennas in Fig. 15, but now all of them are connected to a 50Ω terminal load. The same numerical scheme is used and the coupled noise is generated in the same way as in Fig. 15. After the noise voltages are obtained, they are decoupled as in Eq. 43 using the receiving mutual impedance matrices $\mathbf{Z}$ and $\mathbf{Z}^H$. From the result in Fig. 16, it can be seen that the noise spatial correlation of the antenna elements follows almost exactly as the Bessel function.

To illustrate the variation of the noise power across the elements of an antenna array due to antenna mutual coupling, the noise powers of a seven-element dipole array with different element separations are

DOI 10.1007/978-981-4560-75-7_13-2

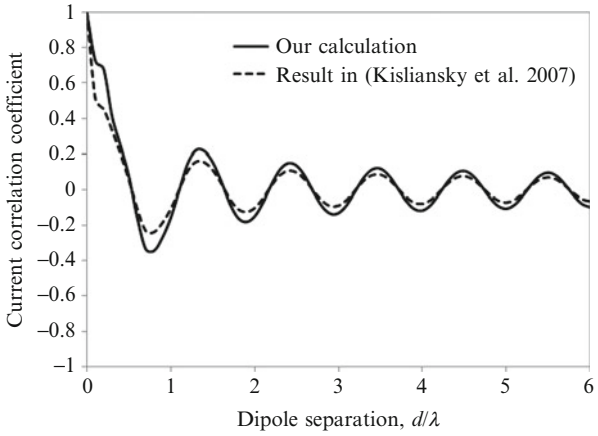


Fig. 15 The magnitude of the noise current correlation coefficient between two dipole antennas at different antenna separations

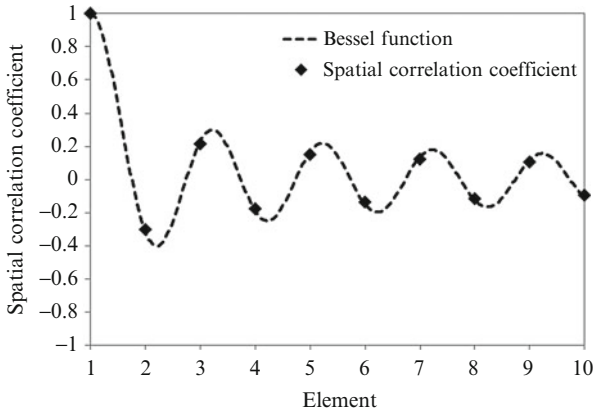


Fig. 16 The normalized spatial correlation of the coupled noise for a ten-element dipole array with element separation $d = 0.5\lambda$

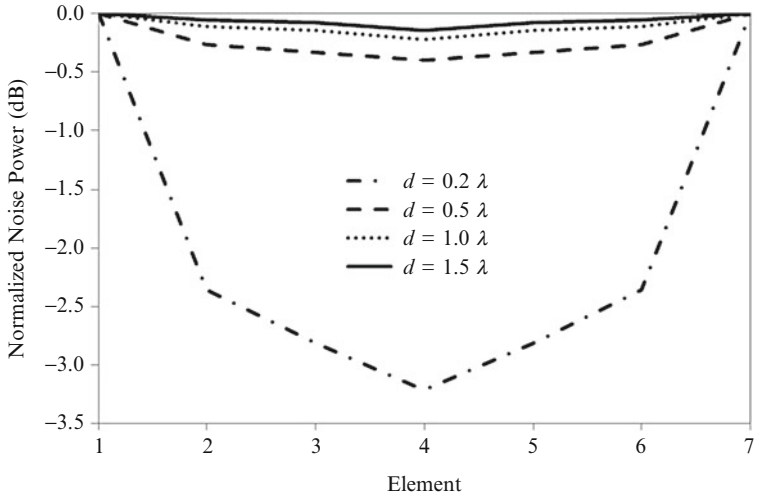


Fig. 17 The normalized noise power for a seven-element dipole array at different element separation

DOI 10.1007/978-981-4560-75-7_13-2

calculated. The dipoles are the same as the dipoles in Fig. 15 and the coupled noise is generated in the same way. The results are shown in Fig. 17 with the noise powers being normalized to those of the edge elements. It can be seen that when the element separation is decreased, the element noise powers differ significantly. At an element separation of $d = 0.2\lambda$, the power difference between the central element and the edge elements is more than 3 dB. On the other hand, when the element separation is increased, the element noise powers tend to be equalized. The results in this figure indicate the importance of the antenna mutual coupling on the distribution of the noise power across an antenna array.

DOA Estimation Experiments In the first DOA estimation experiment, the seven-element array studied in Fig. 17 is used to detect two coherent signals in the presence of coupled and uncoupled noise. The terminal voltages of the dipoles are calculated by FEKO and processed by the MUSIC DOA algorithm. In the literature, Yeh et al. (1989) have provided a DOA study on such a dipole array before. But they considered only the uncoupled noise. In the simulation of the signal environment, the coupled noise is generated as described above. To determine the uncoupled noise, the parameters in Eq. 39 are $Z_{in} = Z_L = 50\Omega$, $Z_{ant} = 70.05 - j8.73\Omega$, $T_0 = 290$ K, $R_n = 50\Omega$, $G_n = 0$, and $Y_\gamma = 0$. That is, the antenna terminal loads, Z_L, are treated as a one-port network. With these noise parameters, the uncoupled noise power dissipated at each antenna terminal load Z_{in} is calculated to be $\sigma_u^2 = 2.76 \times 10^{-21}$ W/Hz, which is equivalent to an effective noise temperature of 200 K. For illustration purpose, it is also assumed that the coupled noise power (when there is no antenna mutual coupling) has the value of $\sigma_c^2 = 2.76 \times 10^{-21}$ W/Hz at a 1 Hz bandwidth. Table 2 shows the coupled and uncoupled noise powers in each element at different element separations under this situation. The coupled noise powers in the third and the fourth columns of Table 2 are the diagonal elements of the matrix $\mathbf{N}_c$ in Eq. 37. The detection results are shown in Fig. 18a and b for element separations of $d = 0.5\lambda$ and $d = 0.2\lambda$, respectively. The MUSIC spectra in these two figures are produced with two cases of noise: (i) coupled noise (with the uncoupled noise removed as in Eq. 41) and (ii) uncoupled noise only (the conventional case). In these two cases, the average noise powers are made equal for the sake of comparison, and the SNR is specified at 3 dB with respect to the average coupled noise power σ_c^2. The two signals are coming from the directions of $\phi = 0°$ and 15°, and both are on the horizontal plane ($\theta = 90°$). In Fig. 18a, the result obtained in Yeh et al. (1989) is also shown. It can be seen that in both Fig. 18a and b, the results using the coupled noise (see Eq. 43) are more accurate and have much higher peaks than the results using only the uncoupled noise. Furthermore, in Fig. 18a, both the coupled and the uncoupled cases produce results more accurately than the result in Yeh et al. (1989). The reason is that they both use the receiving mutual impedances in the mutual coupling compensation, whereas the results in Yeh et al. (1989) were obtained using the conventional mutual impedances for the mutual coupling compensation.

Table 2 Noise powers in each element of the seven-element dipole antenna array

Element	Uncoupled noise power σ_u^2 (W/Hz)	Coupled noise power at $d = 0.2\lambda$ (W/Hz)	Coupled noise power at $d = 0.5\lambda$ (W/Hz)
1	2.76×10^{-21}	1.99×10^{-21}	2.41×10^{-21}
2	2.76×10^{-21}	1.16×10^{-21}	2.26×10^{-21}
3	2.76×10^{-21}	1.05×10^{-21}	2.23×10^{-21}
4	2.76×10^{-21}	9.58×10^{-22}	2.23×10^{-21}
5	2.76×10^{-21}	1.05×10^{-21}	2.23×10^{-21}
6	2.76×10^{-21}	1.16×10^{-21}	2.26×10^{-21}
7	2.76×10^{-21}	1.99×10^{-21}	2.41×10^{-21}

DOI 10.1007/978-981-4560-75-7_13-2

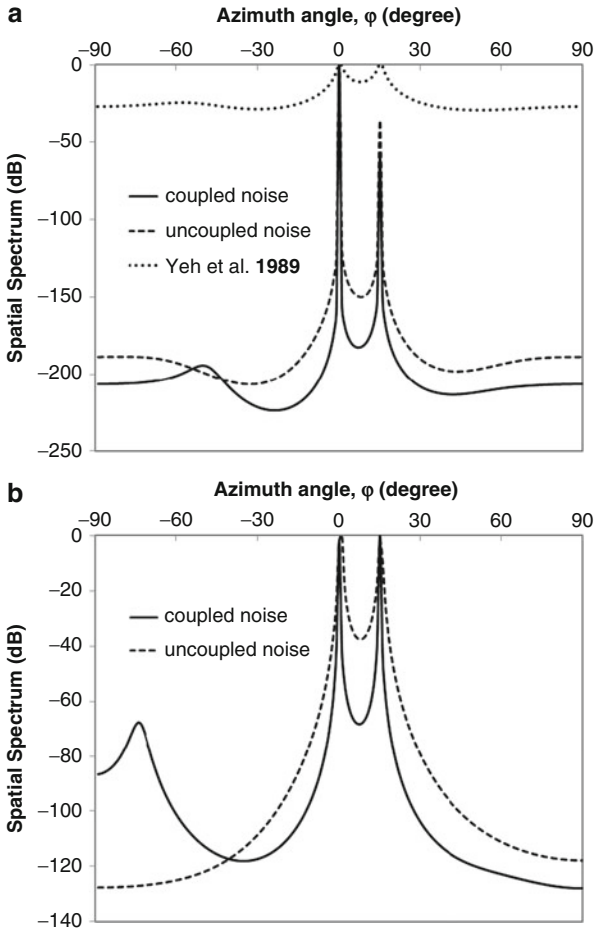


Fig. 18 The MUSIC spectra for the detection of two coherent signals from $\phi = 0°$ and 15° using coupled noise and uncoupled noise with a seven-element dipole array at an SNR = 3 dB; (**a**) $d = 0.5\lambda$, and (**b**) $d = 0.2\lambda$

In the second experiment, the sensitivity of the MUSIC DOA estimation algorithm to the error in the subtraction of the uncoupled noise in Eq. 41 is investigated. As explained earlier, the current method relies on the removal of the uncoupled noise covariance matrix $\mathbf{N}_u$ from the covariance matrix $\mathbf{R}$. Whereas $\mathbf{N}_u$ can be either calculated or measured beforehand, its accurate value may be subjected to errors, leading to the uncoupled noise power being over- or underestimated. The effect of this is shown in Fig. 19a, b, where Fig. 19a is for the case when the uncoupled noise power is overestimated and Fig. 19b is for the case when the uncoupled noise is underestimated. The dipole array and the signal environment are the same as those in the first DOA estimation experiment, and the element separation is 0.5λ. From Fig. 19a, it can be seen that if the uncoupled noise is overestimated, it will have a rather significant effect on the detection results. There are a number of secondary peaks in the MUSIC spectra which indicate wrong signal directions. The heights of these secondary peaks decrease with the increase in the uncoupled noise power overestimation (represented by an increasing percentage of the uncoupled noise power being subtracted from the coupled noise power). From Fig. 19b, it can be seen that an underestimation of the uncoupled noise power has the same significant effect as for the overestimation case. But the secondary peaks tend to move away from the two signal peaks. From these two figures, it can be inferred that an accurate determination of the uncoupled noise power is very important in the proposed method.

DOI 10.1007/978-981-4560-75-7_13-2

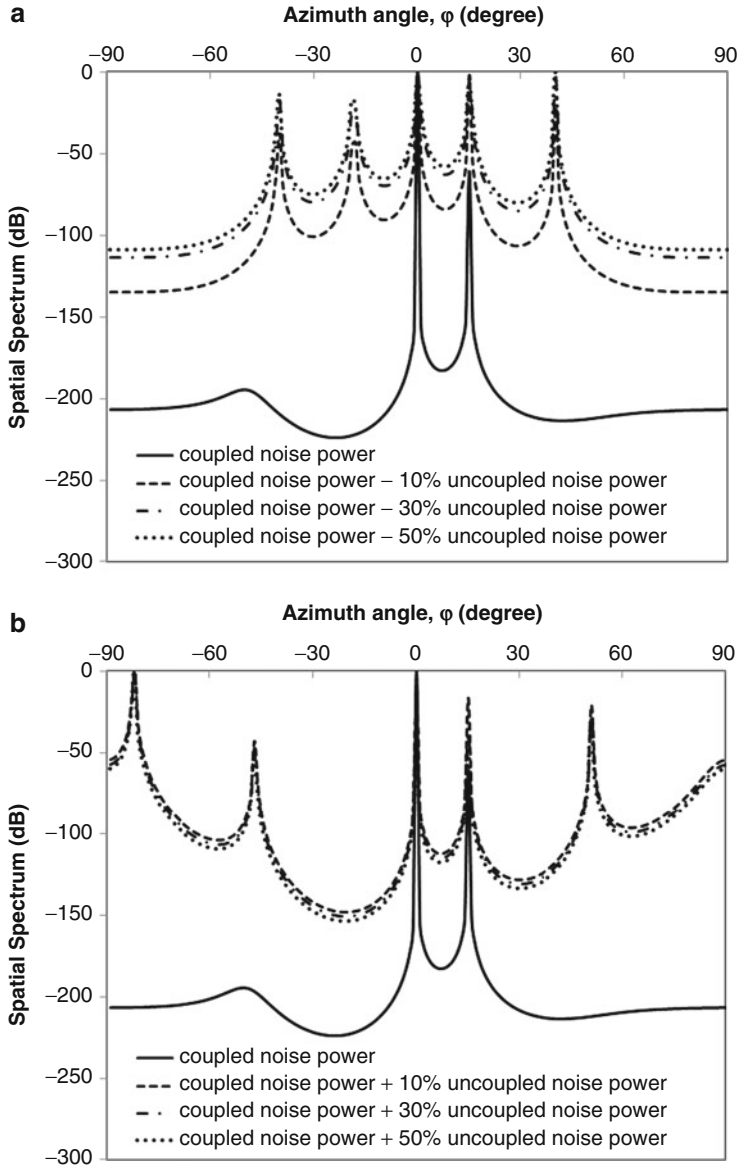


Fig. 19 The MUSIC spectra for the detection of two coherent signals from $\phi = 0°$ and 15° at an SNR = 3 dB; (**a**) uncoupled noise power being overestimated and (**b**) uncoupled noise power being underestimated. The dipole array is the same as that in Fig. 18

The Application of the Receiving Mutual Impedance in Interference Suppression

In adaptive nulling, it has been shown (Adve and Sarkar 2000) that antenna mutual coupling will significantly affect the depths and the accuracy of positioning the nulls of the radiation pattern toward the interferences. In this section, it is to investigate by using computer simulations the performance of a linear dipole array with seven antenna elements employed for adaptive nulling by using the concept of receiving mutual impedance to characterize and compensate for the mutual coupling effect. The performance will be compared with that using the conventional mutual impedances. The dipole array is the one shown in Fig. 1. The dimensions of the array are $n = 7$, $d = 0.5\lambda$, $\ell = 0.5\lambda$, $a = \lambda/200$, and $Z_L = 50\ \Omega$. The adaptive nulling algorithm used in the following examples is the direct data domain (DDD) adaptive

DOI 10.1007/978-981-4560-75-7_13-2

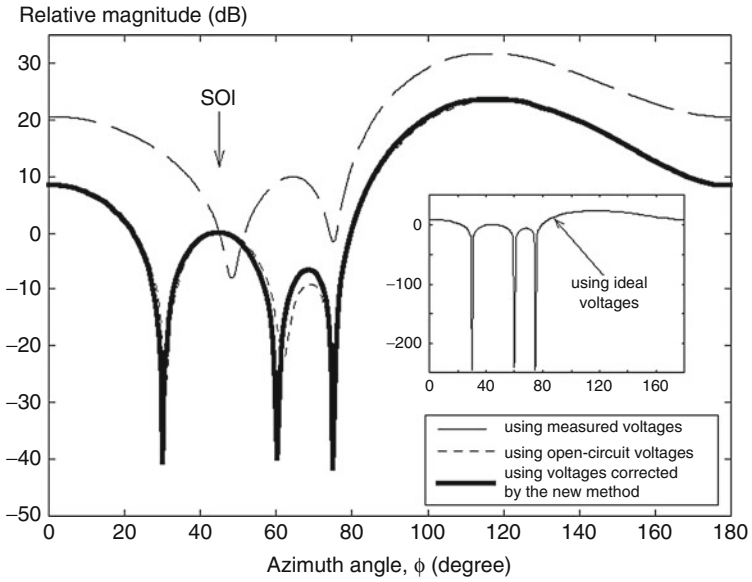


Fig. 20 The adaptive radiation patterns obtained by using the measured voltages, the open-circuit voltages, the voltages corrected by the new method (using receiving mutual impedances), and the ideal voltages. The signal environment is shown in Table 3

Table 3 The signal environment for the adaptive nulling simulation

	Amplitude (V/m)	Azimuth ϕ (degree)	Elevation θ (degree)
Signal of interest (SOI)	1.0 $\hat{z}$	45	90
Interference #1	1.0 $\hat{z}$	75	90
Interference #2	1.5 $\hat{z}$	60	90
Interference #3	2.0 $\hat{z}$	30	90

algorithm (Sarkar et al. 1996), which is simple and easy to implement. A detailed description of this algorithm can be found in Sarkar et al. (1996, 2000). The required inputs to the DDD adaptive algorithm are the voltages on the antenna terminal loads and the knowledge of the direction of arrival (DOA) of the signal of interest (SOI). The received voltages on the antenna terminal loads are first compensated for the mutual coupling effect and then passed to the adaptive algorithm. The signal environment is shown in Table 3. The SOI and interferences are plane waves linearly polarized in the z direction. This example has been studied before in Adve and Sarkar (2000). The result is shown in Fig. 20. Adaptive radiation patterns obtained by using four types of antenna terminal voltages are shown in Fig. 20. The dash-line radiation pattern is the result obtained by using the measured (or received) voltages across the antenna terminal loads which have not been compensated for the mutual coupling effect. The dotted-line radiation pattern is the result obtained by using the open-circuit voltages derived by using the conventional mutual impedances to compensate for the mutual coupling effect (Gupta and Ksienski 1983). The solid-line radiation pattern is the result obtained by using the receiving mutual impedances to compensate for the mutual coupling effect. The small graph inside Fig. 20 shows the radiation pattern obtained by using ideal voltages across the terminal loads which are completely free from any mutual coupling effect. Both the ideal voltages and the measured voltages are calculated by the moment method (Harrington 1993) with the current distribution on each antenna being expanded by 20 sinusoidal basis functions, and the Galerkin method is used in the matching procedure. From Fig. 20, it can be seen that the (*uncompensated*)

DOI 10.1007/978-981-4560-75-7_13-2

measured voltages give a wrong result and therefore can hardly be used to suppress the interferences. The open-circuit voltages can be used to locate the three interferences but the dips are not deep enough. On the other hand, when the compensated (corrected) voltages are obtained by using the receiving mutual impedances (new method), the dips are both more accurate and substantially deeper. For the ideal case, the dips are much deeper than all the other cases because it is a theoretically ideal situation in which the antenna elements are completely isolated from each other.

The above method has also been used to study other signal situations such as different signal elevation angles and different interference strengths. The details can be found in Hui (2004a).

The Application of the Receiving Mutual Impedance in Magnetic Resonance Imaging (MRI)

Phased arrays can be used in magnetic resonance imaging (MRI) to reduce the low signal-to-noise ratio (SNR) problem and the long scanning time problem (Hyde et al. 1986; Sodickson and Manning 1997; Roemer et al. 1990). However, many previous studies have revealed that mutual coupling effect makes it difficult to obtain independent undistorted signals from the MRI phased-array antenna elements. Obviously this problem is inherently more serious in MRI phased arrays than traditional arrays employed for receiving signals in the far-field region. This is because MRI phased arrays are virtually receiving in the near-field region with the element spacing in the order of ~0.1 wavelength, compared with that of the traditional arrays in the order of half wavelength. This is due to the low operation frequencies used in MRI. Thus, some traditional decoupling or compensation methods for the mutual coupling effect may not work effectively for MRI phased arrays. In view of this, some special compensation methods have been designed for MRI phased arrays, especially for coil arrays. Two such popular compensation methods are the overlapping of adjacent coils and the use of low input impedance preamplifiers (Roemer et al. 1990). However, constraints on the use of these two methods have been pointed out by some researchers (Ohliger et al. 2004). Moreover, overlapping of adjacent coils reduces the imaging area, and the use of low input impedance preamplifiers sacrifices the condition of maximum power transfer. Lumped-element decoupling networks have recently been suggested (Zhang and Webb 2004) to tackle this problem, but the decoupling power is still limited.

In this section, the concept of receiving mutual impedance is employed to solve the mutual coupling problem in the design of phased arrays for MRI. For mutual coupling analysis, it is sufficient to consider a two-coil phased array for MRI as shown in Fig. 21. The two coils are square in shape and made of a metallic strip with a side length of L and strip width s. The distance between the centers of the two coils is denoted by d, which can be changed to vary the amount of mutual coupling. The coils are aligned along the y axis on the y-z plane and are equidistant from the origin. The two coils are not exactly at the same plane but offset by a small separation dx in the x direction so that they can overlap with each other. There are four fixed capacitors C_1, C_2, C_3, and C_4 and one tuning capacitor C_t placed along the coils for smoothing the current distributions and making the coils to resonate at the desired frequency f_0, as shown in Fig. 21b. When the coils are placed alone, each of them resonates at the frequency f_0 with a very small self-resistance r. An external matching circuit is thus required to match the small resistance r to the system impedance of 50Ω, as shown in Fig. 21b. The imaged sample is assumed to be a rectangular slice (to simulate the active slice inside an imaged person) and placed in the region with negative x coordinates and close to the phased array as shown in Fig. 22. Its dimensions along the x, y, and z directions are denoted, respectively, by a, b, and c. The relative position of the active slice from the phased array is indicated by the distances of the active slice from the origin which are denoted by px, py, and pz,

DOI 10.1007/978-981-4560-75-7_13-2

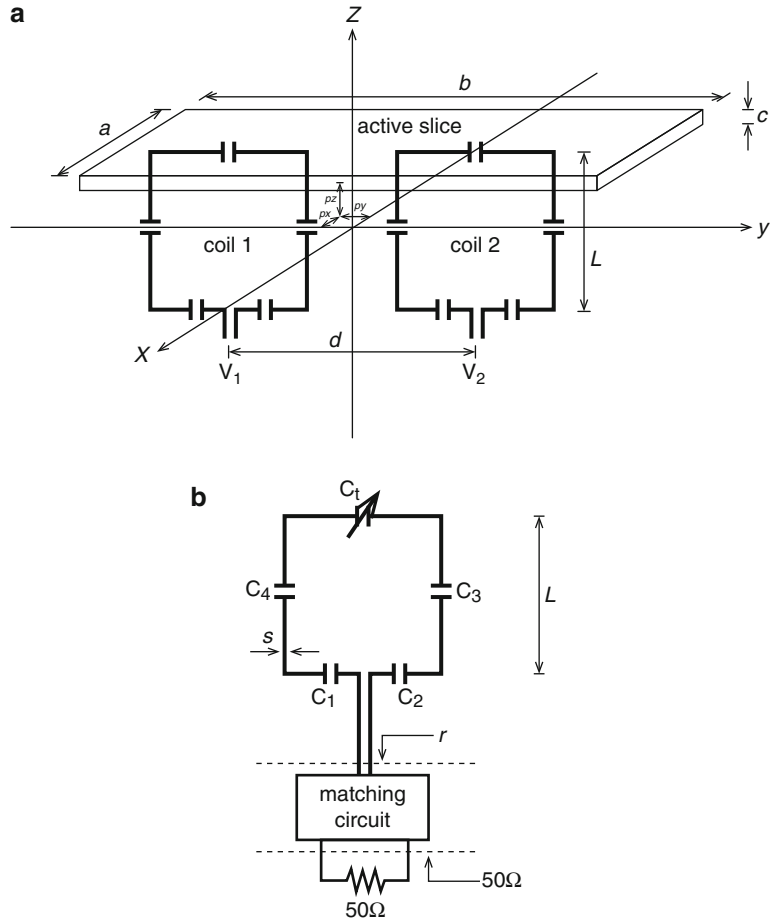


Fig. 21 (**a**) The two-coil phased array for MRI and the active slice, (**b**) the coil configuration and the matching capacitors

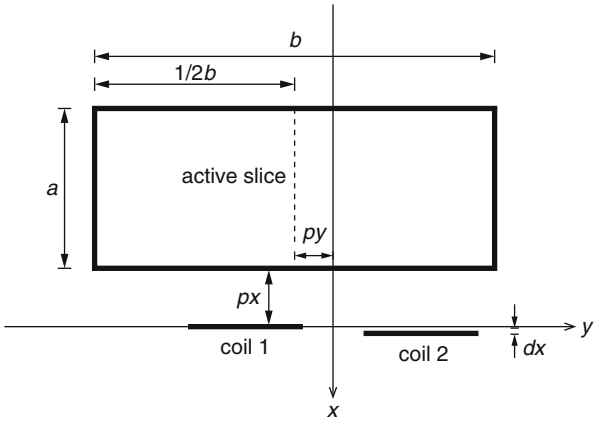


Fig. 22 Top view of the active slice and its relative position to the phased array

respectively, as shown in Figs. 21a and 22. In MRI, the active slice is the signal source whose spatial position can be accurately controlled by the gradient magnetic fields (Jin 1999).

Now assume the received terminal voltages (across the terminal load, $Z_L = r$) are V_1 and V_2, which can be measured or calculated. To simplify the following analysis, the terminal voltages are assumed to be taken before the matching circuit as shown in Fig. 21a. The actual terminal voltages across the 50Ω

DOI 10.1007/978-981-4560-75-7_13-2

resistors differ from V_1 and V_2 by just a multiplicative constant. V_1 and V_2 are corrupted with mutual coupling. If mutual coupling is removed from V_1 and V_2, the uncoupled voltages U_1 and U_2 can be obtained. The relation of V_1 and V_2 to U_1 and U_2 is through the receiving mutual impedances, as explained before in Eq. 22, i.e.,

$$\begin{bmatrix} U_t^1 \\ U_t^2 \end{bmatrix} = \begin{bmatrix} 1 & -\dfrac{Z_t^{12}}{Z_L} \\ -\dfrac{Z_t^{21}}{Z_L} & 1 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \tag{54}$$

where Z_t^{12} and Z_t^{21} are receiving mutual impedances between the two coils. For an example of illustration, the dimensions of the two coils are taken to be $L = 12$ cm, $s = 1$ cm, and $dx = 0.15$ cm. The capacitor values and the resonant frequencies are calculated for four cases of the dc magnetic field, $B_0 = 0.5$ T, 1.0 T, 1.5 T, and 2.0 T. They are shown in Table 4. The dimensions of the active slice are $a = 20$ cm, $b = 40$ cm, and $c = 0.25$ cm, which is approximately the case in spinal cord imaging. In the computer simulations, the active slice is used both as the external source for the calculation of the receiving mutual impedances, and the signal source for the calculation of the signal voltages received by the coils. The method of moments (Harrington 1993) is used in the numerical calculations. The calculated receiving mutual impedances with coil separation d are shown in Table 5. The coil separation is expressed in terms of the side length of the coils, L. The relative position of active slice is at $px = -1$ cm, $py = 0$, and $pz = 0$. It can be seen that at $d = 0.9$ L, both Z_t^{12} and Z_t^{21} reach very low values (real and imaginary parts). This indicates that mutual coupling is small at this separation due to the cancellation of some part of the flux linkage through the overlapping area. This separation corresponds to an ~2 % overlapping area between the two coils. This justifies the use of overlapping method as a means for decoupling adjacent coils in a phased array (Roemer et al. 1990). However, overlapping alone is hardly enough as will be shown later. From Table 5, it can also be seen that the receiving mutual impedances increase rapidly at $d = 1.0\,L$, especially Z_t^{12}. This indicates strong coupling between the two coils as there is no cancellation of flux linkage at this separation.

For the calculation of terminal voltages, three types of terminal voltages are investigated. The first type is the measured (or received) terminal voltages which include all the mutual coupling effect, i.e., V_1 and V_2 in Eq. 46. These voltages can be measured directly. The second type is the compensated terminal voltages U_t^1 and U_t^2 calculated by using Eq. 46. They are calculated from the knowledge of the new mutual impedances (receiving mutual impedances) Z_t^{12} and Z_t^{21}. The last type is the ideal uncoupled terminal voltages, denoted by U'^1_t and U'^2_t, which are calculated when the coils are isolated from each other. U'^1_t and U'^2_t are used as an ideal reference for comparison. To measure U'^1_t and U'^2_t, one coil has to be removed from the array before the measurement of voltage on the other coil and vice versa.

Table 4 The capacitors, self-impedances, and resonant frequencies of the two coils for use at different DC magnetic field strengths. The dimensions of the two coils are $L = 12$ cm and $s = 1$ cm

DC magnetic field, B_0 (Tesla)	0.5 T	1.0 T	1.5 T	2.0 T
Resonant frequency, f_0 (MHz)	21.3	42.6	63.9	85.2
C_1 (pF)	950	240	100	60
C_2 (pF)	950	240	100	60
C_3 (pF)	950	240	100	60
C_4 (pF)	950	240	100	60
C_t (pF)	950.7	228.89	123.7	55.43
Self-impedance (ohm)	0.00016-j0.00001	0.0026-j0.0009	0.013-j0.004	0.042-j0.005

DOI 10.1007/978-981-4560-75-7_13-2

Table 5 The variation of the receiving mutual impedances of the two-coil array with coil separation at $f_0 = 63.9$ MHz. The coil dimensions are $L = 12$ cm, $s = 1$ cm, and $dx = 0.15$ cm. The capacitor values are $C_1 = C_2 = C_3 = C_4 = 100$ pF and $C_t = 123.7$ pF. The self-impedance of the two coils is 0.013-j0.004Ω. The active slice used in the calculation has dimensions $a = 20$ cm, $b = 40$ cm, and $c = 0.25$ cm and is positioned at $px = -1$ cm, $py = 0$, and $pz = 0$. A constant magnetization without phase encoding is assumed

	Receiving mutual impedances	
Separation d	Z_t^{12} (Ω)	Z_t^{21} (Ω)
0.2 L	−4.7928 − j14.9079	21.2884 + j92.5092
0.3 L	−8.9841 − j26.3609	−6.8402 − j19.3298
0.4 L	−5.5937 − j15.6856	−5.0422 − j13.6291
0.5 L	−4.9724 − j14.9149	−4.4424 − j12.7460
0.6 L	−7.5467 − j26.7463	−4.7274 − j19.4993
0.7 L	−2.2733 − j6.5786	−0.1956 + j0.0634
0.8 L	−1.5676 − j4.8052	−1.2157 − j3.7355
0.9 L	0.2473 + j0.8064	0.3163 + j0.9559
1.0 L	138.25 − j22.368	−0.3255 + j35.5022
1.125 L	1.3058 + j4.0361	1.2221 + j3.8111
1.2 L	0.9026 + j2.7978	0.8683 + j2.7132
1.3 L	0.6073 + j1.8916	0.5930 + j1.8582
1.4 L	0.4361 + j1.3658	0.4290 + j1.3493
1.5 L	0.3266 + j1.0293	0.3228 + j1.0200
1.6 L	0.2521 + j0.7999	0.2499 + j0.7941
1.7 L	0.1990 + j0.6364	0.1977 + j0.6326
1.8 L	0.1599 + j0.5159	0.1592 + j0.5133
1.9 L	0.1303 + j0.4248	0.1299 + j0.4230
2.0 L	0.1075 + j0.3544	0.1073 + j0.3530

To characterize the performance of the receiving mutual impedance compensation method, two criteria are employed to indicate the accuracy of this method and to compare it with previous ones. The first is the percentage errors of the compensated terminal voltages and is defined by

$$\%\text{ error in compensated terminal voltage } U_t^i = \frac{\left\|U_t^i - U'^i_t\right\|}{\left\|U'^i_t\right\|} \times 100\%, \quad i = 1,\ 2 \tag{55}$$

where $\|z\|$ represents the norm of the complex number z. It measures the % deviation of the compensated voltages from the ideal uncoupled ones. The second criterion is the isolation between the two coils and is defined by

$$\alpha_{ij} = \left| 20\log\left(\frac{\left\|U_t^i - U'^i_t\right\|}{\left\|U'^{ij}_t\right\|} \right) \right|, \quad i, j = 1,\ 2. \tag{56}$$

It measures the relative amount of signal that couples from one coil to the other after using the receiving mutual impedance compensation method. The larger this value, the smaller is the coupled signal. A similar isolation criterion as this one Eq. 48 was used before in Zhang and Webb (2004) although its meaning may not be exactly the same as Eq. 48.

DOI 10.1007/978-981-4560-75-7_13-2

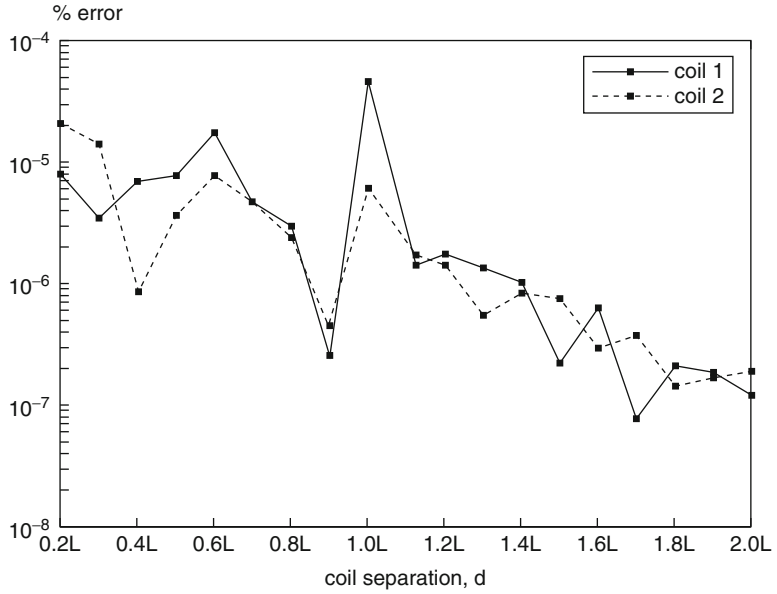


Fig. 23 Percentage errors in the compensated terminal voltages with coil separation for $f_0 = 63.9$ MHz and $px = -1$ cm, $py = 0, pz = 0$

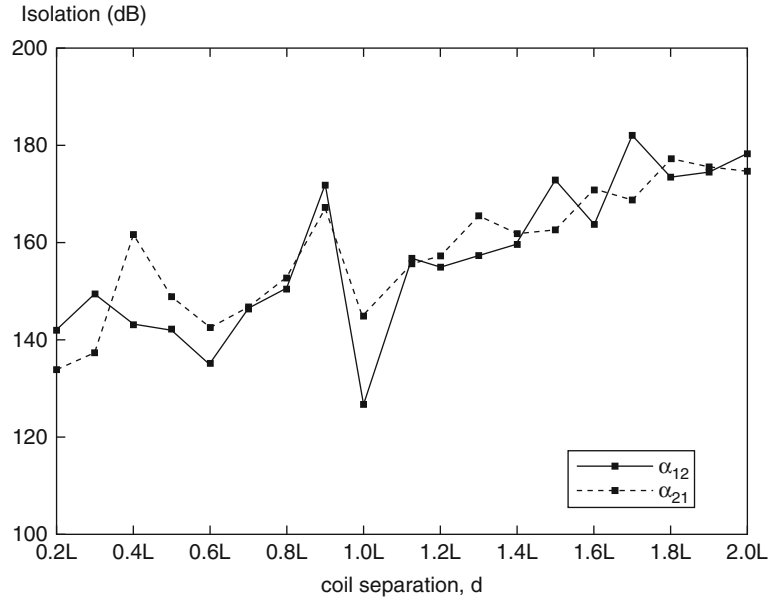


Fig. 24 Isolations of the compensated terminal voltages with coil separation for $f_0 = 63.9$ MHz and px $= -1$ cm, py $= 0$, pz $= 0$

The variation of % errors in the compensated terminal voltages with coil separation is shown in Fig. 23 with $f_0 = 63.9$ MHz ($B_0 = 1.5$ T) and $px = -1$ cm, $py = 0, pz = 0$. It can be seen that the % errors are extremely small, smaller than 10^{-4} % even for $d = 0.2\,L$, i.e., the two coils almost completely overlapped with each other. This indicates that the coupled voltages are almost completely removed. A very low % error (both for coil 1 and coil 2) can be seen at $d = 0.9\,L$, corresponding to the small mutual impedances as shown in Table 5. Figure 24 shows the isolations of the compensated terminal voltages with coil separation. It can be seen that they show a similar trend as the % errors (but turned upside down) and are

DOI 10.1007/978-981-4560-75-7_13-2

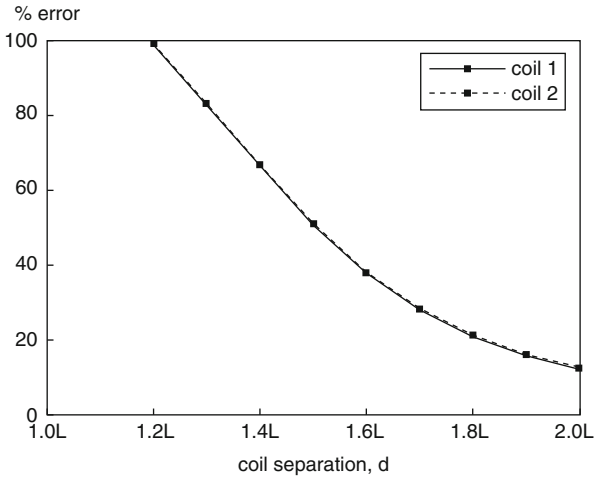


Fig. 25 Percentage errors in the compensated terminal voltages with coil separation using the conventional open-circuit voltage method (Gupta and Ksienski 1983)

generally greater than 120 dB. The corresponding % errors in the compensated terminal voltages using the open-circuit voltage method (Gupta and Ksienski 1983) are shown in Fig. 25. It can be seen that even at a separation of $d = 1.2\ L$, the % errors are as large as about 100 %. In a recent study (Zhang and Webb 2004), it was reported that a capacitive decoupling network can achieve an isolation between two phased-array coils of 20 dB ~ 40 dB. This is much smaller than the isolation that can be achieved by the current compensation method. However, this comparison is based on the assumption that isolation in both cases is taken to be a measure of the relative amount of signal coupled from one coil to other.

As mentioned above, the calculation of the receiving mutual impedances relies on the accurate knowledge of the position of the active slice. In a real situation of MRI, the position of the imaged area (i.e., the active slice) may be under a constant change during the imaging process, whereas the receiving mutual impedances are calculated based only on one position of the active slice. The effect resulted from the change of the active slice position from that used in the calculation of the receiving mutual impedance is important, one that needs to be determined. Figures 26, 27, and 28 show the relations between the % errors and the changes of the position of the active slice along the three axes, indicated by px, py, and pz, respectively. The results are obtained at 63.9 MHz ($B_0 = 1.5$ T) and $d = 1.125\ L$. The receiving mutual impedances used in the calculation are obtained with the slice at $px = -1$ cm, $py = 0$, and $pz = 0$. It can be seen that even when *px*, *py*, or *pz* varies over a substantial distance range, the % errors are still very small. For the change of px and py, the % errors are smaller than 1 %, and for the change of *pz*, the % errors are smaller than 4 %. The isolation between the two coils for the worst case is found to be still greater than 30 dB. That means that even in a practical case when the active slice position is changed to a different position from that used for the calculation of the receiving mutual impedances, the resulting error is still very small. On the other hand, if the change of the active slice position is too large, one may need to reposition the array so that the errors resulting from the change of slice position are still kept within small values as shown above.

More studies on using the receiving mutual impedance concept in the design of phased arrays can be found from the studies in Li et al. (2008), Hui et al. (2006), and Liang et al. (2012, 2013a, b), especially those more recent ones in Liang et al. (2012, 2013a, b).

DOI 10.1007/978-981-4560-75-7_13-2

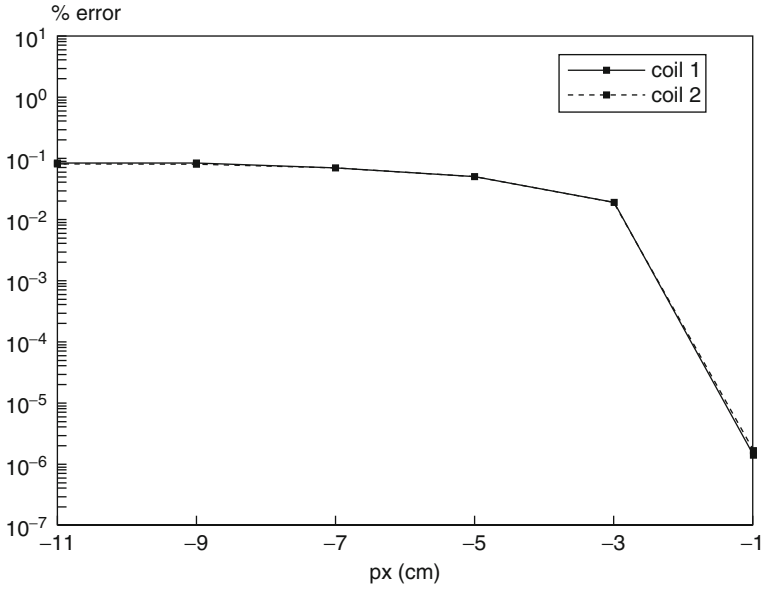


Fig. 26 Percentage error in the compensated terminal voltages with the active slice deviation px by using the receiving mutual impedance compensation method

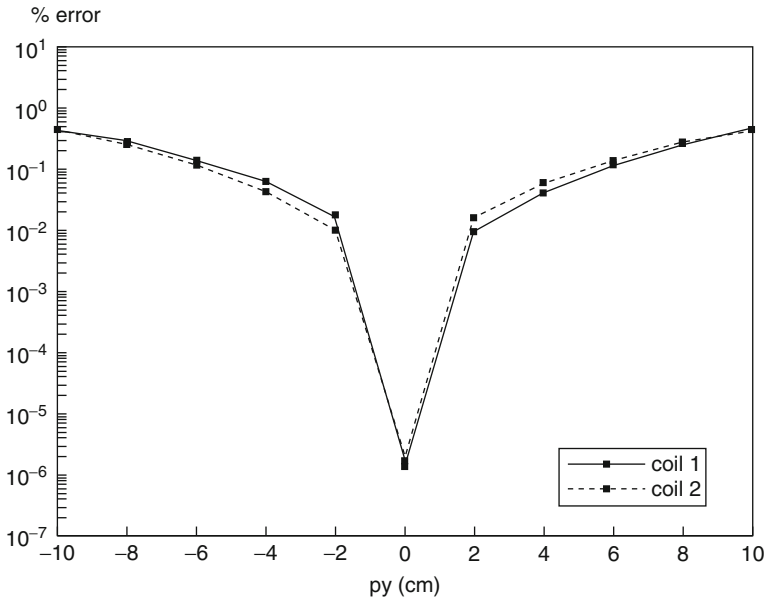


Fig. 27 Percentage error in the compensated terminal voltages with the active slice deviation py by using the receiving mutual impedance compensation method

The Application of the Receiving Mutual Impedance in Signal Correlation Analysis for MIMO Systems

Theory of the Characterization of the Antenna Mutual Coupling Effect in MIMO Systems

Multiple-input and multiple-output (MIMO) systems are well known for achieving a greater system capacity (Foschini and Gans 1998) than the conventional ones employing only single antennas at the two

DOI 10.1007/978-981-4560-75-7_13-2

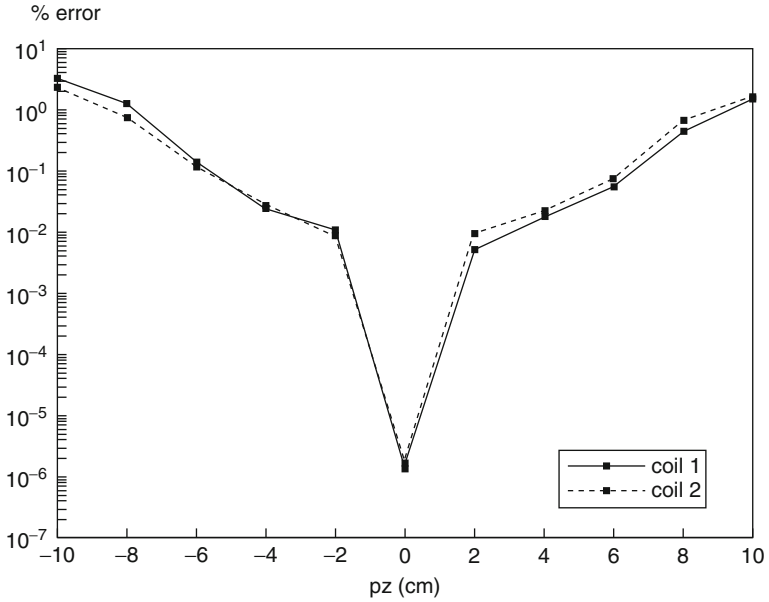


Fig. 28 Percentage error in the compensated terminal voltages with the active slice deviation pz by using the receiving mutual impedance compensation method

sides of a communication link. For MIMO systems, except that spatial correlation due to the multipath scattering effect will contribute to the channel correlation, antenna mutual coupling will also contribute (Svantesson 2002; Jungnickel et al. 2003). Consider the MIMO system shown in Fig. 1 with N transmitting antennas and M receiving antennas. In the transmitter antenna array, antenna mutual coupling causes the input signals to couple to neighboring antennas. This effect can be represented by a mutual coupling impedance matrix as

$$\mathbf{v}_T = \mathbf{Z}_t^{-1}\mathbf{v}_s \tag{57}$$

where $\mathbf{v}_s = [V_{s1}, V_{s2}, \cdots, V_{sN}]^T$ is the excitation voltage vector with mutual coupling not taken into account, $\mathbf{v}_T = [V_{T1}, V_{T2}, \cdots, V_{TN}]^T$ is the excitation voltage vector when mutual coupling is taken into account, and

$$\mathbf{Z}_t = \begin{bmatrix} 1 & \frac{Z_{12}}{Z_{g2}+Z_{22}} & \cdots & \frac{Z_{1N}}{Z_{gN}+Z_{NN}} \\ \frac{Z_{21}}{Z_{g1}+Z_{11}} & 1 & \cdots & \frac{Z_{2N}}{Z_{gN}+Z_{NN}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{Z_{N1}}{Z_{g1}+Z_{11}} & \frac{Z_{N2}}{Z_{g2}+Z_{22}} & \cdots & 1 \end{bmatrix}. \tag{58}$$

In Eq. 50, Z_{ij}, $(i,j = 1, 2, \cdots, N)$ are the conventional mutual impedances used to characterize the mutual coupling effect in transmitting antennas (Gupta and Ksienski 1983), Z_{ij} are the self-impedances of the antennas, and Z_{gi} are the source impedances of the excitation sources driving the antennas. The equivalent circuits for the case of two transmitting antennas for mutual coupling analysis are shown in Fig. 30.

DOI 10.1007/978-981-4560-75-7_13-2

Similarly, for the output signals, they are also modified by the antenna mutual coupling effect in the receiving antenna array. The actual output coupled voltage vector $\mathbf{v}_c$ is related to the uncoupled output voltage vector $\mathbf{v}_u$ as

$$\mathbf{v}_c = \mathbf{Z}_r^{-1}\mathbf{v}_u \tag{59}$$

where $\mathbf{Z}_r$ is the mutual impedance matrix containing the receiving mutual impedances

$$\mathbf{Z}_r = \begin{bmatrix} 1 & -\frac{Z_t^{12}}{Z_L} & \cdots & -\frac{Z_t^{1M}}{Z_L} \\ -\frac{Z_t^{21}}{Z_L} & 1 & \cdots & -\frac{Z_t^{2M}}{Z_L} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{Z_t^{M1}}{Z_L} & -\frac{Z_t^{M2}}{Z_L} & \cdots & 1 \end{bmatrix}. \tag{60}$$

In Eq. 51, $\mathbf{v}_c$ and $\mathbf{v}_u$ are the voltage vectors across the antenna terminal loads. If the total uncoupled output voltages (the open-circuit voltages $\mathbf{v}_{oc}$) are required, then $\mathbf{v}_u$ is related to the open-circuit voltage vector $\mathbf{v}_{oc}$ as

$$\mathbf{v}_u = \frac{Z_L}{Z_{in} + Z_L}\mathbf{v}_{oc}. \tag{61}$$

In Eq. 53, it is assumed that all the antenna elements have the same internal impedance Z_{in} and terminal impedance Z_L. Equation 51 then becomes

$$\mathbf{v}_c = \frac{Z_L}{Z_{in} + Z_L}\mathbf{Z}_r^{-1}\mathbf{v}_{oc}. \tag{62}$$

But for comparison with the performance of the uncoupled system whose output is expressed using the open-circuit voltages, it is required to change the terminal coupled voltage vector $\mathbf{v}_c$ to the open-circuit coupled voltage vector $\mathbf{v}'_{oc}$. That is,

$$\mathbf{v}'_{oc} = \frac{Z_{in} + Z_L}{Z_L}\mathbf{v}_c. \tag{63}$$

Combining Eqs. 49, 54, and 55, the signal model for an MIMO system under both spatial correlation and antenna mutual coupling, as well as channel noise, can be obtained as follows:

$$\mathbf{v}'_{oc} = \mathbf{Z}_r^{-1}\mathbf{H}\mathbf{Z}_t^{-1}\mathbf{v}_s + \mathbf{v}_n \tag{64}$$

where $\mathbf{v}_n$ is the vector of noise voltages which are assumed to be not affected by antenna mutual coupling. Note that spatial correlation is included inside the channel matrix $\mathbf{H}$, while antenna mutual coupling is included inside the matrices $\mathbf{Z}_t$ and $\mathbf{Z}_r$.

Numerical Example

Consider the MIMO system shown in Fig. 29 with three transmitting antennas and three receiving antennas (M=N=3), as well as their equivalent circuits as shown in Fig. 30. The transmitting and receiving

DOI 10.1007/978-981-4560-75-7_13-2

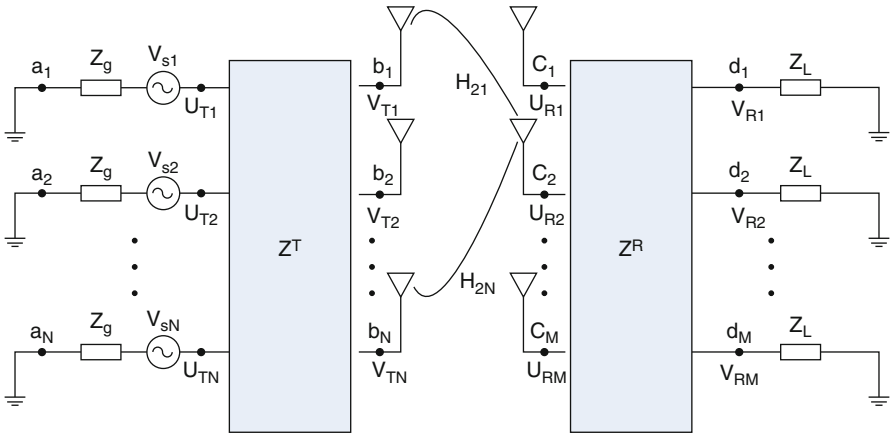


Fig. 29 An MIMO system with antenna mutual coupling effect

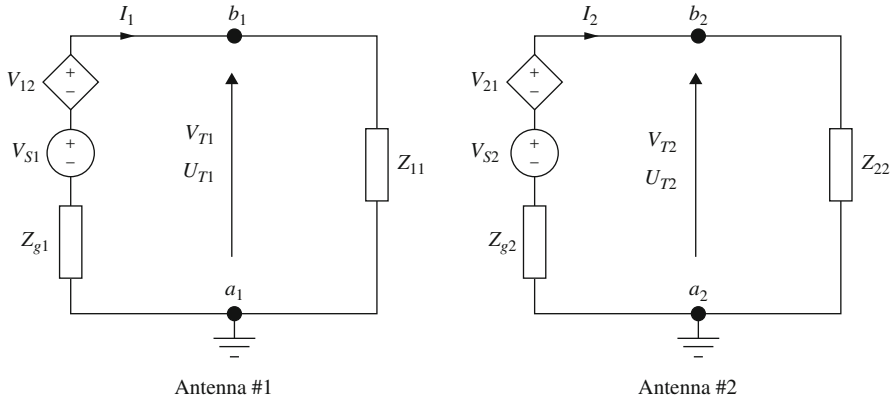


Fig. 30 The equivalent circuits of two transmitting antennas for mutual coupling analysis

antennas are all dipole antennas and that both the transmitting and receiving antennas are aligned to form two uniform linear arrays (ULAs). The antenna separations at the transmitter and receiver are 0.2λ and 0.15λ, respectively. The angle of departure (AOD) at the transmitter and the angle of arrival (AOA) at the receiver of the multipath signals are all 360°. This leads to the spatial correlation coefficients for the antenna elements varying in the form of the Bessel function (Clarke 1968). Assume that the channels are Gaussian random channels with a unit variance and a zero mean. For the transmitting antenna array, the mutual impedances between two transmitting antennas are

$$d_t = 0.2\lambda, Z_{12} = 25.91 - j15.34\Omega, Z_{21} = 25.28 - j15.78\Omega$$

$$d_t = 0.4\lambda, Z_{13} = -0.90 - j20.30\Omega, Z_{31} = -1.42 - j20.11\Omega.$$

For the receiving antenna array, the receiving mutual impedances between two receiving antennas are

$$d_r = 0.15\lambda, Z_t^{12} = -17.73 + j2.75\Omega, Z_t^{21} = -17.48 + j2.94\Omega$$

$$d_r = 0.30\lambda, Z_t^{13} = -8.29 + j10.44\Omega, Z_t^{31} = -7.96 + j10.51\Omega.$$

DOI 10.1007/978-981-4560-75-7_13-2

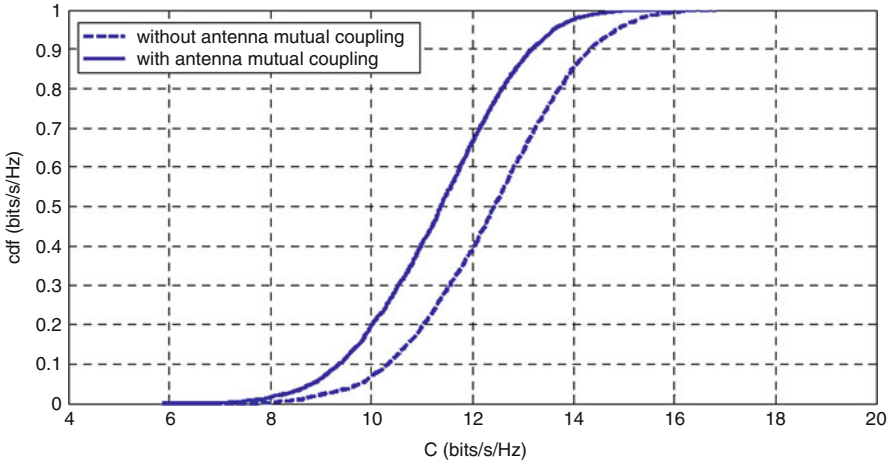


Fig. 31 The channel capacity of an MIMO system with and without antenna mutual coupling

The internal impedance of the dipole antennas is

$$Z_{in} = 39.00 + j7.17\Omega = Z_{11} = Z_{22} = Z_{33}.$$

The terminal load impedance of the dipole antennas is

$$Z_L = 50\Omega.$$

The channel capacity of this MIMO system can be obtained by Monte Carol simulations. The results are shown in Fig. 31 at an SNR = 20d. It can be seen that antenna mutual coupling has reduced the channel capacity substantially.

Summary

In this chapter, the concept of receiving mutual impedance is introduced through derivation from a typical MoM analysis. The theoretical and experimental methods for the determination of the receiving mutual impedance are laid down. The measured and theoretical receiving mutual impedances for typical dipole and monopole antenna arrays are given as examples. The fundamental difference, namely, the truly isolated state between the receiving mutual impedance and the conventional mutual impedance, is explained. Typical examples for the application of the conception of receiving mutual impedance are provided. These include applications in DOA estimation, in interferences suppression, in MRI phased-array design, and in MIMO communication systems.

Cross-References

- ▶ Antennas in MRI Systems
- ▶ Conformal Array Antennas
- ▶ Millimeter-Wave Antennas and Arrays
- ▶ MIMO Systems and Antennas for Terminals
- ▶ Multibeam Antenna Arrays

DOI 10.1007/978-981-4560-75-7_13-2

▶ Phased Arrays
▶ Reconfigurable Antenna Arrays for Wireless Communications
▶ Reflectarray Antennas
▶ Suppression of Mutual Coupling between Elements in Multiple Element Antennas
▶ Waveguide Slot Array Antennas

References

Adve RS, Sarkar TK (2000) Compensation for the effects of mutual coupling on direct data domain adaptive algorithms. IEEE Trans Antennas Propag 48:86–94
Balanis CA (2005) Antenna theory: Analysis and design. 3rd edn. Wiley, Hoboken
Cheng DK (1983) Field and wave electromagnetics. Addison-Wesley Publishing, Reading
Clarke RH (1968) A statistical theory of mobile-radio receptions. Bell Syst Tech J 47:957–1000
Craeye C, Parvais B, Dardenne X (2004) MoM simulation of signal-to-noise patterns in infinite and finite receiving antenna arrays. IEEE Trans Antennas Propag 52:3245–3256
Dandekar KR, Ling H, Xu G (2002) Experimental study of mutual coupling compensation in smart antenna application. IEEE Trans Wirel Commun 1:480–487
Daniel JP (1974) Mutual coupling between antennas for emission or reception-application to passive and active dipoles. IEEE Trans Antennas Propag 22:347–349
Foschini GJ, Gans MJ (1998) On limits of wireless communications in a fading environment when using multiple antennas. Wirel Pers Commun 6:311–335
Goossens R, Rogier H (2007) A hybrid UCA-RARE/Root-MUSIC approach for 2-D direction of arrival estimation in uniform circular arrays in the presence of mutual coupling. IEEE Trans Antennas Propag 43:841–849
Gupta IJ, Ksienski AA (1983) Effect of mutual coupling on the performance of adaptive arrays. IEEE Trans Antennas Propag 31:785–791
Harrington RF (1993) Field computation by moment methods. IEEE Press, New York
Hui HT (2003) Improved compensation for the mutual coupling effect in a dipole array for direction finding. IEEE Trans Antennas Propag 51:2498–2503
Hui HT (2004a) A practical approach to compensate for the mutual coupling effect of an adaptive dipole array. IEEE Trans Antennas Propag 52:1262–1269
Hui HT (2004b) A new definition of mutual impedance for application in dipole receiving antenna arrays. IEEE Antennas Wirel Propag Lett 3:367–364
Hui HT (2007) Decoupling methods for the mutual coupling effect in antenna arrays: a review. Recent Pat Eng 1:187–193
Hui HT, Li BK, Crozier S (2006) A new decoupling method for quadrature coils in magnetic resonance imaging. IEEE Trans Biomed Eng 53:2114–2116
Hyde JS, Jesmanowicz A, Froncisz W, Kneeland JB, Grist TM (1986) Parallel image acquisition from noninteracting local coils. J Magn Reson 70:512–517
Ishimaru A (1991) Electromagnetic wave propagation, radiation, and scattering. Prentice Hall, Englewood Cliffs
Jin JM (1999) Electromagnetic analysis and design in magnetic resonance imaging. CRC Press, Boca Raton
Jordan EC (1968) Electromagnetic waves and radiating systems. Prentice-Hall, Englewood Cliffs, Chap 11

DOI 10.1007/978-981-4560-75-7_13-2

Jungnickel V, Pohl V, Helmolt C (2003) Capacity of MIMO systems with closely spaced antennas. IEEE Commun Lett 7:367–363

Kisliansky A, Shavit R, Tabrikian J (2007) Direction of arrival estimation in the presence of noise coupling in antenna arrays. IEEE Trans Antennas Propag 55:1940–1947

Lau CKE, Adve RS, Sarkar TK (2004) Minimum norm mutual coupling compensation with applications in direction of arrival estimation. IEEE Trans Antennas Propag 52:2034–2040

Li BK, Hui HT, Yang CH, Crozier S (2008) A new decoupling method for phased arrays in magnetic resonance imaging – an experimental approach. IET Proc Sci Meas Technol 2:317–325

Liang D, Hui HT, Yeo TS (2012) Increasing the signal-to-noise ratio by using vertically stacked phased array coils for low-field magnetic resonance imaging. IEEE Trans Inf Technol BioMed 16:1150–1156

Liang D, Hui HT, Yeo TS, Li BK (2013a) Stacked phased array coils for increasing the signal-to-noise ratio in magnetic resonance imaging. IEEE Trans Biomed Circuits Syst 7:24–30

Liang D, Hui HT, Yeo TS (2013b) Improved signal-to-noise ratio performance in magnetic resonance imaging by using a multi-layered surface coil array – a simulation study. IEEE J Biomed Health Inf 17:756–762

Lui HS, Hui HT (2010a) Effective mutual coupling compensation for direction-of-arrival estimation using a new, accurate determination method for the receiving mutual impedance. J Electromagn Waves Appl 24:271–281

Lui HS, Hui HT (2010b) Mutual coupling compensation for direction-of-arrival estimations using the receiving-mutual-impedance method. Int J Antennas Propag Article ID 373061

Lui HS, Hui HT (2010c) Improved mutual coupling compensation in compact antenna arrays. IET Microwave Antennas Propag 4:1506–1516

Lui HS, Hui HT, Leong MS (2009) A note on the mutual coupling problems in transmitting and receiving antenna arrays. IEEE Antennas Propag Mag 51:171–176

Niow CH, Hui HT (2012) Improved noise modeling with mutual coupling in receiving antenna arrays for direction-of-arrival estimation. IEEE Trans Wirel Commun 11:1616–1621

Ohliger MA, Ledden P, McKenzie CA, Sodickson DK (2004) Effects of inductive coupling on parallel MR Image reconstructions. Magn Reson Med 52:628–639

Pasala KM, Friel EM (1994) Mutual coupling effects and their reduction in wideband direction of arrival estimation. IEEE Trans Aerosp Electron Syst 30:1116–1122

Pozar DM (1990) Microwave engineering. Addison-Wesley, New York

Prasad S, Williams RT, Mahalanabis AK, Sibul LH (1988) A transform-based covariance differencing approach for some classes of parameter estimation problems. IEEE Trans Acoust Speech Signal Process 36:631–641

Qi C, Wang Y, Zhang Y, Han Y (2005) Spatial difference smoothing for DOA estimation of coherent signals. IEEE Signal Process Lett 12:800–802

Qi C, Chen Z, Wang Y, Zhang Y (2007) DOA estimation for coherent sources in unknown nonuniform noise fields. IEEE Trans Aerosp Electron Syst 43:1195–1204

Rajagopal R, Rao PR (1991) DOA estimation with unknown noise fields: a matrix decomposition method. IEE Proc 138:495–501

Roemer PB, Edelstein WA, Hayes CE, Souza SP, Mueller OM (1990) The NMR phased array. Magn Reson Med 16:192–225

Roy R, Kailath T (1989) ESPRIT – estimation of signal parameters via rotational invariance techniques. IEEE Trans Acoust Speech Signal Process 37:984–995

Sarkar TK, Park S, Koh J, Schneible RA (1996) A deterministic least squares approach to adaptive antennas. Digital Signal Process 6:185–194

Sarkar TK, Koh J, Adve RS, Schneible RA, Wicks MC, Choi S, Palma MS (2000) A pragmatic approach to adaptive antennas. IEEE Antennas Propag Mag 42:39–55

Schmidt RO (1986) Multiple emitter location and signal parameter estimation. IEEE Trans Antennas Propag 34:276–280

Schmidt RO, Franks RE (1986) Multiple source DF signal processing: an experimental system. IEEE Trans Antennas Propag 34:281–290

Sodickson DK, Manning WJ (1997) Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil array. Magn Reson Med 38:591–603

Svantesson T (2002) Correlation and channel capacity of MIMO systems employing multimode antennas. IEEE Trans Veh Technol 51:1304–1312

Torrieri D, Bakhru K (1997) The effects of nonuniform and correlated noise on superresolution algorithms. IEEE Trans Antennas Propag 45:1214–1218

Werner K, Jansson M (2005) On DOA estimation in unknown colored *noise*-fields using an imperfect estimate of the noise covariance. In: Proceedings of the 2005 IEEE/SP 13th workshop on statistical signal processing, Bordeaux, France, pp 956–961

Wu Y, Nie Z (2008) On the improvement of the mutual coupling compensation in DOA estimation. J Syst Eng Electron 19:1–16

Wu Y, Nie Z (2009) New mutual coupling compensation method and its application in DOA estimation. Front Electr Electron Eng Chin 4:47–51

Yang S, Nie Z (2005) Mutual coupling compensation in time modulated linear antenna arrays. IEEE Trans Antennas Propag 53:4182–4185

Ye Z, Liu C (2008) 2-D DOA estimation in the presence of mutual coupling. IEEE Trans Antennas Propag 56:3150–3158

Yeh CC, Leou ML, Ucci DR (1989) Bearing estimations with mutual coupling present. IEEE Trans Antennas Propag 37:1332–1335

Yu YT, Hui HT (2011) Design of a mutual coupling compensation network for a small receiving monopole array. IEEE Trans Microw Theory Tech 59:2241–2245

Yu YT, Lui HS, Niow CH, Hui HT (2011) Improved DOA estimations using the receiving mutual impedances for mutual coupling compensation: an experimental study. IEEE Trans Wirel Commun 10:2228–2233

Yuan H, Hirasawa K, Zhang Y (1998) The mutual coupling and diffraction effects on the performance of a CMA adaptive array. IEEE Trans Veh Technol 47:728–736

Zhang X, Webb A (2004) Design of a capacitively decoupled transmit/receive NMR phased array for high field microscopy at 14.1 T. J Magn Reson 170:149–155

Zhao Y, Zhang S (2000) Generalised algorithm for DOA estimation in unknown correlated noise. IET Electron Lett 36:1893–1894. http://www.feko.info/

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_14-1

Metamaterials and Antennas

Richard W. Ziolkowski*
Department of Electrical and Computer Engineering, The University of Arizona, Tucson, AZ, USA

Abstract

A variety of antennas have been engineered with metamaterials and metamaterial-inspired constructs to improve their performance characteristics. Interesting examples include electrically small, near-field resonant parasitic (NFRP) antennas that require no matching network and have high radiation efficiencies. Experimental verification of their predicted behaviors has been obtained. This NFRP electrically small paradigm has led to a wide variety of multiband and multifunctional antenna systems. The introduction of active metamaterial constructs further augments the antenna designer's toolbox and leads to systems with many interesting and useful properties.

Keywords

Artificial magnetic conductors; Bandwidth; Directivity; Electrically small antennas; Front-to-back ratio; Matching; Metamaterials; Non-Foster elements; Parasitics

Introduction

There has been a paradigm shift in recent years in the research on the physical properties of electromagnetic materials. This trend can be traced primarily to the development of metamaterials (MTMs) and their exotic physics properties and unusual engineering applications (Engheta and Ziolkowski 2006; Eleftheriades and Balmain 2005; Caloz and Itoh 2005). Metamaterials are artificial media whose basic unit cells are engineered in order that the overall composite possesses predesigned macroscopic physical properties which can be tailored for specific applications. By engineering the unit cell to a scale that is much smaller than the operating wavelength, the MTM exhibits homogenized material properties. MTMs are typically created by designing those inclusions to have resonant properties near a particular wavelength λ of interest, through either planar or volumetric loadings of space. They have provided extremely interesting flexibility in the engineering design processes for a variety of electromagnetic, acoustic, elastic, and thermal wave applications. Several electromagnetic prototypes have been developed for radiating and scattering applications operating from UHF to optical frequencies.

A generally accepted MTM classification scheme in terms of the effective permittivity and permeability of the MTM is shown in Fig. 1 (Engheta and Ziolkowski 2006). Materials normally occurring in nature are usually double positive (DPS). MTM interest in DPS characteristics is commonly connected with desires to achieve permittivity or permeability values that are either large or less than one for either or both parameters. The double negative (DNG) metamaterials (MTMs), which are immediately associated with negative refraction effects, were proposed over 40 years ago (Veselago 1968). They were experimentally demonstrated at the turn of this century. Since then, a large variety of artificial material constructs have

*Email: ziolkowski@ece.arizona.edu

DOI 10.1007/978-981-4560-75-7_14-1

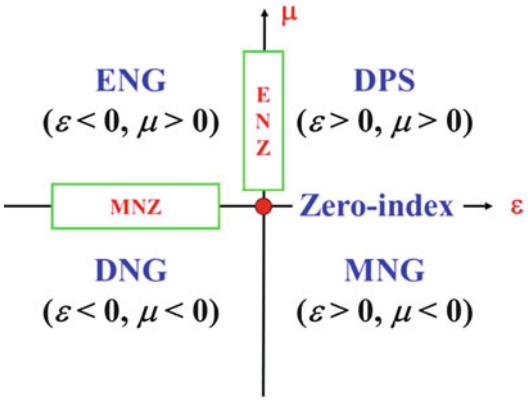


Fig. 1 Metamaterial classifications

been designed, fabricated, and tested, confirming many of the exotic MTM properties (Smith et al. 2000; Ziolkowski 2003; Erentok and Ziolkowski 2007a; Engheta and Ziolkowski 2005).

The epsilon negative (ENG) and mu negative (MNG) MTMs are termed single negative (SNG), in direct contrast to the DNG cases. While the DPS and DNG MTMs allow propagating waves, all waves in SNG MTMs are evanescent. Extreme MTMs exhibit an extreme material or electromagnetic wave parameter. For instance, MTMs that have epsilon-near-zero (ENZ) or mu-near-zero (MNZ) characteristics also have a near-infinite or near-zero wave impedance. Zero-index MTMs have simultaneously zero permittivity and permeability and, thus, a zero index of refraction. Consequently, they exhibit spatially static (dynamic in time), infinite wavelength behaviors (Ziolkowski 2004).

The adaptation of a variety of epsilon-negative (ENG), mu-negative (MNG), and DNG metamaterials or simply metamaterial unit cells to achieve enhanced performance characteristics of antenna systems has since received considerable research attention. This includes studies, for instance, of small antennas (Ziolkowski and Kipple 2003, 2005; Qureshi et al. 2005; Stuart and Tran 2005, 2007; Stuart and Pidwerbetsky 2006; Ziolkowski and Erentok 2006, 2007; Erentok and Ziolkowski 2007a, b, c, 2008; Alici and Ozbay 2007; Arslanagić et al. 2007; Ziolkowski 2008a, b; Antoniades and Eleftheriades 2008; Lee et al. 2008; Greegor et al. 2009; Kim and Breinbjerg 2009; Ziolkowski et al. 2009a, b; Mumcu et al. 2009; Jin and Ziolkowski 2009, 2010a; Lin et al. 2010); multi-functional antennas (Sáenz et al. 2008; Herraiz-Martnez et al. 2009; Antoniades and Eleftheriades 2009; Jin and Ziolkowski 2010b, c, 2011; Lin et al. 2011; Zhu et al. 2010); infinite-wavelength antennas (Sanada et al. 2004; Lai et al. 2007; Park et al. 2007); patch antennas (Buell et al. 2006; Ikonen et al. 2007; Alú et al. 2007; Bilotti et al. 2008); leaky-wave antenna arrays (Eleftheriades and Balmain 2005; Caloz and Itoh 2005; Eleftheriades et al. 2007; Caloz et al. 2008); higher-directivity antennas (Enoch et al. 2002; Wu et al. 2005; Martinez et al. 2006; Franson and Ziolkowski 2009); low-profile antennas achieved with a variety of modified ground planes (Erentok et al. 2005; Erentok et al. 2007d; Yang and Rahmat-Samii 2009), e.g., artificial magnetic conductors (AMCs); and dispersion engineering of time-domain antennas (Caloz et al. 2008; Ziolkowski and Jin 2008). The proliferation of wireless devices for communication and sensor applications has restimulated interest in many different types of antennas. The often conflicting requirements of, for instance, efficiency, bandwidth, directivity, weight, and cost have made the design tasks onerous for antenna engineers with traditional schemes. The metamaterial-inspired engineering of antennas and their performance characteristics has provided an alternative approach to addressing these pressing issues.

DOI 10.1007/978-981-4560-75-7_14-1

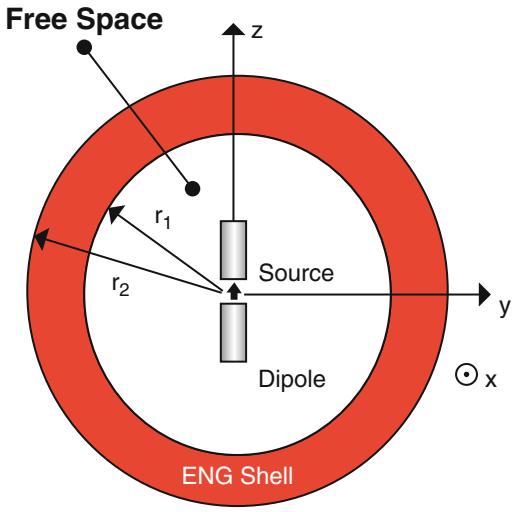


Fig. 2 Original metamaterial-based, efficient electrically small antenna consisting of a center-fed dipole antenna surrounded by an ENG shell

Metamaterial-Based Antennas

The idea of using a resonant metamaterial object in the near field of an electrically small radiator to significantly enhance its performance was introduced in (Ziolkowski and Kipple, 2003; Ziolkowski and Kipple 2005; Ziolkowski and Erentok 2006; Erentok and Ziolkowski 2008) and reviewed in (Ziolkowski et al. 2011). The theoretical models began with enclosing a radiating dipole with double negative or single negative spherical metamaterial shells. For instance, nearly complete matching to a 50Ω source was achieved for a coax-fed dipole (loop) antenna within an ENG (MNG) or DNG shell without any external matching circuit, and high radiation efficiencies were realized giving overall (realized) efficiencies near 100 %. One of the original designs is shown in Fig. 2.

The physical explanation for this configuration is illustrated in Fig. 3. When an electrically small DPS sphere is illuminated by an electromagnetic wave, it responds as an electrically small dipole radiator, which is known to be a highly capacitive element. An electrically small shell also responds in this manner. However, if it is filled with an ENG material, its permittivity is negative, and, hence, its capacitance is negative. This means the shell acts as an inductive element. The combination of the lossy capacitive and inductive elements, i.e., the juxtaposition of the positive and negative material regions, yields a lossy (RLC) resonator.

The driven element, the electrically small dipole antenna, has a large negative reactance, i.e., it too is a capacitive element. Because the lossy resonator is in the extreme near field of the driven element, the fields involved and the subsequent responses are large. It was found that the reactance of this near-field resonant parasitic (NFRP) element, the resonant core-shell structure, can be conjugate matched to the dipole reactance by adjusting their sizes and material properties to achieve an antenna resonance at

$$f_{res} = \frac{1}{2\pi} \frac{1}{\sqrt{L_{eff} C_{eff}}} \tag{1}$$

where L_{eff} and C_{eff} are, respectively, the effective inductance and capacitance of the system, in order to have the total reactance equal to zero (It is noted that in the dual case, a loop antenna and an MNG shell,

DOI 10.1007/978-981-4560-75-7_14-1

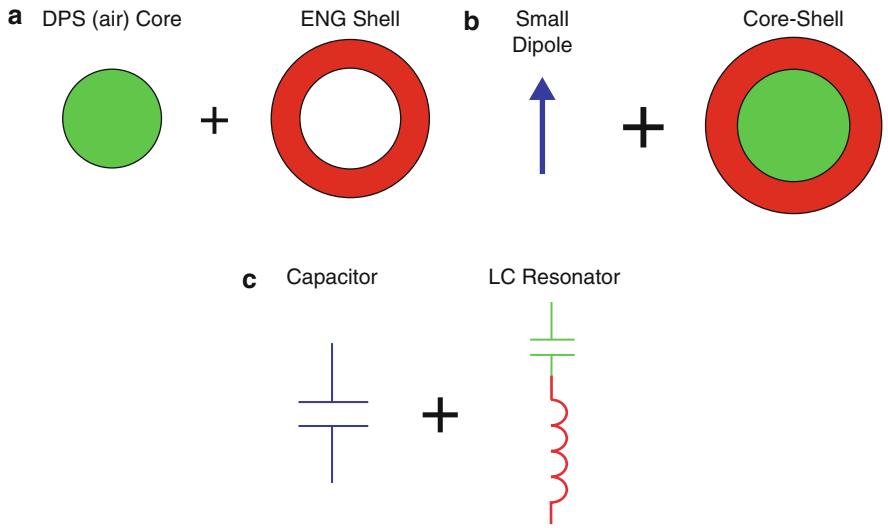


Fig. 3 The physics underlying the electrically small core shell configuration. (**a**) When excited, the DPS core acts as a small dipole element and, hence, is a capacitive element. The ENG shell responds as an electrically small dipole. However, because $\varepsilon < 0$ in the shell, it acts as an inductive element (**b**) Combining the core and shell, he resulting core-shell element acts like an LC resonator that can be matched to a driven electrically small dipole antenna, and (**c**) the circuit elements equivalent to the radiators in (**b**)

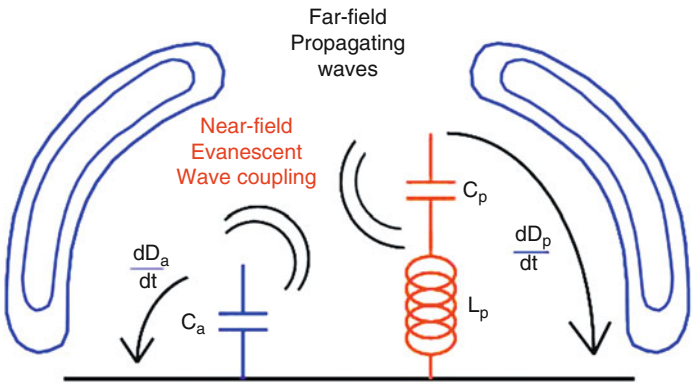


Fig. 4 The basic physics governing the behavior of an electrically small, near-field resonant (NFRP) antenna

the first antenna resonance is generally an antiresonance). Moreover, by tuning the effective capacitances and inductances of both the driven and parasitic elements, the entire antenna can be nearly completely matched to the source. Consequently, it can be said that the NFRP element acts as an impedance transformer.

Furthermore, by arranging the NFRP element so that the currents on it dominate the radiation process, high radiation efficiency and, consequently, very high overall efficiency, i.e., the ratio of the total radiated power to the total input power, can be realized. This basic physics of a NFRP element-based electrically small antenna is depicted in Fig. 4. On the other hand, the bandwidth remains commensurate with the electrical size of the antenna. It was demonstrated (Ziolkowski and Erentok 2007) that with an active ENG shell, the bandwidth could be increased considerably beyond the well-known Chu (1948) and Thal (2006) limits. The introduction of non-Foster elements, which will be discussed later in this chapter, achieves this theoretical construct.

DOI 10.1007/978-981-4560-75-7_14-1

For example, the 300 MHz version of the dipole-ENG shell antenna shown in Fig. 2 was predicted to have an overall efficiency greater than 97 % when $ka = 0.12$ (Ziolkowski and Erentok 2006). It is noted, however, that because the antenna is electrically small, one would expect its directivity to be near that of a small dipole element, i.e., 1.76 dB. Furthermore, the bandwidth will be small. In particular, if a is the radius of the smallest sphere enclosing the entire antenna, $k = 2\pi/\lambda_{res} = 2\pi f_{res}/c$ is the free-space wavenumber at the resonance frequency, and RE is the radiation efficiency of the antenna, then the antenna is electrically small if $ka \leq 0.5$ (1.0) if a (no) ground plane is involved. The Chu lower bound on the quality factor of an electrically small antenna is (Best 2005)

$$Q_{lb} = RE \times \left[\frac{1}{(ka)^3} + \frac{1}{ka}\right] \tag{2}$$

giving $FBW_{ub} \approx 2/Q_{lb}$ as the upper bound of the half-power VSWR fractional bandwidth.

An important practical difficulty with the metamaterial-shell concept is the need to have extremely small unit cell sizes. For instance, if a $ka = 0.10$ antenna is desired, the thickness of the unit cell would be on the order of $\lambda_{res}/100$, thus requiring unit cells at least $\lambda_{res}/300$ in size to have three unit cells across the shell thickness and, consequently, something like a bulk metamaterial. Some of the smallest unit cells fabricated to date are $\lambda/75$ at 400 MHz (Erentok et al. 2007b). Furthermore, when there are losses associated with each unit cell, having many of them can lead to a large cumulative loss value. This behavior has been verified (Greegor et al. 2009) with a dual version of the system shown in Fig. 1. An antenna consisting of an electrically small driven magnetic loop and a mu-negative sphere was designed and fabricated. The measured results demonstrated that the MNG sphere did provide matching of the antenna to the source. However, because of the losses associated with each unit cell and because there were many cells involved in the overall design, it also did not achieve the expected high radiation efficiency.

There have been numerous other metamaterial-based antenna designs considered at RF frequencies. Electrically small antenna designs have been realized with small sets of unit cells and have been shown to have many potential applications (Dong and Itoh 2012). Moreover, the MTM-based NFRP paradigm has been extended to visible wavelengths, leading to active ENZ optical metamaterials (Gordon and Ziolkowski 2008), as well as passive and active nanoantenna (Arslanagić et al. 2007; Geng et al. 2011, 2012, 2013), nanolaser (Gordon and Ziolkowski 2007; Liberal et al. 2014), highly directive nanoantenna (Liberal et al. 2014; Arslanagić and Ziolkowski 2012), nanoamplifier (Arslanagić and Ziolkowski 2010; Arslanagić and Ziolkowski 2014), and quantum jammer (Arslanagić and Ziolkowski 2010, 2013, 2014) designs. The basic configuration is shown in Fig. 5. It results from the fact (Arslanagić et al. 2007) that the driven dipole antenna does not have to be internal to the resonant core-shell element to produce the enhanced total radiated power. As long as the dipole couples strongly with the resonator, the desired enhanced response occurs.

Metamaterial-Inspired Antennas

It was found (Erentok and Ziolkowski 2008) that a NFRP element constructed from a single MTM unit cell is sufficient to achieve the desired matching and high radiation efficiency properties. The resulting radiating systems were termed *metamaterial-inspired* antennas rather than metamaterial-based antennas because only a single MTM unit cell was used and not a bulk medium. One does not need a large resonator around the entire radiating element but rather only a single unit cell – an electrically small resonator – in

DOI 10.1007/978-981-4560-75-7_14-1

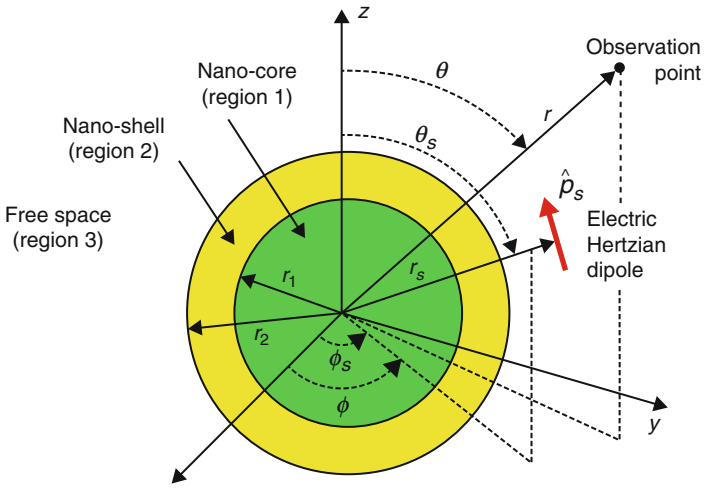


Fig. 5 Nano-antenna based on the NFRP paradigm

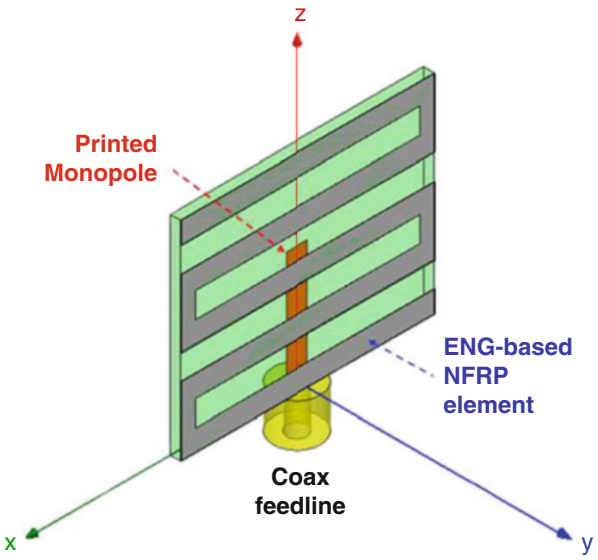


Fig. 6 ENG-based NFRP element, 2D electric EZ antenna

the near field of the driven radiator to achieve nearly complete matching to the source without any matching circuit and nearly 100 % overall efficiency. The initial designs made use of the analytical results that matched the type of driven element with the appropriate type of MTM. Both electric and magnetic coupling mechanisms between the driven and NFRP elements have since been explored. These metamaterial-inspired NFRP elements have led to a variety of interesting electrically small antenna systems. Several of these metamaterial-inspired NFRP designs have been fabricated and tested; the measured results are in good agreement with their simulated values.

Electric NFRP, Electric Coupling

The first NFRP antenna that was tested for its overall efficiency performance was the electric 2D EZ antenna shown in Fig. 6. The term *EZ* was chosen to reflect the fact that these original NFRP antennas were "easy" to design and fabricate. It was a Rogers Duroid^{TM} 5880 design. A monopole was printed on

DOI 10.1007/978-981-4560-75-7_14-1

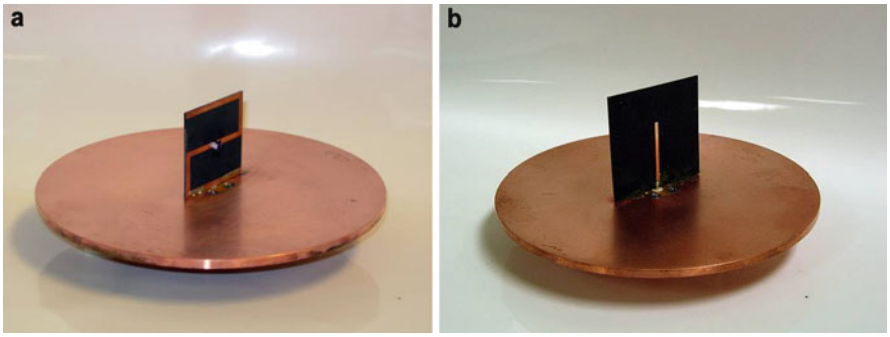


Fig. 7 Fabricated 570 MHz Z antenna on its small circular copper insert. (**a**) Z element side and (**b**) monopole element side

Fig. 8 Physical comparison of the 570 MHz small ground plane Z antenna and the dual-ridged reference horn in the NIST-Boulder reverberation chamber

one side of the Duroid sheet and was coaxially fed through a copper ground plane. The NFRP element was a small meander line connected to the ground plane on the other side of the Duroid sheet. It has been demonstrated (Imhof et al. 2006; Imhof 2006) that this is a unit cell of an ENG MTM. It was demonstrated experimentally (Erentok and Ziolkowski 2008) that a 2D electric EZ antenna with $f_{res} = 1.37$ GHz and $ka \sim 0.49$ was nearly completely matched to the 50Ω source and had an overall efficiency ~94 %, with a 4.1 % fractional bandwidth. The NFRP element is electrically coupled to the driven monopole. The fields generated by the currents on the horizontal lines are nearly cancelled out by their images through the ground plane. The vertical elements radiate coherently and provide the high radiation efficiency.

With an interest to achieve an adjustable and potentially tunable version of this NFRP antenna, the Z antenna, whose 570 MHz, 31 mil, 2 oz Duroid 5880 realization is shown in Figs. 7a and b, was designed, fabricated by Boeing Research and Technology in Seattle, WA, and tested in the reverberation chambers at the National Institute of Science and Technology (NIST) in Boulder, CO (Ziolkowski et al. 2009b). The meander line was reduced to two simple J-elements connected with a lumped element inductor. The bottom J-element is connected to the ground plane. The monopole is coaxially fed through the ground plane. It was a 30 mm × 30 mm design incorporating a CoilCraft 47nH inductor. Measurements were taken with both the small ground plane version (120.6 mm diameter copper disk) shown in Fig. 7 and a larger ground plane version (the small ground plane version was inserted into an 18 in × 18 in = 457.2 mm × 457.2 mm copper ground plane). A physical comparison of the Z antenna and the reference ETS LINGREN 3106 double-ridged waveguide horn, which is about 94 % efficient in its

DOI 10.1007/978-981-4560-75-7_14-1

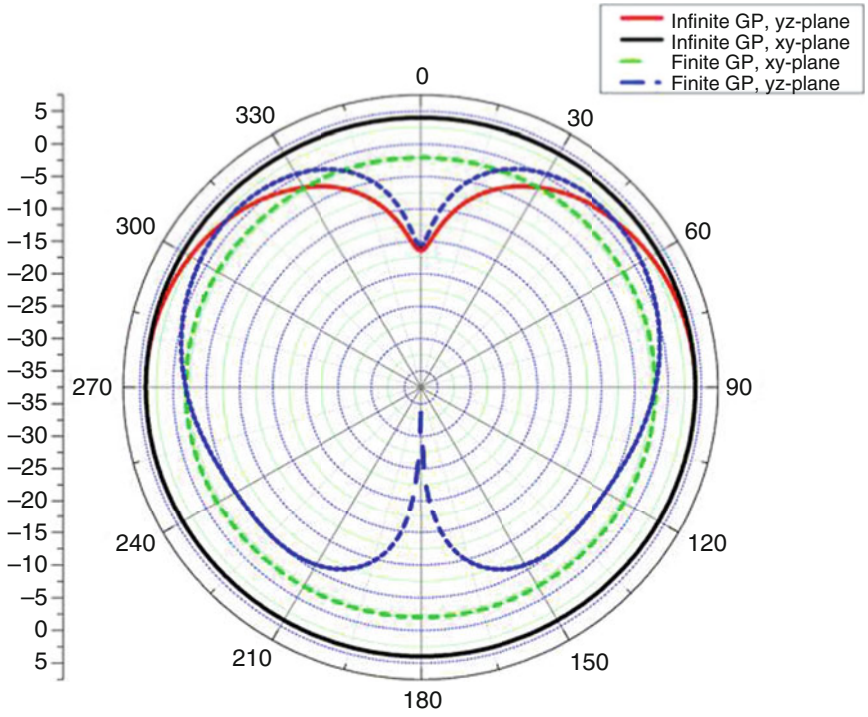


Fig. 9 HFSS-predicted patterns for the Z antenna at f_{res} have the expected electric monopole with a finite ground plane shapes

200 MHz-2GHz frequency band, is shown in Fig. 8. An overall efficiency equal to 80 % was measured at the resonance frequency $f_{res} = 566.2$ MHz ($ka = 0.398$) with a half-power fractional bandwidth $FBW = 3.0$ %, giving $Q = 4.03\ Q_{lb}$. There was little difference between the small ground plane and larger ground plane results. The predicted gain patterns in the small and large ground plane configurations are shown in Fig. 9, confirming that the Z antenna acts like a small vertical monopole with a finite ground plane. A second Z antenna, a 40 mm × 40 mm design incorporating a CoilCraft 169nH maxi inductor, had a 46 % measured overall efficiency at $f_{res} = 294.06$ MHz ($ka = 0.276$). Both sets of experimental results demonstrated, as predicted, the ability to obtain a lower resonance frequency with a simple redesign using a larger lumped element value. These experiments not only confirmed the predicted controllability of the resonance frequency but also provided information on how to treat the lumped element inductor in the electromagnetic simulations. Based on these results, an updated Z antenna was designed with an overall efficiency of OE = 82.3 % ($|S_{11}| = -25.44$ dB) at $f_{res} = 285.6$ MHz ($ka = 0.428$) (Ziolkowski et al. 2009b).

As it was recognized that the NFRP elements were the key to these electrically small designs, several variations have been designed to allow for other functionalities, which will be described below. For instance, the NFRP element associated with the Z antenna can be simplified considerably (Ziolkowski 2008a, b; Jin and Ziolkowski 2009; Jin and Ziolkowski 2011). The version shown in Fig. 10a consists of a split vertical segment connected by a lumped element inductor. The size of the upper horizontal metal rectangle is adjustable to tune the input reactance. HFSS simulations of the currents at the resonance frequency confirm that they are mainly on the vertical segments, which means this antenna radiates as a monopole over the ground plane and that the overall efficiency is above 90 % (Jin and Ziolkowski 2009). The second version in Fig. 10b utilizes a distributed NFRP element. The vertical strip couples directly to the driven monopole, the metal arc provides the inductance, and the horizontal strips provide additional capacitance. In resonance, HFSS simulations predict that the overall efficiency is above 90 % and confirm that the currents, as shown in Fig. 10b, are mainly on the vertical segment of the NFRP element, which again indicates why this antenna radiates as a monopole over the ground plane (Ziolkowski et al. 2011).

DOI 10.1007/978-981-4560-75-7_14-1

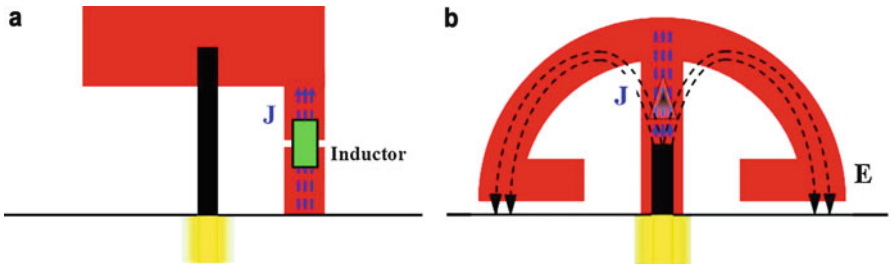


Fig. 10 Electrically small antenna designs incorporating generalized NFRP elements and coax-fed monopoles. (**a**) Rectangular NFRP with lumped element inductor, (**b**) Egyptian axe antenna with its distributed NFRP element. The current densities represent their basic behavior at the first resonance of the antenna

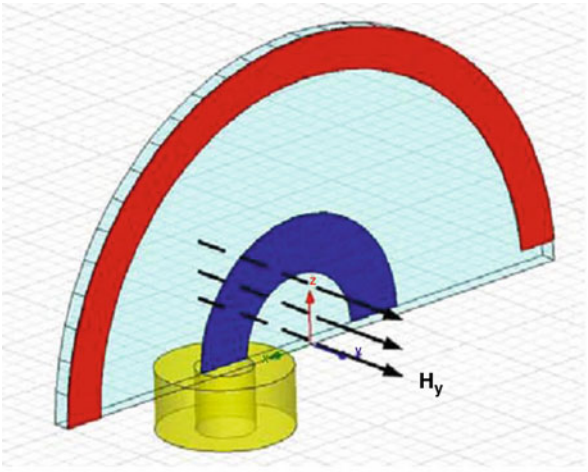


Fig. 11 The magnetic flux of the printed semi-loop antenna coaxially fed through a finite ground plane drives the CLL NFRP element in this variation of the 2D magnetic EZ antenna

Magnetic NFRP, Magnetic Coupling

In a similar fashion, the NFRP element can be driven with the magnetic field of the driven element. In conjunction with the analytical solutions, a driven magnetic semiloop antenna coaxially fed through a finite ground plane was first considered. This led to the magnetic 2D and 3D EZ antennas (Erentok and Ziolkowski 2007a, b; Ziolkowski et al. 2009b; Lin et al. 2010). This feed-coupling scheme is graphically illustrated in Fig. 11, which illustrates a variation of the 2D magnetic EZ antenna. The parasitic element is a capacitively loaded loop (CLL), which was originally used successfully to achieve an artificial magnetic conductor metamaterial (without any ground plane) (Erentok et al. 2005). Both distributed and lumped element versions have been fabricated and tested successfully (Erentok and Ziolkowski 2008).

The 3D magnetic EZ antenna is composed of an electrically small loop antenna that is coaxially fed through a finite ground plane and that is integrated with an extruded CLL element. This 3D CLL structure is designed to be the NFRP element. The measured results (Ziolkowski et al. 2009b) for the 3D magnetic EZ antenna demonstrated that for an electrical size, $ka \sim 0.43$, at 300.96 MHz, nearly complete matching to a 50 Ω source, and a high overall efficiency (>94 %) was achieved. These results and those in Erentok and Ziolkowski (2007b, c) established that the CLL-based elements can work in many frequency bands, e.g., from the ultrahigh-frequency (UHF) band to the X band. While negative permeability cannot be ascribed to either the 2D or 3D CLL elements themselves, a metamaterial constructed with them as its unit cell inclusion would exhibit MNG properties. Nonetheless, as with the electric source–electric coupling

DOI 10.1007/978-981-4560-75-7_14-1

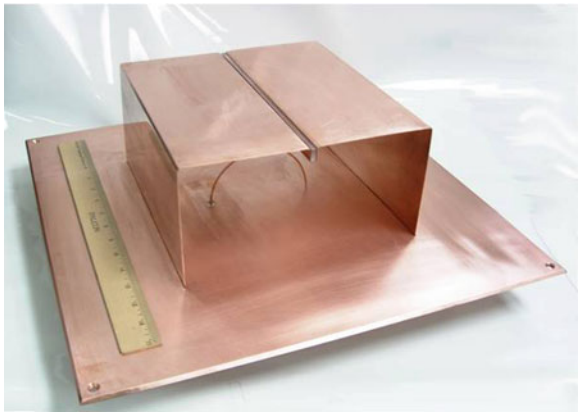


Fig. 12 The fabricated 100 MHz 3D magnetic EZ antenna in its small ground plane configuration

cases, it is the electrically small, magnetic-based NFRP CLL structure that provides the ability to match the electrically small loop antenna to the source. This can be visualized with the configuration dual to that shown in Fig. 4. The NFRP element again enhances the radiation process to achieve high radiation efficiencies. In particular, the CLL element can be engineered to control the strong magnetic flux generated by the small driven loop antenna and convert it into the appropriate currents flowing on the CLL element. Furthermore, this magnetic coupling process between the driven loop and the NFRP CLL element can be adjusted to tune the resonance of the entire antenna system according to Eq. 1 (Erentok and Ziolkowski 2008; Kim and Breinbjerg 2009). These metamaterial-engineered NFRP antennas again help overcome the loss issues associated with an actual metamaterial-based antenna design (Greegor et al. 2009). The wire versions of the CLL NFRP antennas have also been designed and show similar performance characteristics (Lin et al. 2011).

A similar low-profile (height ~ $\lambda_{res}/25$) 3D magnetic EZ antenna was designed for operation at 100 MHz. The fabricated antenna is shown in Fig. 12. The measured and simulated results were in very good agreement. This $ka = 0.46$ antenna was measured to have OE ~ 95 % and a half-power VSWR fractional bandwidth of 1.52 % ($Q = 11.06\ Q_{lb}$) at $f_{res} = 105.2$ MHz. This design utilized a quartz spacer ($\varepsilon_r = 3.78$) to help lower the resonance frequency and to provide mechanical stability during shipping and operation. Similar antennas have been designed for operation at 20 MHz using simply a $\varepsilon_r = 100$ spacer (Lin et al. 2010).

The corresponding HFSS-predicted gain patterns are shown in Fig. 13; the maximum gain value is 5.94 dB. Because the gain patterns are symmetric, it can be immediately inferred that the surface currents induced on the electrically small CLL element by the flux of the driven semiloop antenna are uniform and symmetric. This behavior is verified with the HFSS-predicted vector surface current distributions. This current distribution also demonstrates that this electrically small antenna system is radiating as a magnetic dipole over a finite ground plane (Lin et al. 2010).

Magnetic NFRP, Electric Coupling

Several important features distinguish the electric-based and the magnetic-based cases. The electric-based cases generally exhibit about half the Q values of the magnetic-based cases. Simply, the variation of the input impedance with frequency associated with a resonance is smaller than that associated with an antiresonance (Yaghjian and Best 2005). The electric-based cases radiate a vertical electric monopole type pattern; the magnetic-based cases radiate a horizontal magnetic dipole type pattern. The rate of decrease in the radiation efficiency as *ka* decreases is slower for the electric-based cases than for the magnetic-based ones. Because for many applications one would like broadside radiators and high radiation efficiencies for

DOI 10.1007/978-981-4560-75-7_14-1

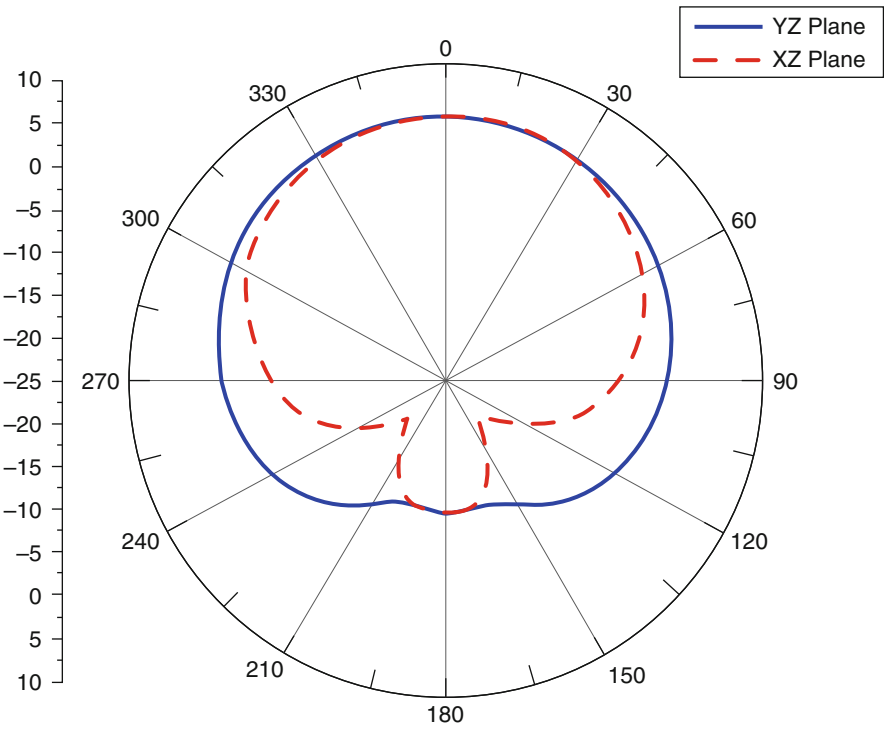


Fig. 13 E- and H-plane gain patterns for the 100 MHz 3D magnetic EZ antenna

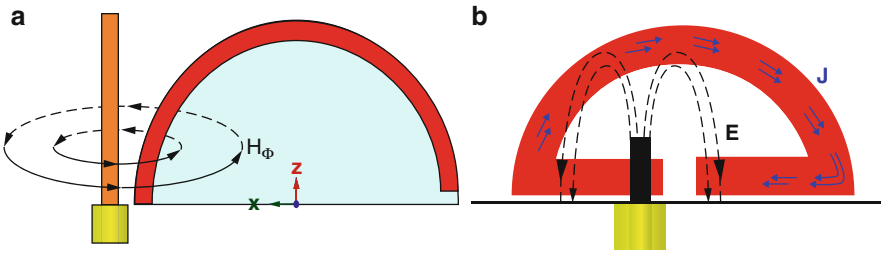


Fig. 14 Antenna composed of a driven electric source that is (**a**) magnetically coupled and (**b**) electrically coupled to a (magnetic) CLL-based NFRP element

yet smaller *ka* values, antennas with a driven electric source have been investigated that are either magnetically or electrically coupled to a magnetic NFRP element. At first, because it deviated from the MTM-based analytical solution paradigm, this cross-design was not expected to yield a successful antenna. Fortunately it was, and it was found that it provides some interesting flexibilities for the multifunctional designs to be discussed below.

The driven electric source, magnetically coupled design is illustrated in Fig. 14a. The magnetic field flux generated by the driven monopole is directly coupled to the CLL element. A realization of this design has been reported (Alici and Ozbay 2007). While it does produce a pattern with its maximum in the direction orthogonal to the plane, this magnetically coupled design is more sensitive and requires larger ka values to achieve nearly complete matching to the source than the corresponding electrically coupled design shown in Fig. 11b. The protractor antenna (Jin and Ziolkowski 2011) illustrated in Fig. 14b has a thin gap between the horizontal legs of the CLL-based NFRP element and the ground plane. Because this NFRP element is in the extreme near field of the driven element, the electric fields across this gap are very large. Note that if the monopole were centered on the CLL element, the magnetic flux contributions would

DOI 10.1007/978-981-4560-75-7_14-1

completely cancel because of symmetry; and the radiation process would be very weak from the electric field-generated currents along either side of the NFRP protractor element because a majority of them would be canceled by their images in the ground plane. However, by offsetting the monopole, the electric driven currents can form the desired loop mode around the CLL element. This loop mode radiates the desired magnetic dipole over a ground plane pattern. The coupling is strong, and consequently, the radiation efficiency can be large. Over 85 % overall efficiencies have been demonstrated for several protractor antennas (Jin and Ziolkowski 2010b; Jin and Ziolkowski 2011).

Multifunctional Electrically Small Antennas

Because of the known need for more functionality in a given antenna, the single antenna designs have been extended to multiantenna ones while trying to maintain the same footprint (overall size). By properly combining and phasing their effective electric and magnetic dipoles electrically small multiband, circular polarized (CP), elemental Huygens source and broader bandwidth antennas were obtained that are nearly completely matched to a 50Ω source and have high radiation efficiencies. These various designs are reviewed below.

Multiband Antennas

Several multiband antenna systems which consist of multiple NFRP elements in the very near field of the driven element, each being resonant at a specified, unique frequency, have been developed (Jin and Ziolkowski 2010c; Lin and Ziolkowski 2010). At each antenna resonance, only one or majorly one parasitic element efficiently radiates. It is worth pointing out that the directly driven monopole remains essentially nonradiating at all of the resonant frequencies. Impedance matching to the source at all multiband frequencies is achieved by adjusting the dimensions of each parasitic element and its distance from (and, hence, coupling to) the driven monopole. When there are multiple parasitic elements, the couplings among the parasitics also impact the radiation and matching processes. Efficient multiband operation results from properly understanding the effects of these multiple resonances and how to minimize their couplings. This NFRP approach is a distinct difference from the directly driven multiband dipole antennas proposed in Herraiz-Martnez et al. (2009) and related works that use parasitic split resonant rings to achieve the multiband capability. At each resonant frequency, the directly driven dipole plays a significant part in the radiation process, and the SRRs act as different loads at different resonant frequencies. Using the simplified NFRP element shown in Fig. 10a, the dual-band GPS L1 (1575.4 MHz) and L2 (1227.6 MHz) antenna shown in Fig. 15 was designed (Jin and Ziolkowski 2010c). In the same manner, a quad-band 430.0 MHz and GPS L1 (1575.42 MHz), L2 (1227.60 MHz), and L5 (1176.45 MHz) antenna was also designed (Jin and Ziolkowski 2010c).

The dual-band, asymmetric split EZ antenna system shown in Fig. 16 was designed to achieve a low-frequency dual-band system for low-earth-orbit satellite (LEOS) communications: 137.475 MHz uplink and 149.15 MHz downlink bands (Lin and Ziolkowski 2010). Note that these frequencies are separated by only 11 MHz. The radiation efficiencies for those resonant frequencies are 96.4 % and 71.5 %, respectively.

The reason why the higher resonant frequency has lower radiation efficiency is due to the currents being in opposition in these two CLL elements, i.e., they are out of phase. At the lower frequency, they are in phase. These EZ design results have provided us with the insight into the need for careful recognition of the relative current flows on the set of NFRP elements to achieve the highest possible radiation efficiencies. By controlling the relative phases of the currents, a related dual-band 429.8 MHz (SATCOM) and 1575.6 MHz (GPS L1) design, in which the two 3D CLL NFRP elements are nested

DOI 10.1007/978-981-4560-75-7_14-1

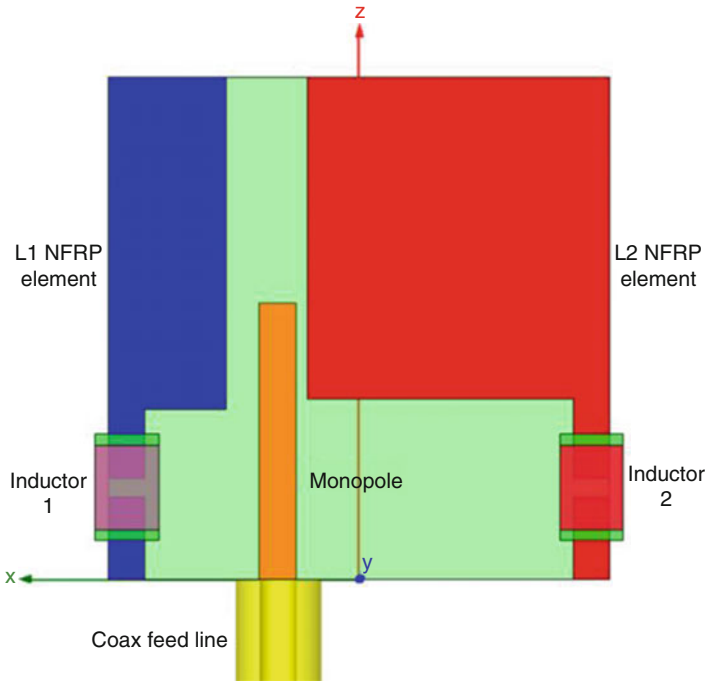


Fig. 15 Duroid 5880-based, single feed, dual-band NFRP antenna. This system has been designed, e.g., for the GPS L1 and L2 frequencies

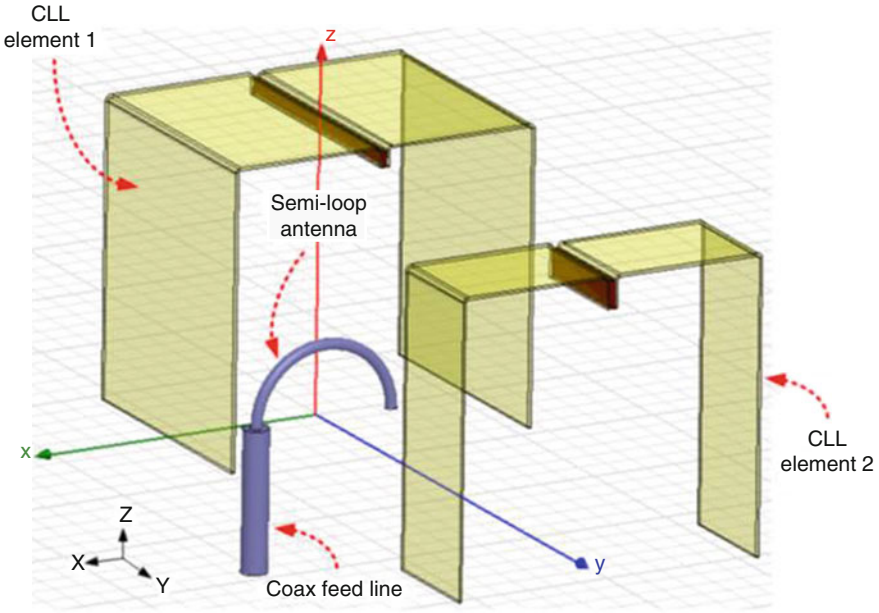


Fig. 16 A dual-band, asymmetric split, 3D magnetic EZ antenna for LEOS communications

and the semiloop antenna is diagonally oriented to excite each element, has radiation efficiencies of 83.8 % and 97.5 %, respectively (Lin and Ziolkowski 2010).

Another dual-band application is the desire to have both satellite positioning (GPS) and communications. Highly compact, multifunctional antennas that can communicate with satellites to accommodate both voice and data exchanges, while providing GPS functionality, are of great interest for portable

DOI 10.1007/978-981-4560-75-7_14-1

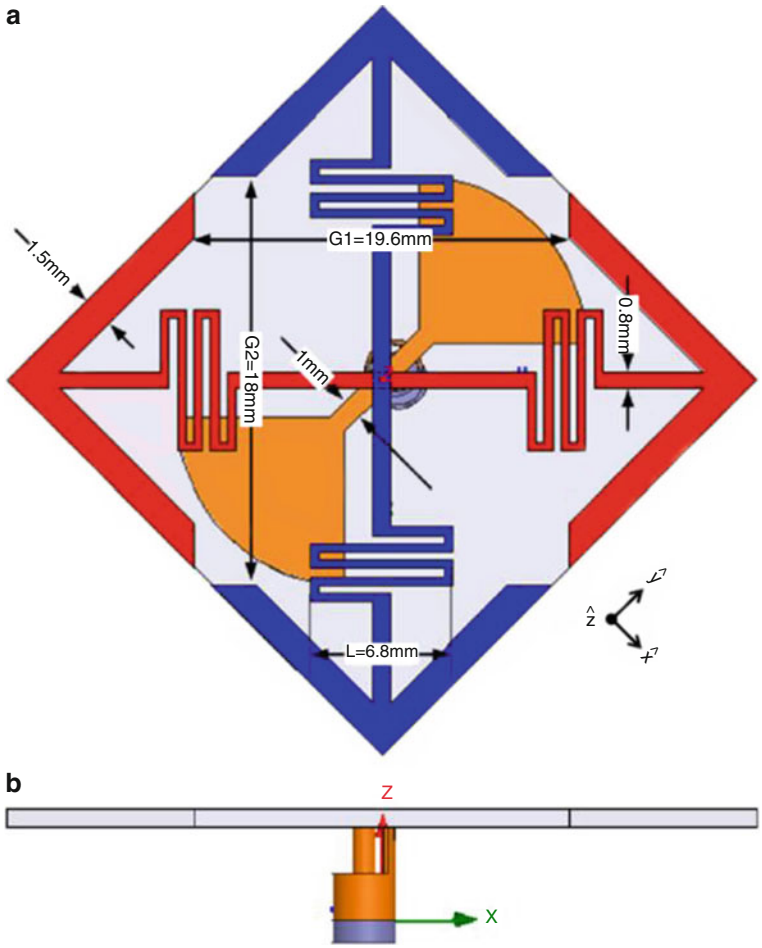


Fig. 17 Dual-band GPS L1/GS NFRP antenna with dimensions (in millimeters). (**a**) Top view. (**b**) Side view

communication device applications. For this reason, a dual-band GPS L1 (1575.42 MHz) and Global Star (GS, 1610–1621 MHz) electrically small, planar NFRP antenna shown in Fig. 17 was designed (Jin et al. 2012). This antenna has, by choice (Jin et al. 2012), a square footprint of 1.0×1.0 in^2. It was designed to be fabricated with the Rogers Duroid 4350 board material, which has a relative permittivity equal to 3.66, a loss tangent equal to 0.004, a 0.762-mm substrate thickness, and a 0.017-mm (0.5 oz) copper thickness.

To achieve the desired dual-band performance, two NFRP elements were incorporated with a single driven element. The two NFRP elements lie on one side of the Duroid sheet; the driven element, which is coaxially fed, lies on the other side. The NFRP elements are both top-loaded meander-line dipoles, which are rectangular variations of the circular Egyptian-axe dipole NFRP elements introduced in Jin and Ziolkowski (2010b). Both dipoles are the same except for the gaps between the capacitive loads, i.e., the capacitances formed by the gaps between the ends of the arrows. Because of the differences in the resulting capacitances, two independent resonant frequencies were obtained. However, to avoid any unnecessary structural overlaps and to keep the NFRP elements in a one-layer layout, the meander lines were introduced on their legs to increase the overall inductance while allowing a decrease in the requisite capacitances provided by their gaps and still maintaining the desired operating frequency. Furthermore, the NFRP elements are oriented orthogonal to each other. This necessitates the overlapping of these elements at the center of the square. It also required the introduction of the driven bowtie antenna (orange), which facilitates electric coupling to each NFRP element. Because the two dipoles are perpendicularly

DOI 10.1007/978-981-4560-75-7_14-1

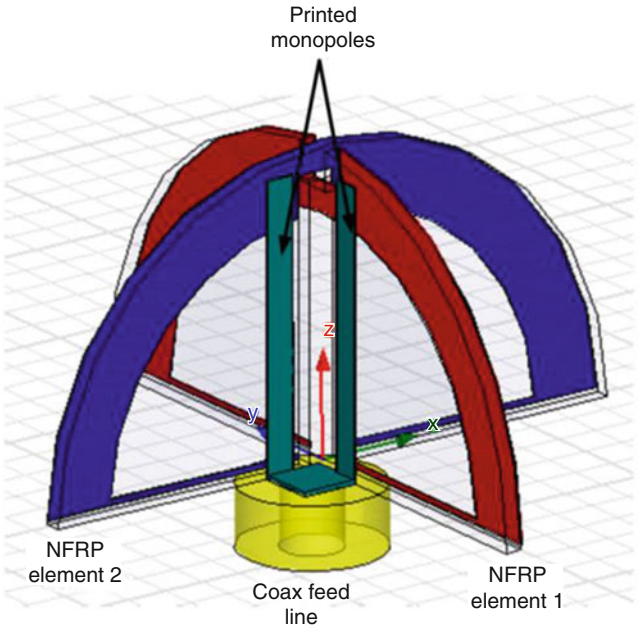


Fig. 18 CP protractor antenna. The CP behavior is achieved at a frequency between two closely spaced resonances through a 90° phase shift obtained between their input reactances

oriented, they operate essentially independently even though they are connected. The red (horizontal) dipole in Fig. 17a has a larger gap (G1) and is thus resonant at the GS band. The blue (vertical) dipole has a smaller gap (G2) and thus operates at the GPS L1 band.

CP Antennas

Another functionality that has been added to the NFRP antenna paradigm is circular polarization (CP). In all of the cases noted above, the antennas are linearly polarized (LP). To achieve a CP system with a single feed in an electrically small package, one needs two orthogonal dipole radiators with an effective 90° phase shift between them. This behavior was realized with the protractor antenna system shown in Fig. 18. It is a GPS L1 CP design (Jin and Ziolkowski, 2011) with $a = 15$ mm; $f_{res} = 1575.4$ MHz giving $ka = 0.495$; a 89.2 % overall efficiency; a 29.3 MHz (−10 dB) bandwidth; a 7.2 MHz CP bandwidth; and an axial ratio equal to 0.26. Each protractor element produces a magnetic dipole parallel to the ground plane. One is tuned to be resonant below the desired operating frequency and thus creates an inductive reactance at it. The other is tuned to be resonant above the desired operating frequency and thus creates a capacitive reactance at it. Because these protractor NFRP elements are orthogonal to each other, their resonances can be tuned independently to generate the 90° phase shift between them necessary for the CP behavior.

Similarly, the planar (no ground plane) GPS L1 design shown in Fig. 19 was obtained (Jin and Ziolkowski 2010b). It has $ka = 0.539$ at 1575.4 MHz, an overall efficiency of 85 %, a 31.4 MHz (−10 dB) bandwidth, a 7.2 MHz CP bandwidth, and an axial ratio of 0.66. It makes use of the same phase shift paradigm but with two of the Egyptian axe NFRP elements shown in Fig. 14b and a driven printed dipole antenna. On the other hand, an electrically small wire, GPS L1 CP antenna, shown in Fig. 16, has been obtained (Lin et al. 2011) using two CLL-based wire NFRP elements driven by a coaxially fed semiloop antenna. It has $a = 15$ mm giving ka = 0.495 at $f_{res} = 1575.4$ MHz, a 96.9 % overall efficiency, a 30.7 MHz (−10 dB) bandwidth, a 7.9 MHz CP bandwidth, and an axial ratio of 0.6. The requisite phase shift is obtained by tuning the angle between the semiloop antenna and the NFRP elements (Fig. 20).

DOI 10.1007/978-981-4560-75-7_14-1

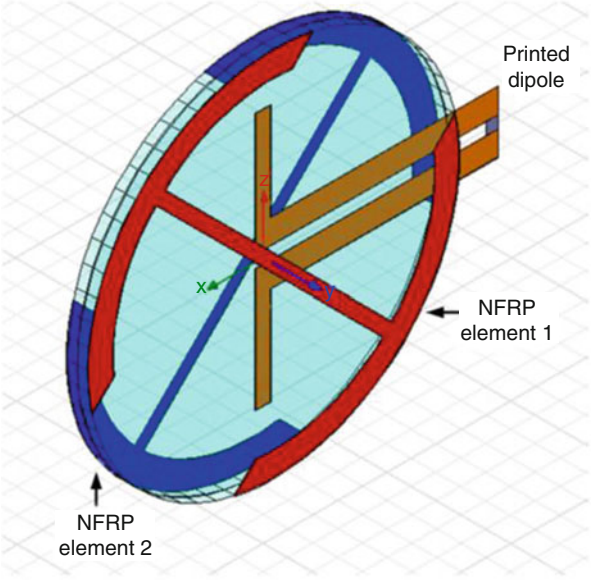


Fig. 19 A GPS L1 CP antenna consisting of two dipole NFRP elements and a driven printed dipole antenna. The CP behavior is achieved at a frequency between two closely spaced resonances through a 90° phase shift obtained between their input reactances. This Rogers Duroid™ 5880 design consists of three metal and two dielectric layers. The NFRP elements are on the outside faces of the dielectric layers; the printed dipole lies between those dielectric layers

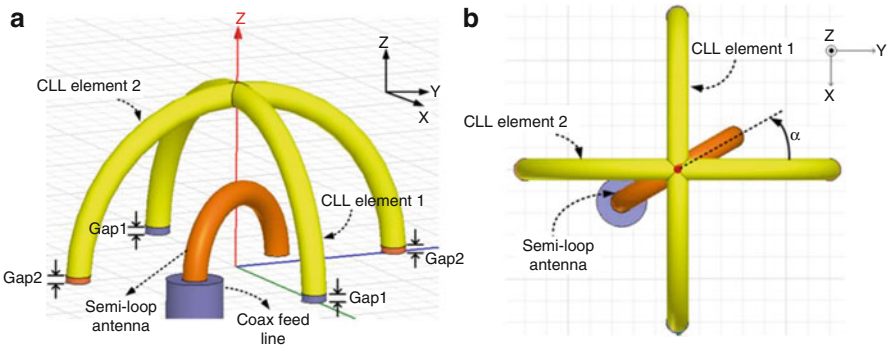


Fig. 20 Geometry of a GPS L1, CLL-based NFRP CP wire antenna whose coax-fed semi-loop is rotated α° with respect to the ZY-plane. The two CLL NFRP elements have two closely spaced resonances; the requisite 90° phase shift between their input reactances is obtained by tuning the angle α between the driven semi-loop antenna and the NFRP elements

In the same manner, the GS/GPS antenna shown in Fig. 17 was modified to achieve the CP GPS-L1 version shown in Fig. 21, i.e., by simply tuning the resonances of each NFRP element to the requisite values. The axial ratio is minimum at 1.580 GHz with a value AR = 0.84. The overall efficiency at this frequency was 73.41 %. The corresponding peak directivity was (1.88 dB).

Multifunctional Designs

The multiband and CP aspects have been combined together (Jin and Ziolkowski 2011) into the dual-band GPS L1/L2 protractor antenna whose design is shown in Fig. 22a and whose fabricated prototype is shown in Fig. 22b. It was found possible to interleave two pairs of protractor NFRP elements with the appropriate phasing between the resonances to produce CP at both the GPS L1 and L2 frequencies with a single feed structure. As depicted in Fig. 22a, one can see in Fig. 22b the two pairs of protractor elements

DOI 10.1007/978-981-4560-75-7_14-1

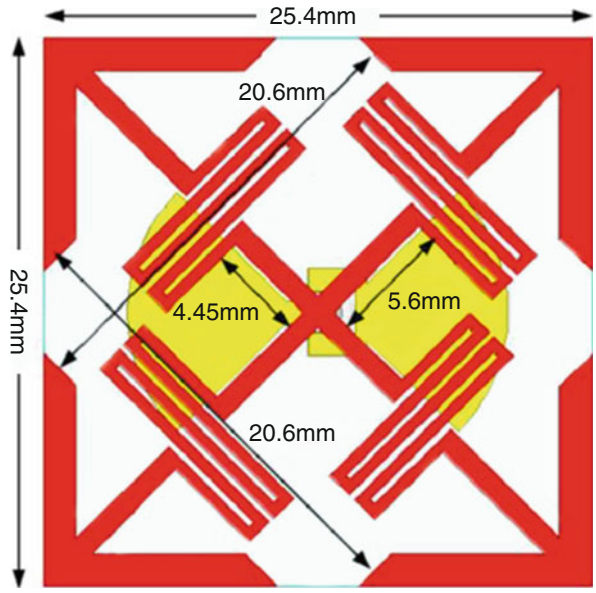


Fig. 21 Planar GPS L1 CP NFRP antenna

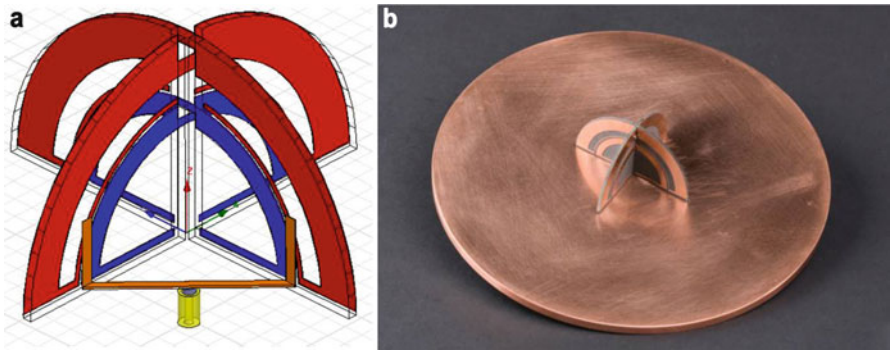


Fig. 22 Dual-band CP GPS L1/L2 NFRP antenna. (**a**) HFSS design (the L1 and L2 bands are associated, respectively, with the blue inside and red outside NFRP elements), (**b**) fabricated prototype

lying on two pieces of Rogers Duroid^{TM} 5880, which are oriented orthogonal to each other and to the finite copper ground plane. The initial experiments have demonstrated the basic operating principles, including the high overall efficiencies and the CP behavior (Jin and Ziolkowski 2011).

Planar multiband, CP GPS antennas have been obtained (Ta et al. 2012, 2013a, b, c) as variations of the designs shown in Figs. 17 and 21. The version shown in Fig. 23 covers the GPS L1–L5 bands. The primary radiating elements are two crossed printed dipoles, which incorporate a 90 phase delay line realized with a vacant-quarter printed ring to produce the CP radiation and broadband impedance matching. To achieve multiple resonances, each dipole arm is divided into four branches with different lengths, and a printed inductor with a barbed end is inserted in each branch to reduce the radiator size. An inverted, pyramidal, cavity-backed reflector is incorporated with the crossed dipoles to produce a unidirectional radiation pattern with a wide 3-dB axial ratio (AR) beamwidth and a high front-to-back ratio. This multiband CP antenna has broad impedance matching and 3-dB AR bandwidths covering the GPS L1–L5 bands.

DOI 10.1007/978-981-4560-75-7_14-1

Elemental Huygens Sources

Using the same NFRP element principles, an electrically small, elemental Huygens source can be obtained (Jin and Ziolkowski 2010b). It requires the combination of an electric and a magnetic dipole. A GPS L1 planar design that combines the Egyptian axe NFRP element for the electric dipole and two NFRP protractor elements to achieve a similar amplitude magnetic dipole is shown in Fig. 24a. It is a Rogers Duroid^{TM} 5880 design consisting of three metal and two dielectric layers. The NFRP elements lie on the outside faces of the dielectric layers; the driven printed dipole antenna lies between those dielectric layers. It has $ka = 0.46$ at $f_{res} = 1475$ MHz, a 85.9 % overall efficiency, and a 23.2 MHz (−10 dB)

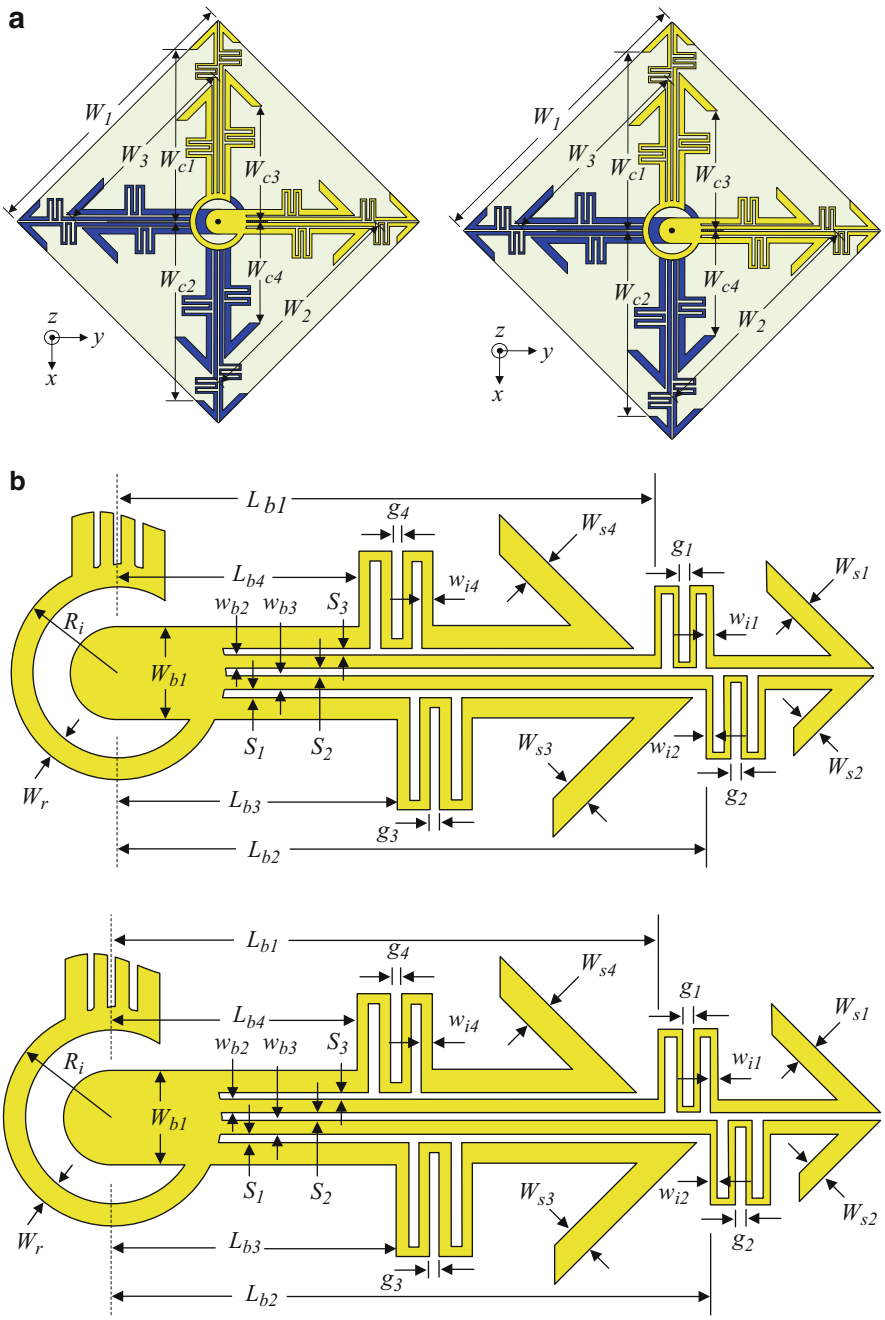


Fig. 23 (continued)

DOI 10.1007/978-981-4560-75-7_14-1

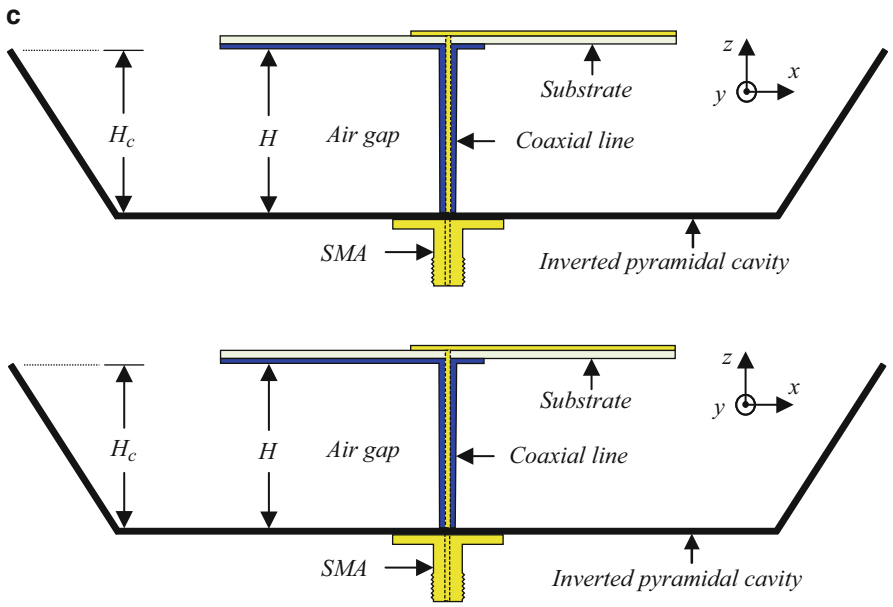


Fig. 23 Geometry of the multi-branch, asymmetrically barbed, crossed dipoles. (**a**) *Top view*; (**b**) vacant-quarter printed-ring and dipole arm; and (**c**) *side view*, including the cavity-backed reflector

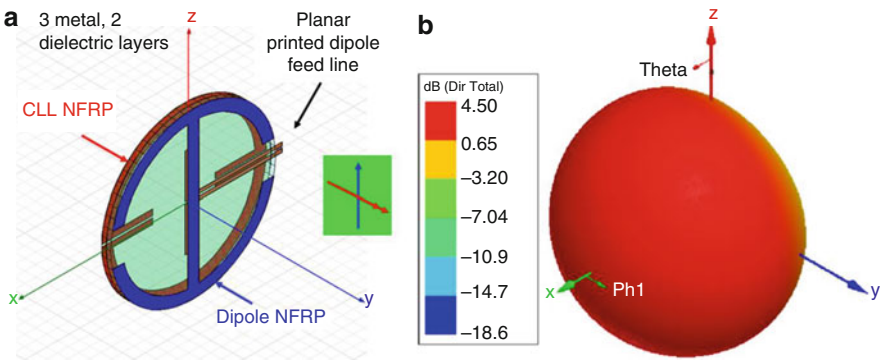


Fig. 24 An electrically small Huygens source that combines equal amplitude electric and magnetic dipoles and nearly achieves the maximum possible directivity. (**a**) HFSS design (the effective electric and magnetic dipoles are depicted in the insert), and (**b**) HFSS-predicted 3D directivity pattern

bandwidth. The HFSS-predicted 3D directivity pattern at this resonance frequency is shown in Fig. 18b. It has a maximum directivity of 4.5 dB and a 17.1 dB front-to-back ratio. The ideal maximum is 4.77 dB (i.e., a directivity of 3).

Using these concepts, a Huygens source nanoparticle laser has been designed (Liberal et al. 2014). The design is a three-layer, 120 nm outer-radius, nanoparticle consisting of a gain impregnated silicon core, a silver (ENG) shell surrounding it, and a gain impregnated silicon outer coating. At visible wavelengths, the high permittivity of silicon allows the appearance of a magnetic dipole mode in addition to the electric one. The former is associated with the outer shell; the latter with the core. By tuning the radius and gain constants, a proper balance of the electric and magnetic dipole modes is obtained. Lasing with a directivity of 3 in the forward scattering direction with respect to the exciting field occurs at 673.1 nm.

DOI 10.1007/978-981-4560-75-7_14-1

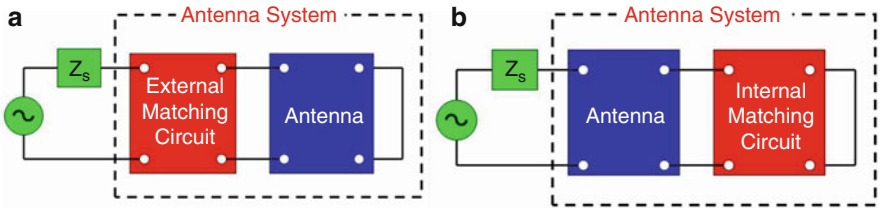


Fig. 25 Introducing active elements into an antenna system allows one to overcome the passive bounds on its bandwidth. (**a**) External active matching circuit, and (**b**) internal active matching element

Non-Foster Electrically Small Antennas

As noted previously, there is an upper bound on the bandwidth of a passive electrically small antenna. As noted, recent efforts (Gustafsson et al. 2009; Yaghjian and Stewart 2010) have led to improved versions of the original Chu limit. Nevertheless, as Eq. 2 indicates, the fractional bandwidth basically decreases as $(ka)^3$ when ka becomes smaller. Moreover, it also indicates that one can obtain more bandwidth simply by reducing the radiation efficiency, i.e., by introducing more loss into the antenna. Another more complicated approach is to combine together multiple radiating elements with slightly different but overlapping resonance frequencies. It was found that some care must be exercised in this strategy. If the NFRP elements are too tightly coupled, the resonances tend to merge together and produce an even narrower bandwidth response. Moreover, it is usually difficult physically to squeeze several elements into an electrically small footprint. A variety of metamaterial-inspired electrically small antennas have been designed with Q factors that reach the fundamental bound (Best 2004, 2005, 2009; Kim et al. 2010; Kim 2010). Note, however, that even if an antenna reaches the upper bound, that bound is extremely small and the bandwidth cannot surpass it.

A method to overcome the passive bound is to introduce active elements into the antenna system. As shown in Fig. 25a, the standard approach (Aberle and Loepsinger-Romak 2007; Sussman-Fort and Rudish 2009) is to introduce an active matching circuit that produces the necessary resistance and reactance variations to maintain matching of the antenna to the source over a large bandwidth. Because of the sensitivity of active circuits, this becomes both a rather difficult design task and realization issue. In contrast, a metamaterial-inspired paradigm is to introduce the active element internally, integrating it with the NFRP element. This internal matching circuit approach, illustrated in Fig. 25b, has led to several successful designs (Jin and Ziolkowski 2010a; Zhu and Ziolkowski 2011, 2012a, b, 2013; White et al. 2012; Mirzaei and Eleftheriades 2013).

Consider the NFRP antenna shown in Fig. 26. This canopy antenna was designed (Jin and Ziolkowski 2010a) to approach the Chu-Thal limit. The legs of the canopy are lumped element inductors; the canopy is a copper spherical shell cap. With passive inductors, $ka = 0.047$ designs with $Q = 1.75\ Q_{Chu} = 1.17\ Q_{Thal}$ have been obtained with high overall efficiencies.

It was demonstrated further, as with the Z antennas, that the antenna resonances can be tuned over a frequency band with the same physical structure by varying only the value of the inductors. It was found that with this frequency-agile behavior, the inductor value decreases with an increasing resonance frequency, i.e., to achieve a larger instantaneous bandwidth, the inductor would have to be an active non-Foster element. Such an active inductor was designed with a Linvill-based negative impedance convertor (NIC) (Linvill 1953). HFSS simulations demonstrated a 10 % fractional bandwidth at 300 MHz was possible when $ka = 0.047$, nearly a thousandfold improvement over the passive design. This active design provides a different paradigm from the traditional non-Foster approaches. In principle, it requires only non-Foster reactive matching without any resistive matching. This simplifies the requirements on the active element considerably. This design is analogous to the original active analytical solution that

DOI 10.1007/978-981-4560-75-7_14-1

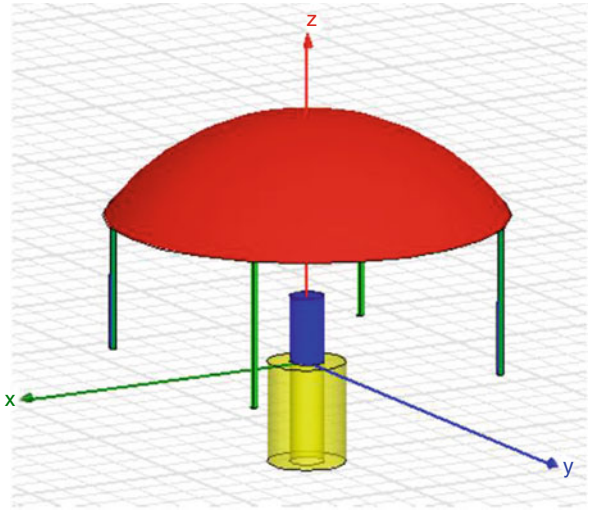


Fig. 26 Canopy antenna design. The NFRP element consists of four vertical legs (*green*), each being a lumped element inductor, and a copper spherical cap (*red*). The driven element is a monopole (*blue*) coaxially fed (*yellow*) through the ground (xy) plane. Efficient designs with $ka = 0.047$ at 300 MHz when the inductor is passive have bandwidths that approach the Chu-Thal limits and exhibit over 10 % bandwidth when an active inductor is used

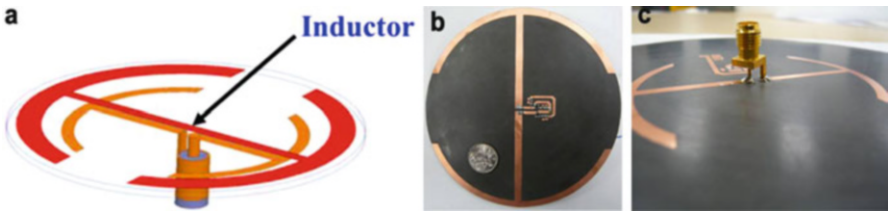


Fig. 27 Non-Foster-augmented Egyptian axe dipole (EAD) antenna to realize an efficient, electrically small antenna with large instantaneous impedance bandwidth. (**a**) Isometric view of the HFSS model, and (**b**) top and (**c**) bottom views of the fabricated and measured prototype

produced a large fractional bandwidth for the idealized metamaterial-based, gain impregnated core-shell NFRP element, dipole-driven antenna system (Ziolkowski and Erentok 2007).

To test these non-Foster-augmented, metamaterial-inspired antenna concepts, the Egyptian axe dipole antenna shown in Fig. 27a was designed, fabricated, and tested. A lumped element inductor, which is placed across the gap centered in the NFRP element, is treated as a tunable element. The $ka = 0.444$ EAD design for 300 MHz was obtained initially. Its frequency-agile behavior was then determined by varying the inductor value. The resulting non-Foster reactance was matched with a two-BJT Linvill NIC circuit. This circuit was then incorporated into the NFRP element as shown in Figs. 27b and c. The measured instantaneous impedance bandwidth was improved by a factor of 6.2 and was about a factor of 4 greater than the passive upper bound (Zhu and Ziolkowski 2012b, 2013). Similar performance improvements were obtained for a non-Foster-augmented protractor antenna (Zhu and Ziolkowski 2012a).

As with any active element, the stability of a non-Foster element is a significant practical issue. Component and fabrication tolerances can adversely affect the system enough that it will become unstable. It has been demonstrated that a key test of stability is to consider the time-domain response of the entire system – the non-Foster circuit and the antenna system augmented with it. Growth in time of any frequency component is an indicator of an instability. The stability of the NIC circuits and their integration

DOI 10.1007/978-981-4560-75-7_14-1

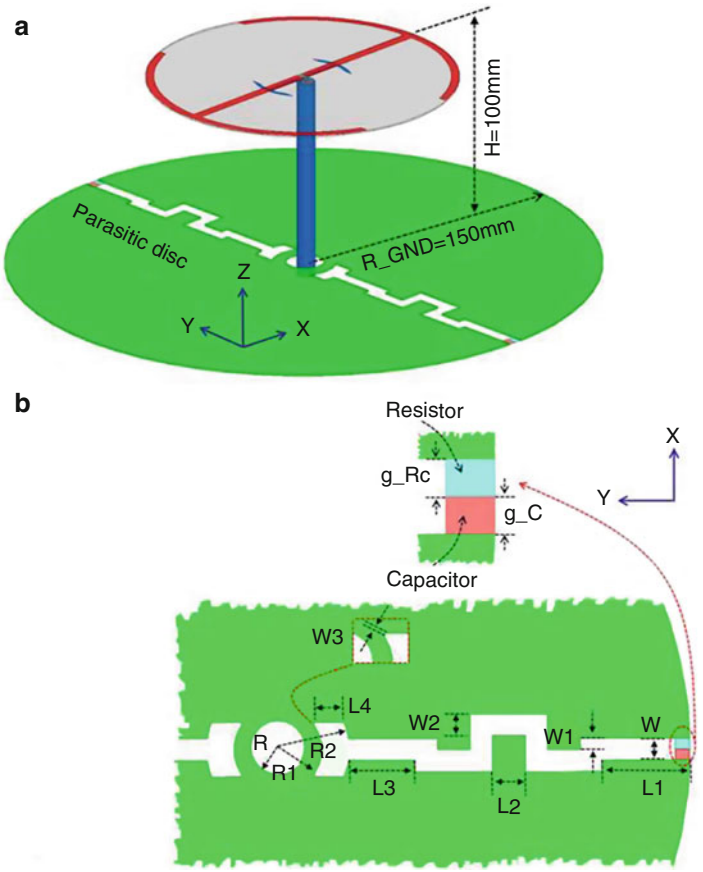


Fig. 28 EAD antenna integrated with a slot-modified, parasitic copper disk. (**a**) Isometric view, and (**b**) zoom-in view of one of the capacitor-augmented slot regions

with the EAD's radiating elements was confirmed in this manner before it was verified by experiment (Zhu and Ziolkowski 2013).

The internal non-Foster design paradigm has been extended to enhance not only the impedance bandwidth but also the directivity bandwidth (Tang et al. 2013). This was accomplished by introducing yet another NFRP element, augmenting it with a non-Foster element, and tuning it to act like an AMC surface. This approach also provided a solution to the holy grail of electrically small antennas: a design that exhibits simultaneously high efficiency, high directivity, impedance matching, and large front-to-back ratio (FTBR) over a broad instantaneous bandwidth. This antenna configuration is shown in Fig. 28 (Ziolkowski et al. 2013). The EAD antenna is augmented with an inductor-based NIC; it is integrated with a slot-modified parasitic disk that is augmented with a capacitor-based NIC.

The HFSS-simulated performance characteristics of this electrically small, $ka = 0.94$, system include a radiation efficiency > 81.63 %, directivity > 6.25 dB, FTBR > 26.71 dB over a 10.0 % fractional bandwidth. This yields a directivity-to-quality factor ratio which is greater than 10 times the fundamental upper bound: $(D/Q) > 10 \times (D/Q)_{\text{bare EAD}}$. These results are illustrated in Fig. 29. The directivity patterns in Fig. 29a have the characteristic cardioid pattern. This behavior persists over the bandwidth of the system. Consequently, it confirms that the parasitic disk acts as a broad-bandwidth AMC element. The other performance characteristics in Fig. 29b demonstrate a rather uniform behavior of the various performance characteristics over the entire bandwidth of the system.

DOI 10.1007/978-981-4560-75-7_14-1

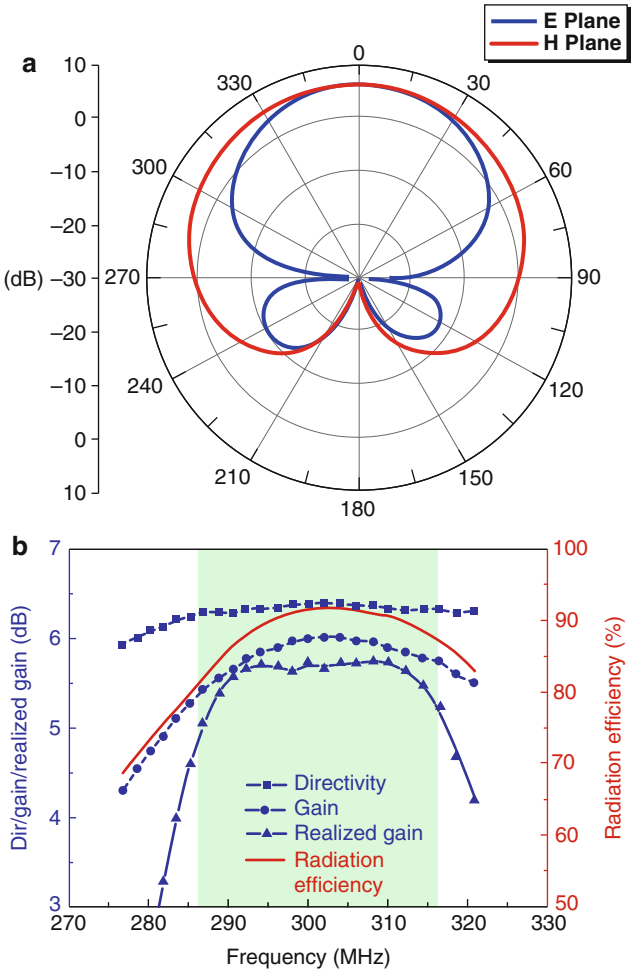


Fig. 29 Performance characteristics of the EAD antenna augmented with an inductive NIC and with a slot-modified, parasitic copper disk integrated with capacitive NICs. (**a**) E- and H-plane directivity patterns for the fully NIC-augmented EAD antenna at $f_{res} = 300$ MHz, and (**b**) simulated D, G, $G_{realized}$, and RE values versus the resonance frequency. The cyan shaded region represents, for easy reference, the instantaneous $BW_{10\ dB}$ region.

Summary

As more has been learned about metamaterials and their electromagnetic properties, a variety of metamaterial-based and metamaterial-inspired antenna systems have been developed from RF to optical frequencies. Their metamaterial constructs have led to enhanced performance characteristics. An electrically small antenna paradigm, the metamaterial-inspired near-field resonant parasitic (NFRP) antenna, is an example. It has provided realizations of a variety of interesting multifrequency, multifunctional designs whose input impedance can be nearly completely matched to the source and that exhibit high radiation efficiencies. The different coupling mechanisms between the driven and NFRP elements provide extra degrees of freedom to tailor the currents that produce the radiated fields. Both planar and volumetric loadings of space are possible with the NFRP elements, leading to more flexibility in antenna designs. The introduction of internal non-Foster elements provides a means to overcome the physical constraints associated with passive systems, e.g., a path to large instantaneous bandwidths. Moreover, the recognition that non-Foster-based elements can also be used to modify other antenna properties provides yet more design degrees of freedom. In particular, it allows one to design an electrically small antenna with simultaneous high efficiency, high directivity, large front-to-back ratio, and nearly complete impedance

DOI 10.1007/978-981-4560-75-7_14-1

matching over a large instantaneous bandwidth, i.e., active metamaterial constructs have provided the means to overcome the usually assumed performance characteristic trade-off approach to small antenna design.

At the heart of any wireless system is an antenna. Metamaterials, whether explicitly introduced into an antenna system as physical constructs or implicitly introduced using their often exotic properties to guide specifics of a design, have provided exciting new opportunities to antenna engineers to meet the needs of a continually increasing demand for wireless products. Given slightly more than a decade of experience with the engineering applications of metamaterials, there is a great hope that they may be able to provide additional degrees of freedom for the design of many different forms of electromagnetic systems from DC to light.

Cross-References

► Physical Bounds of Antennas
► Theory of Transformation Optics in Antenna Design

References

Aberle JT, Loepsinger-Romak R (2007) Antenna with non-foster matching networks. Morgan & Claypool Publishers, San Rafael

Alici KB, Ozbay E (2007) Radiation properties of a split ring resonator and monopole composite. Phys Stat Sol (b) 244:1192–1196

Alú A, Bilotti F, Engheta N, Vegni L (2007) Subwavelength, compact, resonant patch antennas loaded with metamaterials. IEEE Trans Antennas Propag 55:13–25

Antoniades MA, Eleftheriades GV (2008) A folded-monopole model for electrically small NRI-TL metamaterial antennas. IEEE Ant Wireless Propag Lett 7:425–428

Antoniades MA, Eleftheriades GV (2009) A broadband dual-mode monopole antenna using NRI-TL metamaterial loading. IEEE Ant Wireless Propag Lett 8:258–261

Arslanagić S, Ziolkowski RW, Breinbjerg O (2007) Radiation properties of an electric Hertzian dipole located near-by concentric metamaterial spheres. Rad Sci 42:RS6S16

Arslanagić S, Ziolkowski RW (2010) Active coated nano-particle excited by an arbitrarily located electric Hertzian dipole – resonance and transparency effects. J Opt 12:024014

Arslanagić S, Ziolkowski RW (2012) Directive properties of active coated nano-particles. Advance Electromagn 1:57–64

Arslanagić S, Ziolkowski RW (2013) Jamming of quantum emitters by active coated nano-particles. IEEE J Sel Topics Quantum Electron 19:4800506

Arslanagić S, Ziolkowski RW (2014) Influence of active nano particle size and material composition on multiple quantum emitter enhancements: their enhancement and jamming effects. Prog Electromagn Res 149:85–99

Best SR (2004) The radiation properties of electrically small folded spherical helix antennas. IEEE Trans Antennas Propag 52:953–960

Best SR (2005) Low Q electrically small linear and elliptical polarized spherical dipole antennas. IEEE Trans Antennas Propag 53:1047–1053

Best SR (2009) A low Q electrically small magnetic (TE mode) dipole. IEEE Ant Wireless Propag Lett 8:572–575

DOI 10.1007/978-981-4560-75-7_14-1

Bilotti F, Alú A, Vegni L (2008) Design of miniaturized metamaterial patch antennas with μ-negative loading. IEEE Trans Antennas Propag 56:1640–1647

Buell K, Mosallaei H, Sarabandi K (2006) A substrate for small patch antennas providing tunable miniaturization factors. IEEE Trans Microw Theor Tech 54:135–146

Caloz C, Itoh T (2005) Electromagnetic metamaterials: transmission line theory and microwave applications. Wiley-IEEE, Hoboken

Caloz C, Itoh T, Rennings A (2008) CRLH traveling-wave and resonant metamaterial antennas. IEEE Ant Propag Mag 50:25–39

Chu LJ (1948) Physical limitations of omni-directional antennas. J Appl Phys 19:1163–1175

Dong Y, Itoh T (2012) Metamaterial-based antennas. Proc IEEE 100:2271–2285

Eleftheriades GV, Balmain KG (eds) (2005) Negative-refraction metamaterials fundamental principles and applications. Wiley-IEEE, Hoboken

Eleftheriades GV, Antoniades MA, Qureshi F (2007) Antenna applications of negative-refractive-index transmission-line structures. IET Microw Ant Propag 1:12–22

Engheta N, Ziolkowski RW (2005) A positive future for double negative metamaterials. IEEE Microwav Theor Tech 53:1535–1556

Engheta N, Ziolkowski RW (eds) (2006) Metamaterials: physics and engineering explorations. Wiley-IEEE Press, Hoboken, NJ

Enoch S, Tayeb G, Sabouroux Gúerin PN, Vincent P (2002) A metamaterial for directive emission. Phys Rev Lett 89:213902

Erentok A, Luljak PL, Ziolkowski RW (2005) Characterization of a volumetric metamaterial realization of an artificial magnetic conductor for antenna applications. IEEE Trans Ant Propag 53:160–172

Erentok A, Ziolkowski RW (2007a) A hybrid optimization method to analyze metamaterial-based electrically small antennas. IEEE Trans Ant Propag 55:731–741

Erentok A, Ziolkowski RW (2007b) An efficient metamaterial-inspired electrically-small antenna. Microw Opt Tech Lett 49:1287–1290

Erentok A, Ziolkowski RW (2007c) Two-dimensional efficient metamaterial-inspired electrically-small antenna. Microw Opt Tech Lett 49:1669–1673

Erentok A, Lee D, Ziolkowski RW (2007d) Numerical analysis of a printed dipole antenna integrated with a 3D AMC block. IEEE Ant Wireless Propag Lett 6:134–136

Erentok A, Ziolkowski RW, Nielsen JA, Greegor RB, Parazzoli CG, Tanielian MH, Cummer SA, Popa BI, Hand T, Vier DC, Schultz S (2007b) Low frequency lumped element-based negative index metamaterial. Appl Phys Lett 91:184104

Erentok A, Ziolkowski RW (2008) Metamaterial-inspired efficient electrically-small antennas. IEEE Trans Ant Propag 56:691–707

Franson SJ, Ziolkowski RW (2009) Confirmation of zero-N behavior in a high gain grid structure at millimeter-wave frequencies. IEEE Ant Wireless Propag Lett 8:387–390

Geng J, Ziolkowski RW, Jin R, Liang X (2011) Numerical study of active open cylindrical coated nano-particle antennas. IEEE Photon 3:1093–1110

Geng J, Ziolkowski RW, Jin R, Liang X (2012) Detailed performance characteristics of vertically polarized, cylindrical, active coated nano-particle antennas. Rad Sci 47, RS2013

Geng J, Jin R, Liang X, Ziolkowski RW (2013) Active cylindrical coated nano-particle antennas: polarization-dependent scattering properties. J Electromagnet Wave Appl (JEMWA). doi:10.1080/09205071.2013.809669

Gordon JA, Ziolkowski RW (2007) The design and simulated performance of a coated nano-particle laser. Opt Express 15:2622–2653

Gordon JA, Ziolkowski RW (2008) CNP optical metamaterials. Opt Express 16:6692–6716

DOI 10.1007/978-981-4560-75-7_14-1

Greegor RB, Parazzoli CG, Nielsen JA, Tanielian MH, Vier DC, Schultz S, Holloway CL, Ziolkowski RW (2009) Demonstration of impedance matching using a mu-negative (MNG) metamaterial. IEEE Ant Wireless Propag Lett 8:92–95

Gustafsson M, Sohl C, Kristensson G (2009) Illustrations of new physical bounds on linearly polarized antennas. IEEE Trans Antennas Propag 57:1319–1327

Herraiz-Martnez J, Garca-Muoz LE, Gonzlez-Ovejero D, Gonzlez-Posadas V, Segovia-Vargas D (2009) Dual-frequency printed dipole loaded with split ring resonators. IEEE Ant Wireless Propag Lett 8:137–140

Ikonen PMT, Alitalo P, Tretyakov SA (2007) On impedance bandwidth of resonant patch antennas implemented using structures with engineered dispersion. IEEE Ant Wireless Propag Lett 6:186–190

Imhof PD, Ziolkowski RW, Mosig JR (2006) Highly isotropic, low loss epsilon negative (ENG) unit cells at UHF frequencies. In: Proc European conference on antennas and propagation, EuCAP2006, European Space Agency, Noordwijk, The Netherlands, ESA SP-626, pp 552

Imhof PD (2006) Metamaterial-based epsilon negative (ENG) media: analysis and designs. Ecole Polytechnique Fédérale de Lausanne (EPFL) Master Thesis, Lausanne, Switzerland

Jin P, Ziolkowski RW (2009) Low Q, electrically small, efficient near field resonant parasitic antennas. IEEE Trans Ant Propag 57:2548–2563

Jin P, Ziolkowski RW (2010a) Broadband, efficient, electrically small metamaterial-inspired antennas facilitated by active near-field resonant parasitic elements. IEEE Trans Ant Propag 58:318–327

Jin P, Ziolkowski RW (2010b) Metamaterial-inspired, electrically small Huygens sources. IEEE Ant Wireless Propag Lett 9:501–505

Jin P, Ziolkowski RW (2010c) Multiband extensions of the electrically small metamaterial-engineered Z antenna. IET Microwav Ant Propag 4:1016–1025

Jin P, Ziolkowski RW (2011) Multi-frequency, linear and circular polarized, metamaterial-inspired near-field resonant parasitic antennas. IEEE Trans Ant Propag 59:1446–1459

Jin P, Lin CC, Ziolkowski RW (2012) Multifunctional, electrically small, conformal near-field resonant parasitic antennas. IEEE Ant Wireless Propag Lett 11:200–204

Kim OS, Breinbjerg O (2009) Miniaturized self-resonant split-ring resonator antenna. Electronics Lett 45:196–197

Kim OS, Breinbjerg O, Yaghjian AD (2010) Electrically small magnetic dipole antennas with quality factors approaching the Chu lower bound. IEEE Trans Ant Propag 58:1898–1906

Kim OS (2010) Low-Q electrically small spherical magnetic dipole antennas. IEEE Trans Ant Propag 58:2210–2217

Lai A, Leong MKH, Itoh T (2007) Infinite wavelength resonant antennas with monopolar radiation pattern based on periodic structures. IEEE Trans Ant Propag 55:868–876

Lee DH, Chauraya A, Vardaxoglou Y, Park WS (2008) A compact and low-profile tunable loop antenna integrated with inductors. IEEE Ant Wireless Propag Lett 7:621–624

Liberal I, Ederra I, Gonzalo R, Ziolkowski RW (2014) Induction theorem analysis of resonant nanoparticles: design of a Huygens source nanoparticle laser. Phys Rev Appl 1:044002

Lin CC, Ziolkowski RW, Nielsen JA, Tanielian MH, Holloway CL (2010) An efficient, low profile, electrically small, VHF 3D magnetic EZ antenna. Appl Phys Lett 96:104102

Lin CC, Ziolkowski RW (2010) Dual-band 3D magnetic EZ antenna. Microw Opt Tech Lett 52:971–975

Lin CC, Jin P, Ziolkowski RW (2011) Electrically small dual-band and circularly polarized magnetically-coupled near-field resonant parasitic wire antennas. IEEE Trans Ant Propag 59:714–724

Linvill JG (1953) Transistor negative-impedance converters. Proc IRE 41:725–729

Martinez A, Piqueras MA, Marti J (2006) Generation of highly directional beam by k-space filtering using a metamaterial flat slab with a small negative index of refraction. Appl Phys Lett 89:131111

Mirzaei H, Eleftheriades GV (2013) A resonant printed monopole antenna with an embedded non-Foster matching network. IEEE Trans Ant Propag 61:5363–5371

Mumcu G, Sertel K, Volakis JL (2009) Miniature antenna using printed coupled lines emulating degenerate band edge crystals. IEEE Trans Ant Propag 57:1618–1624

Park J-H, Ryu YH, Lee JG, Lee JH (2007) Epsilon negative zeroth-order resonator antenna. IEEE Trans Ant Propag 55:3710–3712

Qureshi F, Antoniades MA, Eleftheriades GV (2005) Compact and low-profile metamaterial ring antenna with vertical polarization. IEEE Ant Wireless Propag Lett 4:333–336

Sáenz E, Gonzalo R, Ederra I, Vardaxoglou JC, de Maagt P (2008) Resonant meta-surface superstrate for single and multifrequency dipole antenna arrays. IEEE Trans Ant Propag 56:951–960

Sanada A, Kimura M, Awai I, Caloz C, Itoh T (2004) A planar zeroth-order resonator antenna using a left-handed transmission line. In: Proc. 34th European Microwave Conference, Amsterdam, The Netherlands, pp 1341–1344

Smith DR, Padilla WJ, Vier DC, Nemat-Nasser SC, Schultz S (2000) Composite medium with simultaneously negative permeability and permittivity. Phys Rev Lett 84:4184–4187

Stuart HR, Tran C (2005) Subwavelength microwave resonators exhibiting strong coupling to radiation modes. Appl Phys Lett 87:151108

Stuart HR, Pidwerbetsky A (2006) Electrically small antenna elements using negative permittivity resonators. IEEE Trans Ant Propag 54:1664–1653

Stuart HR, Tran C (2007) Small spherical antennas using arrays of electromagnetically coupled planar elements. IEEE Ant Wireless Propag Lett 6:7–10

Sussman-Fort SE, Rudish RM (2009) Non-Foster impedance matching of electrically-small antennas. IEEE Trans Ant Propag 57:2230–2241

Ta SX, Park I, Ziolkowski RW (2012) Dual-band wide-beam crossed asymmetric dipole antenna for GPS applications. Electronic Lett 48:1580–1581

Ta SX, Park I, Ziolkowski RW (2013a) Circularly polarized crossed dipole on an HIS for 2.4/5.2/5.8-GHz WLAN applications. IEEE Ant Wireless Propag Lett 12:1464–1467

Ta SX, Park I, Ziolkowski RW (2013b) Multi-band, wide-beam, circularly polarized, crossed asymmetrically barbed arrowhead dipole antenna for GPS applications. IEEE Trans Ant Propag 61:5771–5775

Ta SX, Han JJ, Park I, Ziolkowski RW (2013c) Wide-beam circularly polarized crossed scythe-shaped dipoles for global navigation satellite systems. J Electromagn Eng Sci 13:224–232

Tang MC, Zhu N, Ziolkowski RW (2013) Augmenting a modified Egyptian axe dipole antenna with non-Foster elements to enlarge its directivity bandwidth. IEEE Ant Wireless Propag Lett 12:421–424

Thal H (2006) New radiation Q limits for spherical wire antennas. IEEE Trans Ant Propag 54:2757–2763

Veselago VG (1968) Experimental demonstration of negative index of refraction. Sov Phys Usp 47:509–514

White CR, Colburn JS, Nagele RG (2012) A non-Foster VHF monopole antenna. IEEE Ant Wireless Propag Lett 21:584–587

Wu BI, Wang W, Pacheco J, Chen X, Grzegorczyk T, Kong JA (2005) A study of using metamaterials as antenna substrate to enhance gain. Progress in Electromagnetics Research, PIER 51, EMW Publishing, Cambridge, MA, pp 295–328

Yaghjian AD, Best SR (2005) Impedance, bandwidth, and Q of antennas. IEEE Trans Ant Propag 53:1298–1324

Yaghjian AD, Stewart HR (2010) Lower bounds on the Q of electrically small dipole antennas. IEEE Trans Ant Propag 58:3114–3121

Yang F, Rahmat-Samii Y (2009) Electromagnetic band gap structures in antenna engineering. Cambridge University Press, New York

DOI 10.1007/978-981-4560-75-7_14-1

Zhu J, Antoniades MA, Eleftheriades GV (2010) A compact tri-band monopole antenna with single-cell metamaterial loading. IEEE Trans Ant Propag 244:1031–1038

Zhu N, Ziolkowski RW (2011) Active metamaterial-inspired broad bandwidth, efficient, electrically small antennas. IEEE Ant Wireless Propag Lett 10:1582–1585

Zhu N, Ziolkowski RW (2012a) Design and measurements of an electrically small, broad bandwidth, non-Foster circuit-augmented protractor antenna. Appl Phys Lett 101:024107

Zhu N, Ziolkowski RW (2012b) Broad bandwidth, electrically small antenna augmented with an internal non-Foster element. IEEE Ant Wireless Propag Lett 11:1116–1120

Zhu N, Ziolkowski RW (2013) Broad bandwidth, electrically small, non-Foster element-augmented antenna designs, analyses, and measurements. IEICE Trans Commun E96-B:2399–2409

Ziolkowski RW (2003) Design, fabrication, and testing of double negative metamaterials. IEEE Trans Ant Propag 51:1516–1529

Ziolkowski RW, Kipple A (2003) Application of double negative metamaterial to increase the power radiated by electrically small antennas. IEEE Trans Ant Propag 51:2626–2640

Ziolkowski RW (2004) Propagation in and scattering from a matched metamaterial having a zero index of refraction. Phys Rev E 70:046608

Ziolkowski RW, Kipple A (2005) Reciprocity between the effects of resonant scattering and enhanced radiated power by electrically small antennas in the presence of nested metamaterial shells. Phys Rev E 72:036602

Ziolkowski RW, Erentok A (2006) Metamaterial-based efficient electrically small antennas. IEEE Trans Ant Propag 54:2113–2130

Ziolkowski RW, Erentok A (2007) At and beyond the Chu limit: passive and active broad bandwidth metamaterial-based efficient electrically small antennas. IET Microwav Ant Propag 1:116–128

Ziolkowski RW (2008a) An efficient, electrically small antenna designed for VHF and UHF applications. IEEE Ant Wireless Propag Lett 7:217–220

Ziolkowski RW (2008b) Efficient electrically small antenna facilitated by a near-field resonant parasitic. IEEE Ant Wireless Propag Lett 7:580–583

Ziolkowski RW, Jin P (2008) Metamaterial-based dispersion engineering to achieve phase center compensation in a log-periodic array. IEEE Trans Ant Propag 56:3619–3629

Ziolkowski RW, Lin CC, Nielsen JA, Tanielian MH, Holloway CL (2009a) Design and experimental verification of a 3D magnetic EZ antenna at 300 MHz. IEEE Ant Wireless Propag Lett 8:989–993

Ziolkowski RW, Jin P, Nielsen JA, Tanielian MH, Holloway CL (2009b) Design and experimental verification of Z antennas at UHF frequencies. IEEE Ant Wireless Propag Lett 8:1329–1333

Ziolkowski RW, Jin P, Lin CC (2011) Metamaterial-inspired engineering of antennas. Proc IEEE 99:1720–1731

Ziolkowski RW, Tang MC, Zhu N (2013) An efficient, broad bandwidth, high directivity, electrically small antenna. Microw Opt Technol Lett 55:1430–1434

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_15-1

Optimization Methods in Antenna Engineering

Douglas Werner*, Micah Gregory, Zhi Hao Jiang and Donovan E. Brocker
The Pennsylvania State University, University Park, PA, USA

Abstract

Optimization strategies have been heavily used in the antenna design community for many years. Oftentimes, the details of the strategies and optimization configurations are left out of the designs in open literature. The goal of this chapter is to inform the reader of many of the algorithms currently being used in the antenna design community, give a detailed operation of four of the most commonly used algorithms, and compare their performance on test functions and an example antenna design problem. Additionally, several recent examples of complex antenna designs which made use of these algorithms will be examined.

Keywords

Optimization; Evolutionary strategy; Antennas; Covariance matrix adaptation; Particle swarm; Differential evolution; Genetic algorithm; Function minimization

Introduction

Due to the complex and ever-changing requirements for antenna capabilities, global optimization strategies have seen extensive and increasing use in the field of antenna design. Antennas and systems which require ultra-wideband, multiband, and multifrequency operation are prime candidates for making use of optimization to find a suitable set of geometric and material properties which fulfill the requisite performance criteria. These strategies are often invaluable tools for the design of antenna structures which have no closed-form solutions to determine their impedance and gain characteristics. Oftentimes, an antenna must be designed in a constrained fashion due to the profile of the supporting device as well as other practical considerations. This can present a particularly challenging problem to the antenna engineer, one well suited for global optimization.

Optimization strategies are iterative algorithms that aim to minimize a user-determined cost function given a set of parameters and settings. The optimizable parameters are a set of binary or real-coded values that are passed to the cost function which is constructed such that low-cost values represent designs that are more desirable than those with higher values. For antenna design, the cost function evaluation process often includes some form of simulation or computation that determines the performance characteristics of the device generated using the associated set of input parameters. The cost function can return one or more values for single-objective or multi-objective optimization, respectively. Since these simulations or computations are usually very time-consuming processes, it is highly desirable to use an optimization strategy which requires the fewest cost function calls to achieve the desired cost function goal.

There exists many different types of optimization strategies, and it is not unusual that some are better suited for certain classes of antenna design than others. In this chapter, an overview of the operation,

*Email: dhw@psu.edu

DOI 10.1007/978-981-4560-75-7_15-1

functionality, and usage of several algorithms commonly used in antenna design will be provided. Then, some performance comparisons involving test functions and concept antenna designs will be presented. Lastly, several detailed examples of interesting antenna designs that have been produced with the aid of global optimization strategies will be covered.

Introduction to the Algorithms

Four algorithms are chosen here in this chapter for close examination of their performance on an assortment of test functions and antenna design problems. They are the genetic algorithm (GA), particle swarm optimization (PSO), differential evolution (DE), and the covariance matrix adaptation evolutionary strategy (CMA-ES). The GA is most commonly implemented in a binary scheme; however, the binary strings can also be used to represent discretized real values according to the number of bits chosen to represent each value. The remaining three algorithms are naturally real coded and lend themselves to optimizations with continuously variable parameters.

The GA has had a significant impact on the field of antenna design since it was conceived in the early 1970s (Holland 1973). It is an iterative, population-based algorithm with operation based on the theory of natural selection and survival of the fittest. The population members have their chromosomes (binary strings that form the set of input parameters to the cost function) blended and mutated in a way that lends itself to constantly forming better-performing (lower cost) future population members, while still effectively exploring the given search region. The binary operation makes it well suited for decision- or on/off-based optimization, but it has been applied effectively to many real-valued problems as well by representing real parameters with a finite number of chromosome bits. It has seen extensive application in antenna design as well as many other electromagnetic problems due to its ease of use, simple implementation, and global search abilities (Haupt and Werner 1995, 2007; Rahmat-Samii and Michielssen 1999; Weile and Michielssen 1997; Johnson and Rahmat-Samii 1997). For example, it has been used in the design of wire antennas (Altshuler 2002; Boag et al. 1996; Altshuler and Linden 1997; Werner et al. 2008), array synthesis (Bray et al. 2002; Gregory and Werner 2009; Gregory and Werner 2010; Gregory et al. 2010a; Petko and Werner 2008; Spence and Werner 2008; Boeringer and Werner 2004, 2005; Haupt 1994; Ares-Pena et al. 1999; Weile and Michielssen 1996), planar and patch antenna design (Kerkhoff and Ling 2007; Johnson and Rahmat-Samii 1999; Villegas et al. 2004), and many others (Jones and Joines 1997; Mosallaei and Rahmat-Samii 2001; Villegas et al. 2004; Santarelli et al. 2006; Akhoondzadeh-Asl et al. 2007; Godi et al. 2007; Werner et al. 2001).

PSO was introduced in the mid-1990s as an algorithm based on the swarm behavior of insects such as bees as they search for food (Eberhart and Kennedy 1995; Kennedy and Eberhart 1995, 2001). Like the GA, PSO is a population-based iterative algorithm; however, it tends to operate more cooperatively than competitively. Each population member moves around the search space based on recorded memory of its personal best-found location and the best-found location (lowest cost function value) of all of the population members (global best). Rather than binary strings that form chromosomes, each population member is comprised of two vectors of real-valued parameters, one forming its position and the other its velocity. Particle swarm has been a very popular algorithm in the antenna design community due to its ease of use, easy implementation, and reasonably good optimization performance (Robinson and Rahmat-Samii 2004; Gies 2004; Jin and Rahmat-Samii 2007, 2008). It has been successfully applied to a wide variety of antenna design problems such as multiband and wideband antennas (Jin and Rahmat-Samii 2005; Lizzi et al. 2007, 2008) as well as phased array design (Boeringer and Werner 2004; Boeringer et al. 2005) and many others.

DE appeared in the late 1990s as a new powerful global optimization algorithm (Storn and Price 1997). Being a real-valued strategy like PSO, it is well suited to many of the same design problems (Rocca et al. 2011; Goudos et al. 2011). It is an iterative, population-based algorithm that operates by generating

DOI 10.1007/978-981-4560-75-7_15-1

mutant vectors (candidate solution sets) based on current population members, testing to see if they perform better and replacing members if they do. Some examples of where DE has been applied include low-RCS antennas (Wang et al. 2010), circularly polarized microstrip antennas (Deb et al. 2014), monopulse antennas (Massa et al. 2006), slot antennas (Li et al. 2014), Yagi-Uda design (Goudos et al. 2010), and array synthesis (Chen and Wang 2012; Zhang et al. 2013; Lin et al. 2010; Panduro et al. 2009; Kurup et al. 2003).

The last algorithm to be considered is the covariance matrix adaptation evolutionary strategy (CMA-ES), introduced in the early 2000s by Hansen and Ostermeier (2001; Hansen et al. 2003). It has seen increasing interest coupled with a growing number of applications in the electromagnetic and antenna design community (Gregory et al. 2010b, 2011; Gregory and Werner 2011) due to its ease of use and fast, robust optimization characteristics. Like PSO and DE, it is also a global population-based, real-valued iterative algorithm. It operates by intelligently moving a Gaussian search distribution of varying shape and size throughout the search space. The population is formed by sampling from the search distribution at each iteration; the population size is simply the number of samples taken from this distribution. It has been successfully applied to many different electromagnetic and antenna design problems including planar antennas (Gregory et al. 2011; Gregory and Werner 2013), dielectric resonator antennas (Fang et al. 2011; Fang and Leung 2012; Leung et al. 2013; Pan et al. 2014), phased arrays (Gregory et al. 2011, 2013; Zhang et al. 2012), and electromagnetic band-gap materials (Martin et al. 2014).

These algorithms are all excellent choices for antenna designers seeking enhanced performance characteristics from their design criteria. In the following section, they will be discussed and analyzed on various problems, giving the reader a better understanding of how they operate, and which algorithms might be a better choice for a particular design problem.

Alternative Algorithms

Although not discussed in detail or applied in this chapter, there are many other algorithms that have been used in the design of electromagnetic structures such as antennas. Some are stochastic, and others are based on natural phenomena similar to the GA or PSO. Some examples include simulated annealing (Kirkpatrick et al. 1983; Murino et al. 1996; Coleman et al. 2004; Martínez-Fernández et al. 2007), ant colony optimization (Rajo-Iglesias and Quevedo-Teruel 2007; Rocca et al. 2008; Lewis et al. 2009; Dorigo et al. 2006), evolutionary programming (Hoorfar 2007; Jamnejad and Hoorfar 2004; Hoorfar and Chellapilla 1998), invasive weed optimization (Karimkashi and Kishk 2010), wind-driven optimization (WDO) (Bayraktar et al. 2011, 2013), clonal selection algorithm (CLONALG) (Castro and Von Zuben 2002; Campelo et al. 2005; Bayraktar et al. 2010, 2012), and many others. In addition to these, there are often variations of the aforementioned algorithms (including the GA, PSO, DE, and CMA-ES) to include additional features such as multi-objective optimization, additional or different operators (such as mutation, selection, etc.), mixed-parameter optimization (i.e., real-valued and/or binary parameters), and so on. In general, most algorithms are very flexible and customizable and can be modified as needed to better suit certain types of design problems.

The Genetic Algorithm

The genetic algorithm (GA), primarily due to its binary nature, has been extensively used in electromagnetics for a variety of reasons including ease of implementation, easy modification for additional or different features, robust global optimization, and suitability to many different types of design problems, including most notably to antennas. In this chapter, the basic single-objective GA will be discussed and utilized in the proposed test function and antenna design problems.

DOI 10.1007/978-981-4560-75-7_15-1

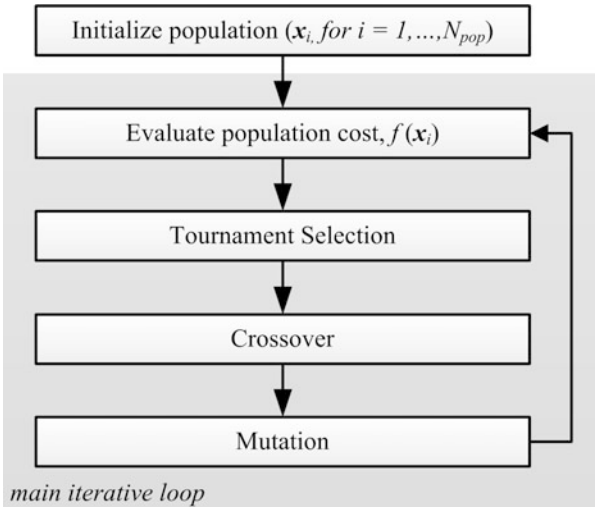


Fig. 1 Operation of the genetic algorithm

	On/Off Decisions (15 bits)															Parameter 1					Parameter 2				
Member 1	0	1	1	1	0	0	1	0	1	1	0	0	1	1	0	0	0	1	1	1	1	0	1	0	0
Member 2	1	1	1	0	1	0	1	0	0	1	0	1	0	1	1	1	1	0	0	1	0	1	1	0	0

Fig. 2 Binary chromosome examples (randomly configured) which combine decisions with discretized real values

GA Operation

The GA is an iterative, population-based algorithm with basic operation outlined in Fig. 1 that aims to refine a set of parameters such that the lowest possible cost function value is obtained (Holland 1973). It is initialized by randomly configuring the chromosomes of a user-determined number of population members. Prior to this, the chromosome must usually be configured according to the optimization task at hand. On occasion, all of the design parameters are natively binary and hence the chromosome configuration is straightforward. The GA can be configured to include discretized, real-valued parameters by defining them by a simple binary representation or Gray code. It is up to the user to determine the number of bits to include to define each parameter, with fewer bits making the optimization task easier at the cost of reduced parameter resolution. Typically, the design problem will lend some insight into the desired resolution (i.e., antenna fabrication tolerance, etc.). To describe the operation of the GA, an example problem will be considered here that has 15 decisions (binary 0 s and 1 s) and two 5-bit parameters for a total chromosome length of 20 bits as shown in Fig. 2. In this case, the 5-bit parameters have 2^5 possible values, and the binary sequences must be mapped to the available parameter ranges before being passed to the cost function. Usually, most freely available GA tools will automatically make these mappings.

After the chromosomes are initialized, the algorithm continues by evaluating each population member based on the user-defined cost function. Once the cost values for each population member are computed, the process that generates new population members can begin. The first step is selection, where parent population members are chosen for reproduction based on their cost. There are several different types of GA selection methods such as roulette wheel, rank-based, and tournament, with the most commonly used being the tournament variety. In this case, small tournaments (usually two or three members) are held between population members that are drawn at random; their cost values are compared and the lowest cost member is passed on for crossover. This is done until a pool of better-performing members are obtained

DOI 10.1007/978-981-4560-75-7_15-1

Crossover Point	1																2							
Parent 1	0	1	1	1	0	0	1	0	1	1	0	0	1	1	0	0	0	1	1	1	1	0	1	0	0
Parent 2	1	1	1	0	1	0	1	0	0	1	0	1	0	1	1	1	1	0	0	1	0	1	1	0	0
Offspring 1	0	1	1	1	0	0	1	0	1	1	0	0	1	1	0	0	0	0	0	1	0	1	1	0	0
Offspring 2	1	1	1	0	1	0	1	0	0	1	0	1	0	1	1	1	1	1	1	1	1	0	1	0	0

Fig. 3 Example of the single-point crossover being applied to the sample population members. The crossover is performed here between bits 17 and 18 of the chromosome

Crossover Mask	1	1	2	2	2	2	2	1	1	1	1	1	1	2	2	1	2	2	1	1	2	1	2	1	1
Parent 1	0	1	1	1	0	0	1	0	1	1	0	0	1	1	0	0	0	1	1	1	1	0	1	0	0
Parent 2	1	1	1	0	1	0	1	0	0	1	0	1	0	1	1	1	1	0	0	1	0	1	1	0	0
Offspring 1	0	1	1	0	1	0	1	0	1	1	0	0	1	1	1	0	1	0	1	1	0	1	1	0	0
Offspring 2	1	1	1	1	0	0	1	0	0	1	0	1	0	1	0	1	0	1	0	1	1	1	1	0	0

Fig. 4 Example of the uniform crossover being applied to the sample population members

that is of sufficient size for the number of children desired from each crossover pair. In the example GA to be considered, one child is obtained per set of parents, therefore a crossover pool of $PopSize \times 2$ is required.

In crossover, traits from two (and sometimes more) parents are combined to (hopefully) form an offspring with lower cost. In the binary GA, this is done by blending the chromosomes, which can be accomplished in different fashions such as a single-point or uniform crossover as shown in Figs. 3 and 4, respectively. The GA can be configured to yield one or two children per set of parents; in the case of one child, the second offspring is simply not generated. The uniform crossover is employed here as it is generally recommended for best optimization performance. A user-determined crossover probability (p_{cross}) determines the likelihood of crossover for the single-point type, or the ratio of one parent's genes to another for the uniform type. For the single-point type, if the crossover is to occur (i.e., $U_{\text{rand}}[0,1] < p_{\text{cross}}$), then the crossover location is determined at random. For uniform crossover, $p_{\text{cross}} = 0.5$ is generally recommended to evenly combine the genes of both parents. A $p_{\text{cross}} \neq 0.5$ causes the offspring to favor the characteristics of one parent or another. Each mask bit is determined by generating a random number in [0,1] and choosing the first parent if it is less than p_{cross} and the second if greater than p_{cross}.

Once enough offspring are generated to refill the population, they are then mutated to create new genetic traits and explore new portions of the search space. There are several ways to implement mutation, the two most common being creep mutation and jump mutation. Jump mutation is where single bits are inverted based on a mutation probability p_{mutate}. Creep mutation is where parameter values are incremented or decremented by a small amount (usually ± 1 LSB). For example, a positive creep mutation of parameter 1 in member 1 would change from 00111 to 01000. Although many bits are changed here, the parameter value itself has only been changed by one least-significant-bit. On occasion, both forms of mutation may be used simultaneously. In the example GA employed in this chapter, only jump mutations are implemented.

After mutation, the iterative cycle repeats with cost function evaluation of the new population members. The algorithm can be terminated if a maximum number of iterations or cost function evaluations are exceeded, time is expired, or the cost function goal is met. For problems where there is no defined goal, indefinite minimization is commonly performed where convergence criteria are monitored (i.e., the best cost value has not been exceeded for a certain number of iterations) or the algorithm is simply terminated after a certain number of iterations.

DOI 10.1007/978-981-4560-75-7_15-1

Comments on Usage of the GA

Typically, the GA will require a large population compared to the other algorithms, especially if discretized parameters are used with a large number of bits. Tradeoffs invariably occur in terms of optimization performance versus number of bits used to define parameters. If time and computational resources are available, it is sometimes advantageous to perform multiple optimizations on the problem with various, increasing population sizes. If the best achieved cost function values are not significantly different upon convergence, then the population size is likely adequate for the problem and algorithm configuration at hand. If this is undesirable, it is usually advisable to choose a large population to ensure that an adequate solution is found. Population sizes on the order of chromosome length are generally sufficient for most problems. Jump mutation rates are typically $1/n_{\text{chrome}}$, where n_{chrome} is the number of bits in the encoded chromosomes. This leads to a high probability that single bit in the chromosome will be flipped per population member during the mutation process.

Unlike the remaining PSO, DE, and CMA-ES algorithms, the GA natively includes parameter boundaries due to the nature of its binary chromosome. Therefore, no boundary handling needs to be implemented, except where creep mutations cause an overflow or underflow of bits (e.g., positive creep with 1111 or negative creep with 0000), which can be easily prevented in the code. Generally, the GA is most suited to and has been successfully applied to many electromagnetic design problems where decision-based parameters are primarily implemented. One popular and compelling example is a thinned linear array on a periodic lattice as demonstrated in Haupt (1994), where each GA chromosome bit controls the amplitude or presence of an associated element (i.e., on or off). In Thors et al. (2005) and Gregory and Werner (2009), the GA was used to configure the pixellized metal that comprises the radiating portion of planar antenna elements. A 0 or 1 in the GA chromosome directly controls the absence or presence of a metal "pixel," while other discretized GA parameters determine the remaining properties such as dielectric constants and thicknesses of a substrate/superstrate material.

Particle Swarm Optimization

The first of the real-coded optimization strategies covered here is PSO, a very popular algorithm in the electromagnetic and antenna design community. This is due to its simple operation, making implementation straightforward, along with very capable optimization performance. Many antenna optimization problems naturally lend themselves to real-valued design parameters, making strategies such as PSO an excellent choice for these tasks.

PSO Operation

As previously mentioned, PSO is a cooperative swarm-based optimization strategy that is based on natural phenomena such as bees searching for food. Each particle moves about the search space in an effort to find function values of lower cost. It uses knowledge of its personal best-found location (set of parameter values that are the inputs to the cost function) and the global best location to determine the direction to travel in future iterations. Social and nostalgia constants, as well as particle inertia, determine the specific velocity and direction of the particles as they move about. A flowchart of the operation of PSO is shown in Fig. 5. The algorithm is initialized with the particles randomly distributed in the desired search region and velocities randomly configured such that $v_{ij}^{(g=0)} = U_{\text{rand}}\left(\text{range}_j^{\min}, \text{range}_j^{\max}\right)$, where j is the parameter index and i is the population member index.

DOI 10.1007/978-981-4560-75-7_15-1

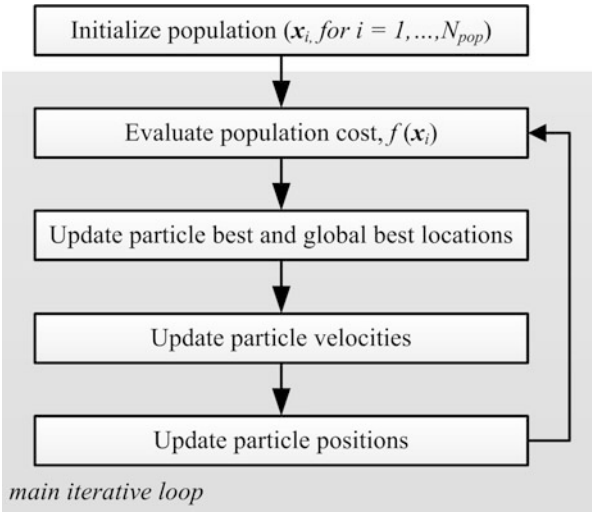


Fig. 5 Operation of particle swarm optimization

The particles are first evaluated for cost, and then the cost values are recorded; the particle with the lowest cost has its position recorded as the global best location, and the cost values of each particle are recorded as their current best location. Particle velocities are then updated according to

$$\boldsymbol{v}_i^{(g+1)} = \omega(g) \cdot \boldsymbol{v}_i^{(g)} + c_1 \cdot U_{\text{rand}}(0,1) \cdot \left(\boldsymbol{x}_i^{\text{best}} - \boldsymbol{x}_i^{(g)}\right) + c_2 \cdot U_{\text{rand}}(0,1) \cdot \left(\boldsymbol{x}_{\text{global}}^{\text{best}} - \boldsymbol{x}_i^{(g)}\right), \quad (1)$$

where g is the iteration number, $\boldsymbol{x}_i$ is the current particle position, $\omega(g)$ is the particle inertia at iteration g, c_1 is the cognitive constant, c_2 is the social constant, $\boldsymbol{x}_i^{\text{best}}$ is the best location found by particle i, $\boldsymbol{x}_{\text{global}}^{\text{best}}$ is the lowest cost location found in the optimization thus far, and $U_{\text{rand}}(0,\ 1)$ is a random number uniformly distributed between 0 and 1. At this point, the velocities are examined for limiting purposes in order to prevent particles from crossing over an excessive portion of the search space during a single iteration. This is accomplished though a conditional statement

$$\text{if } \|\boldsymbol{v}_i\| > v_{\max} \text{ then } \boldsymbol{v}_i = v_{\max} \cdot \boldsymbol{v}_i, \quad (2)$$

where $\|\boldsymbol{v}_i\|$ is the parameter-normalized velocity (i.e., lies between 0 and 1) and $v_{\max}$ is a user-determined maximum velocity constant. Once new velocities are computed, the particles' new positions can then be updated using

$$\boldsymbol{x}_i^{(g+1)} = \boldsymbol{x}_i^{(g)} + \boldsymbol{v}_i^{(g+1)}. \quad (3)$$

The particle inertia $\omega(g)$ is a term that varies throughout the optimization, initially starting at $\omega_{\max}$ for iteration 1 and then tapering linearly to $\omega_{\min}$ at iteration g^{exp} and then holding constant thereafter. This is done to encourage a more global search at the beginning of the optimization and gradually move to a more local search near the end. One potential issue with this is the need to predict how many iterations that the optimization will require to find a suitable cost value, as using different values of g^{exp} can have a significant impact on the convergence of the algorithm.

In order to constrain the particles to a desired search space, boundary conditions must be applied at each iteration in order to limit particle positions in such a way that does not significantly disturb the effectiveness of the optimization. This can become especially important when regions of low cost lie

DOI 10.1007/978-981-4560-75-7_15-1

on the boundaries of parameter spaces, a situation that can frequently occur in many design problems. There are several different types of boundary conditions in use. Incorporated here is the absorbing boundary condition, where if a particle strikes a boundary, the associated parameter is set to that boundary value, and the velocity of a particle in that direction is set to zero.

Comments on Usage of PSO

There are several strategy parameters to configure with PSO; however, there are commonly available suggested values that tend to give good performance on a wide range of problems. The suggested values are usually $c_1 = c_2 = 2.0$ and $v_{\max} = 0.5$. Typically, the inertial parameters $\omega_{\max}$ and $\omega_{\min}$ are set to 0.9 and 0.4, respectively. The expected number of iterations $g^{\exp}$ is highly problem dependent. Using too small of a value will lead to early convergence, potentially resulting in a poor cost value. Using too large of a value will lead to slow convergence and cause the algorithm to take more iterations than may normally be required. To illustrate the potential performance differences, a ten-dimensional Ackley test function is selected as a test case for PSO with four different choices of $g^{\exp}$. Twenty seeds are run for each case with a population size of 20 members to obtain a statistical performance comparison. It can be seen that there is not a significant difference between the results with $g^{\exp} = 20$ and $g^{\exp} = 100$, except that one of the seeds converged before reaching the desired function value of 10^{-10}. With a larger $g^{\exp}$, however, the optimization times grow without any significant benefit. Another issue is that for the same problem, a good choice of $g^{\exp}$ can vary with population size (Fig. 6).

It is easy to see how choosing parameters such as $g^{\exp}$ is very important to achieving an efficient and effective optimization. Unlike test functions, in most realistic optimization scenarios, the user does not know the global cost function minimum or how many iterations are needed to obtain a good solution (if one is even possible). A poor choice of $g^{\exp}$ can lead to either a bad solution or an optimization that takes much longer than needed, often without the designer being aware. Becoming familiar with the design problem and operation of the algorithm is then a key factor in obtaining fruitful, worthwhile optimizations.

Differential Evolution

DE is another popular real-valued global optimization algorithm that has found use in solving many challenging electromagnetic design problems. It does not have a strong foundation in nature such as PSO; however, it has several attractive properties from a functional standpoint. One is that unlike traditional

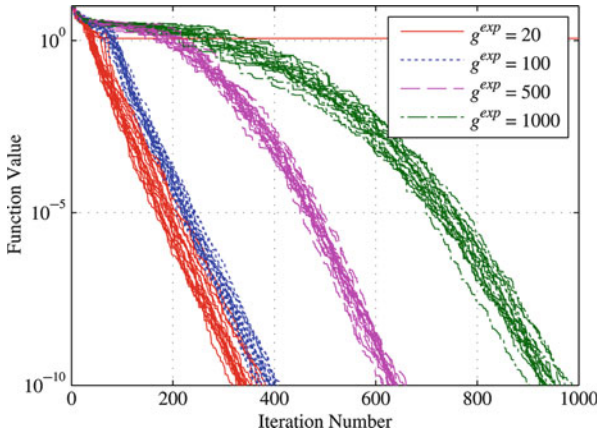


Fig. 6 Comparison of PSO on the ten-dimensional Ackley test function for different values of $g^{\exp}$

DOI 10.1007/978-981-4560-75-7_15-1

gradient-based methods, only cost function comparisons are used instead of cost derivatives. Another is that information is shared among individuals in the population in an effort to create future members with lower cost. Differential evolution can be a powerful tool when properly applied to challenging design problems.

Operation of Differential Evolution

The iterative operation of DE is shown in Fig. 7. It is initialized in the same way as many other real-valued optimization strategies such as PSO, where a population is distributed randomly inside the search domain. The population is then evaluated to obtain cost values for each member. The iterative process then begins, starting by choosing the first population member and three other distinct population members (with index p, q, r). A test vector ($\boldsymbol{y}$) is generated using these three members and two evolutionary parameters, along

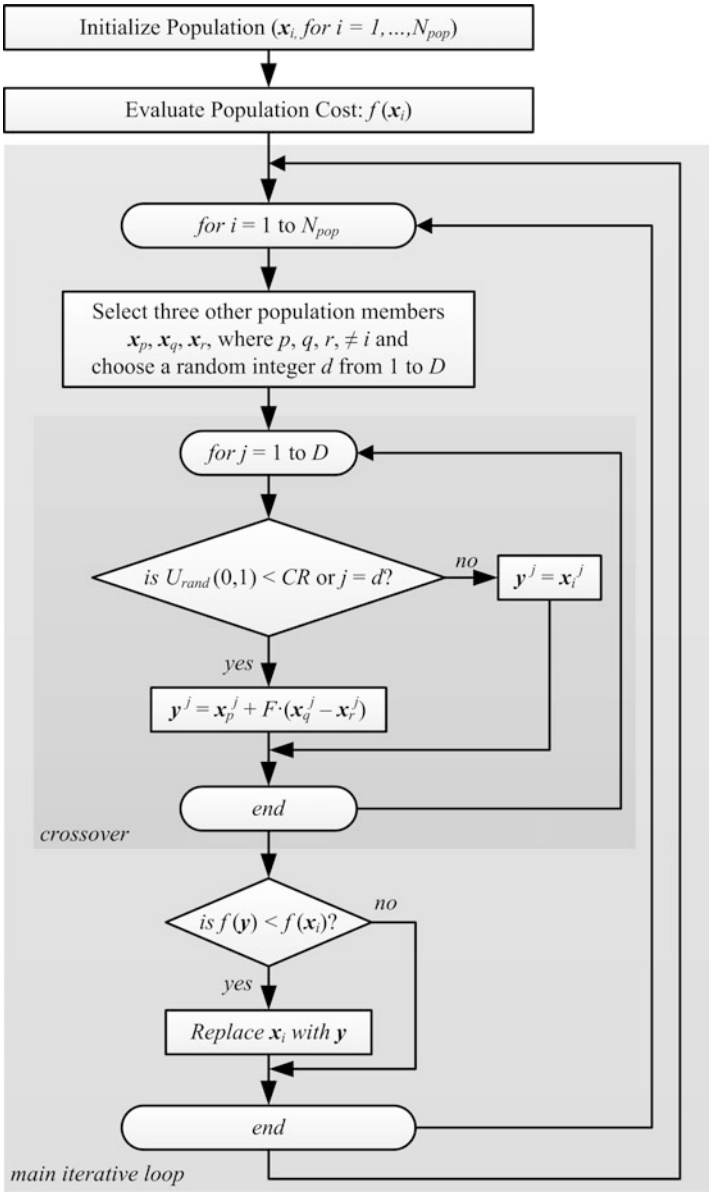


Fig. 7 Operation of DE on a D-dimensional problem. The superscript j represents the parameter index of the given population vector

DOI 10.1007/978-981-4560-75-7_15-1

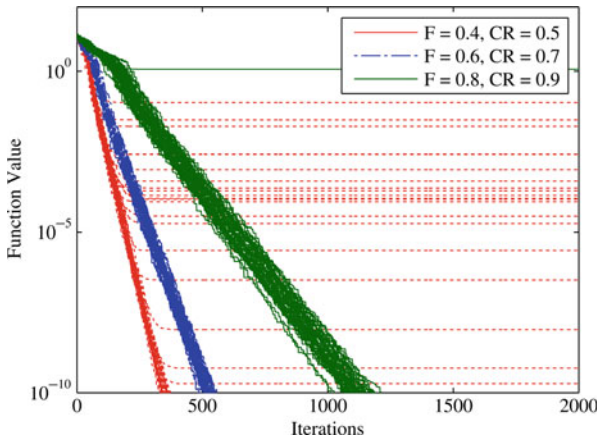


Fig. 8 Operation of DE on the ten-dimensional Ackley test function with various settings of *CR* and *F* and a population size of 20. Fifty seeds are used for each pair of algorithm settings

with several random variables as shown in Fig. 7. The test vector is a blend of the properties of the three vectors and the original vector, with proportions determined by the crossover process. Much like the GA, the crossover rate determines the proportions; at least one parameter in the test vector is used in order to provide a unique individual and avoid wasting a cost function evaluation (i.e., when $j = d$). The test vector is then evaluated for cost and if it is better than the population member initially considered, it replaces that member.

This process repeats for each population member. Ideally, the test vectors provide a continuous improvement to the population. At any time in the process, the algorithm can terminate if the cost function goal is achieved, the maximum number of function evaluations or iterations are achieved, or time is expired.

The two evolutionary parameters can play an important role on the performance of the optimization. The crossover rate (*CR*) determines the rate at which new test parameters are injected into the trial vector. A large crossover rate will lead to test vectors that are very different from the initial population member, creating a very exploratory search. The weighting constant (*F*) determines the proportion of distance between the second and third test vectors that is added to the first. An example of how these parameters can affect the performance of DE is shown in Fig. 8, where the algorithm is applied with various selections of *CR* and *F*. The limits of ($F = 0.4$, $CR = 0.5$) to ($F = 0.8$, $CR = 0.9$) are those generally recommended for common usage. Cost function variable dependence, noise, and other properties determine which set of evolutionary parameters are most effective for the problem at hand. It can be seen that with ($F = 0.4$, $CR = 0.5$), DE performs fast, but with limited success as evidenced by the large number of seeds that converge before reaching the cost function goal of 10^{-10}. With ($F = 0.6$, $CR = 0.7$), DE is both reasonably fast and reliable. With ($F = 0.8$, $CR = 0.9$), DE is reliable yet tends to be slower than needed, taking almost twice as long on average to reach the cost function goal.

Comments on Usage of DE

Like PSO, boundary conditions must be applied to prevent population members from exiting the desired or feasible search space. In this case, when any parameter in a trial vector is location out of the boundaries, it is snapped to the corresponding boundary. This method is fairly effective and requires little in terms of computational resources.

Because of its operation, it is easily observed that DE requires at least a population size of four members to function. In general, population sizes must be sufficiently large to ensure robust optimization, while not being so large that optimization time is excessive. Like the other algorithms, this behavior is highly

DOI 10.1007/978-981-4560-75-7_15-1

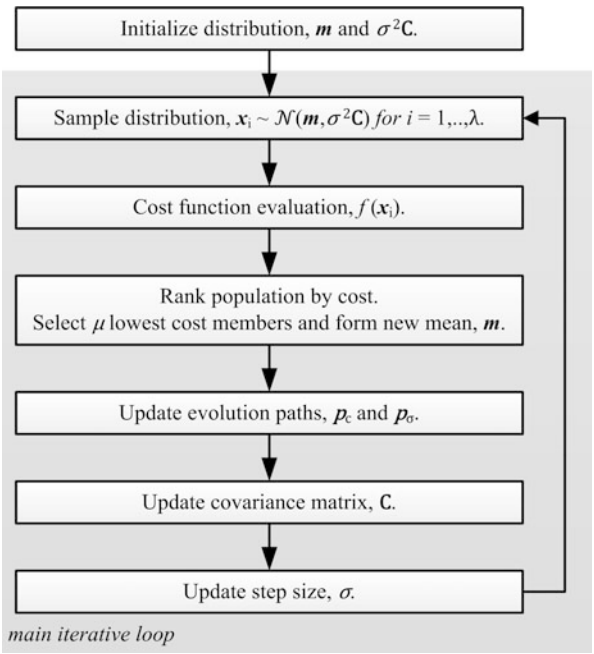


Fig. 9 Operational diagram of CMA-ES

problem dependent. It has been suggested to choose a population size of between five and ten times the number of optimizable parameters if unsure of the problem difficulty; however, it will be shown that smaller populations can be successful on easier problems (Storn and Price 1997). For difficult problems, further optimizations with larger populations can be carried out if poor results are initially achieved.

In practical use, the effectiveness of DE can be somewhat sensitive to the choice of the crossover rate and weighting constant, as is shown in Fig. 8. The parameters used for the test function in Fig. 8 will not always have the same relative outcomes for all problems (i.e., $F = 0.8$, $CR = 0.9$ may prove to be more effective than $F = 0.4$, $CR = 0.5$ in some cases). Because of this, two set of parameters for DE will be used for the analysis presented in this chapter. It is up to the user to determine the parameter values that are most appropriate for the problem at hand, although some insight is given in Storn and Price (1997).

The Covariance Matrix Adaptation Evolutionary Strategy

The covariance matrix adaptation evolutionary strategy (CMA-ES) is a relative newcomer to the field of evolutionary computation as well as electromagnetic design. Like PSO and DE, it is a real-valued population-based optimization algorithm that operates on an iterative foundation. Internally, the algorithm is much more complex than either PSO or DE; however, it is self-adaptive and only requires the user to choose a population size before beginning an optimization rather than population size plus several other evolutionary parameters.

Operation of CMA-ES

A functional diagram of CMA-ES can be found in Fig. 9. The internal operating constants, vectors, and arrays are given in Tables 1 and 2. Superscripts with parenthesis indicate the parameter value at the indicated iteration (i.e., iteration g or $g + 1$). Since CMA-ES is a distribution-based algorithm, it differs slightly in its population initialization than the other algorithms. Rather than typical uniform distribution of the population in the search domain, the initial distribution position is chosen randomly, and its size is

DOI 10.1007/978-981-4560-75-7_15-1

Table 1 Scalars, vectors, and matrices used internally by CMA-ES

N	Problem dimension
λ	Population size
μ	Number of selected children
g	Iteration number
σ	Step size
$\boldsymbol{m} \in \mathbb{R}^N$	Distribution mean
$\boldsymbol{x}_k \in \mathbb{R}^N$	Parameter set of the k^{th} population member
$\mathbf{C} \in \mathbb{R}^{N \times N}$	Covariance matrix
$\mathbf{B} \in \mathbb{R}^{N \times N}$	Eigenvectors of the covariance matrix
$\mathbf{D} \in \mathbb{R}^{N \times N}$	Eigenvalues of the covariance matrix
$\boldsymbol{p}_c \in \mathbb{R}^N$	Evolution path
$\boldsymbol{p}_\sigma \in \mathbb{R}^N$	Conjugate evolution path

Table 2 Additional scalars used internally by CMA-ES and their definitions

$w_i = \frac{\log_2(\mu+0.5)-\log_2(i)}{\sum_{j=1}^{\mu}(\log_2(\mu+0.5)-\log_2(j))}$	Selection weights
$\mu_{\text{eff}} = \left(\sum_{i=1}^{\mu} w_i^2\right)^{-1}$	Variance effective selection mass
$c_\sigma = \frac{\mu_{\text{eff}}+2}{N+\mu_{\text{eff}}+5}$	Learning rate for step size control
$c_1 = \frac{2}{(N+1.3)^2+\mu_{\text{eff}}}$	Learning rate for rank-one update of covariance matrix
$c_\mu = \min\left(1 - c_1, \frac{2\mu_{\text{eff}}-4+2/\mu_{\text{eff}}}{(N+2)^2+\mu_{\text{eff}}}\right)$	Learning rate for rank-μ update of covariance matrix
$c_c = \frac{4+\mu_{\text{eff}}/N}{N+4+2\mu_{\text{eff}}/N}$	Learning rate for cumulation for the rank-one update of covariance matrix
$d_\sigma = 1 + c_\sigma + 2 \cdot \max\left(0, \sqrt{\frac{\mu_{\text{eff}}-1}{N+1}} - 1\right)$	Damping for step size control

configured such that the standard deviation of the distribution is (typically) one third of the parameter range for each parameter. Like the other algorithms, if there is some insight on where to search initially, it can be used to alter the initial distribution to improve optimization performance.

After the initial distribution is configured, the population is created by drawing samples according to

$$\boldsymbol{z}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{4}$$

where 0 is a vector of length N and the identity matrix, $\mathbf{I}$ is of size $N \times N$. This is done since most computer systems will natively generate random values from the standard normal distribution (rather than one that is skewed and with covariances). The samples are then transformed to the desired mean location and distribution shape through

$$\boldsymbol{x}_k^{(g)} = \boldsymbol{m}^{(g)} + \sigma \boldsymbol{y}_k^{(g)}, \tag{5}$$

where

DOI 10.1007/978-981-4560-75-7_15-1

$$\boldsymbol{y}_k^{(g)} = \mathbf{B}\mathbf{D}\boldsymbol{z}_k. \tag{6}$$

Note that $\mathbf{B}$ and $\mathbf{D}$ are generated using principle component analysis (eigen-decomposition), where $\mathbf{C} = \mathbf{B}(\mathbf{D})^2\mathbf{B}^T$. The first round of cost function evaluations are now carried out using $\boldsymbol{x}_k^{(g)}$. Afterwards, the population members are sorted according to cost, indicated by $\boldsymbol{y}_{i:\lambda}^{(g)}$ and $\boldsymbol{x}_{i:\lambda}^{(g)}$, where the lowest cost members are the first entries. Then, a weighted average of the positions of the best performing μ members (typically $\lfloor \lambda/2 \rfloor$) is used to reposition the new distribution mean according to

$$\langle \boldsymbol{y} \rangle_w = \sum_{i=1}^{\mu} w_i \boldsymbol{y}_{i:\lambda}^{(g)} \tag{7}$$

and

$$\boldsymbol{m}^{(g+1)} = \boldsymbol{m}^{(b)} + \sigma^{(g)} \boldsymbol{y}_w. \tag{8}$$

Next, the evolution paths are updated according to

$$\boldsymbol{p}_\sigma^{(g+1)} = (1 - c_\sigma)\, \boldsymbol{p}_\sigma^{(g)} + \sqrt{c_\sigma(2 - c_\sigma)\mu_{\text{eff}}} \left(\mathbf{C}^{(g)}\right)^{-1/2} \frac{\boldsymbol{m}^{(g+1)} - \boldsymbol{m}^{(g)}}{\sigma^{(g)}} \tag{9}$$

and

$$\boldsymbol{p}_c^{(g+1)} = (1 - c_c)\, \boldsymbol{p}_c^{(g)} + \sqrt{c_c(2 - c_c)\mu_{\text{eff}}}\, \frac{\boldsymbol{m}^{(g+1)} - \boldsymbol{m}^{(g)}}{\sigma^{(g)}}, \tag{10}$$

where $\mathbf{C}^{-1/2} = \mathbf{B}\mathbf{D}^{-1}\mathbf{B}^T$ and, since $\mathbf{D}$ is diagonal only, results in a simple matrix inversion. These evolution paths serve to keep track of the movement of the distribution and are used to give the algorithm its self-adaptive properties.

The covariance matrix is then updated, where three separate contributions are employed which are termed the rank-μ update (with portion c_μ), rank-one update (with portion c_1), and the cumulative update (with portion $1 - c_1 - c_\mu$). The cumulative update is simply the amount of the historical portion of the previous covariance matrix to use. The rank-one update is formed using the evolution path, which serves to elongate the covariance matrix in the direction of the movement of the mean, where lower cost regions of the search space are likely to be found. The rank-μ update is formed from a scaled estimation of distribution between the surviving population members and the previous distribution mean. The complete covariance update is given by

$$\mathbf{C}^{(g+1)} = \left(1 - c_1 - c_\mu\right)\mathbf{C}^{(g)} + c_1 \left(\boldsymbol{p}_c^{(g+1)}\right)\left(\boldsymbol{p}_c^{(g+1)}\right)^T + c_\mu \sum_{i=1}^{\mu} w_i \left(\boldsymbol{y}_{i:\lambda}^{(g+1)}\right)\left(\boldsymbol{y}_{i:\lambda}^{(g+1)}\right)^T, \tag{11}$$

where

$$\boldsymbol{y}_{i:\lambda}^{(g+1)} = \frac{\left(\boldsymbol{x}_{i:\lambda}^{(g+1)} - \boldsymbol{m}^{(g)}\right)}{\sigma^{(g)}}. \tag{12}$$

DOI 10.1007/978-981-4560-75-7_15-1

With small populations, cumulation allows the algorithm to operate efficiently by smoothing out the reshaping of the covariance matrix over several iterations, avoiding sporadic movement and reshaping that could be caused by small sample sizes. The final part of the iterative process is the step size update, where the conjugate evolution path is compared with the expected value of a normal distribution, and the size of the distribution is appropriately scaled via

$$\sigma^{(g+1)} = \sigma^{(g)} e^{\frac{c_\sigma}{d_\sigma}\left(\frac{\|\boldsymbol{p}_\sigma^{(g+1)}\|}{\mathrm{E}\|\mathcal{N}(\mathbf{0},\mathbf{I})\|}-1\right)}. \tag{13}$$

This allows the algorithm to increase speed (enlarge the distribution) or slow down (shrink the distribution) as needed throughout the course of the optimization. Large steps seen in the conjugate evolution path lead to an increase in the step size, ideally reducing the number of iterations needed to move to the cost function goal. Small steps lead to a reduction in the step size, allowing for a more focused local search and convergence on an optimum.

To illustrate the operation of CMA-ES, a plot of the movement and reshaping of the distribution, along with population sampling and selection is given in Fig. 10. Here, a relatively large population is used on a two-dimensional rotated hyperellipse test function to more clearly illustrate the behavior of CMA-ES across several iterations. In Fig. 10a, the selection process is illustrated at the first iteration. The μ members that have the lowest cost (those closest to the focal point of the test function contour lines) are filled in and used to generate the new mean location through a weighted average. The empty circles are members that were in the sample but discarded from the selection pool. In Fig. 10b, the movement of the mean is shown by the arrows connecting the centers of the distributions. The use of covariance matrix (rather than a vector of variances) allows the distribution to elongate in different dimensions diagonally, a particularly powerful feature for problems with inseparable relationships between variables in the cost function. That is, problems which cannot be adequately minimized by optimizing each variable independently. The long dashed lines represent the single standard deviation likelihood of selection of the distribution, while the short dashed lines represent the covariance matrix principle axes.

Comments on Usage of CMA-ES

Given that CMA-ES is much more internally complex than any of the other aforementioned algorithms, it must possess some other benefits that make it an attractive choice for engineers and designers. Indeed, the self-adaptive properties make the algorithm easy to use and an efficient choice for most kinds of design problems as no evolutionary parameters other than population size need to be chosen. In general, difficult problems require larger population sizes to obtain satisfactory optimization results; the usual drawback to using larger populations is an increased number of cost function evaluations. In cases where a population size that is much more than sufficient is used on a problem, the optimization can take significantly longer than needed; however, this potential drawback must be balanced with the possibility of using too small a population size and not achieving a good end result. In many cases, the amount of cost function evaluations required to achieve a certain cost function goal does not grow significantly for larger-than-needed population sizes with CMA-ES, especially for difficult problems (Hansen and Ostermeier 2001). That is, there is not a significant time penalty for choosing a population size that is large, as will be shown for several of the example problems in the following sections.

In general, the only evolutionary parameter that must be chosen for CMA-ES is the population size. Hansen recommends a population size of

$$\lambda = 4 + \lfloor 3 \cdot \ln(N) \rfloor. \tag{14}$$

DOI 10.1007/978-981-4560-75-7_15-1

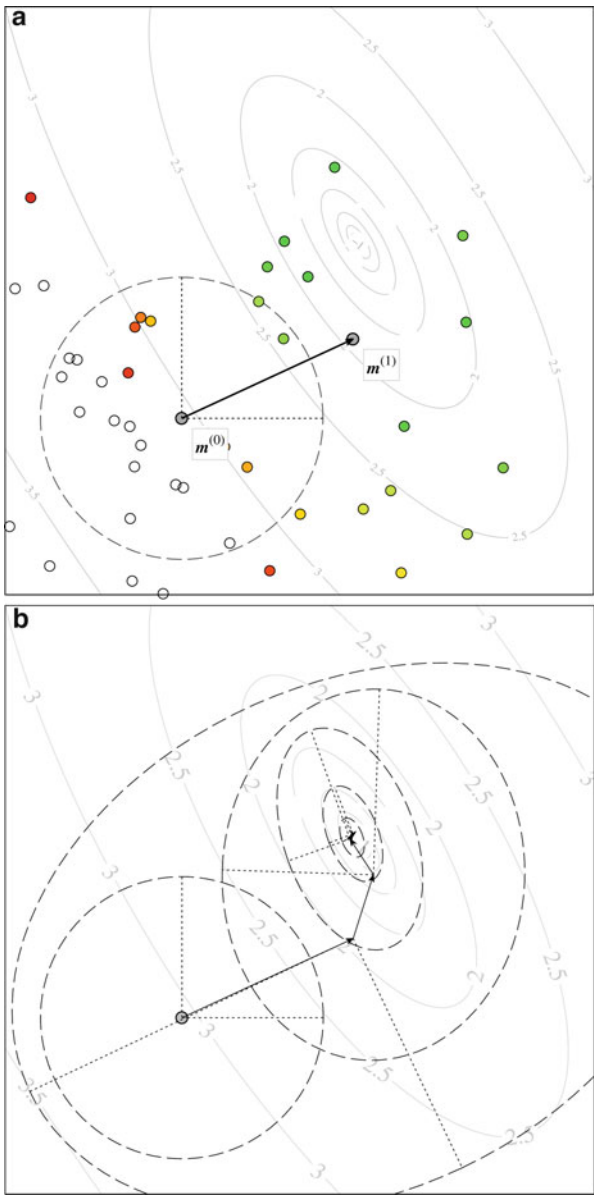


Fig. 10 Illustration of the population sampling and selection process with generation of the new mean (**a**) and movement and reshaping of the search distribution of CMA-ES across several iterations (**b**) using a population size of 50 members on a two-dimensional rotated hyperellipse test function

More difficult problems such as those with noisy, discontinuous, or multimodal cost landscapes may require larger population sizes. As with the other optimization algorithms, if unsatisfactory results are obtained with small population sizes, optimizations can be restarted with larger populations (Auger and Hansen 2005).

Due to the collapsing distribution size as CMA-ES begins to converge on a low-cost region, it is fairly easy to determine when CMA-ES has converged and is no longer producing improved cost values. If the average cost value of the population is very close to the lowest cost value, this is generally a good indicator that the algorithm is sufficiently converged. In some cases, this can be used to automate algorithm termination and minimize wasted CPU time.

Boundary conditions must be handled differently than the other algorithms covered in this chapter. Since the algorithm operates with assumptions made about the sampling distribution, samples that are out

DOI 10.1007/978-981-4560-75-7_15-1

of boundaries cannot simply be snapped to the appropriate bounds with no consequences. In the case of Hansen's CMA-ES publicly available code (Hansen 2014), a method is used that applies an internal fitness penalty to members that leave the feasible range, with progressively increasing penalties if the mean continues to move away from the associated boundary. Sampled values are always truncated to the search domain before being passed to the cost function, however, in order to prevent non-feasible solutions.

Algorithm Performance Comparisons

In this section, the four algorithms that have been covered thus far will be applied to several test functions and antenna design task. Test functions are a fast, commonly used method for determining the relative performance of algorithms on problems of various dimensions. They are particularly convenient since the function minimum values and their associated parameter sets are known a priori for any number of dimensions. Single-objective test functions are often classified into unimodal or multimodal types, with unimodal having only one global and local optimum and multimodal functions having more than one local optima but still typically only one global minimum. They will be used here to gain further understanding of the algorithms and how well they perform on different types of problems.

Test Function Comparisons

Four different test functions will be utilized here as a basis for comparison of the algorithms. One is a unimodal and real valued; two are multimodal, real valued; and one is a binary decision-based test function. The functions, their minimum values, and the set of parameters that yield these values are given in Table 3. For the GA to use the real-valued test functions, the binary chromosomes must be configured to translate to real-valued numbers that span the range of the allowed input parameters ($\boldsymbol{x}$). Because of this, the allowable ranges $x_i \in [x_{\min}, x_{\max}]$ are shifted from their traditional ranges such that the discrete values translated from the GA result in sets that contain $\boldsymbol{x}^{\text{best}}$. For example, a 4-bit chromosome can realize 16 values. For the Ackley test function, this results in $b_i \in \{-28, -24, \ldots, -4, 0, 4, \ldots, 28, 32\}$, a set which contains a zero, allowing the algorithm to reach the desired function value of 10^{-10}. The same is done for the other two real-valued test functions and these same ranges will work for the 8 bits per parameter trials as well. In actual realistic optimization scenarios, the global best ($\boldsymbol{x}^{\text{best}}$) is not known a priori, and the user must choose bit lengths sufficiently large for the problem, but not so large such that it hinders GA performance. The number of bits that define each parameter can be different and are often determined based on knowledge of the problem.

In the case of the real-valued test functions, ten dimensions are used ($N = 10$). For the binary test function, 40 dimensions are used to allow for a beneficial comparison between the GA 4 bits per parameter case. Here, the GA will still use a 40-bit chromosome, but the real-valued algorithms will require 40 parameters. Since the real values are rounded in the function to binary values, this creates a discontinuous cost landscape which can be challenging for the algorithms but nevertheless appears in some realistic optimization problems. For particle swarm optimization, the expected number of iterations must be chosen before each optimization; values are given in Table 4. Values are adjusted and usually decreased as population size increases. If the algorithm is generally unsuccessful on a particular function, then the values are not decreased as drastically or at all. For all of the algorithms, population sizes start at ten members and are increased on a 1-2-5 basis until either a reasonable success rate is achieved, a maximum population size of 500 is reached, or the optimization times become excessive (which can sometimes occur even with smaller population sizes). Trials are typically run until convergence. For differential evolution, two sets of evolutionary parameters are used, indicated by DE1 ($F = 0.8$,

DOI 10.1007/978-981-4560-75-7_15-1

Table 3 Test functions used for comparison of the optimization algorithms

Function	Notes
$F_{\mathrm{ACKLEY}}(\boldsymbol{x}) = 20 + e - 20e^{-0.2\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}} - e^{\frac{1}{N}\sum_{i=1}^{N}\cos(2\pi x_i)}$	$x_i \in [-28, 32]$ for all $i = 1, \ldots, N$ $F(\boldsymbol{x}^{\mathrm{best}}) = 0$ where $\boldsymbol{x}^{\mathrm{best}} = 0$
$F_{\mathrm{LEVY}}(\boldsymbol{x}) = \sin^2(\pi w_1) + (w_N - 1)^2\left[1 + \sin^2(\pi w_N + 1)\right] + \sum_{i-1}^{N-1}(w_i - 1)^2\left[1 + 10\sin^2(\pi w_i + 1)\right]$, where $w_i = 1 + \frac{x_i - 1}{4}$	$x_i \in [-7, 13]$ for all $i = 1, \ldots, N$ $F(\boldsymbol{x}^{\mathrm{best}}) = 0$ where $\boldsymbol{x}^{\mathrm{best}} = 1$
$F_{\mathrm{ROSENBROCK}}(\boldsymbol{x}) = \sum_{i=1}^{N-1}\left[(1 - x_i)^2 + 100(x_{i+1} - x_i^2)^2\right]$	$x_i \in [-7, 13]$ for all $i = 1, \ldots, N$ $F(\boldsymbol{x}^{\mathrm{best}}) = 0$ where $\boldsymbol{x}^{\mathrm{best}} = 1$
$F_{\mathrm{PATTERN}}(\boldsymbol{b}) = \sum_{i=1}^{N}\left\{\begin{array}{ll} b_i & \text{for odd } i \\ 1 - b_i & \text{for even } i \end{array}\right\}$	$b_i \in \{0, 1\}$ for all $i = 1, \ldots, N$ $F(\boldsymbol{b}^{\mathrm{best}}) = 0$ where $\boldsymbol{b}^{\mathrm{best}} = [0\ 1\ 0\ 1\ 0\ 1 \ldots]$ Real-coded input values are rounded to the nearest integer: $b_i = [x_i]$, where $x_i \in [0, 1]$

DOI 10.1007/978-981-4560-75-7_15-1

Table 4 Expected number of iterations (g^{exp}) used with PSO for the test function analysis

Pop. size	Ackley	Levy	Rosenbrock	Pattern
10	1,000	500	10,000	2,000
20	750	300	10,000	2,000
50	500	200	10,000	1,500
100	–	–	10,000	1,250
200	–	–	–	1,000
500	–	–	–	750

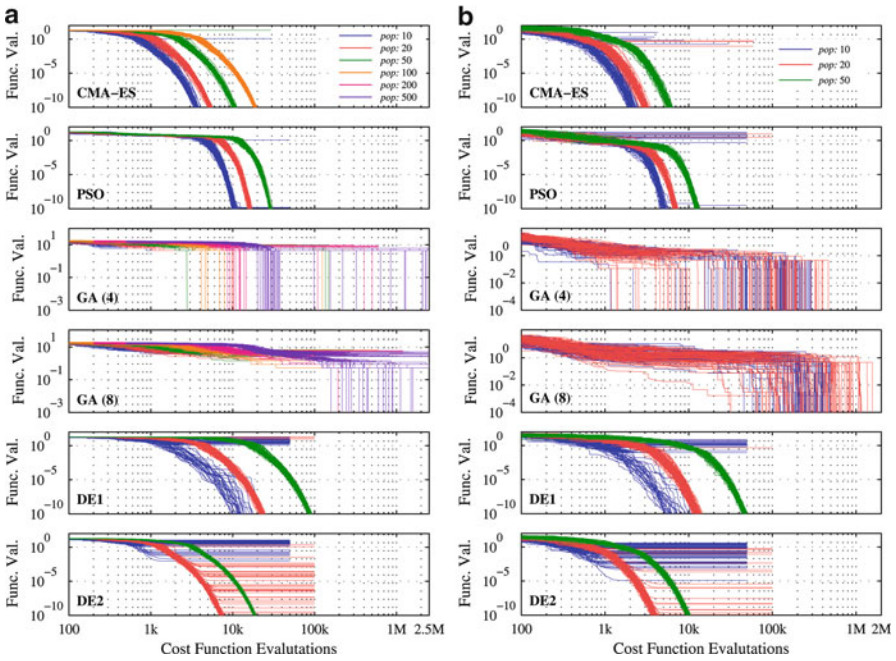


Fig. 11 Optimization results for the Ackley (**a**) and Levy (**b**) test functions. Each line represents the best function value obtained for the optimization seed at the given number of function evaluations

$CR = 0.9$) and DE2 ($F = 0.4$, $CR = 0.5$). The maximum number of iterations for all algorithm and test function combinations is chosen such that the optimizations are well converged before termination, or an excessive number of function evaluations is reached. Lastly, 100 trials are run for each population size, algorithm, and function case in order to obtain some statistical significance to the data.

The results of the optimizations for each of the test functions are given in Figs. 11 through 14. Data is arranged such that the number of function evaluations is maintained on the bottom axes rather than iteration counts to give a better comparison between algorithms and population sizes. Note that the vertical axes for the GA cases are different since the very small function values (i.e., between 0 and 10^{-6}) do not exist due to the parameter discretization (i.e., the closest function value to the minimum is 10^{-3} for the Levy function, realized by the GA chromosome being one bit off). In this case, the ranges are adjusted to show more details of the progress of the optimizations. A summary of the success and average number of function evaluations (NFE) of the algorithms for each of the test functions is given in Table 5. For the Pattern test function, 40 bits were used such that the GA (4) case resulted in the same length chromosome as for the real-valued test functions. Although not comparable to the other results, 80 bits were used for the GA (8) case.

DOI 10.1007/978-981-4560-75-7_15-1

Table 5 Summary of results for the four algorithms applied to test functions. Success rates and mean number of function evaluations required are listed for the seeds that were successful. Asterisks for the Pattern test function indicate that the number of parameters (bits) was 80 instead of 40 for the other cases

Function	Pop. size	CMA-ES	PSO	GA (4)	GA (8)	DE1	DE2
Ackley	10	86 % | 3625	94 % | 10549	0 %	0 %	28 % | 13141	0 %
	20	99 % | 5316	100 % | 16049	1 % | 9840	0 %	97 % | 23083	59 % | 7324
	50	99 % | 10457	100 % | 29078	3 % | 95617	0 %	100 % | 86753	100 % | 18551
	100	100 % | 18784	–	7 % | 23343	0 %	–	–
	200	–	–	12 % | 93733	1 % | 193800	–	–
	500	–	–	62 % | 325403	18 % | 645778	–	–
Levy	10	67 % | 2112	86 % | 5104	99 % | 111101	74 % | 246617	27 % | 6761	0 %
	20	95 % | 3036	98 % | 6910	100 % | 112174	100 % | 408626	99 % | 12638	79 % | 3903
	50	100 % | 5923	100 % | 12511	–	–	100 % | 47147	100 % | 9835
Rosenbrock	10	93 % | 7134	0 %	69 % | 11473	0 %	58 % | 808854	0 %
	20	97 % | 9028	26 % | 10675058	69 % | 16390	0 %	96 % | 28477	0 %
	50	100 % | 14759	17 % | 5043853	85 % | 30524	1 % | 214500	100 % | 93258	1 % | 115250
	100	–	15 % | 5579840	95 % | 54395	0 %	–	0 %
	200	–	21 % | 6794857	100 % | 111484	5 % | 3921480	–	100 % | 19129362
Pattern	10	100 % | 1514	1 % | 390	100 % | 450	100 % | 1025*	5 % | 718	30 % | 466
	20	100 % | 777	4 % | 5135	100 % | 496	100 % | 1156*	55 % | 1901	97 % | 648
	50	100 % | 1072	17 % | 41568	100 % | 933	100 % | 1631*	100 % | 5201	100 % | 1248
	100	–	31 % | 42290	–	–	–	–
	200	–	53 % | 21128	–	–	–	–
	500	–	80 % | 147406	–	–	–	–

For the Ackley test function, CMA-ES, PSO, and DE1 performed fairly well with small population sizes. With slightly larger population sizes, DE2 was also relatively reliable. The GA had difficulty with this test function for both parameter bit lengths, achieving only reasonable success with the 4-bit parameters and a population size of 50, albeit with large NFE requirements. For the Levy function, all of the algorithms were able to competently reach the function goal, especially with population sizes of 20 and 50. The GA did require on average more than an order of magnitude more function evaluations than the real-coded algorithms, however. The Rosenbrock test function presented a particularly difficult problem for the optimization strategies. Optimizations were truncated at 3×10^7 NFE due to the extreme time requirements. For PSO, some additional seeds may have reached the function goal value; however, the NFE requirement for these seeds would not be very practical when compared to the performance of the other algorithms. CMA-ES fared the best here, with reliable optimization achieved at a population size of 50 and with a relatively small NFE requirement. DE1 and GA (4) performed reasonably well with population sizes above 20; however, the 4-bit discretization of the input parameters for the GA may have significantly reduced the difficulty of the test function by eliminating some of the deceptive features of its landscape that the other real-coded algorithms must contend with. For the 8-bit GA case, these features reappear, and the difficulty of the problem is restored, leading to a much more significant challenge. Population sizes larger than 200 were not used due to the already very large NFE requirements of the smaller population sizes. The Pattern test function presented an interesting problem for the real-valued optimization strategies. As expected for the GA cases, the function was relatively easy and required very few function evaluations to achieve success. Reliability was also excellent, even for the 80-bit pattern

DOI 10.1007/978-981-4560-75-7_15-1

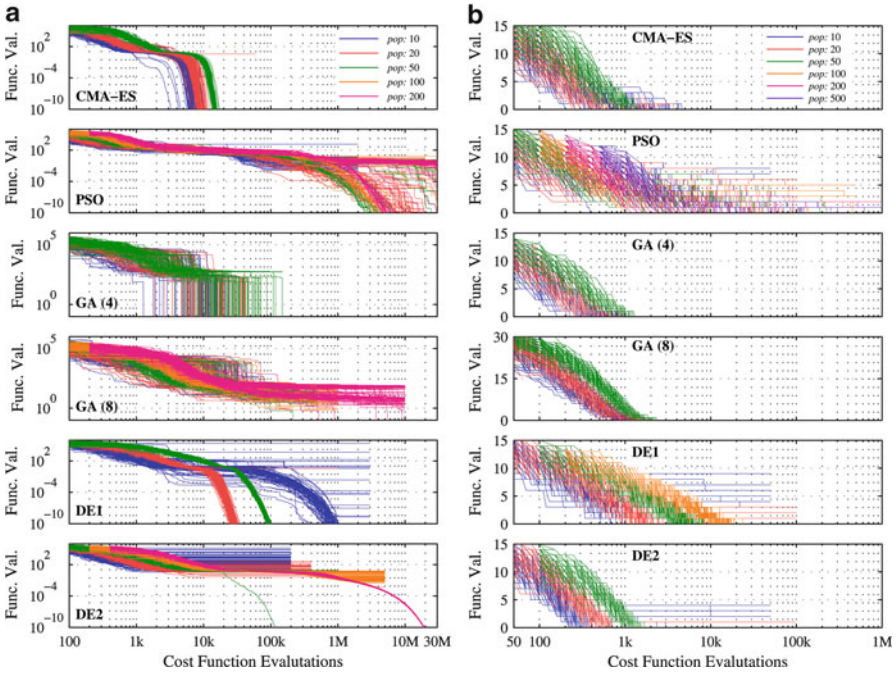


Fig. 12 Optimization results for the Rosenbrock (**a**) and Pattern (**b**) test functions. Each line represents the best function value obtained for the optimization seed at the given number of function evaluations. For the GA (8) case, 80 bits in the pattern function are used instead of 40 bits as with the GA (4) and real-coded algorithms

length used in the GA (8) case. Of the real-valued strategies, CMA-ES and differential evolution were effective at finding the function minimums. CMA-ES and DE2 are surprisingly competitive to the GA for population sizes of 20 and 50, indicating that they fare well for problems where rounding of the input parameters is performed or decisions are required (Fig. 12).

The four test functions that have been used with the algorithms here illustrate some of their strengths and weaknesses that the user must keep in mind when an optimization is to be performed. In addition, it is shown how the algorithms behave with differing population sizes. For example, it is common for many of the algorithms to optimize at a fast pace for small population sizes; however, it can be seen that the drawback is potential unreliability. For realistic design problems, a best guess is usually made as to the difficulty of the problem, and the user must decide on a population size that will likely give good results with a reasonable optimization time.

Antenna Design Task

In this section, the algorithms will be applied to a simple antenna design task to compare their performance on a realistic problem that may be encountered. The antenna is a simple planar slotted patch with eight optimizable parameters. The design is based on the antenna considered in Maci et al. (1995), where a standard patch antenna has two symmetrical rectangular cutouts. A drawing and details for the patch design are given in Fig. 13 with parameter ranges provided in Table 6. To avoid unrealizable designs (e.g., the probe feed in the middle of a slot), some parameters (lengths) are based on fractions of other parameters. The patch antenna is optimized for S_{11} and gain at two target frequencies of 900 MHz and 1.6 GHz, well within the ratio prescribed by Maci et al. (1995) for this type of antenna design. For this design task, population sizes of 15 members will be used for all of the algorithms. The optimizations are truncated at 400 iterations (6,000 cost function evaluations), with 100 iterations expected for the PSO algorithm ($g^{\exp}$). The cost function corresponding to the design is given by

DOI 10.1007/978-981-4560-75-7_15-1

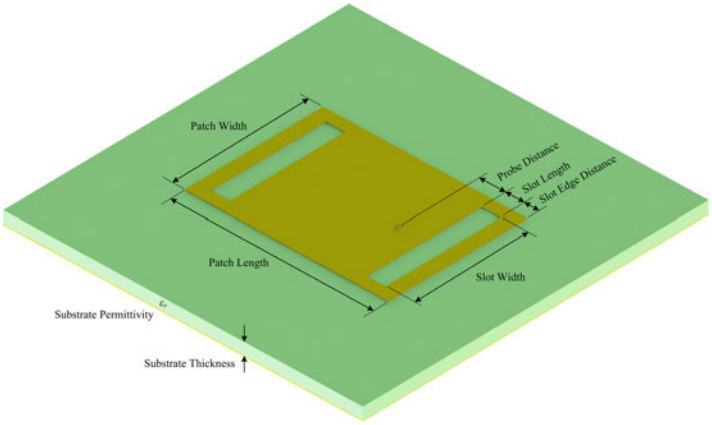


Fig. 13 Slotted patch antenna design with the dimensions and optimizable parameters indicated

Table 6 Optimizable parameters and ranges for the slotted patch antenna design problem

Parameter	Range
Patch length	5–14 cm
Patch width	(0.6–1.1) × patch length
Slot width	(0.6–0.95) × patch width
Slot length	(0.04–0.1) × patch length
Slot edge distance	(0.02–0.1) × patch length
Probe distance	(0.05–0.3) × patch length
Substrate thickness	0.81–3.18 mm
Substrate permittivity	(2.0–4.5) × ε_0

$$F_{\mathrm{COST}}(\boldsymbol{x}) = \sum_{n=1,2} \max\left(0, S_{11}(\boldsymbol{x}, f_n) - S_{11}^{\mathrm{goal}}\right)^2 + \max\left(0, G^{\mathrm{goal}} - G(\boldsymbol{x}, f_n)\right)^2, \tag{15}$$

where scattering coefficients and gain values are expressed in dB and given by $S_{11}^{\mathrm{goal}} = -10\mathrm{dB}$ and $G^{\mathrm{goal}} = 5\mathrm{dBi}$. Squared terms are used here to balance performance across the multiple objectives and frequency points.

The method-of-moments simulation tool FEKO is employed to efficiently evaluate the candidate antenna designs to determine their scattering coefficients and gain. Approximately 20 s are required per cost function evaluation (on eight cores of an Intel Xeon 2.6 GHz processor); however, this can vary based on the geometry of the design (i.e., larger patches require more triangles and longer simulation times). Since this is many orders of magnitude longer than what is required for test functions, it is of considerable interest to be able to find suitable antenna designs in the fewest number of cost function evaluations. Ten seeds are used for each of the four algorithms as a basis for comparison.

The results of the antenna design analysis are shown in Fig. 14 and Table 7. An example of the performance that can be expected from an optimized seed (taken from the pool of CMA-ES results) is shown in Fig. 15. The optimized design clearly meets the specified S_{11} and gain requirements, and the radiation patterns at the two frequencies of interest possess no irregularities, indicating that the optimization is producing designs as intended. Since the cost function contains square terms, the values are

DOI 10.1007/978-981-4560-75-7_15-1

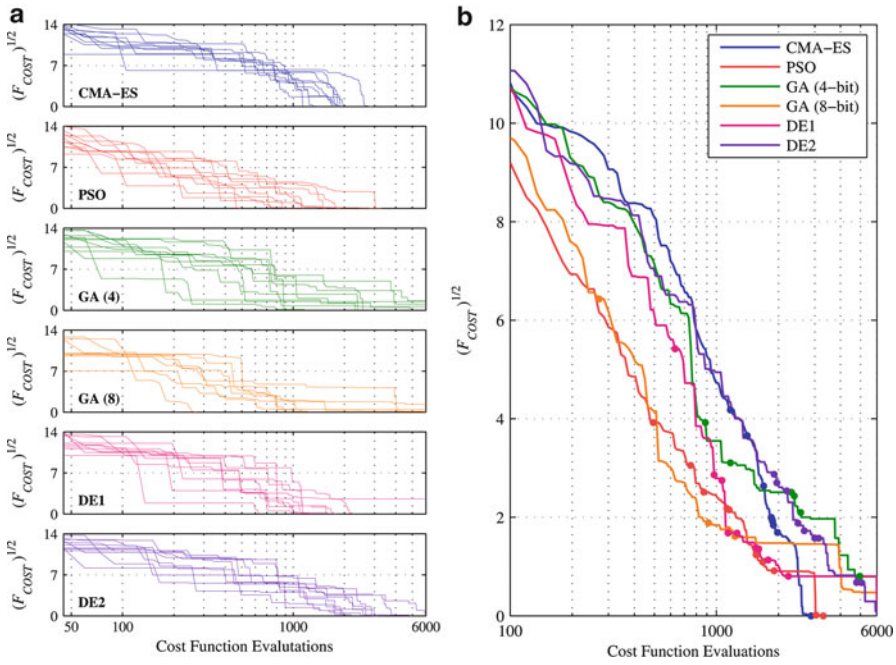


Fig. 14 Results of the slotted patch antenna design comparison. In (**a**), each optimization seed is given by an individual line that signifies the best result of that optimization at the given number of cost function evaluations. In (**b**), each line is the average of the 10 seeds' best cost value achieved at the given iteration. Each circle represents the occurrence of a seed reaching the cost goal ($F_{\mathrm{COST}} = 0$) for the associated algorithm

Table 7 Summary of results for the slotted patch antenna design problem

Algorithm	Successful seeds	Mean NFE of successful seeds	Mean cost value of unsuccessful seed(s)
CMA-ES	10	1,830	–
PSO	10	1,589	–
GA (4)	6	2,380	1.595
GA (8)	5	765	0.447
DE1	9	1,362	6.475
DE2	9	3,022	0.086

plotted as the square root of cost to make them more clearly visible. Circles are placed along the mean cost lines where seeds achieved success. As may be expected since this is a real-valued optimization problem, the real-valued algorithms fared the best. CMA-ES and PSO achieved a 100 % success rate with low NFE requirements. Differential evolution was also fairly competitive at 90 % success rates for both parameter settings. The first parameter setting, however, yielded on average much faster optimization times. The genetic algorithm was somewhat effective on this problem for both cases. Given that six seeds reached the cost function goal, it is clearly evident that the parameters had sufficient resolution with 4 bits per parameter; however, it is interesting to note that the case with 8 bits per parameter fared better. It is possible that the higher-resolution case offered more possible solutions that satisfied the cost requirement and performance goals, allowing the algorithm to more effectively reach the cost goal. Given that the mean cost value of the unsuccessful seeds is much lower than the 4-bit scenario, this may indeed be the

DOI 10.1007/978-981-4560-75-7_15-1

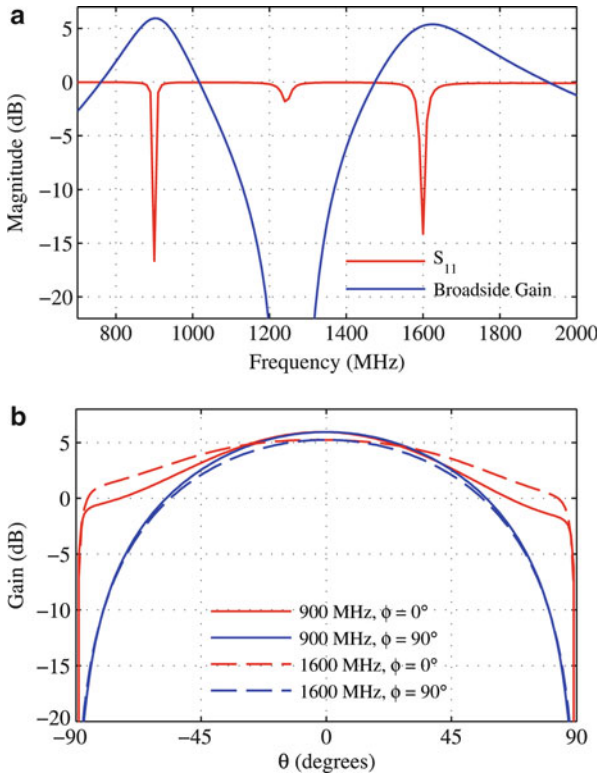


Fig. 15 Scattering coefficient and broadside gain for one of the seeds optimized with CMA-ES

case. In all situations, however, useful designs will likely have been produced for the antenna designer, even if the ideal cost value of 0 has not been achieved. This is indeed a great benefit of these global optimization strategies, where suitable antenna designs can be produced autonomously with minimal user interaction and overhead.

Practical Design Examples

In addition to the test functions and designs that have been shown in the previous section, some detailed examples of where these algorithms have been recently applied in the antenna design community are given here. The first example is a more complex version of the slotted patch antenna discussed in the previous section. The second is a simple monopole antenna with enhanced performance through the addition of a two-layer cylindrical anisotropic high-index metamaterial coating. The last is a low-profile substrate-integrated waveguide antenna with significantly enhanced directivity achieved through the application of an anisotropic low-index metamaterial coating.

Folded, Shorted Meander-Slot Patch Antenna

Miniaturization has been a popular field in antenna design due in large part to the desire for portability in mobile and wireless communications devices (Volakis et al. 2010; Fujimoto and James 2001). Microstrip patch antennas are particularly attractive due to their low profile and ease of fabrication; however, a conventional patch antenna has a length of ~λ/2 making it too large for many wireless applications. As a result, several design methodologies have been developed to reduce the size of the patch antenna. Typically, these techniques utilize some combination of meandered conductors, notches, and shorting (Li et al. 2004; Kan and Waterhouse 1999; Wang and Yang 1998). Generally speaking, the main

DOI 10.1007/978-981-4560-75-7_15-1

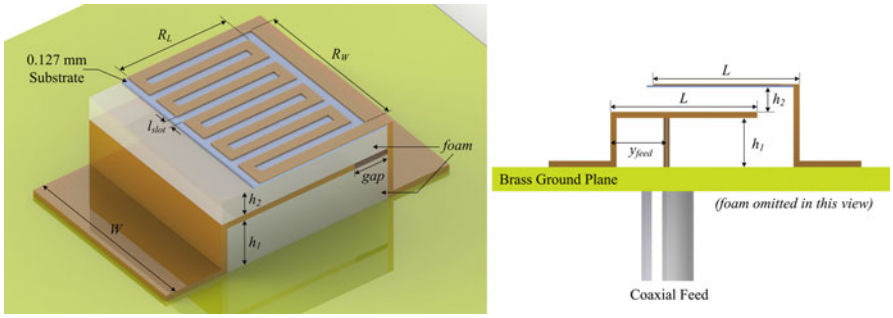


Fig. 16 Parameterized antenna geometry used for optimization: side view (*left*) and *top view* (*right*)

drawbacks of employing these size-reducing techniques are reduced gain, increased cross-polarization, and reduced bandwidth. In addition to size reduction, many wireless systems also require multiband operation. Several techniques to achieve dual- or multiband operation with patch antennas have been reported; two common methods include a stacked patch configuration (Long and Walton 1979) or placing slots on the patch surface (Maci et al. 1995; Lee et al. 1997). Employing a combination of the aforementioned techniques creates the possibility of developing reduced size multiband patch antennas (Chen et al. 2003; Ollikainen et al. 1999; Guo et al. 2000; Brocker et al. 2014), which will be explored here.

In Maci et al. (1995), the authors demonstrated that dual-frequency operation with a frequency ratio of less than 2 could be achieved by placing thin slots near the radiating edges of a conventional patch antenna. As outlined in Guo et al. (2000), the two operational modes of the slotted patch antenna are TM_{10} and TM_{30}, which have minimum potential at the center plane of the patch. Hence, a short can be implemented in this plane at the cost of maximum broadside gain. As a result, the length of the slotted patch can be reduced by a factor of about 2. In addition to reduced gain, it is noted that the shorted slotted patch (SSP) antenna will exhibit increased cross-polarization, specifically in the H-plane. In Maci et al. (1995), the authors showed that the dual-band frequency ratio is a characteristic of the slot lengths. Therefore, even after the short is applied (subsequently reducing the patch length), the width of the patch and slot is fixed by the targeted frequencies, presenting a challenge for further patch size reduction. To overcome this limitation, the straight slot can be replaced with a meandered slot, creating a longer path for slot perimeter currents while at the same time allowing for a reduced patch width. The footprint of the antenna element is further reduced by employing the techniques presented in Li et al. (2004), where the authors demonstrated that the shorted-patch antenna can be folded while still maintaining suitable radiation and input properties. The proposed design here is folded once, reducing the footprint of the antenna by half, with the resulting geometry shown in Fig. 16. To demonstrate the effectiveness of the design procedure, the Wi-Fi bands of 2.4 and 5.0 GHz were targeted for optimization and fabrication.

The composition and construction of the antenna element shown in Fig. 16 were based on ease of fabrication. For example, the metal patch sections are fairly easy to bend and form without substrates. As a result, foam was inserted between patch layers for structural support. The meander pattern was created by chemically etching a 0.127-mm-thick Rodgers RT/duroid® 5880 copper-plated substrate. A 1.57-mm-thick brass square ground plane was used that was 0.3 λ at 2.4 GHz (~3.75 cm) on an edge. The lower patch layer of the antenna element was fed by a 12.7-cm-long semirigid RG402 coaxial cable. This relatively long feed was not considered during optimization since it would dramatically increase the size of the simulation domain and associated simulation time. Instead, the coax waveguide considered during optimization was only 3 mm long. When coupled with the small ground plane, different lengths of coaxial feed line did not significantly affect the input impedance of the antenna element; however, radiation patterns were affected. Therefore, the longer coax was integrated into the simulation after the

DOI 10.1007/978-981-4560-75-7_15-1

optimization, allowing for better agreement between simulated and measured radiation patterns. If a larger ground plane were utilized, however, the effect of the coax length on the radiation characteristics would be reduced.

The antenna element was optimized using CMA-ES for better than (or equal to) −12-dB reflection coefficient and better than (or equal to) 6 dBi of broadside gain at each frequency, requiring a relatively simple cost function given by

$$F_{\mathrm{SSPA}} = \sum_{i=1}^{n} \left(12 + \max\left\{S_{11}^{\mathrm{dB}}(f_i),\ -12\right\}\right)^2 + \sum_{i=1}^{n} \left(6 - \min\left\{\mathrm{Gain}^{\mathrm{dB}}(f_i), 6\right\}\right)^2. \tag{16}$$

In this case, the $n = 2$ frequencies are 2.4 and 5.0 GHz. The squaring operations help to evolve designs with balanced performance for all goals and frequencies. Throughout the optimization procedure, the cost value of each candidate design was evaluated using Ansoft HFSS.

The optimization configuration and parameters are listed in Table 8. A crucial step in any optimization is the initial parameterization; that is, specifying the optimization variables and their respective ranges. In general, a large enough search space (i.e., parameter ranges) should be specified such that a design with a suitable cost value can be generated; however, if the search space is too vast, the optimization algorithm can be overwhelmed and may fail due to pre-convergence on a local minimum that does not meet the cost goal. In addition, for designs which must be physically realized, geometries and components (i.e., capacitor values) must be specified such that only feasible solutions are generated in the cost function. Alternatively, optimized designs can have their geometric or component values rounded to the nearest feasible solution after the optimization, and then fine tuning can be applied if the performance is significantly degraded.

Antenna geometric parameters and ranges are determined here such that resulting designs can be feasibly constructed with the available fabrication technologies. For example, if the patch width is limited to 0.1 λ at 2.4 GHz (12.5 mm) and the meander is composed of five back-and-forth segments with traces approximately as wide as their spacing, the feature sizes would be approximately 0.6 mm, a reasonable value for a photolithography fabrication setup. Hence, l_{slot} was permitted to vary from 0.4 to 0.6 mm with the described configuration. Next, the substrate thicknesses h_1 and h_2 were restricted to discrete multiples of the available foam sheet thicknesses. Specifically, the thicknesses were only permitted to take one of two values, indicated by the *minimum* and *maximum* in Table 8. Because CMA-ES uses continuous variables, binary selection was achieved through rounding as in the previous section using the Pattern test function. The ranges for variable R_w were chosen based on findings from Maci et al. (1995) where it was shown that the slot must stretch the width of the path in order to achieve a radiation maximum at

Table 8 Optimized antenna geometry parameters

Parameter	Minimum value	Maximum value	Optimized value	Fabricated value
L	8.00 mm	9.50 mm	9.50 mm	9.50 mm
W	11.50 mm	12.50 mm	12.48 mm	**12.70** mm
gap	2.50 mm	3.00 mm	2.73 mm	**2.31** mm
R_L	0.70*L	0.80*L	0.80*L	0.80*L
R_W	0.80*W	0.90*W	0.87*W	0.87*W
h_1 (discrete)	1.59 mm	3.18 mm	3.18 mm	**3.22** mm
h_2 (discrete)	1.59 mm	3.18 mm	1.59 mm	**1.65** mm
l_{slot}	0.40 mm	0.60 mm	0.60 mm	0.60 mm
y_{feed}	0.20*L	0.80*L	0.34*L	0.34*L

DOI 10.1007/978-981-4560-75-7_15-1

broadside. Note that R_w, as well as R_L, was specified in terms of fractions of W and L, respectively. The *gap* parameter should be small in order to keep the overall antenna size to a minimum, however, during fabrication *gap* is set when the support tabs shown in Fig. 16 are soldered to the ground plane. Since this is expected to be a large source of fabrication error, *gap* was configured to be at least 2–3 mm to avoid any detrimental effects due to manufacturing imperfections.

CMA-ES was utilized to carry out the antenna optimization of the nine variables listed in Table 8. A population size of 16 was used and converged to the optimized values by the 24th iteration. The optimized antenna element achieved a cost function value of $F_{\mathrm{SSPA}} = 9.58$. In terms of antenna performance, this cost equated to better than −12 dB reflection coefficients at both frequencies and 2.9 and 5.9 dBi broadside gain at 2.4 and 5.0 GHz, respectively. Once the 12.7-cm-long coax feed was included in the optimized antenna design, the calculated broadside gains changed to 3.4 and 5.5 dBi at 2.4 and 5.0 GHz, respectively. The antenna impedance was not significantly affected by including the longer coax feed model.

The design was fabricated and measured in order to validate the simulated results of the optimized design. The fabricated prototype is shown in the inset of Fig. 17, where a comparison of the simulated and measured reflection coefficients is given. Although the fabricated design agreed well with the simulated results of the optimized design, further improvement in agreement was achieved by measuring the dimensions of the fabricated antenna element and modifying the simulation model accordingly. The parameters that required adjustment in the simulation are given in bold in the last column of Table 8. Figure 18 compares the radiation pattern of the simulation model and the fabricated antenna element, where excellent agreement was again found.

It has been shown here that optimization strategies such as CMA-ES are a powerful tool for antenna design. It was used in this case to optimize the parameters of a folded, shorted-patch antenna with a meander slot in order to achieve significant size reduction and multiband operation. An antenna with this amount of complexity generally does not lend itself to closed-form solutions such as with standard patch or folded-patch structures, hence a simulation tool was required that allowed the antenna parameters to be accurately determined. Because the design had a significant number of geometric parameters that needed to be optimized, a parametric optimization or simple hand tuning would be impractical; however, CMA-ES proved to be a capable strategy for finding a suitable set of parameters that met the performance criteria. The proposed antenna design methodology and optimization yielded a miniaturized dual-band patch antenna with dimensions all smaller than 0.1 λ (at 2.4 GHz). The resulting design achieved −10-dB

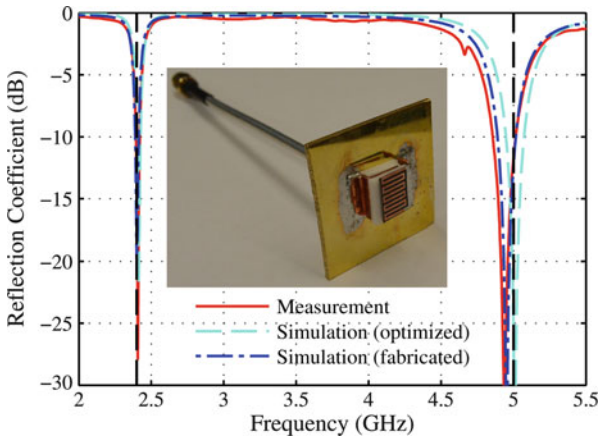


Fig. 17 Simulated and measured reflection coefficient of optimized antenna element using a $0.3\lambda_{2.4\mathrm{GHz}}$ (≈3.75 cm) square ground plane along with the fabricated structure (*inset*). Vertical *dashed lines* indicate the targeted frequencies of the optimization

DOI 10.1007/978-981-4560-75-7_15-1

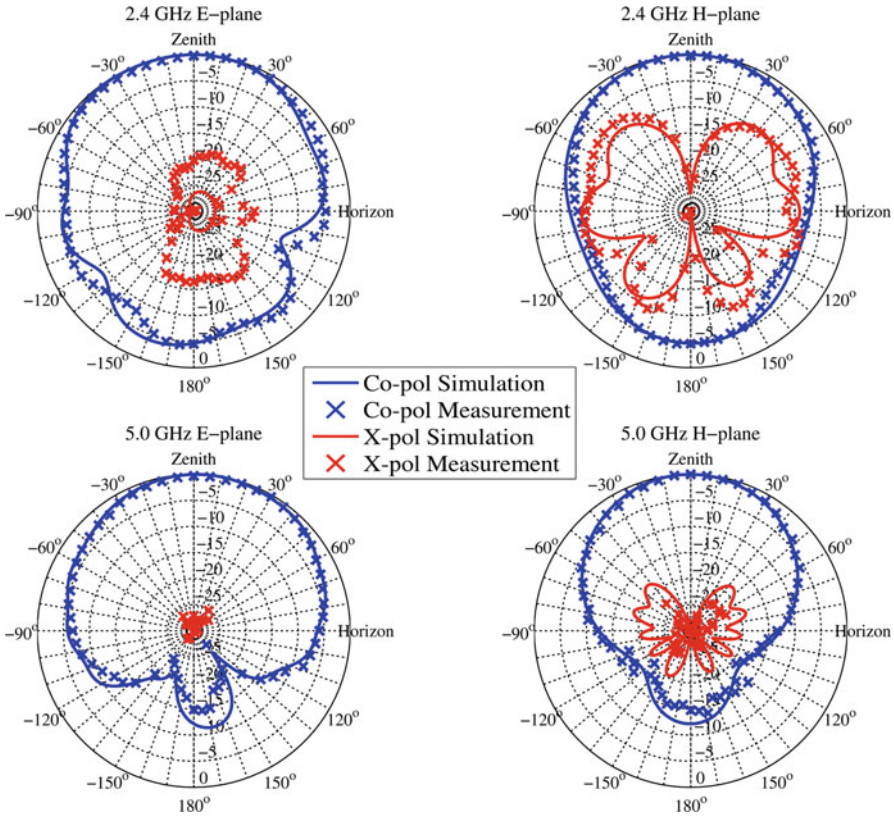


Fig. 18 Radiation pattern of optimized antenna element using a 0.3 λ (~3.75 cm) square ground plane. *Solid lines* represent simulated radiation patterns and ×'s represent measured radiation patterns. Broadside gains were simulated and calculated to be 3.4 dBi and 5.5 dBi at 2.4 GHz and 5.0 GHz, respectively

impedance matching bandwidths of 1.0 % and 2.5 % at 2.4 and 5.0 GHz, respectively. The broadside gains of the optimized antenna element were 3.4 and 5.5 dBi at 2.4 and 5.0 GHz, respectively. Furthermore, the small antenna size was leveraged to include a ground plane size of only 0.3 λ (at 2.4 GHz), an exceedingly small figure compared to typical antenna ground plane requirements.

Broadband Monopole Antenna Enabled by a Metamaterial Coating

Broadband antennas have been widely used in various wireless communication systems for high data transfer rates and in radar systems for achieving shorter pulse durations. Several techniques have been proposed since the early 1970s to broaden the impedance bandwidth of wire and planar antennas while maintaining their vertical polarization. Conventional approaches include placing conducting sleeves around the main radiator to provide a second resonance at a higher frequency (Volakis 2007; King and Wong 1972) or inserting lumped series resistor-inductor circuits into an electrically long wire monopole (Volakis 2007; Lo and Lee 1988). The planar version of the sleeve monopole has been realized by using planar sleeves on both sides of an end-loaded microstrip monopole on the same printed circuit substrate (Spence and Werner 2006). In addition to using conducting sleeves to produce a second resonance, bulk dielectric resonators of various shapes have also been utilized to enhance the impedance bandwidth of wire antennas (Chang and Kiang 2007; Guha et al. 2009). Other techniques involving reshaping the periphery of the planar monopole to achieve functionality similar to a traveling wave antenna have also been reported (Liang et al. 2005; Chen et al. 2007). The gradually tapered slot between the edge of the planar monopole and the ground plane enables impedance matching over a broad bandwidth. More recently, artificial electromagnetic metamaterials have been shown to extend the bandwidth of planar

DOI 10.1007/978-981-4560-75-7_15-1

monopoles and microstrip antennas by properly loading split ring resonators or negative refractive index transmission lines (Palandoken et al. 2009; Antoniades and Eleftheriades 2009).

Presented in this section are the design optimization and experimental validation of an octave bandwidth quarter-wave wire-type monopole antenna surrounded by an ultrathin flexible anisotropic metamaterial coating. In contrast to previously reported broadband vertical polarized planar monopoles which typically develop multiple lobes in their radiation patterns as frequency increases (Liang et al. 2005; Chen et al. 2007), the new octave bandwidth metamaterial-enabled monopole has stable vertically polarized radiation patterns over the entire frequency band of operation. Moreover, compared to broadband open-sleeve dipoles and monopoles (King and Wong 1972) as well as broadband dielectric resonator antennas fed by monopoles (Chang and Kiang 2007; Guha et al. 2009), the proposed metamaterial-coated monopole antenna is more compact and extremely lightweight, suggesting possible applications ranging from broadband arrays to portable wireless devices.

Unit Cell Design

The unit cell of the metamaterial coating is comprised of two identical I-shaped copper patterns printed on both sides of a Rogers ULTRALAM® 3850 substrate (see Fig. 19a). The thicknesses of the substrate (d_s) and the copper (d_c) are 51 and 17 μm, respectively. Using this thin flexible substrate, the nominally planar metamaterial structure can be easily formed into a cylindrical configuration. The effective medium

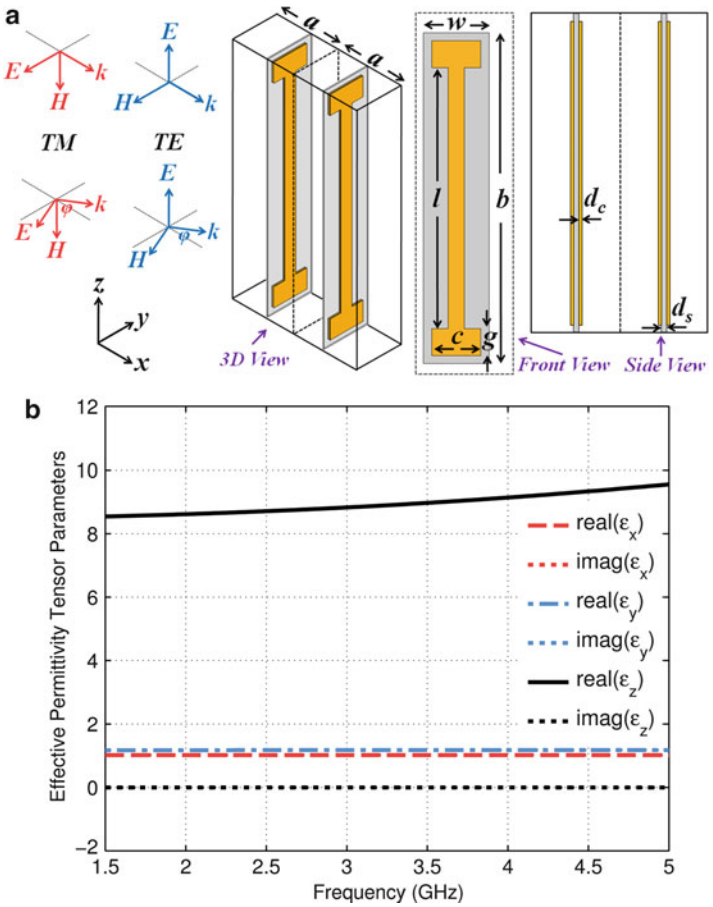


Fig. 19 (**a**) Geometry and dimensions of the unit cells of the anisotropic metamaterial coating. All dimensions are in millimeters: $a = 2.5$, $d_s = 0.051$, $d_c = 0.017$, $w = 2$, $b = 10$, $c = 1.5$, $g = 0.8$, and $l = 8$. (**b**) Real and imaginary parts of the retrieved effective anisotropic permittivity tensor parameters (ε_x, ε_y, ε_z)

DOI 10.1007/978-981-4560-75-7_15-1

properties of the metamaterial are obtained according to the diagram shown in Fig. 19a, where periodic boundary conditions are assigned to the walls in the *y*- and *z*-directions. A TE/TM polarized plane wave, with the *E*-field/*H*-field oriented along the *z*-direction, is incident from the left half-space at an angle of φ ($0^\circ \leq \varphi \leq 90^\circ$) with respect to the *x*-axis. An anisotropic inversion technique (Jiang et al. 2011a) was employed to extract all six effective permittivity and permeability tensor quantities from the scattering parameters calculated at different angles of incidence using HFSS™.

The retrieved effective permittivity tensor parameters are shown in Fig. 19b. It can be seen that none of the parameters exhibit a resonant response in the band of interest as a result of the subwavelength I-shaped elements, which corroborates previously reported results on I-shaped unit cells utilized for broadband microwave metamaterial devices (Liu et al. 2009; Jiang et al. 2011b). The retrieved ε_x and ε_y parameters have nondispersive values near unity, whereas ε_z exhibits a large value that can be attributed to the inductance provided by the central microstrip in the I-shaped elements and capacitance associated with the gaps between the stubs of adjacent unit cells in the *z*-direction. Controlling the series inductance and capacitance enables manipulation of the value of ε_z across the band, which will be optimized during the design process. The three effective permeability tensor parameters (not shown here) have nondispersive values equal to unity with very low loss, indicating that the metamaterial does not have any effect on the radiated magnetic field.

Optimization of the Monopole Antenna Coated with Metamaterial

The configuration of the monopole antenna with and without the metamaterial coating is shown in Fig. 20. The monopole is 28.5 mm long and resonates at 2.5 GHz. The cylindrical metamaterial coating is composed of two concentric layers of printed metamaterial cells as illustrated in Fig. 20b. The inner and outer layers contain 8 and 16 unit cells along their circumference, respectively, in order to approximate a circular outer periphery to minimize its impact on the monopole's omnidirectional radiation patterns in the *H*-plane. A genetic algorithm (GA) (Haupt and Werner 2007) was employed during the design process to optimize the dimensions of the I-shaped metamaterial unit cell and the periodicities of the metamaterial in both the *z*- and φ-directions, which determine the height and radius of the coating, respectively. In this case, the GA that is included with the MATLAB optimization toolbox was employed with 8 bits per parameter. Single-point crossover was used with a random mutation probability of 0.02 applied only at even generations. The optimization used a cost function given by

$$F_{\text{MONOPOLE}} = \sum_{i=1}^{n} \max\{VSWR(f_i) - 2, 0\} + \sum_{i=1}^{n} \max\{0.95 - \text{Efficiency}(f_i), 0\}, \tag{17}$$

with $n = 12$ frequency points evenly distributed from 2.2 to 4.4 GHz (0.2 GHz increments). The six parameters to be optimized are the monopole length (h_a), the radius of the inner functional layer (d_i), the height of each unit cell (b), the length of the central bar of the I-shaped cell (l), the length of the end bars of the I-shaped cell (c), and the width of the end bars of the I-shaped cell (g). A population size of 32 was used, with the GA requiring 30 generations before obtaining the desired cost goal of 0.01, yielding an approximate performance of VSWR $\leq$ 2 over the specified frequency range. The resulting metamaterial coating shown in Fig. 20 has an outer radius of 5 mm (about $\lambda/24$ at 2.5 GHz), ensuring a compact footprint in the radial direction. The resulting optimized monopole performance is given in Fig. 21, along with the VSWR of the uncoated monopole and that of a monopole coated with a perfect homogenous anisotropic material with identical material tensors.

The monopole alone yields a VSWR $\leq$ 2 bandwidth of 0.4 GHz (2.3~2.7 GHz) with a single resonance at 2.5 GHz, whereas with the GA-optimized metamaterial coating present, the VSWR $\leq$ 2 bandwidth is remarkably broadened to 2.3 GHz (2.1~4.4 GHz). The main resonance shifts down slightly to 2.35 GHz,

DOI 10.1007/978-981-4560-75-7_15-1

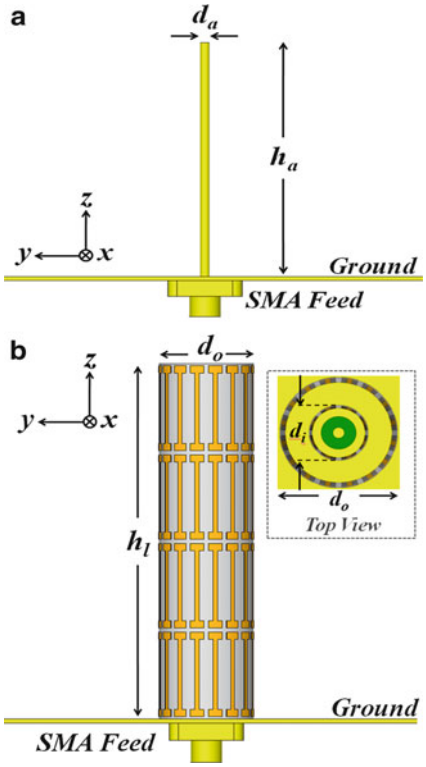


Fig. 20 Configuration of (**a**) the quarter-wave monopole antenna and (**b**) the same monopole with the ultrathin flexible anisotropic metamaterial coating. All dimensions are in millimeters: $d_a = 1$, $h_a = 28.5$, $d_i = 5$, $d_o = 2d_i$, and $h_l = 40$. The dielectric is 51 μm thick Rogers ULTRALAM® 3850 ($\varepsilon_r = 2.9$, $\delta_{\tan} = 0.0025$)

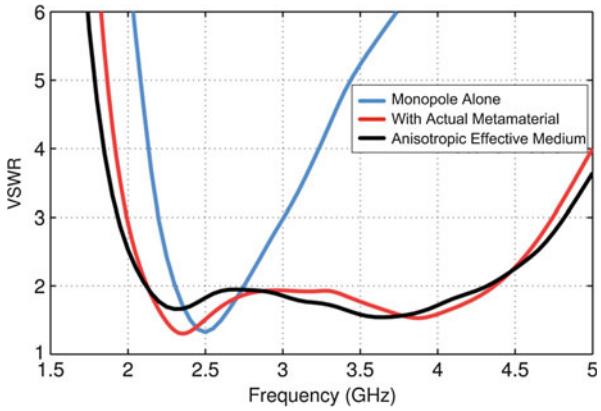


Fig. 21 Simulated VSWR of the monopole alone, the monopole with actual metamaterial coating, and the monopole with a homogeneous anisotropic effective medium coating. Identical ground plane sizes (32 × 32 cm) were used in all three simulations

and a new resonance is enabled at 3.85 GHz. To further examine the efficacy of the anisotropic effective medium model, a homogeneous anisotropic effective medium coating was also simulated, showing a VSWR with a behavior similar to that of the actual optimized metamaterial coating. The VSWR ≤ 2 bandwidth is 2.2 GHz (2.1~4.3 GHz) with the first and the second resonance located at 2.32 and 3.65 GHz, respectively, indicating that the assumed homogeneous anisotropic effective medium model is a valid approximation for the actual curved metamaterial. This is primarily because a sufficient number

DOI 10.1007/978-981-4560-75-7_15-1

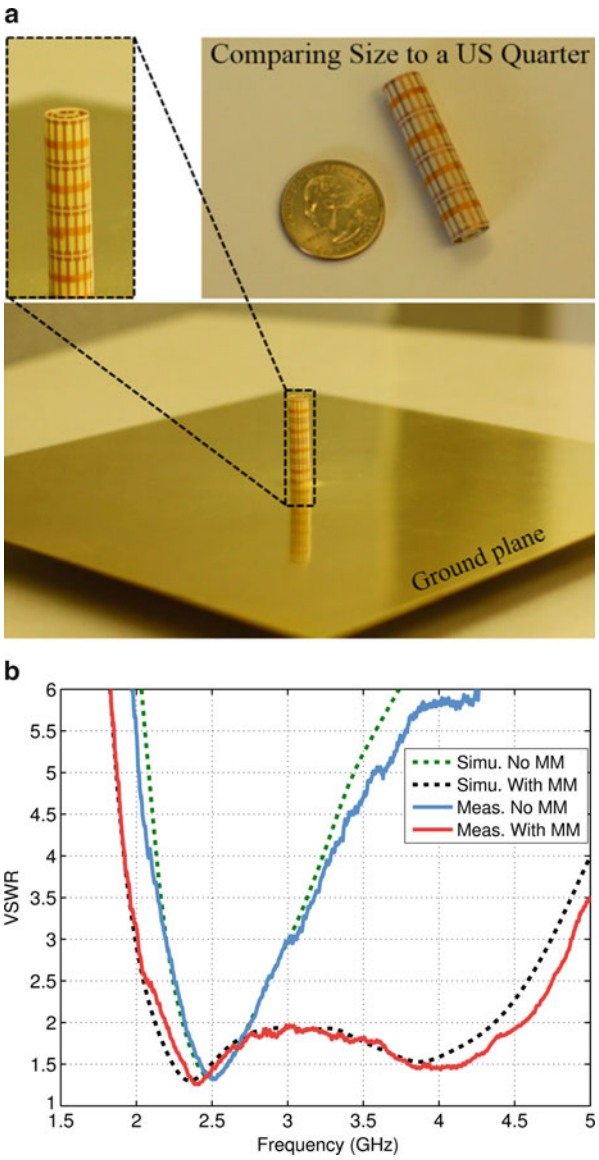


Fig. 22 Photographs of the fabricated metamaterial-coated monopole (**a**) and simulated and measured VSWR of the monopole antenna with and without the metamaterial coating (**b**)

of small unit cells are used to form the cylindrical coating such that the metamaterial still possesses a reasonably good local flatness.

Experimental Verification

The metamaterial coating was fabricated and assembled with the monopole as shown in Fig. 22a. VSWR measurements were carried out using an Agilent E8364B network analyzer. Figure 22b compares the simulated and measured VSWRs of the monopole with and without the metamaterial coating on a 32×32 cm ground plane. The measured VSWR of the monopole alone is almost identical to the simulated results with VSWR ≤ 2 from 2.3 to 2.7 GHz. With the metamaterial present, a 2.14:1 ratio bandwidth (2.15~4.6 GHz) of VSWR ≤ 2 is obtained. Frequency shifts of 0.05 and 0.2 GHz were found in the lower and higher ends of the band, respectively, the likely result of a slight tilt between the monopole and the coating, as well as fabrication imperfections.

DOI 10.1007/978-981-4560-75-7_15-1

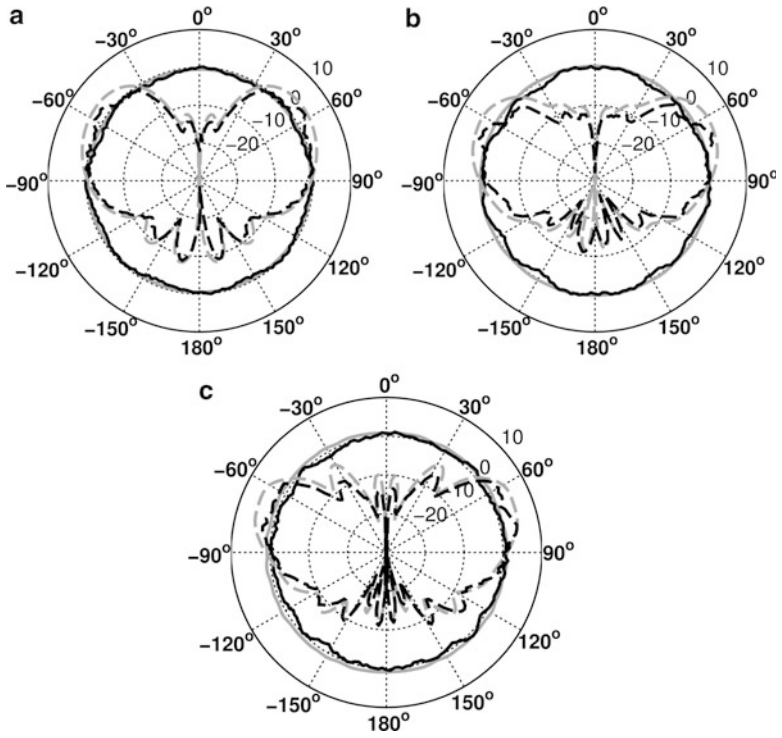


Fig. 23 Simulated and measured *H*-plane (*x*-*y* plane) and *E*-plane (*y*-*z* plane) radiation patterns of the metamaterial-coated monopole at (**a**) 2.2 GHz, (**b**) 3.3 GHz, and (**c**) 4.4 GHz. *Gray lines*: simulated *H*-plane patterns. *Black lines*: measured *H*-plane patterns. *Dashed gray lines*: simulated *E*-plane patterns. *Dashed black lines*: measured *E*-plane patterns

The radiation patterns of the metamaterial-coated monopole were also measured in an anechoic chamber. Figure 23 presents the simulated and measured *E*-plane and *H*-plane patterns at 2.2, 3.3, and 4.4 GHz. The *H*-plane patterns exhibit stable omnidirectional radiation characteristics throughout the entire band. The gain variations are around 0.5 and 1.2 dB for simulation and measurement, respectively. The increased measured gain variation as a function of the azimuthal angle is primarily caused by the imperfection of assembly and noise as well as the antenna rotation platform. In the *E*-plane, characteristic rabbit-ear-shaped patterns can be observed which are very similar to the patterns for the monopole without the metamaterial coating, indicating that the added coating has negligible impact on the spatial distribution of the radiated energy of the monopole. The maximum gain of the metamaterial-coated monopole varies from 3.75 to 5.46 dBi in the VSWR ≤ 2 band with the direction of maximum gain moving from 32° to 26° off horizon due to the finite-sized ground plane used in both simulation and measurement. The measured gain is 0.3–0.8 dB smaller than the simulated values. The overall very good agreement between simulation and measurement confirms the expected performance of the proposed metamaterial antenna coating.

Comparison to Sleeve Monopole Antenna

In order to prove the utility of the anisotropic metamaterial coating and demonstrate that the additional complexity is worthwhile, simulations for monopole antennas surrounded with simple parasitic conducting sleeves are carried out as shown in Fig. 24a. When the parasitic conducting sleeves are employed, a monopole-like resonance mode can be excited on the sleeves, thereby extending the impedance bandwidth of the original antenna. As a fair comparison, the footprint of the sleeve monopole is fixed to be the same as that of the optimized metamaterial-coated monopole, and the design is optimized for the largest possible bandwidth. It can be seen from Fig. 24b that the sleeve monopole achieves a

DOI 10.1007/978-981-4560-75-7_15-1

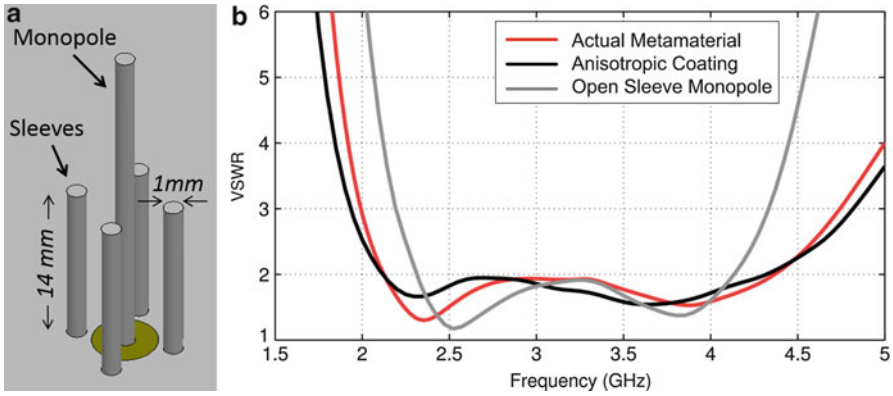


Fig. 24 (**a**) Configuration of sleeve monopole with the same horizontal footprint. (**b**) Simulated VSWR of the sleeve monopole, monopole with actual metamaterial coating, and monopole with homogeneous anisotropic effective medium coating

VSWR ≤ 2 bandwidth from 2.3 to 4.15 GHz, which is about 21 % narrower than that of the metamaterial-coated monopole.

Low-Profile SIW-Fed Slot Antenna with High-Gain Enabled by a Metamaterial Coating

It is of particular interest to capably generate unidirectional radiation with high-gain at broadside in a single compact antenna for applications such as point-to-point communications, wireless power transfer, radar systems, and various other wireless systems (Balanis 2008; Chen and Luk 2009). Conventional approaches have utilized methods such as high-Q Fabry-Pérot (FP) cavities (Guérin et al. 2006; Volakis 2007) to obtain narrowband high directivity with a profile of a half of the operating wavelength. Recent advanced FP cavity designs employing metamaterial surfaces allow the total device thickness to be reduced to around $\lambda/9$ (Feresidis et al. 2005; Sun et al. 2012). However, most of these FP cavity antennas exhibit an enhanced directivity and matched input impedance over very narrow bandwidths, which considerably limits their utility. Apart from the FP cavity-related techniques, another metamaterial approach for increasing directivity has recently been theoretically proposed and experimentally demonstrated by embedding the electromagnetic source within a volumetric isotropic and anisotropic zero-index material/low-index material (ZIM/LIM) lens (Enoch et al. 2002; Ziolkowski 2004; Turpin et al. 2010; Jiang et al. 2012a; Cheng and Cui 2011). However, these ZIM or anisotropic ZIM (AZIM) lenses are usually electrically large in all three dimensions, unavoidably resulting in increased net size and weight of the device.

Recently, it has been proposed that thin single- and bilayer-grounded metamaterial slabs with either isotropic negative or zero index of refraction can give rise to novel electromagnetic properties such as surface wave guidance (Baccarelli et al. 2005a) and suppression (Baccarelli et al. 2003) as well as frequency-dependent directive radiation (Lovat et al. 2006; Baccarelli et al. 2005b; Alu et al. 2007; Shahvarpour et al. 2011). The structures considered thus far are generally infinite or finite but electrically long and support leaky waves with phase velocities slightly higher than that in free space. As a result, the direction of peak radiation drifts away from broadside and varies as a function of frequency, rendering them unsuitable for maintaining broadband high-gain unidirectional radiation at broadside (Liu et al. 2002; Caloz and Itoh 2005; Lovat et al. 2006; Baccarelli et al. 2005a; Alu et al. 2007; Shahvarpour et al. 2011). In this section, a low-profile, high-gain, and compact-footprint antenna that is

DOI 10.1007/978-981-4560-75-7_15-1

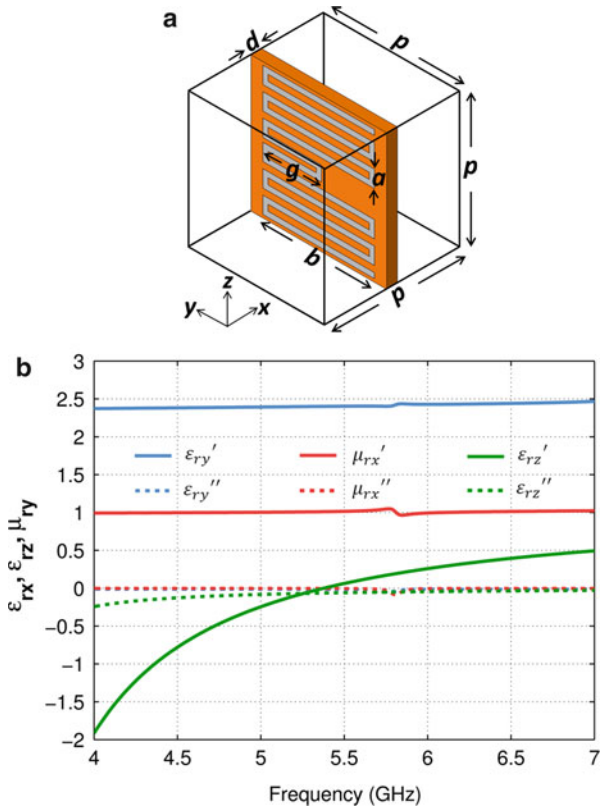


Fig. 25 (**a**) Geometry of the end-loaded dipole unit cell for constructing an anisotropic ZIM. The dimensions are $p = 6.5$ *mm*, $b = 5.35$ mm, $a = 0.7$ mm, $d = 0.508$ mm, and $g = 2.8$ mm. The substrate material is Rogers RT/duroid 5880 with a dielectric constant of 2.2 and a loss tangent of 0.009. (**b**) The retrieved effective medium parameters μ_{rx}, ε_{ry}, and ε_{rz}

composed of a subwavelength thick AZIM coating and a substrate-integrated waveguide (SIW) (Deslandes and Wu 2003) fed slot antenna is proposed and experimentally realized.

Metamaterial Unit Cell Design

To realize materials with anisotropic zero-/low-index properties, periodic end-loaded dipole resonators (ELDRs) are employed (Jiang et al. 2012b). The self-inductance of the wire and the capacitance provided by the gaps between the meandered end-loaded arm traces give rise to an electric resonance. The unit cell geometry and dimensions are shown in Fig. 25a. To retrieve the effective medium parameters (ε_{ry}, ε_{rz}, and μ_{rx}), an infinite doubly periodic array of ELDRs was simulated in HFSS, with periodic boundary conditions and Floquet ports properly assigned to the boundaries of the simulation domain. Scattering results are then used with an anisotropic parameter inversion algorithm (Jiang et al. 2011a). The retrieved ε_{ry}, ε_{rz}, and μ_{rx} are shown in Fig. 25b, with ε_{ry} and μ_{rx} being virtually nondispersive with values around 2.4 and 1, respectively. ε_{rz} exhibits a Lorentz-shaped electric resonance with an effective plasma frequency of 5.38 GHz. Furthermore, because the resonance tail is weakly dispersive, the value of ε_{rz} remains positive and below 0.15 within the broad frequency range of 5.4–6.1 GHz. In this frequency band, the practical dispersive metamaterial has β_y^{TM} that varies from a near-zero value to 0.35. This geometry yields a broader low-index bandwidth compared to that of the commonly used subwavelength electric LC resonator (Schurig et al. 2006) due to the lower quality factor provided by the large capacitance inherent in the meandered arms.

DOI 10.1007/978-981-4560-75-7_15-1

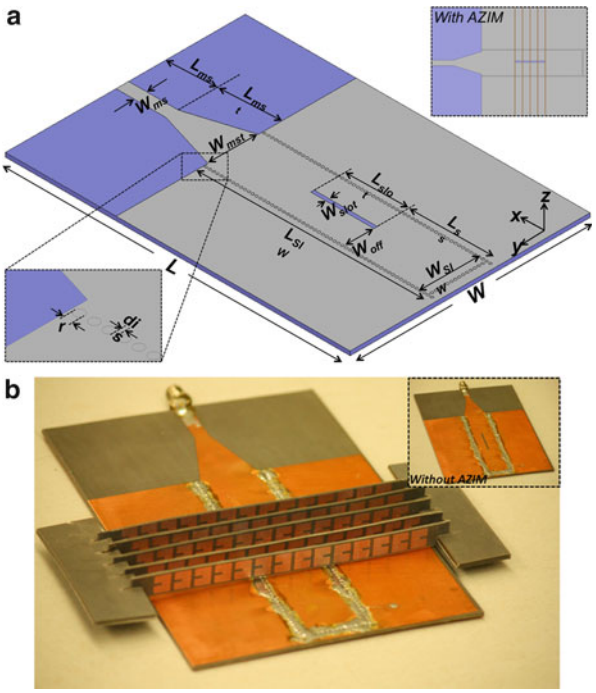


Fig. 26 (**a**) Schematic view of the SIW-fed slot antenna. The dimensions are $L = 133$, $W = 92.5$ mm, $W_{ms} = 4.83$ mm, $L_{ms} = 19$ mm, $W_{mst} = 19.3$ mm, $L_{mst} = 23.5$ mm, $r = 1$ mm, $dis = 0.5$ mm, $L_{SIW} = 86.75$ mm, $W_{SIW} = 22.5$ mm, $W_{\text{slot}} = 1.33$ mm, $L_{\text{slot}} = 24.4$ mm, $L_{ss} = 32.55$ mm, and $W_{\text{off}} = 11.24$ mm. The substrate material is Rogers RT/duroid 5880 with a dielectric constant of 2.2 and a loss tangent of 0.009. The substrate thickness is 1.575 mm. The *inset* in the *top right* corner shows the SIW-fed slot antenna symmetrically covered by the AZIM coating with five rows in the *y*-direction each containing 14 unit cells. (**b**) Photograph of the fabricated SIW-fed slot antenna covered by the AZIM coating. The inset shows the SIW-fed slot antenna alone

Slot Antenna Integrated with AZIM Coating

A half-wave slot fed by an SIW is adopted here due to its low profile compared to cavity backed or conventional waveguide fed slots. The SIW is another emerging technology that has been widely applied in microwave and millimeter wave components (Hao et al. 2005; Liu et al. 2007) and antennas (Zhang et al. 2011; Wu et al. 2012) due to its planar topology which can be readily connected to other planar guided-wave transmission lines. A schematic view of the SIW-fed slot antenna is shown in Fig. 26a. It is comprised of a 50 Ω microstrip and a shorted SIW with a longitudinal slot cut on its broad wall. The SIW is equivalent to a conventional rectangular waveguide filled with dielectric that only supports TE_{n0} modes due to its subwavelength height. A tapered microstrip is used for impedance matching between the 50 Ω feed line and the SIW. In order to achieve a magnetic dipole mode in the slot for efficient radiation, the length of the slot is chosen to be about $\lambda/2$ at 5.8 GHz. The distance between the center of the slot and the shorted wall of the SIW in the *x*-direction is set to be about $3\lambda/4$ at 5.8 GHz, which allows the standing wave peak to be located at the center of the slot. The slot also has a slight offset from the central axis of the SIW in the *y*-direction, following design considerations for the slot antenna fed by a conventional rectangular waveguide. The AZIM coating is then placed directly on top of the SIW-fed slot antenna. The single-layer metamaterial coating consists of five strips in the *x*-direction, each having 14 unit cells in the *y*-direction. The 14 cells are oriented symmetrically in the *y*-direction to maintain a mirror symmetry, which helps to reduce beam squinting of the AZIM-coated slot.

The geometrical properties of the AZIM-coated SIW antenna were optimized using CMA-ES (Gregory et al. 2011) for suitable impedance matching and radiation properties. The ten parameters to be optimized are the SIW length (L_{SIW}) and width (W_{SIW}), the slot length (L_{slot}), the width of the slot (W_{slot}), the

DOI 10.1007/978-981-4560-75-7_15-1

distance between the slot and the sidewall of the SIW (W_{offset}), the distance between the slot and the end of the SIW (L_{ss}), the width (W_{mst}) and length (L_{mst}) of the tapered microstrip, the length of the ELDRs (b), and the length of the central line of the ELDRs (g). The cost function includes contributions from the input and radiation properties via

$$F_{\text{SIW-AZIM}} = \sum_{i=1}^{n} \max\left\{S_{11}^{dB}(f_i) + 10, 0\right\} + \sum_{i=1}^{n} \max\left\{10 - \text{Gain}^{dB}(f_i), 0\right\}, \tag{18}$$

with a goal of 0.01. The $n = 5$ frequency samples include 5.6, 5.7, 5.8, 5.9, and 6.0 GHz. The cost function targets S_{11} of less than −10 dB and broadside gain higher than 10 dBi in the band of interest. For the optimization, a population of 25 was used, and 30 generations were required to obtain a design with suitable properties. Again, Ansoft HFSS was employed to simulate the structure to gather the requisite electromagnetic properties. The resulting thickness of the optimized AZIM coating is only 6.5 mm (~0.12 λ at 5.8 GHz), much thinner than conventional FP cavities (Guérin et al. 2006; Feresidis et al. 2005), or recently proposed bulky ZIM lenses (Enoch et al. 2002; Ziolkowski 2004; Turpin et al. 2010; Jiang et al. 2012a; Cheng and Cui 2011).

The simulated S_{11} of the SIW-fed slot antenna is shown in Fig. 27a. The S_{11} is below −10 dB from 5.58 to 6.03 GHz. The normalized E-plane (y-z plane) and H-plane (x-z plane) radiation patterns at 5.6, 5.8, and 6.0 GHz are presented in the top portions of Fig. 28a through f. As a comparison, the normalized radiation patterns for the same antenna on a ground plane infinite in the y-direction are also plotted. It can be seen that, for the infinite ground plane case, omnidirectional radiation exists in the E-plane, while the beam

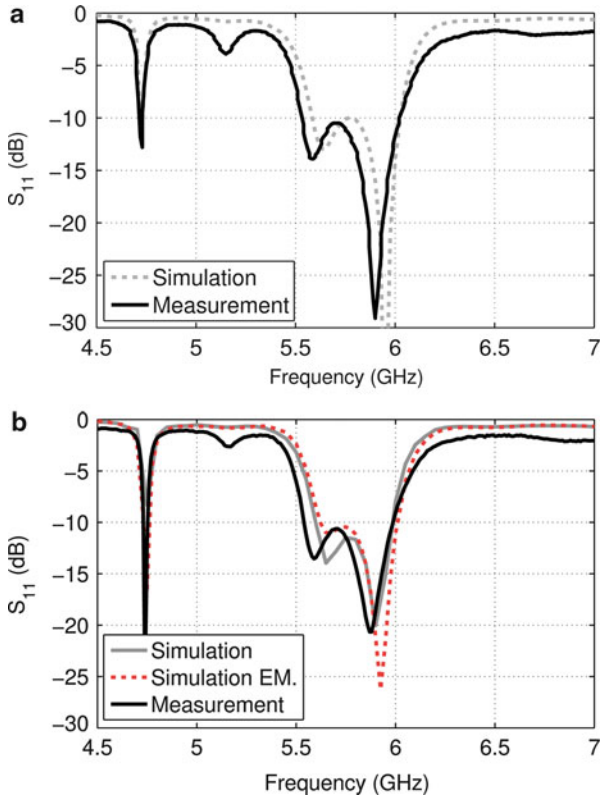
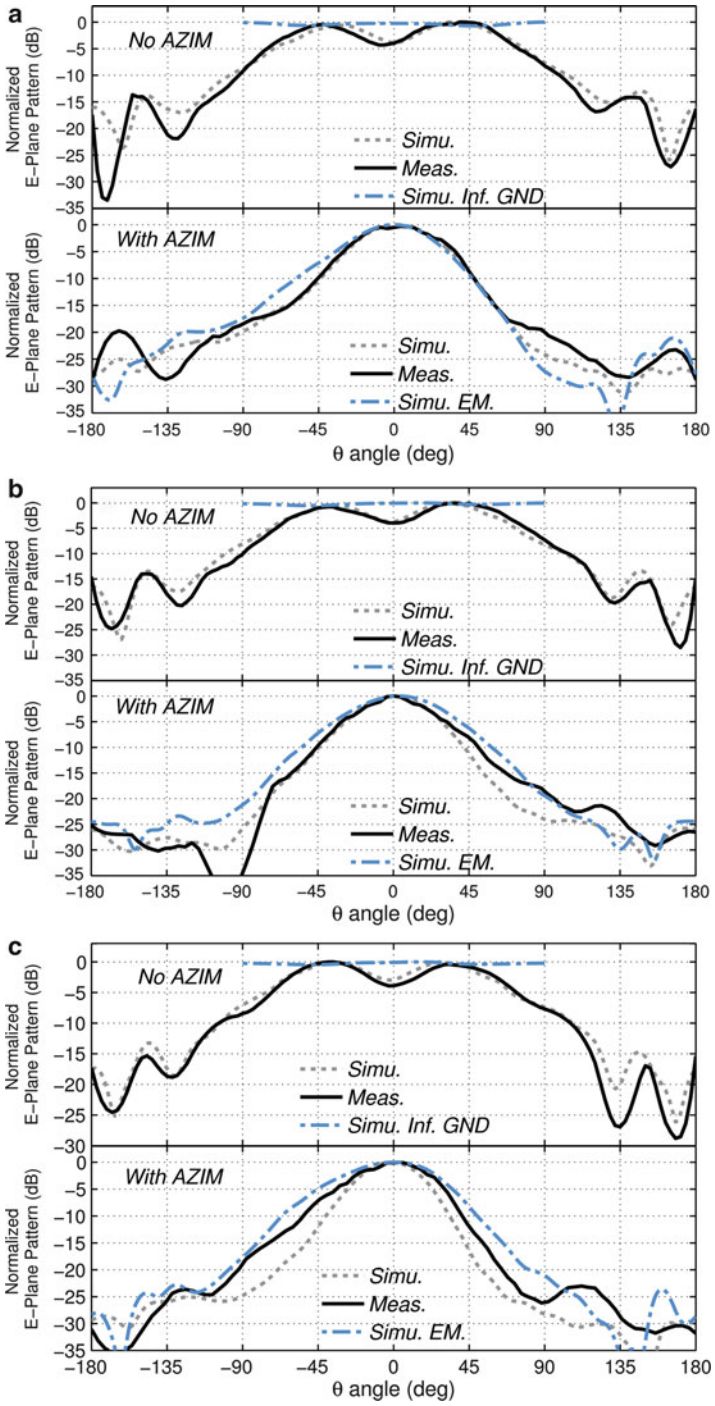


Fig. 27 (**a**) Simulated and measured S_{11} of the SIW-fed slot antenna without the AZIM coating. (**b**) Simulated and measured S_{11} of the SIW-fed slot antenna with the AZIM coating, including the dispersive homogeneous effective medium AZIM coating

DOI 10.1007/978-981-4560-75-7_15-1

maximum is directed at broadside in the *H*-plane. However, for a finite ground plane with a size of 1.7λ in the *y*-direction, significant edge diffraction can be observed in the *E*-plane, showing two peaks located at around 40–45° off broadside. The double-peak radiation in the *E*-plane caused by the finite ground plane is undesirable for unidirectional radiation due to the gain drop at broadside. The *H*-plane patterns are similar for the two cases since the ground plane size in the *x*-direction is identical. The simulated broadside gain of the antenna on the finite and infinite (in the *y*-direction) ground planes are shown in Fig. 29a. For the infinite ground plane case, the broadside gain is maintained at around 5 dBi, while for the finite ground plane case, the gain varies from 3 to 3.8 dBi.



DOI 10.1007/978-981-4560-75-7_15-1

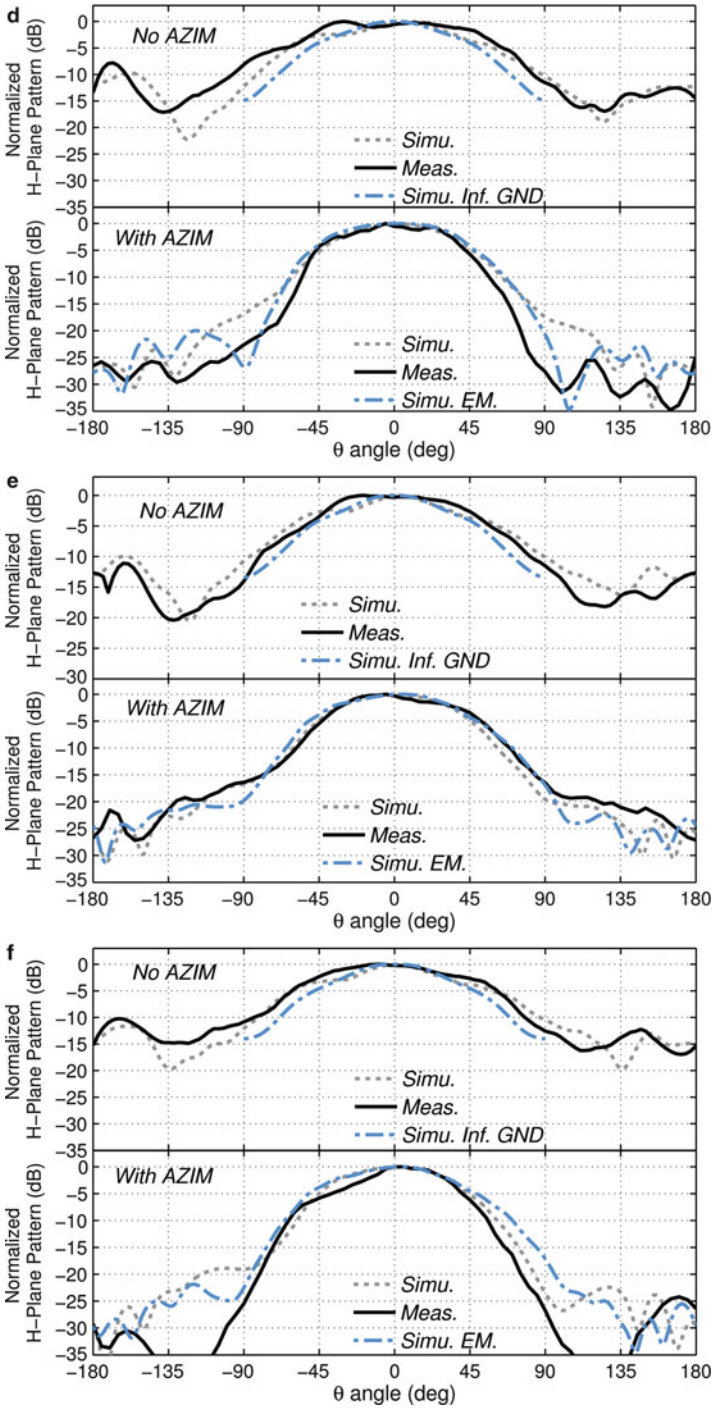


Fig. 28 Simulated and measured normalized radiation patterns of the SIW-fed slot antenna with and without the AZIM coating in the *E*-plane (*y*-*z* plane) at (**a**) 5.6 GHz, (**b**) 5.8 GHz, and (**c**) 6.0 GHz. Simulated and measured normalized radiation patterns of the SIW-fed slot antenna with and without the AZIM coating in the *H*-plane (*x*-*z* plane) at (**d**) 5.6 GHz, (**e**) 5.8 GHz, and (**f**) 6.0 GHz. The simulated *E*- and *H*-plane patterns for the SIW-fed slot antenna without the AZIM coating on a ground plane infinite in the *y*-direction and the simulated *E*- and *H*-plane patterns for the SIW-fed slot antenna with the homogeneous dispersive effective medium AZIM coating are also shown for comparison

DOI 10.1007/978-981-4560-75-7_15-1

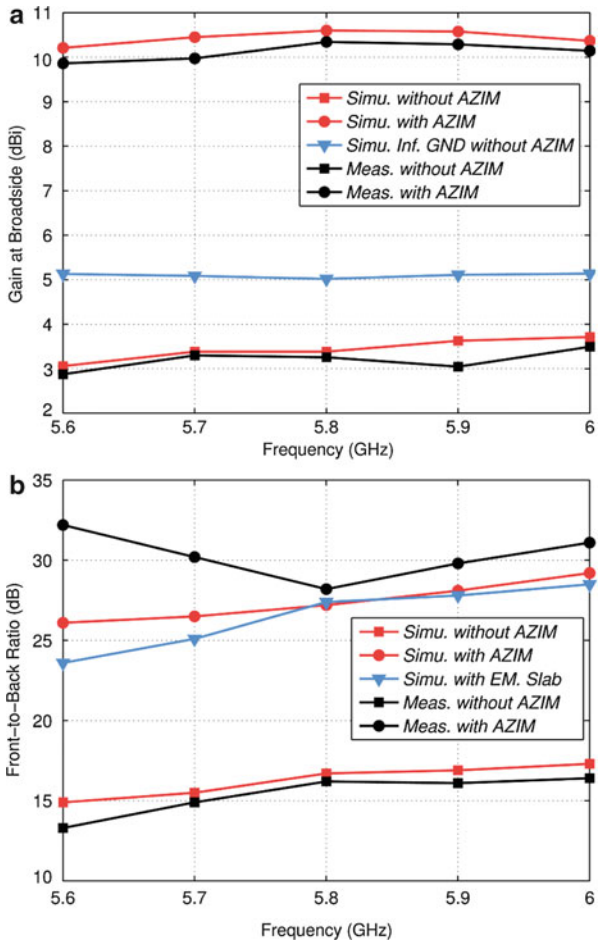


Fig. 29 (**a**) Simulated and measured gain at broadside ($\theta = 0°$) for the SIW-fed slot antenna with and without the AZIM coating. (**b**) Simulated and measured front-to-back ratio for the SIW-fed slot antenna with and without the AZIM coating (*EM*: an effective medium with properties equivalent to the AZIM coating)

As shown in Fig. 27b, the simulated S_{11} of the integrated antenna is below −10 dB from 5.6 to 6 GHz which is very similar to that of the slot alone. This robust input impedance behavior ensures that the AZIM coating can be readily added onto or taken away from the slot to achieve different radiation properties without any additional modification to the slot antenna itself. The normalized *E*-plane (*y*-*z* plane) and *H*-plane (*x*-*z* plane) radiation patterns at 5.6, 5.8, and 6.0 GHz are presented in Fig. 28a through f (bottom sections). In this case, the combined system produces a well-defined single beam at broadside observed in the *E*-plane with a beam squint less than 2° off broadside and a 3 dB beamwidth of about 35–40°. In contrast to the radiation properties in the *E*-plane, the *H*-plane patterns have beamwidths similar to the slot without the AZIM coating. Most notable is that the broadside gain is significantly increased to 10.2–10.6 dBi with the coating, an improvement of about 7 dB over the slot alone. It should also be noted that the front-to-back ratio is greatly reduced by about 10 dB throughout the entire frequency range, which is shown in Fig. 29b. Both the broadside gain increase and the front-to-back ratio drop are primarily attributed to the reduction of fields at the edges of the ground plane, a result of the presence of the AZIM coating. As an alternative perspective, the ultrathin AZIM coating can be considered a highly efficient radiating aperture comprised of an array of subwavelength radiators fed by the slot. The simulated aperture efficiency is greater than 87 % throughout the band.

DOI 10.1007/978-981-4560-75-7_15-1

As a comparison, the actual metamaterial coating structure is replaced by a homogeneous slab with the dispersive effective anisotropic material parameters. It can be observed that the S_{11} and the front-to-back ratio of the slot covered with the effective medium slab correspond well to those of the slot covered with the actual discrete metamaterial. The radiation patterns in both the *E*-plane and *H*-plane also show good agreement, especially in the main beam; the greatest disagreement can be found in the angular range near the back lobes. The strong correlation here justifies the effective medium approximation that was employed, which is assumed to be valid due to the subwavelength size of the unit cells.

Experimental Verification

The SIW-fed slot antenna and the AZIM coating structure were fabricated and assembled as shown in Fig. 26b. An Agilent E8364B network analyzer was used to characterize the S_{11} of the slot with and without the AZIM coating. As shown in Fig. 27a, b, good agreement can be found between simulations and measurements not only in terms of the −10-dB bandwidth but also the resonance positions. The measured S_{11} of the slot with and without the AZIM coating has a −10-dB band from 5.52 to 6.03 GHz and from 5.54 to 6.01 GHz, respectively. They are both slightly broader than the simulations predict due to minor resonant frequency shift at 5.6 GHz and the lower quality factors of both resonances within the −10-dB bandwidth.

The radiation patterns and the gain of the slot antenna with and without the AZIM coating, shown in Fig. 28a through f, were characterized in an anechoic chamber using an automated antenna movement platform. The measured radiation patterns in both the *E*-plane and *H*-plane agree well with the simulated results. Specifically, patterns exhibiting a double peak can be seen for the slot alone, while a single sharp beam pointing at broadside can be observed for the slot with the AZIM coating. The measured 3-dB beamwidth in the *E*-plane is about 40–50°, slightly broader than the simulated beamwidth, especially in the high-frequency band. This is mainly attributed to fabrication and assembly imperfections, which result in a nonideal symmetry in the actual metamaterial structure. The measured broadside gain for the slot with and without the AZIM coating is in the range of 9.8–10.4 dBi and 2.9–3.5 dBi, respectively, indicating an improvement of about 6.9 dBi. The measured front-to-back ratio is displayed in Fig. 29b, showing more than a 10-dB reduction throughout the band which is in good agreement with simulation.

Conclusion

In addition to these three examples, optimization strategies have been successfully applied on countless occasions to other antenna design problems not only in the open literature but in the manufacturing sector as well. These algorithms are extremely useful and powerful if properly applied to the design task at hand. A detailed overview of the usage and operation of several popular optimization algorithms that are well suited for antenna design has been given here. Through comparisons with test functions and an example antenna design task, a good understanding of how the parameterization (coding of the values that require optimization), choice of evolutionary parameters (such as the weighting constant and crossover rate of differential evolution), and even the choice of algorithm itself has been conveyed. For more information, the reader is encouraged to access the websites of the algorithm creators as well as other freely available online materials not only for further reading but to obtain code samples and application examples in order to facilitate their own antenna designs.

DOI 10.1007/978-981-4560-75-7_15-1

Cross-References

- Broadband and Multiband Planar Antennas
- Circularly Polarized Antennas
- Conformal Array Antennas
- Dielectric Resonator Antennas
- Frequency Selective Surfaces
- Low-Profile Antennas
- Metamaterials and Antennas
- Microstrip Patch Antennas
- Phased Arrays
- Reconfigurable Antenna Arrays for Wireless Communications
- Small Antennas
- Substrate-Integrated Waveguide Antennas
- Ultra-wideband Antennas

References

Akhoondzadeh-Asl L, Kern DJ, Hall PS, Werner DH (2007) Wideband dipoles on electromagnetic bandgap ground planes. IEEE Trans Antennas Propag 55(9):2426–2434

Altshuler EE (2002) Electrically small self-resonant wire antennas optimized using a genetic algorithm. IEEE Trans Antennas Propag 50(3):297–300

Altshuler EE, Linden DS (1997) Wire-antenna designs using genetic algorithms. IEEE Trans Antennas Propag 39(2):33–43

Alu A, Bilotti F, Engheta N, Vegni L (2007) Subwavelength planar leaky-wave components with metamaterial bilayers. IEEE Trans Antennas Propag 55(3):882–891

Antoniades MA, Eleftheriades GV (2009) A broadband dual-mode monopole antenna using NRI-TL metamaterial loading. IEEE Antennas Wirel Propag Lett 8:258–261

Ares-Pena FJ, Rodriguez-Gonzalez JA, Villanueva-Lopez E, Rengarajan SR (1999) Genetic algorithms in the design and optimization of antenna array patterns. IEEE Trans Antennas Propag 47(3):506–510

Auger A, Hansen N (2005) A restart CMA evolution strategy with increasing population size. In: Proceedings of the IEEE congress on evolutionary computation, Edinburgh, Scotland, UK, pp 1769–1776

Baccarelli P, Burghignoli P, Lovat G, Paulotto S (2003) Surface-wave suppression in a double-negative metamaterial grounded slab. IEEE Antennas Wirel Propag Lett 2:269–272

Baccarelli P, Burghignoli P, Frezza F, Galli A, Lampariello P, Lovat G, Paulotto S (2005a) Fundamental modal properties of surface waves on metamaterial grounded slabs. IEEE Trans Microwave Theory Tech 53(4):1431–1442

Baccarelli P, Burghignoli P, Frezza F, Galli A, Lampariello P, Lovat G, Paulotto S (2005b) Effects of leaky-wave propagation in metamaterial grounded slabs excited by a dipole source. IEEE Trans Microwave Theory Tech 53(1):32–43

Balanis CA (2008) Modern antenna theory: analysis and design. Wiley, New York

Bayraktar Z, Bossard JA, Wang X, Werner DH (2010) Real-valued parallel clonal selection algorithm for design optimization in electromagnetics. In: Proceedings of the 2010 IEEE international symposium on antennas and propagation, Toronto

Bayraktar Z, Komurcu M, Jiang ZH, Werner DH, Werner PL (2011) Stub-loaded inverted-F antenna synthesis via wind driven optimization. In: Proceedings of the 2011 IEEE international symposium on antennas and propagation, Spokane, Washington, pp 2920–2923

Bayraktar Z, Bossard JA, Wang X, Werner DH (2012) A real-valued parallel clonal selection algorithm and its application to the design optimization of multi-layered frequency selective surfaces. IEEE Trans Antennas Propag 60(4):1831–1843

Bayraktar Z, Komurcu M, Bossard JA, Werner DH (2013) The wind driven optimization technique and its application in electromagnetics. IEEE Trans Antennas Propag 61(5):2745–2757

Boag A, Boag A, Michielssen E, Mittra R (1996) Design of electrically loaded wire antennas using genetic algorithms. IEEE Trans Antennas Propag 44(5):687–695

Boeringer DW, Werner DH (2004) Particle swarm optimization versus genetic algorithms for phased array synthesis. IEEE Trans Antennas Propag 52(3):771–779

Boeringer DW, Werner DH (2005) Efficiency-constrained particle swarm optimization of a modified Bernstein polynomial for conformal array excitation amplitude synthesis. IEEE Trans Antennas Propag 53(8):2662–2673

Boeringer DW, Werner DH, Machuga DW (2005) A simultaneous parameter adaptation scheme for genetic algorithms with application to phased array synthesis. IEEE Trans Antennas Propag 53(1):356–371

Bray MG, Werner DH, Boeringer DW, Machuga DW (2002) Optimization of thinned aperiodic linear phased arrays using genetic algorithms to reduce grating lobes during scanning. IEEE Trans Antennas Propag 50(12):1732–1742

Brocker DE, Werner DH, Werner PL (2014) Dual-band shorted patch antenna with significant size reduction using a meaner slot. In: Proceedings of the 2014 IEEE international symposium on antennas and propagation, Memphis, Tennessee, pp 289–290

Caloz C, Itoh T (2005) Electromagnetic metamaterials: transmission line theory and microwave applications. Wiley, New York

Campelo F, Guimarães FG, Igarashi H, Ramírez JA (2005) A clonal selection algorithm for optimization in electromagnetics. IEEE Trans Magn 41(5):1736–1739

Chang T-H, Kiang J-K (2007) Broadband dielectric resonator antenna with metal coating. IEEE Trans Antennas Propag 55(5):1254–1259

Chen ZN, Luk K-M (2009) Antennas for base stations in wireless communications. McGraw-Hill Professional, New York

Chen Y, Wang C (2012) Synthesis of reactively controlled antenna arrays using characteristic modes and DE algorithm. IEEE Antennas Wirel Propag Lett 11:385–388

Chen HT, Wong KL, Chio TW (2003) PIFA with a meandered and folded patch for the dual-band mobile phone application. IEEE Trans Antennas Propag 51(9):2468–2471

Chen ZN, See TSP, Qing X (2007) Small printed ultrawideband antenna with reduced ground plane effect. IEEE Trans Antennas Propag 55(2):383–388

Cheng Q, Cui TJ (2011) Multi-beam generations at pre-designed directions based on anisotropic zero-index metamaterials. Appl Phys Lett 99:131913(1)–131913(3)

Coleman CM, Rothwell EJ, Ross JE (2004) Investigation of simulated annealing, ant-colony optimization, and genetic algorithms for self-structuring antennas. IEEE Trans Antennas Propag 52(4):1007–1014

de Castro LN, Von Zuben FJ (2002) Learning and optimization using the clonal selection principle. IEEE Trans Evol Comput 6(3):239–251

Deb A, Roy JS, Gupta B (2014) Performance comparison of differential evolution, particle swarm optimization, and genetic algorithm in the design of circularly polarized microstrip antennas. IEEE Trans Antennas Propag 62(8):3920–3928

Deslandes D, Wu K (2003) Single-substrate integration technique of planar circuits and waveguide filters. IEEE Trans Microwave Theory Tech 51(2):593–596

Dorigo M, Birattari M, Stützle T (2006) Ant colony optimization. IEEE Comput Intell Mag 1(4):28–39

Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. In: IEEE proceedings of the sixth international symposium on micro machine and human science, Nagoya, Japan, pp 39–43

Enoch S, Tayeb G, Sabouroux P, Guérin N, Vincent P (2002) A metamaterial for directive emission. Phys Rev Lett 89(21):213902(1)–213902(4)

Fang XS, Leung KW (2012) Linear-/circular-polarization designs of dual-/wide-band cylindrical dielectric resonator antennas. IEEE Trans Antennas Propag 60(6):2662–2671

Fang XS, Chow CK, Leung KW, Lim EH (2011) New single-/dual-mode design formulas of the rectangular dielectric resonator antenna using covariance matrix adaptation evolutionary strategy. IEEE Antennas Wirel Propag Lett 10:734–737

Feresidis AP, Goussetis G, Wang S, Vardaxoglou JC (2005) Artificial magnetic conductor surfaces and their application to low-profile high-gain planar antennas. IEEE Trans Antennas Propag 53(1):209–215

Fujimoto K, James JR (eds) (2001) Mobile antenna systems handbook, 2nd edn. Artech House, Boston

Gies D (2004) Particle swarm optimization: applications in electromagnetic design. M. Eng. thesis, UCLA, Los Angeles

Godi G, Sauleau R, Le Coq L, Thouroude D (2007) Design and optimization of three-dimensional integrated lens antennas with genetic algorithm. IEEE Trans Antennas Propag 55(3):770–775

Goudos SK, Siakavara K, Vafiadis EE, Sahalos JN (2010) Pareto optimal Yagi-Uda antenna design using multi-objective differential evolution. Prog Electromagn Res 105:231–251

Goudos SK, Siakavara K, Samaras T, Vafiadis EE, Sahalos JN (2011) Self-adaptive differential evolution applied to real-valued antenna and microwave design problems. IEEE Trans Antennas Propag 59(4):1286–1298

Gregory MD, Werner DH (2009) Optimization of broadband antenna elements in a periodic planar infinite array. In: Proceedings of the 2009 IEEE international symposium on antennas and propagation, Charleston, South Carolina

Gregory MD, Werner DH (2010) Ultrawideband aperiodic antenna arrays based on optimized raised power series representations. IEEE Trans Antennas Propag 58(3):756–764

Gregory MD, Werner DH (2011) Next generation electromagnetic optimization with the covariance matrix adaptation evolutionary strategy. In: Proceedings of the 2011 IEEE international symposium on antennas and propagation, Spokane, Washington

Gregory MD, Werner DH (2013) Multi-band and wideband antenna design using port substitution and CMA-ES. In: Proceedings of the 2013 IEEE international symposium on antennas and propagation and USNC/URSI national radio science meeting, Orlando

Gregory MD, Petko JS, Spence TG, Werner DH (2010a) Nature-inspired design techniques for ultra-wideband aperiodic antenna arrays. IEEE Antennas Propag Mag 52(3):28–45

Gregory MD, Bayraktar Z, Werner DH (2010b) Fast optimization of electromagnetics design problems through the CMA evolutionary strategy. In: Proceedings of the 2010 IEEE international symposium on antennas and propagation, Toronto, Ontario

Gregory MD, Bayraktar Z, Werner DH (2011) Fast optimization of electromagnetic design problems using the covariance matrix adaptation evolutionary strategy. IEEE Trans Antennas Propag 59(4):1275–1285

Gregory MD, Namin FA, Werner DH (2013) Exploiting rotational symmetry for the design of ultra-wideband planar phased array layouts. IEEE Trans Antennas Propag 61(1):176–184

Guérin N, Enoch S, Tayeb G, Sabouroux P, Vincent P, Legay H (2006) A metallic Fabry-Perot directive antenna. IEEE Trans Antennas Propag 54(1):220–224

Guha D, Gupta B, Antar YMM (2009) New pawn-shaped dielectric ring resonator loaded hybrid monopole antenna for improved ultrawide bandwidth. IEEE Antennas Wirel Propag Lett 8:1178–1181

Guo YX, Luk KM, Lee KF (2000) Dual-band slot-loaded short-circuited patch antenna. IEEE Electron Lett 36(4):289–291

Hansen N (2014) The CMA evolutionary strategy. Website https://www.lri.fr/~hansen/cmaesintro.html. Accessed 12 Nov 2014

Hansen N, Ostermeier A (2001) Completely derandomized self-adaptation in evolution strategies. Evol Comput 9(2):159–195

Hansen N, Müller SD, Koumoutsakos P (2003) Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). Evol Comput 11(1):1–18

Hao Z-C, Hong W, Chen J-X, Chen X-P, Wu K (2005) Single-compact super-wide bandpass substrate integrated waveguide (SIW) filters. IEEE Trans Microwave Theory Tech 53(9):2968–2977

Haupt RL (1994) Thinned arrays using genetic algorithms. IEEE Trans Antennas Propag 42(7):993–999

Haupt RL (1995) An introduction to genetic algorithms for electromagnetics. IEEE Antennas Propag Mag 37(2):7–15

Haupt RL, Werner DH (2007) Genetic algorithms in electromagnetics. Wiley, Hoboken

Holland JH (1973) Genetic algorithms and the optimal allocation of trials. SIAM J Comput 2(2):88–105

Hoorfar A (2007) Evolutionary programming in electromagnetic optimization: a review. IEEE Trans Antennas Propag 55(3):523–537

Hoorfar A, Chellapilla K (1998) Gain optimization of a multi-layer printed dipole array using evolutionary programming. In: Proceedings of the 1998 IEEE international symposium on antennas and propagation, Atlanta, Georgia, vol 1, pp 46–49

Jamnejad V, Hoorfar A (2004) Design of corrugated horn antennas by evolutionary optimization techniques. IEEE Antennas Wirel Propag Lett 3:276–279

Jiang ZH, Bossard JA, Wang X, Werner DH (2011a) Synthesizing metamaterials with angularly independent effective medium properties based on an anisotropic parameter retrieval technique coupled with a genetic algorithm. J Appl Phys 109:013515(1)–013515(11)

Jiang ZH, Gregory MD, Werner DH (2011b) A broadband monopole antenna enabled by an ultrathin anisotropic metamaterial coating. IEEE Antennas Wirel Propag Lett 10:1543–1546

Jiang ZH, Gregory MD, Werner DH (2012a) Broadband high directivity multibeam emission through transformation optics-enabled metamaterial lenses. IEEE Trans Antennas Propag 60(11):5063–5074

Jiang ZH, Wu Q, Werner DH (2012b) Demonstration of enhanced broadband unidirectional electromagnetic radiation enabled by a subwavelength profile leaky anisotropic zero-index metamaterial coating. Phys Rev B 86(12):125131(1)–125131(7)

Jin N, Rahmat-Samii Y (2005) Parallel particle swarm optimization and finite difference time-domain (PSO/FDTD) algorithm for multiband and wide-band patch antenna designs. IEEE Trans Antennas Propag 53(11):3459–3468

Jin N, Rahmat-Samii Y (2007) Advances in particle swarm optimization for antenna designs: real-number, binary, single-objective and multi-objective implementations. IEEE Trans Antennas Propag 55(3):556–567

Jin N, Rahmat-Samii Y (2008) Particle swarm optimization for antenna designs in engineering electromagnetics. J Artif Evol Appl 2008(9):1–10

Johnson JM, Rahmat-Samii Y (1997) Genetic algorithms in engineering electromagnetics. IEEE Antennas Propag Mag 39(4):7–21

Johnson JM, Rahmat-Samii Y (1999) Genetic algorithms and method of moments (GA/MOM) for the design of integrated antennas. IEEE Trans Antennas Propag 47(10):1606–1614

Jones EA, Joines WT (1997) Design of Yagi-Uda antennas using genetic algorithms. IEEE Trans Antennas Propag 45(9):1386–1392

Kan HK, Waterhouse RB (1999) Size reduction technique for shorted patches. IEEE Electron Lett 35(12):948–949

Karimkashi S, Kishk AA (2010) Invasive weed optimization and its features in electromagnetics. IEEE Trans Antennas Propag 58(4):1269–1278

Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of the ninth international conference on neural networks, Perth, Western Australia, vol 4, pp 1942–1948

Kennedy J, Eberhart R (2001) Swarm intelligence. Morgan Kaufmann/Academic, San Francisco

Kerkhoff AJ, Ling H (2007) Design of a band-notched planar monopole antenna using genetic algorithm optimization. IEEE Trans Antennas Propag 55(3):604–610

King HE, Wong JL (1972) An experimental study of a balun-fed open-sleeve dipole in front of a metallic reflector. IEEE Trans Antennas Propag 20:201–204

Kirkpatrick S, Gelatt CD Jr, Vecchi MP (1983) Optimization by simulated annealing. Science 220(4598):671–680

Kurup DG, Himdi M, Rydberg A (2003) Synthesis of uniform amplitude unequally spaced antenna arrays using the differential evolution algorithm. IEEE Trans Antennas Propag 51(9):2210–2217

Lee KF, Luk KM, Tong KF, Shwn SM, Huynh T, Lee RQ (1997) Experimental and simulation studies of the coaxially fed U-slot rectangular patch antenna. IEEE Proc Microw Antennas Propag 144(5):354–358

Leung KW, Fang XS, Pan YM, Lim EH, Luk KM, Chan HP (2013) Dual-function radiating glass for antennas and light covers–part II: dual-band glass dielectric resonator antennas. IEEE Trans Antennas Propag 61(2):587–597

Lewis A, Weis G, Randall M, Galehdar A, Thiel D (2009) Optimising efficiency and gain of small meander line RFID antennas using ant colony system. In: Proceedings of the IEEE congress on evolutionary computation, Trondheim, Norway, pp 1486–1492

Li R, DeJean G, Tentzeris MM, Laskar J (2004) Development and analysis of a folded shorted-patch antenna with reduced size. IEEE Trans Antennas Propag 52(2):555–562

Li W, Liu B, Zhao H (2014) The U-shaped structure in dual-band circularly polarized slot antenna design. IEEE Antennas Wirel Propag Lett 13:447–450

Liang J, Chiau CC, Chen X, Parini CG (2005) Study of a printed circular disc monopole antenna for UWB systems. IEEE Trans Antennas Propag 53(11):3500–3504

Lin C, Qing A, Feng Q (2010) Synthesis of unequally spaced antenna arrays by using differential evolution. IEEE Trans Antennas Propag 58(8):2553–2561

Liu L, Caloz C, Itoh T (2002) Dominant mode leaky-wave antenna with backfire-to-endfire scanning capability. IEEE Electron Lett 38(23):1414–1416

Liu B, Hong W, Zhang Y, Tang HJ, Yin X, Wu K (2007) Half mode substrate integrated waveguide 180° 3-dB directional couplers. IEEE Trans Microwave Theory Tech 55(12):2586–2592

Liu R, Ji C, Mock JJ, Chin JY, Cui TJ, Smith DR (2009) Broadband ground-plane cloak. Science 323(5912):366–369

Lizzi L, Viani F, Azaro R, Massa A (2007) Optimization of a spline-shaped UWB antenna by PSO. IEEE Antennas Wirel Propag Lett 6:182–185

Lizzi L, Viani F, Azaro R, Massa A (2008) A PSO-driven spline-based shaping approach for ultrawideband (UWB) antenna synthesis. IEEE Trans Antennas Propag 56(8):2613–2621

Lo YT, Lee SW (1988) Antenna handbook: theory, applications, and design. Van Nostrand Reinhold, New York

Long SA, Walton WD (1979) A dual-frequency stacked microstrip circular-disc antenna. IEEE Trans Antennas Propag 27(2):270–273

Lovat G, Burghignoli P, Capolino F, Jackson DR, Wilton DR (2006) Analysis of directive radiation from a line source in a metamaterial slab with low permittivity. IEEE Trans Antennas Propag 54(3):1017–1030

Maci S, Gentili GB, Piazzesi P, Salvador C (1995) Dual-band slot-loaded patch antenna. IEEE Proc Microw Antennas Propag 142(3):225–232

Martin SH, Martinez I, Turpin JP, Werner DH, Lier E, Bray MG (2014) The synthesis of wide- and multi-bandgap electromagnetic surfaces with finite size and nonuniform capacitive loading. IEEE Trans Microwave Theory Tech 62(9):1962–1972

Martínez-Fernández J, Gil JM, Zapata J (2007) Ultrawideband optimized profile monopole antenna by means of simulated annealing algorithm and the finite element method. IEEE Trans Antennas Propag 55(6):1826–1832

Massa A, Pastorino M, Randazzo A (2006) Optimization of the directivity of a monopulse antenna with a subarray weighting by a hybrid differential evolution method. IEEE Antennas Wirel Propag Lett 5:155–158

Mosallaei H, Rahmat-Samii Y (2001) Nonuniform Luneberg and two-shell lens antennas: radiation characteristics and design optimization. IEEE Trans Antennas Propag 49(1):60–69

Murino V, Trucco A, Regazzoni CS (1996) Synthesis of unequally spaced arrays by simulated annealing. IEEE Trans Signal Process 44(1):119–123

Ollikainen J, Fischer M, Vainikainen P (1999) Thin dual-resonant stacked shorted patch antenna for mobile communications. IEEE Electron Lett 35(6):437–438

Palandoken M, Grede A, Henke H (2009) Broadband microstrip antenna with left-handed metamaterials. IEEE Trans Antennas Propag 57(2):331–338

Pan YM, Zheng SY, Hu BJ (2014) Design of dual-band omnidirectional cylindrical dielectric resonator antenna. IEEE Antennas Wirel Propag Lett 12:710–713

Panduro MA, Brizuela CA, Balderas LI, Acosta DA (2009) A comparison of genetic algorithms, particle swarm optimization, and the differential evolution method for the design of scannable circular antenna arrays. Prog Electromagn Res B 13:171–186

Petko JS, Werner DH (2008) The Pareto optimization of ultrawideband polyfractal arrays. IEEE Trans Antennas Propag 56(1):97–107

Rahmat-Samii Y, Michielssen E (eds) (1999) Electromagnetic optimization by genetic algorithms. Wiley, New York

Rajo-Iglesias E, Quevedo-Teruel O (2007) Linear array synthesis using an ant-colony-optimization-based algorithm. IEEE Antennas Propag Mag 49(2):70–79

Robinson J, Rahmat-Samii Y (2004) Particle swarm optimization in electromagnetics. IEEE Trans Antennas Propag 52(2):397–407

Rocca P, Manica L, Stringari F, Massa A (2008) Ant colony optimisation for tree-searching-based synthesis of monopulse array antenna. IEEE Electron Lett 44(13):783–785

Rocca P, Oliveri G, Massa A (2011) Differential evolution as applied to electromagnetics. IEEE Antennas Propag Mag 53(1):38–49

Santarelli S, Yu T, Goldberg DE, Altshuler E, O'Donnell T, Southall H, Mailloux R (2006) Military antenna design using simple and competent genetic algorithms. J Math Comput Model 43:991–1022

Schurig D, Mock JJ, Smith DR (2006) Electric-field-coupled resonators for negative permittivity metamaterials. Appl Phys Lett 88(4):041109(1)–041109(3)

Shahvarpour A, Caloz C, Alvarez-Melcon A (2011) Broadband and low-beam squint leaky radiation from a uniaxially anisotropic grounded slab. Radio Sci 46:4006(1)–4006(13)

Spence TG, Werner DH (2006) A novel miniature broadband/multiband antenna based on an end-loaded planar open-sleeve dipole. IEEE Trans Antennas Propag 54(12):3614–3620

Spence TG, Werner DH (2008) Design of broadband planar arrays based on the optimization of aperiodic tilings. IEEE Trans Antennas Propag 56(1):76–86

Storn R, Price K (1997) Differential evolution – a simple and effective heuristic for global optimization over continuous spaces. J Global Optim 11(4):341–359

Sun Y, Chen ZN, Zhang Y, Chen H, See TSP (2012) Subwavelength substrate-integrated Fabry-Pérot cavity antennas using artificial magnetic conductor. IEEE Trans Antennas Propag 60(1):30–35

Thors B, Steyskal H, Holter H (2005) Broad-band fragmented aperture phased array element design using genetic algorithms. IEEE Trans Antennas Propag 53(10):3280–3287

Turpin JP, Massoud AT, Jiang ZH, Werner PL, Werner DH (2010) Conformal mappings to achieve simple material parameters for transformation optics devices. Opt Express 18(1):244–252

Villegas FJ, Cwik T, Rahmat-Samii Y, Manteghi M (2004) A parallel electromagnetic genetic-algorithm optimization (EGO) application for patch antenna design. IEEE Trans Antennas Propag 52(9):2424–2435

Volakis J (ed) (2007) Antenna engineering handbook, 4th edn. McGraw-Hill Professional, New York

Volakis JL, Chen C-C, Fujimoto K (2010) Small antennas: miniaturization techniques and applications, 1st edn. McGraw-Hill, New York

Wang KL, Yang KP (1998) Modified planar inverted F antenna. IEEE Electron Lett 34(1):7–8

Wang W, Gong S, Wang X, Jiang W (2010) Differential evolution algorithm and method of moments for the design of low-RCS antenna. IEEE Antennas Wirel Propag Lett 9:295–298

Weile DS, Michielssen E (1996) Integer coded Pareto genetic algorithm design of constrained antenna arrays. IEEE Electron Lett 32(19):1744–1745

Weile DS, Michielssen E (1997) Genetic algorithm optimization applied to electromagnetics: a review. IEEE Trans Antennas Propag 45(3):343–353

Werner DH, Werner PL, Church KH (2001) Genetically engineered multiband fractal antennas. IEEE Electron Lett 37(19):1150–1151

Werner PL, Bayraktar Z, Rybicki B, Werner DH, Schlager KJ, Linden D (2008) Stub-loaded long-wire monopoles optimized for high gain performance. IEEE Trans Antennas Propag 56(3):639–644

Wu K, Cheng YJ, Hong W (2012) Substrate-integrated millimeter-wave and terahertz antenna technology. Proc IEEE 100(7):2219–2232

Zhang Y, Chen ZN, Qing X, Hong W (2011) Wideband millimeter-wave substrate integrated waveguide slotted narrow-wall fed cavity antennas. IEEE Trans Antennas Propag 59(5):1488–1496

Zhang L, Jiao Y-C, Chen B, Wang Z-B (2012) Optimization of non-uniform circular arrays with covariance matrix adaptation evolutionary strategy. Prog Electromagn Res C 28:113–126

Zhang F, Jia W, Yao M (2013) Linear aperiodic array synthesis using differential evolution algorithm. IEEE Antennas Wirel Propag Lett 12:797–800

Ziolkowski RW (2004) Propagation in and scattering from a matched metamaterial having a zero index of refraction. Phys Rev E 70(10):046608(1)–046608(12)

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_18-1

Physical Bounds of Antennas

Mats Gustafsson*, Doruk Tayli and Marius Cismasu
Department of Electrical and Information Technology, Lund University, Lund, Sweden

Abstract

Design of small antennas is challenging because fundamental physics limits the antennas performance. Physical bounds provide basic restrictions on the antenna performance solely expressed in the available antenna design space. These bounds offer antenna designers a priori information about the feasibility of antenna designs and a figure of merit for different antenna designs. Here, an overview of physical bounds on antennas and the development from circumscribing spheres to arbitrary shaped regions and embedded antennas are presented. The underlying assumptions for the methods based on circuit models, mode expansions, forward scattering, and current optimization are illustrated and their pros and cons are discussed. The physical bounds are compared with numerical data for several antennas.

Keywords

Physical bounds; Limitations; Stored energy; Q-factor; Small antennas; Circuit models; Mode expansions; Forward scattering; Sum rules; Convex optimization

Introduction

Physical bounds provide information about the maximum achievable performance of antennas (Hansen 1981, 2006; Hansen and Collin 2011; Volakis et al. 2010; Skrivervik and Gustafsson 2012). Bounds are derived, in general, independently of antenna material and type. This makes the physical bounds a powerful tool for developing and analyzing all types of antennas. The bounds are expressed in parameters that describe the geometry and electromagnetic properties of the antennas. Moreover, the bounds themselves and their underlying theory provide insight to antenna design and antenna theory. The bounds offer a priori information of the feasibility of a specific antenna, they give a measure for the figure of merit of an antenna, and they can be used as stopping criteria in heuristic optimization algorithms and for the optimization of the antenna position in wireless terminals.

Antenna performance deteriorates with decreasing electrical size, i.e., physical size measured in wavelengths (Skrivervik et al. 2001; Best and Hanna 2010). The bounds are therefore particularly important for the design of electrically small antennas. There are several definitions of the electrical size of antennas. The classical definition is the radius, measured in wavelengths, of the smallest sphere circumscribing the antenna (Chu 1948; Hansen 1981, 2006; Hansen and Collin 2011; Volakis et al. 2010; Skrivervik and Gustafsson 2012); see also Fig. 1. This is most suitable for antennas with a spherical shape (Best 2004). A generalized measure of electrical size is used for antennas of arbitrary shape and antennas embedded in devices, e.g., mobile phones, tablets, and laptops; see Fig. 1. It should be pointed out that for antennas embedded in devices, the devices may be electrically large, whereas the antennas are electrically small.

*Email: mats.gustafsson@eit.lth.se

DOI 10.1007/978-981-4560-75-7_18-1

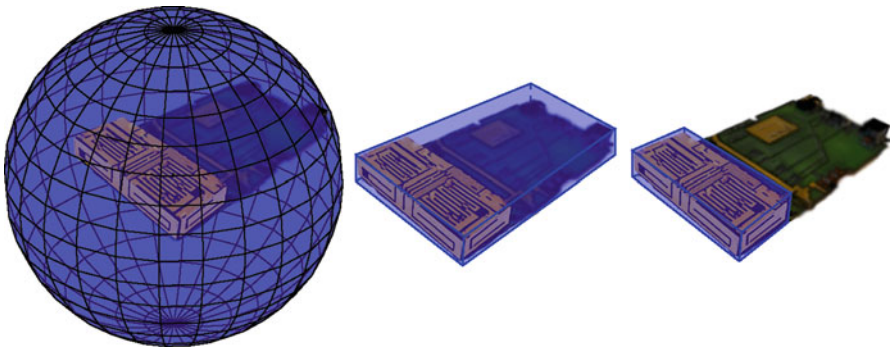


Fig. 1 Illustration of circumscribing geometries used for physical bounds. (*Left*) Circumscribing sphere introduced by Wheeler (1947) and Chu (1948) and used for spherical mode expansions. (*Middle*) Box circumscribing the antenna structure used in the forward scattering approach (Gustafsson et al. 2007a, 2009b, 2010). (*Right*) Box circumscribing the antenna region used in antenna current optimization (Gustafsson and Nordebo 2013)

In small antenna theory, matching bandwidth and efficiency are two performance parameters often considered for analysis (e.g., Wheeler 1947; Chu 1948; Harrington 1961; Collin and Rothschild 1964; Fante 1969; McLean 1996; Foltz and McLean 1999; Sten et al. 2001; Geyi 2003a; Karlsson 2004; Yaghjian and Best 2005; Gustafsson and Nordebo 2006; Thal 2006; Pozar 2009; Nyberg et al. 2010; Glazunov et al. 2011; Hansen et al. 2012). The radiation pattern and the polarization parameters are also of importance in some studies (Gustafsson et al. 2007a, 2009b; Gustafsson and Nordebo 2013). The bandwidth of an antenna depends on the threshold level of the reflection coefficient (or equivalently the standing wave ratio) (Pozar 1998; Volakis et al. 2010; Hansen and Collin 2011) and it is customarily expressed in terms of the corresponding Q-factor (or antenna Q). This latter parameter is defined as the ratio of the energy stored in the fields created by an antenna, to the energy dissipated by radiation and ohmic heating (Pozar 1998; Volakis et al. 2010). The Q-factor is inversely proportional to the fractional bandwidth (Yaghjian and Best 2005; Gustafsson and Nordebo 2006).

The concept of physical bounds for electrically small antennas was introduced by Wheeler (1947) and Chu (1948) at the end of the 1940s. Wheeler used two-lumped-element circuit models to estimate antenna bandwidth and practical efficiency of small antennas. The classical results by Chu (1948) express the lower bound on the Q-factor of antennas in terms of the antenna radius and wave number; see Fig. 1. Although these results are useful and provide a first step in analyzing small antennas, they underestimate the lower bound on the Q-factor of non-spherical antennas. More recently, the forward scattering sum rule presented in Gustafsson et al. (2007a, 2009b, 2010) expresses antenna performance and bounds in terms of the polarizability of the antenna structure or any circumscribing shape; see Fig. 1. These bounds are more realistic than those derived for spherical geometries and are beneficial for analyzing many small antennas that utilize the circumscribing geometry efficiently. On the other hand, the forward scattering sum-rule bounds are less useful for antennas that are embedded in or placed in the proximity of larger structures. Instead, these situations can be studied using antenna current optimization (Gustafsson and Nordebo 2013; Gustafsson et al. 2014b). In this technique the structure is decomposed, into an antenna region with controllable currents and a surrounding structure with induced currents.

The remainder of this paper is organized as follows. Section "Background and Overview" presents an overview of physical bounds. Section "Antenna Quantities for Physical Bounds" reviews the antenna quantities used in the physical bounds. Physical bounds for small antennas based on circuit models and mode expansions are discussed in section "Circuit Models and Mode Expansions," forward scattering in section "Forward Scattering Sum Rule," and current optimization in section "Antenna Current

DOI 10.1007/978-981-4560-75-7_18-1

Fig. 2 Physical bounds on antennas answer the *fundamental question: how good can an antenna be*? Bounds can be determined using circuit models (1947), mode expansions (1948), sum rules (2007), and current optimization (2012). Antenna current optimization is the most flexible method (yet) to obtain physical bounds appropriate for many practical antenna design situations

Optimization." Bandwidth bounds for array antennas are discussed in section "Periodic Array Antennas." The paper is concluded in section "Conclusions."

Background and Overview

Physical bounds on antennas have been analyzed in different ways since the work of Wheeler (1947) and Chu (1948), and some of them have been mentioned previously. An attempt to categorize the physical bound approaches and methods is made in Fig. 2, even though not all can be described and characterized. Four main techniques have been classified to derive the physical bounds. These are circuit models, mode expansions, forward scattering, and antenna current optimization.

Circuit models consisting of an inductor or capacitor and a radiation resistance were used by Wheeler (1947) to approximate the input impedance of antennas. Based on these circuits, he estimated the radiation power factor (related to the inverse of the Q-factor) and bandwidth of antennas. The results obtained by Wheeler, even though approximate, provide physical insight into the dependence of performance on the shape and material loading of antennas. Chu (1948) used circuit models of spherical modes to compute the stored energy and Q-factor for spherical geometries. The results obtained from circuit models have been generalized to mixed modes and nonmagnetic sources by Thal (2006, 2009).

Physical bounds on antennas can be obtained by expanding the fields in spherical modes, also referred to as spherical vector waves, outside a circumscribing sphere of the antenna structure. Chu (1948) used

DOI 10.1007/978-981-4560-75-7_18-1

this mode expansion together with circuit models to compute the minimum stored energy and to derive the Q and D/Q bounds for single spherical modes, where D is the directivity. He showed that the dipole modes have the lowest Q-factor and that the Q-factor of antennas can be reduced by letting the antenna excite a combination of electric (transverse magnetic TM) and magnetic (transverse electric TE) modes. Mode expansions have dominated the research on physical bounds since Chu's work (see e.g., Harrington 1961; Collin and Rothschild 1964; Fante 1969; McLean 1996; Foltz and McLean 1999; Sten et al. 2001; Geyi 2003a; Karlsson 2004; Kwon 2005; Yaghjian and Best 2005; Gustafsson and Nordebo 2006; Thal 2006; Pozar 2009; Nyberg et al. 2010; Glazunov et al. 2011; Davis et al. 2011; Hansen et al. 2012) and the historical expose in Volakis et al. (2010). The mode expansion approach to physical limitations provides simple analytic formulas for the lower bound on the Q-factor for single modes (Collin and Rothschild 1964; McLean 1996). In addition, this approach provides insight into the physics of combinations of different modes (e.g., TE + TM). Thal extended the mode expansions by restricting the electromagnetic fields to originate from electric current densities (Thal 2006).

Physical bounds were generalized to arbitrary shapes using the forward scattering sum rule in Gustafsson et al. (2007a, 2009b, 2010). In this approach, antennas are analyzed in receive mode. It was shown that the interaction over all frequencies between an incident electromagnetic plane wave and an antenna is related to the polarizability of the radiating structure. The bounds derived with this approach are valid for arbitrarily shaped antennas made of reciprocal, linear, and time-translational materials. These bounds are formulated in terms of the quotient between the directivity and Q-factor D/Q, and of the product of the realized gain and bandwidth. The forward scattering sum-rule bounds have been verified for several antennas with electrical sizes up to $ka \approx 1.5$, where k is the wave number and a is the radius of the smallest sphere circumscribing the antenna. Yaghjian and Stuart derived similar bounds on the Q-factor in the limit of small antennas $ka \ll 1$ using a different technique (Yaghjian and Stuart 2010). These latter bounds were generalized to electric and magnetic sources in Yaghjian et al. (2013).

Antenna current optimization can be used to derive physical bounds on antennas from optimal current distributions on the radiating structure (Gustafsson et al. 2012a; Gustafsson and Nordebo 2013). This approach is widely applicable to practical antenna design situations and has the potential to include many different requirements on, e.g., performance, size, etc. One of the challenges of this method is to express the antenna quantities involved in defining the parameters of interest, in terms of the current density on the antenna. For example, here the energy expressions introduced by Vandenbosch (Vandenbosch 2010) for current densities in free space (see also Geyi 2003b; Gustafsson and Jonsson 2015b; Gustafsson et al. 2014a; Harrington 1972) are used to express the Q-factor of antennas. These energy expressions are further used in minimization of the stored energy to determine current densities optimum in the sense of maximum gain-Q ratio for a fixed radiated field. These and many other antenna design situations result in convex optimization problems (Gustafsson and Nordebo 2013), which are easily solvable by common numerical methods (Grant and Boyd 2011).

There are alternative approaches to analyze electrically small or large antennas. A few of these approaches are briefly discussed in the following. Thiele et al. (2003) propose an approach to physical limitations based on the concepts of superdirectivity and visible and invisible spatial regions of the fields created by an antenna (Taylor 1955). A method based on the capacity of the free space communication channel is described in Nordebo et al. (2006) and Gustafsson and Nordebo (2007). Considerations about directivity limitations of antennas are presented in Kildal and Best (2008). Bandwidth limitations for infinite array antennas are analyzed in Doane et al. (2013), Jonsson et al. (2013), and Doane et al. (2014) using the results in Rozanov (2000) and Gustafsson and Sjöberg (2011); see section "Periodic Array Antennas".

DOI 10.1007/978-981-4560-75-7_18-1

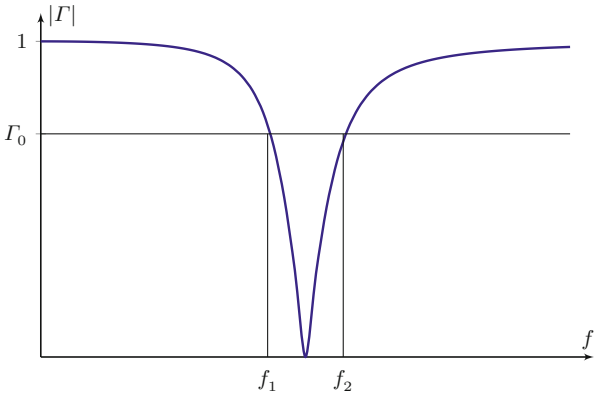


Fig. 3 Notations used in the definition of bandwidth and fractional bandwidth in Eq. 2 Magnitude of the reflection coefficient Γ as a function of the frequency f

Antenna Quantities for Physical Bounds

Antenna Parameters

Antenna parameters commonly used in the study of small antennas include the bandwidth, Q-factor, input impedance, reflection coefficient, etc. The parameters used in this section are introduced in the following.

The reflection coefficient, Γ, of an antenna is determined from its input impedance, $Z_a = R_a + jX_a$, as

$$\Gamma = \frac{Z_a - Z_0}{Z_a + Z_0} = \frac{R_a - Z_0 + jX_a}{R_a + Z_0 + jX_a}, \tag{1}$$

where Z_0 is a normalization impedance, usually the characteristic impedance of the transmission line that feeds the antenna. The mismatch depends on the difference between the real part and the commonly real-valued normalization impedance, $R_a - Z_0$, and the reactance, X_a.

The *bandwidth* of an antenna is defined as "the range of frequencies within which the performance of the antenna, with respect to some characteristic, conforms to a specified standard" by the IEEE in (IEEE145-1993). Usual performance metrics are matching, efficiency, radiation pattern, polarization, etc. In this chapter, the focus is on the matching bandwidth defined as the frequency interval $[f_1, f_2]$ for which the amplitude of the reflection coefficient $|\Gamma|$ is smaller than a given threshold Γ_0; see Fig. 3. The bandwidth of narrowband antennas is usually expressed as fractional bandwidth, B, defined as the bandwidth normalized with the center frequency, $f_0 = (f_1 + f_2)/2$, i.e.,

$$B = \frac{f_2 - f_1}{f_0} = \frac{\omega_2 - \omega_1}{\omega_0}, \tag{2}$$

where the angular frequency is $\omega = 2\pi f$.

The Q-factor for an antenna is defined as the ratio between the maximum of the stored electric, W_e, and magnetic, W_m, energies and the dissipated power $P_r + P_\Omega$ (Yaghjian and Best 2005), i.e.,

$$Q = \max\{Q_e, Q_m\} = \frac{2\omega \max\{W_e, W_m\}}{P_r + P_\Omega} = \eta \frac{2\omega \max\{W_e, W_m\}}{P_r}, \tag{3}$$

where the electric and magnetic Q-factors are $Q_e = 2\omega W_e/(P_r + P_\Omega)$ and $Q_m = 2\omega W_m/(P_r + P_\Omega)$, respectively, and the efficiency $\eta = P_r/(P_r + P_\Omega)$, where P_r denotes the radiated power and P_Ω the

DOI 10.1007/978-981-4560-75-7_18-1

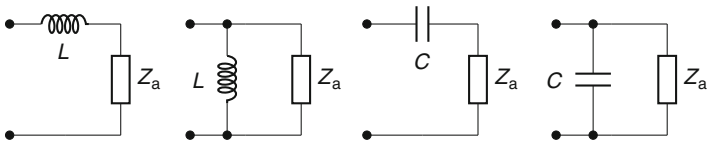


Fig. 4 Tuning of an antenna input impedance Z_a with series or parallel lumped capacitors or inductors

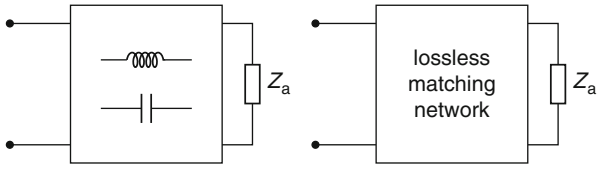


Fig. 5 *Right* – matching networks containing two lumped elements used for estimating the bandwidth potential (Rahola 2009). *Left* – black-box-type matching network, i.e., with arbitrary complexity and topology, that may reach some design requirements, e.g., reach close to the Bode-Fano limit (Fano 1950; Hansen 2006; Gustafsson and Nordebo 2006)

ohmic and dielectric power losses. The electric and magnetic Q-factors correspond to the stored energy in the capacitors and inductors, respectively, normalized with the dissipated power in the resistors for lumped circuit networks, see section "Antenna Tuning and Matching".

Antenna Tuning and Matching

Tuning is used to eliminate the reactance mismatch in Eq. 1. An antenna with input impedance $Z_a = R_a + jX_a$ is capacitive if $X_a < 0$ and inductive if $X_a > 0$. The simplest way to tune an antenna which is not self-resonant is by connecting a lumped element as in one of the situations depicted in Fig. 4. The tuning capacitance and inductance are chosen such that the tuned input impedance, Z_a, is purely resistive at the resonance frequency.

More complex matching networks can be used to increase the bandwidth of antennas, as illustrated in Fig. 5. The bandwidth potential is a concept that gives information about the bandwidth achievable with a two-lumped-element matching network (Rahola 2009). The upper bound on the fractional bandwidth for lossless matching networks is given by the Bode-Fano limit (Fano 1950; Hansen 2006; Gustafsson and Nordebo 2006)

$$B \leq \frac{27}{Q|\Gamma_{0,\,\mathrm{dB}}|}, \tag{4}$$

for an RLC antenna input impedance with the threshold level $\Gamma_{0,\mathrm{dB}} = 20 \log_{10} \Gamma_0$.

The bandwidth is inversely proportional to the Q-factor, i.e., a high Q-factor implies a narrow bandwidth. The precise proportionality is determined by the shape of the reflection coefficient, i.e., by its dependence on frequency, that is often quantified with the distribution of the resonances. The simplest case of a single resonance corresponds to series or parallel RLC circuits (see right inset of Fig. 6), where the fractional bandwidth for a single resonance is (Yaghjian and Best 2005)

$$B \approx \frac{2}{Q} \quad \frac{\Gamma_0}{\sqrt{1-\Gamma_0^2}} = \frac{2}{Q} \quad \text{for} \quad \Gamma_0 = 1/\sqrt{2}\,. \tag{5}$$

The reflection coefficients for single resonance RLC circuits with $Q = 5$, 10, and 40 are depicted in Fig. 6.

DOI 10.1007/978-981-4560-75-7_18-1

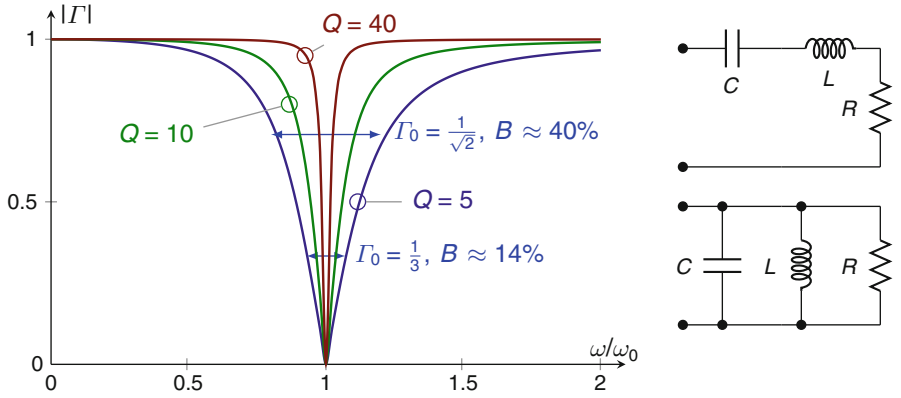


Fig. 6 Reflection coefficient, $|\Gamma|$, for RLC circuits (*inset right*) with $Q = \{5,\ 10,\ 40\}$. The corresponding fractional bandwidths for $\Gamma_0 = \{1/\sqrt{2}, 1/3\}$ are depicted

The estimate Eq. 5 is very accurate for $Q \gg 2$ for the RLC circuit. The special case of the half-power bandwidth $B \approx 2/Q$ predicts an infinite bandwidth for $Q = 1$. This suggests that the Q-factor is most useful for $Q \gg 1$. Reasonable accuracy in Eq. 5 is obtained for $Q > 5$ or $Q > 10$, in practice. The assumption of a single resonance is however essential for Eq. 5. The use of multiple resonances to increase the bandwidth (Hansen 2006) is described less accurately by this expression.

Differentiation of the input impedance offers an alternative method to estimate the Q-factor of antennas (Yaghjian and Best 2005; Gustafsson and Nordebo 2006; Kajfez and Wheless 1986):

$$Q_{z'} = \omega|\Gamma'| = \frac{\omega|Z'_{\mathrm{m}}|}{2R} = \frac{\omega|Z' + \mathrm{j}|X||}{2R} = \frac{\sqrt{(\omega R')^2 + (\omega X' + |X|)^2}}{2R}, \tag{6}$$

where Z_{m} denotes the input impedance tuned to resonance with a series capacitor or inductor, as in Fig. 4 (see also Gustafsson et al. 2014a). Equation 6 is exact for the RLC series circuit and often very accurate for antennas with $Q \gg 1$. However, this expression can underestimate the Q-factor for multiple resonances (Gustafsson and Nordebo 2006; Stuart et al. 2007; Gustafsson and Jonsson 2015a; Gustafsson et al. 2014a).

The $Q_{z'}$ estimate Eq. 6 can be interpreted as a local Padé approximation of the reflection coefficient (or input impedance) with a lumped-element circuit (Gustafsson and Nordebo 2006). This expression estimates the bandwidth Eq. 5 for $B \ll 1$ or similarly $\Gamma_0 \ll 1$. Accurate estimates of the fractional bandwidth Eq. 5 using $Q_{z'}$ require that the first-order derivative $|Z'_{\mathrm{m}}|$ (linear term) dominates over the second- and higher-order derivatives. This requirement is not always met and there are synthesized cases with $|Z'_{\mathrm{m}}| \approx 0$ where $Q_{z'}$ overestimates the fractional bandwidth using Eq. 5 (Gustafsson and Nordebo 2006; Gustafsson and Jonsson 2015a; Gustafsson et al. 2014a).

Stored Electric and Magnetic Energies

Physical bounds on antennas are often derived using the electromagnetic fields that surround the antenna (Harrington 1961; Collin and Rothschild 1964; Fante 1969; McLean 1996; Foltz and McLean 1999; Sten et al. 2001; Geyi 2003a; Karlsson 2004; Yaghjian and Best 2005; Thal 2006; Volakis et al. 2010). Consider an antenna confined to the region V as depicted in Fig. 7. The antenna current density is denoted $\boldsymbol{J}(\boldsymbol{r})$ and the radiated electric field is $\boldsymbol{E}(\boldsymbol{r})$ that simplifies to

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{\mathrm{e}^{-\mathrm{j}kr}}{r}\boldsymbol{F}(\hat{\boldsymbol{r}}) \quad \text{as} \quad r \to \infty\,, \tag{7}$$

DOI 10.1007/978-981-4560-75-7_18-1

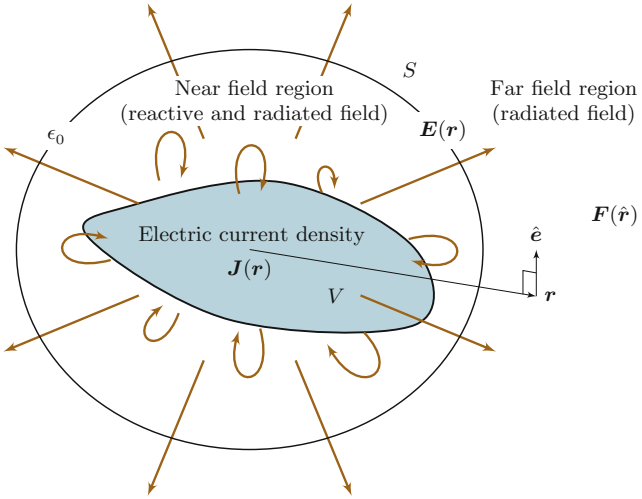


Fig. 7 Reactive and radiated fields from a current density $\boldsymbol{J}(\boldsymbol{r})$ in the volume V. Note that, in the far field, $\hat{\boldsymbol{e}}$ is perpendicular to $\boldsymbol{r}$

where $\boldsymbol{F}$ is the electric far field, $r = |\boldsymbol{r}|$, and $\hat{\boldsymbol{r}} = \boldsymbol{r}/r$. The time average electric and magnetic energy densities in free space are $w_\mathrm{e} = \epsilon_0|\boldsymbol{E}|^2/4$ and $w_\mathrm{m} = \mu_0|\boldsymbol{H}|^2/4$, respectively, where $\boldsymbol{H}(\boldsymbol{r})$ is the magnetic field. These energy densities resemble the time average energies stored in capacitors and inductors, $W_\mathrm{e} = C|V|^2/4$ and $W_\mathrm{m} = L|I|^2/4$, respectively. The energy in a region is obtained by integrating the energy density in that region. The $1/r$ decay of the field, see Eq. 7, implies that the total electromagnetic energy for antennas in free space is unbounded and hence dominated by the contributions from the radiated field far away from the antenna.

The stored electromagnetic energy is the part of the energy that is not radiated. This energy is confined to the near-field region around the antenna. The stored energy has been evaluated in the history of antenna technology by different methods. Chu (1948) and Thal (2006) used circuit models of the spherical vector waves to determine the stored energy for spherical regions. Synthesized lumped-circuit models give an alternative method to estimate the Q-factor from the input impedance of antennas (Gustafsson and Jonsson 2015a). Collin and Rothschild (1964) calculate the stored energy by subtraction of the energy density of the power flow giving

$$W_\mathrm{e}^{(\mathrm{P})} = \frac{\epsilon_0}{4}\int_{\mathbb{R}^3_\mathrm{r}} |\boldsymbol{E}(\boldsymbol{r})|^2 - \eta_0 \mathrm{Re}\{\boldsymbol{E}(\boldsymbol{r}) \times \boldsymbol{H}(\boldsymbol{r})^* \cdot \hat{\boldsymbol{r}}\}\ \mathrm{dV}, \tag{8}$$

for the electric stored energy, where the asterisk* denotes complex conjugation and $\mathbb{R}^3_\mathrm{r} = \{\boldsymbol{r} : \lim_{r0\to\infty}|\boldsymbol{r}| \le r_0\}$ denotes an infinitely large spherical volume (see also Fante 1969; McLean 1996; Geyi 2003a; Karlsson 2004; Yaghjian and Best 2005; Gustafsson and Jonsson 2015b). The magnetic energy is expressed in terms of a magnetic energy density analogous to Eq. 8. The integral expressions by Vandenbosch (2010) represent the stored energy as quadratic forms in terms of current densities. These expressions are particularly useful as radiated fields are generated by current densities on the antenna structure. Furthermore, these integral expressions are directly applicable to antenna current optimization (Gustafsson et al. 2012a; Gustafsson and Nordebo 2013). They are identical to subtraction of the energy density of the radiated field for many cases (Gustafsson and Jonsson 2015b) and reduce to the stored energy in Geyi (2003b) and Carpenter (1989) in the limit of small antennas. Here, the stored electric and magnetic energies given by the integral expressions (Vandenbosch 2010; Gustafsson and Jonsson 2015b)

DOI 10.1007/978-981-4560-75-7_18-1

$$W_{\mathrm{e}} = \frac{\eta_0}{4\omega} \int_V \int_V \nabla_1 \cdot \boldsymbol{J}(\boldsymbol{r}_1) \ \nabla_2 \cdot \boldsymbol{J}(\boldsymbol{r}_2)^* \frac{\cos\left(k|\boldsymbol{r}_1 - \boldsymbol{r}_2|\right)}{4\pi k|\boldsymbol{r}_1 - \boldsymbol{r}_2|} - \left(k^2 \boldsymbol{J}(\boldsymbol{r}_1) \cdot \boldsymbol{J}(\boldsymbol{r}_2)^* - \nabla_1 \cdot \boldsymbol{J}(\boldsymbol{r}_1) \ \nabla_2 \cdot \boldsymbol{J}(\boldsymbol{r}_2)^*\right) \frac{\sin\left(k|\boldsymbol{r}_1 - \boldsymbol{r}_2|\right)}{8\pi} \mathrm{dV}_1 \mathrm{dV}_2 \tag{9}$$

and

$$W_{\mathrm{m}} = \frac{\eta_0}{4\omega} \int_V \int_V k^2 \boldsymbol{J}(\boldsymbol{r}_1) \cdot \boldsymbol{J}(\boldsymbol{r}_2)^* \frac{\cos\left(k|\boldsymbol{r}_1 - \boldsymbol{r}_2|\right)}{4\pi k|\boldsymbol{r}_1 - \boldsymbol{r}_2|} - \left(k^2 \boldsymbol{J}(\boldsymbol{r}_1) \cdot \boldsymbol{J}(\boldsymbol{r}_2)^* - \nabla_1 \cdot \boldsymbol{J}(\boldsymbol{r}_1) \ \nabla_2 \cdot \boldsymbol{J}(\boldsymbol{r}_2)^*\right) \frac{\sin\left(k|\boldsymbol{r}_1 - \boldsymbol{r}_2|\right)}{8\pi} \mathrm{dV}_1 \mathrm{dV}_2, \tag{10}$$

respectively, are used, where it is noted that $\eta_0/\omega = \mu_0/k$. The corresponding radiated power is (Geyi 2011; Vandenbosch 2010; Gustafsson and Jonsson 2015b)

$$P_{\mathrm{r}} = \frac{\eta_0}{2} \int_V \int_V \left(k^2 \boldsymbol{J}(\boldsymbol{r}_1) \cdot \boldsymbol{J}(\boldsymbol{r}_2)^* - \nabla_1 \cdot \boldsymbol{J}(\boldsymbol{r}_1) \ \nabla_2 \cdot \boldsymbol{J}(\boldsymbol{r}_2)^*\right) \frac{\sin\left(k|\boldsymbol{r}_1 - \boldsymbol{r}_2|\right)}{4\pi k|\boldsymbol{r}_1 - \boldsymbol{r}_2|} \mathrm{dV}_1 \mathrm{dV}_2. \tag{11}$$

These energy expressions resemble the impedance matrices obtained from the electric field integral equation (EFIE) (Harrington 1972, 1975; Gustafsson et al. 2014a). The stored energies are also generalized to electric and magnetic currents (Jonsson and Gustafsson 2015) and lossy media (Gustafsson et al. 2014a).

Sum Rules

Sum rules have been used to derive physical bounds on electromagnetic systems such as: matching networks (Fano 1950), radar absorbers and array antennas (Rozanov 2000; Doane et al. 2013; Jonsson et al. 2013), antennas (Gustafsson et al. 2007a, 2009b; Gustafsson 2010c), scattering (Sohl et al. 2007; Gustafsson 2010b), high-impedance surfaces (Gustafsson and Sjöberg 2011), and metamaterials (Sohl et al. 2008; Gustafsson and Sjöberg 2010; Gustafsson et al. 2012b). The sum rules are integral identities that often relate the parameter of interest integrated over all frequencies with some low-frequency quantity.

The sum rules presented here are based on integral identities for Herglotz functions (Nussenzveig 1972; Bernland et al. 2011b) or similarly positive real (PR) functions (Zemanian 1987). Positive real functions (Zemanian 1987), $Z(s)$, are analytic and $\mathrm{Re}\{Z(s)\} \geq 0$ for $\mathrm{Re}\, s > 0$. They are often found in linear, passive, and causal systems (Bernland et al. 2011b; Zemanian 1987). The identities

$$\frac{1}{\pi} \int_{\mathbb{R}} \frac{\mathrm{Re}\, Z(\mathrm{j}\omega)}{\omega^2} \, \mathrm{d}\omega \stackrel{\mathrm{def}}{=\!=} \lim_{\varepsilon \to 0^+} \lim_{y \to 0^+} \frac{1}{\pi} \int_{-1/\varepsilon}^{1/\varepsilon} \frac{\mathrm{Re}\, Z(\mathrm{j}\omega + y)}{\omega^2} \mathrm{d}\omega = a_1 - b_1 \leq a_1 \tag{12}$$

are valid for all PR functions having the asymptotic expansions

$$Z(s) = a_1 s + \mathrm{o}(s) \quad \text{as} \quad s \hat{\to} 0 \quad \text{and} \quad Z(s) = b_1 s + \mathrm{o}(s) \quad \text{as} \quad s \hat{\to} \infty, \tag{13}$$

DOI 10.1007/978-981-4560-75-7_18-1

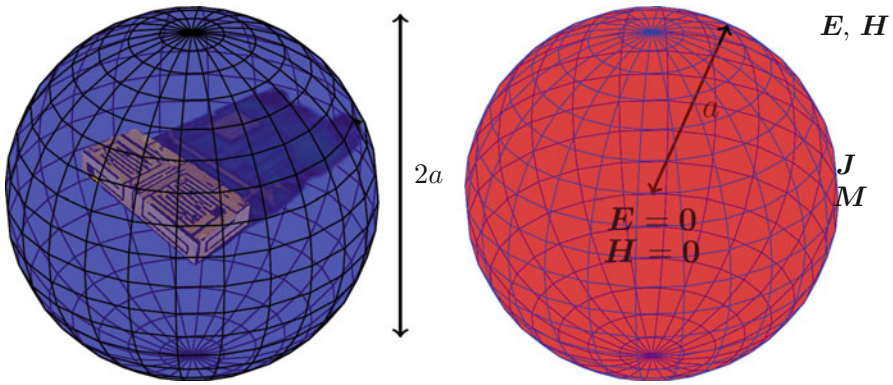


Fig. 8 (*Left*) Sphere with radius a that circumscribes the antenna. (*Right*) The Chu model (Chu 1948) with vanishing field in the interior of the sphere or similar electric $\boldsymbol{J}$ and magnetic $\boldsymbol{M}$ surface currents

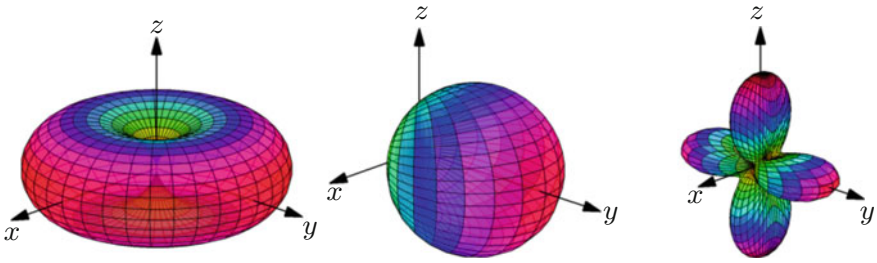


Fig. 9 Radiation patterns of spherical modes. (*Left*) Electric, $TM_{0,1}$, and magnetic, $TE_{0,1}$, dipole modes. (*Middle*) Huygens source $TM_{0,1}$ + $TE_{1,1}$. (*Right*) Quadrupole-mode $TM_{1,2}$

where $\hat{\to}$ means limits in some sector $0 < \alpha < |\arg s| < \pi/2 - \alpha$ and the principal value integral Eq. 12 is interpreted as a limit from the complex valued right half plane; see Bernland et al. (2011b) for details.

Circuit Models and Mode Expansions

Wheeler used circuit models to analyze small antennas (Wheeler 1947). The electric antenna consists of a lumped capacitance in parallel with a radiation resistance. The corresponding magnetic antenna is a lumped inductance in series with a radiation resistance. These results provide intuition and rules of thumb.

Chu (1948) determined the minimum stored energy around an antenna in three steps:

1. Circumscribe the antenna by a sphere with radius a; see Fig. 8.
2. Expand the radiated electromagnetic field outside the circumscribing sphere in spherical modes (Van Bladel 2007; Harrington 1961; Hansen 1988; Thal 1978); see Fig. 9.
3. Compute the stored energy outside the sphere from a lumped circuit representation of the spherical modes (Thal 1978).

The lumped-element circuit models of the electric ($TM_{m,1}$) and magnetic dipole ($TE_{m,1}$) modes (Chu 1948; Thal 1978; McLean 1996) are depicted in Fig. 10. The electric and magnetic cases having identical order (n-index) are dual such that the corresponding Q-factors are identical. The electric dipole has the stored electric and magnetic energies and radiated power

DOI 10.1007/978-981-4560-75-7_18-1

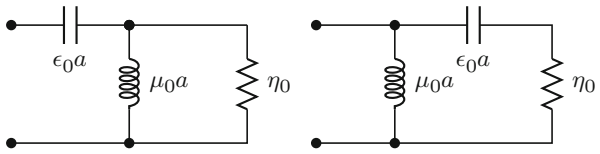


Fig. 10 Lumped-element circuit models of the electric ($TM_{m,1}$) and magnetic dipole ($TE_{m,1}$) modes (Chu 1948; Harrington 1961; Thal 1978)

$$W_{\mathrm{e}} = \frac{\eta_0 |I|^2}{4\omega} \frac{1}{ka}, \quad W_{\mathrm{m}} = \frac{\eta_0 |I|^2}{4\omega} \frac{ka}{1+k^2a^2}, \quad \text{and} \quad P_{\mathrm{r}} = \frac{\eta_0 |I|^2}{2} \frac{k^2a^2}{1+k^2a^2}, \tag{14}$$

respectively, for a current source I. The electric energy dominates over the magnetic energy, $W_{\mathrm{e}} > W_{\mathrm{m}}$, i.e., the electric dipole mode is capacitive. The stored energy and radiated power give the Chu bound for single mode antennas, stating that the Q-factor of a single mode lossless antenna circumscribed by a sphere with radius a satisfies

$$Q \geq Q_{\mathrm{Chu}} = \frac{2\omega \max\{W_{\mathrm{e}}, W_{\mathrm{m}}\}}{P_{\mathrm{r}}} = \frac{2\omega W_{\mathrm{e}}}{P_{\mathrm{r}}} = \frac{1}{(ka)^3} + \frac{1}{ka}. \tag{15}$$

The inequality $Q \geq Q_{\mathrm{Chu}}$ is interpreted as the contribution to the Q-factor from the stored energy within the spherical volume. This means that antenna designs reaching the Chu lower bound should have negligible electric field inside the sphere (Chu 1948).

Circuit models and closed-form expressions are convenient for dipole modes. However, these methods become increasingly complicated for higher-order modes; see also Fig. 9. The input impedance of the circuits modeling spherical modes offers an approximation of the Q-factor. Chu approximated the Q by differentiation of the reactance (Chu 1948) to show that $Q_n \sim (ka)^{-2n-1}$ as $ka \to 0$ for modes of order n (see also Gustafsson and Nordebo 2006) for the corresponding results from the differentiation of the input impedance (Eq. 6). Chu also discussed the case of mixed TM and TE modes, e.g., the Huygens source in Fig. 9, and showed that mixed modes can lower the bound on Q by a factor of two for $ka \ll 1$ (Chu 1948; Harrington 1961) by, e.g., using an inductive TE mode instead of a lumped inductor as the tuning element in Fig. 4. The resulting Q-factor is bounded as

$$Q \geq \frac{1}{2(ka)^3} + \frac{1}{ka} \tag{16}$$

for larger ka (McLean 1996; Pozar 2009).

Mode expansions have dominated the research on physical bounds since these expansions were introduced by Chu (see, e.g., Harrington 1961; Collin and Rothschild 1964; Fante 1969; McLean 1996; Foltz and McLean 1999; Sten et al. 2001; Geyi 2003a; Karlsson 2004; Thal 2006; Nyberg et al. 2010; Yaghjian and Best 2005; Gustafsson and Nordebo 2006; Pozar 2009 and the historical expose in Volakis et al. (2010)). In particular Collin and Rothschild (Collin and Rothschild 1964) used spherical mode expansions and analytic evaluation of Eq. 8 to derive closed-form expressions of the Q-factor for spherical modes of arbitrary order, e.g.,

$$Q_1 = \frac{1}{(ka)^3} + \frac{1}{ka}, \quad Q_2 = \frac{18}{(ka)^5} + \frac{6}{(ka)^3} + \frac{3}{ka}, \quad \text{and} \quad Q_n \sim \frac{1}{(ka)^{2n+1}}, \tag{17}$$

DOI 10.1007/978-981-4560-75-7_18-1

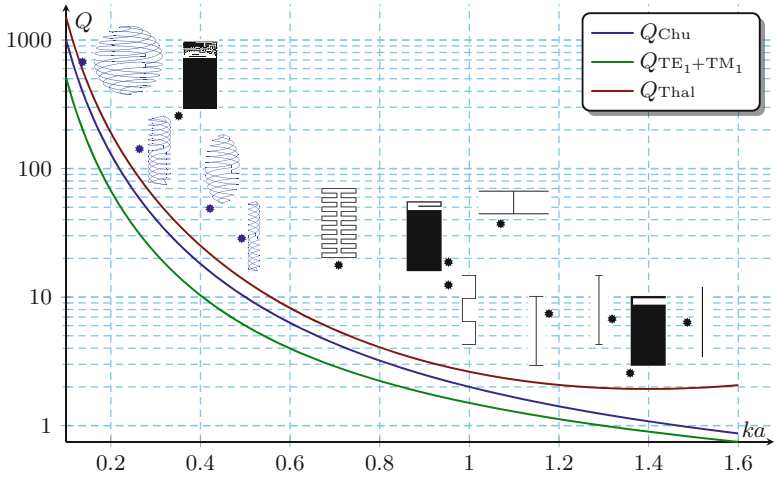


Fig. 11 The Chu Eq. 15, Thal Eq. 18, and mixed-mode bounds on Q compared with numerical estimates using $Q_{Z'_{\mathrm{in}}}$ in Eq. 6; see Tables 1 and 2 (Altair Development 2014). Wire antennas (*blue*) and planar antennas (*black*) are depicted. The different types of dipoles (*black*) are constructed with planar strips

where the rapid increase in the Q-factor for small antennas $ka \ll 1$ and higher-order modes $n > 1$ are noted.

Equivalent (surface) currents (Van Bladel 2007) offer a powerful interpretation of physical bounds. The Chu bound in Eq. 15 is derived under the assumption of negligible stored energy in the interior of a sphere. This requires electric, $\boldsymbol{J}$, and magnetic, $\boldsymbol{M}$, surface currents to exist on this sphere in free space; see Fig. 8. Antennas without magnetic currents (or magnetic material) have internal fields and hence an internal stored energy. Thal (2006) extended the Chu bound to the case of only electric surface currents by adding the energy stored in the interior of a sphere. The resulting bound for the electric dipole mode is

$$Q \geq \frac{3}{2(ka)^3} = \frac{3}{2} Q_{\mathrm{Chu}} \quad \text{for } ka \ll 1 \tag{18}$$

(see also Hansen and Collin 2009; Hansen et al. 2012). The closed-form expressions for arbitrary spherical modes are (Gustafsson and Jonsson 2015b)

$$Q^{(\mathrm{P})}_{\tau n,\mathrm{e}}(\kappa) = \kappa - \frac{\left(\kappa \mathrm{R}^{(1)}_{\tau n}(\kappa) \mathrm{R}^{(2)}_{\tau n}(\kappa)\right)'}{2\left(R^{(1)}_{\tau n}(\kappa)\right)^2} \quad \text{and} \quad Q^{(\mathrm{P})}_{\tau n,\mathrm{m}} = Q^{(\mathrm{P})}_{\tau n,\mathrm{e}}(\kappa) - \frac{\mathrm{R}^{(2)}_{\tau n}(\kappa)}{\mathrm{R}^{(1)}_{\tau n}(\kappa)} \tag{19}$$

for the electric and magnetic Q-factors, respectively, with $\kappa = ka$. Here the expressions for the TE ($\tau = 1$) and TM ($\tau = 2$) modes are written in identical forms by using radial functions (Hansen 1988); see Appendix A "Radial Functions."

The Chu (Eq. 15), Thal (Eq. 18), and mixed mode bounds are compared with several antennas in Fig. 11 (see also Sievenpiper et al. 2012). All antennas have Q-factors above the bounds, as expected. The folded spherical helix antenna fills the spherical surface and performs close to the Thal bound (Eq. 18), while the dipole and planar antennas do not utilize the spherical volume as efficiently as spherical helices. This implies that the former have Q-factors closer to the bounds than the latter. On the other hand, for the cylindrical structures, cylindrical helix and spheroidal helix, the antenna performance is between the performance of spherical and planar structures. It should be noted that increasing the thickness of the wire and strip antennas decreases their Q-factors. There are several suggestions for antennas with Q-factors

DOI 10.1007/978-981-4560-75-7_18-1

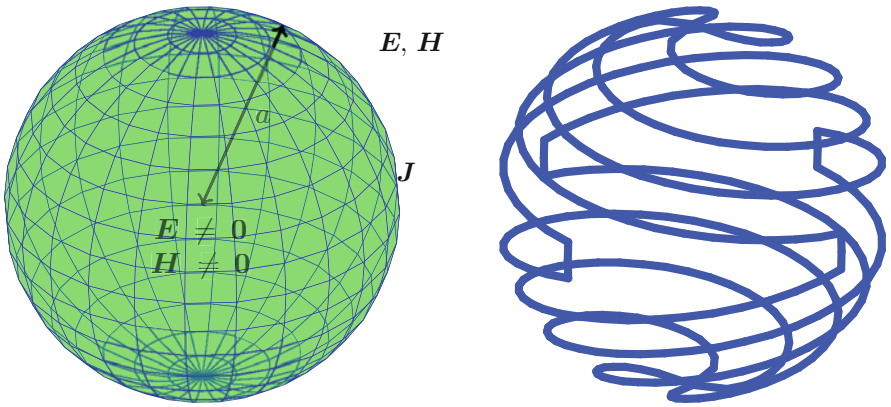


Fig. 12 (*Left*) The Thal model (Thal 2006) with an electromagnetic field in the interior of the sphere and electric surface currents $\boldsymbol{J}$. (*Right*) The folded spherical helix (Best 2004) with electric surface currents and Q-factor $Q \approx 1.5Q_{\text{Chu}}$

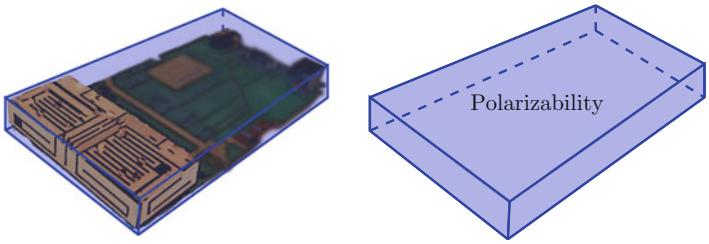


Fig. 13 (*Left*) Circumscribing box for an antenna. (*Right*) Polarizability of the circumscribing structure to determine the physical bounds Eq. 23

approaching the Chu lower bound or corresponding bounds for TE and mixed-mode cases. The folded spherical helix (see Fig. 11) investigated by Best (2004) has a Q-factor close to the Thal bound (see also Adams et al. 2011). Kim proposed a few designs reaching close to the physical bounds on Q (Kim 2012; Kim et al. 2010; Figs. 12 and 13).

Most antennas are not spherical and, as seen in Fig. 11, their Q-factors are far from the bound. This has encouraged researchers to extend the mode expansions to non-spherical regions. Collin and Rothschild (1964) used cylindrical waves to derive bounds for infinite cylinders and Foltz and McLean (1999) and Sten et al. (2001) used expansions in spheroidal coordinates to derive bounds for antennas confined to spheroidal volumes. However, it turns out to be difficult to extend the results from spherical regions using mode expansions. An alternative approach based on sum rules for spherical waves is investigated in Bernland et al. (2011a) and Bernland (2012), where the bandwidth is shown to be related to the polarizability. The case with spherical regions in the vicinity of a ground plane is analyzed in Sten et al. (2001). There are also extensions to antennas embedded in lossy media (Karlsson 2004).

Forward Scattering Sum Rule

The forward scattering bounds in Gustafsson et al. (2007a, 2009b) and Derneryd et al. (2009) are solely based on the assumptions of linearity, time-translational invariance, causality, and reciprocity. These are generally accepted assumptions in the antenna community and valid for a large class of antennas, e.g., antennas with metallic and dielectric components, antennas fed by a single transmission line, impedance matched with a matching network, etc. Some of the main advantages of these forward scattering bounds

DOI 10.1007/978-981-4560-75-7_18-1

are that they are simple to use, provide physical insight into small antennas given by expressions in terms of polarizability dyadics, hold for arbitrary enclosing geometries, and are formulated in terms of the realized gain (the parameter quantifying the system gain) and bandwidth or the directivity and Q-factor.

Physical bounds on arbitrarily shaped antennas were introduced in Gustafsson et al. (2007a) and further developed in Gustafsson et al. (2009b), Derneryd et al. (2009), Gustafsson et al. (2010), Sohl and Gustafsson (2008), Gustafsson and Sohl (2009), and Gustafsson (2011). These results are based on the forward scattering sum rule (Sohl et al. 2007; Gustafsson 2010b) that states that the all-spectrum interaction between objects and electromagnetic fields is proportional to the electro-, $\boldsymbol{\gamma}_{\mathrm{e}}$, and magneto-static, $\boldsymbol{\gamma}_{\mathrm{m}}$, polarizability dyadics, i.e.,

$$\frac{1}{\pi}\int_{\mathbb{R}} \frac{\sigma_{\mathrm{a}}(k)+\sigma_{\mathrm{s}}(k)}{k^2}\,\mathrm{d}k = \hat{\boldsymbol{e}}^{*}\cdot\boldsymbol{\gamma}_{\mathrm{e}}\cdot\hat{\boldsymbol{e}} + (\hat{\boldsymbol{r}}\times\hat{\boldsymbol{e}}^{*})\cdot\boldsymbol{\gamma}_{\mathrm{m}}\cdot(\hat{\boldsymbol{r}}\times\hat{\boldsymbol{e}}) \tag{20}$$

for all objects composed of linear, passive, and time-translational invariant media (Sohl et al. 2007; Gustafsson and Sohl 2009; Gustafsson 2010b). Here, σ_{a} and σ_{s} are the absorption and scattering cross sections, respectively. This identity was derived for dielectric spheroids in Purcell (1969) and generalized to arbitrary objects and polarizations in Sohl et al. (2007), Gustafsson and Sohl (2009), and Gustafsson (2010b), see also Eq. 12.

An antenna identity is obtained from the forward scattering sum rule by using reciprocity and the relation between the effective aperture (or absorption cross section) and partial directivity for lossless antennas:

$$\sigma_{\mathrm{a}}(k,\hat{\boldsymbol{r}},\hat{\boldsymbol{e}}) = \frac{\pi}{k^2}\Big(1-|\Gamma(k)|^2\Big)D(k,-\hat{\boldsymbol{r}},\hat{\boldsymbol{e}}). \tag{21}$$

Combining Eqs. 20 and 21 gives

$$\int_0^{\infty} \frac{\Big(1-|\Gamma(k)|^2\Big)D(k,\hat{\boldsymbol{r}},\hat{\boldsymbol{e}})}{k^4}\,\mathrm{d}k = \frac{\eta}{2}(\hat{\boldsymbol{e}}\cdot\boldsymbol{\gamma}_{e}\cdot\hat{\boldsymbol{e}} + (\hat{\boldsymbol{r}}\times\hat{\boldsymbol{e}})\cdot\boldsymbol{\gamma}_{\mathrm{m}}\cdot(\hat{\boldsymbol{r}}\times\hat{\boldsymbol{e}})), \tag{22}$$

for the linear polarization $\hat{\boldsymbol{e}}$, where the generalized (or all spectrum) absorption efficiency η (Gustafsson et al. 2007a, 2009b; Gustafsson 2010b) is introduced. This is further transformed to a bound on the D/Q (directivity bandwidth product) expressed in terms of the high-contrast polarizability dyadic, $\boldsymbol{\gamma}_{\infty} \geq \boldsymbol{\gamma}_{\mathrm{e}}$ by assuming a resonance model (Gustafsson et al. 2007a, 2009b) and

$$\frac{D}{Q} \leq \frac{\eta k^3}{2\pi}\hat{\boldsymbol{e}}\cdot\boldsymbol{\gamma}_{\infty}\cdot\hat{\boldsymbol{e}} \leq \frac{\eta k^3}{2\pi}\max\operatorname{eig}\boldsymbol{\gamma}_{\infty} \tag{23}$$

for nonmagnetic media, $\boldsymbol{\gamma}_{\mathrm{m}} = \mathbf{0}$ (see Gustafsson et al. 2007a, 2009b for the case of electric and magnetic media). The lower bound on the Q-factor

$$Q \geq \frac{6\pi}{k^3 \max\operatorname{eig}\ \boldsymbol{\gamma}_{\infty}} \tag{24}$$

is based on the fact that small electric dipoles have directivity $D = 3/2$ and generalized absorption efficiency $\eta \leq 1/2$ (Gustafsson et al. 2010). Results similar to those introduced in this paragraph are

derived by Yaghjian and Stuart (2010) using antenna current optimization (Gustafsson et al. 2012a) (see also Vandenbosch 2011; Thal 2012; Yaghjian et al. 2013). This bound is identical to that derived by Thal (2006) for small spherical structures in Eq. 18 with electric currents radiating as an electric dipole, i.e.,

$$Q \geq \frac{3}{2k^3a^3} = \frac{3}{2}Q_{\text{Chu}} \quad \text{for} \quad ka \ll 1\,, \tag{25}$$

where the high-contrast polarizability dyadic $\boldsymbol{\gamma}_\infty = 4\pi a^3\mathbf{I}$ for a sphere with the radius a is used.

The computed bounds Eqs. 23 and 24 for spheroids (Gustafsson et al. 2007a), cylinders (Gustafsson et al. 2009b), planar rectangles (Gustafsson et al. 2009b), and planar ellipses (Gustafsson 2011) are depicted in Fig. 14. The Chu (Eq. 15) and Thal (Eq. 18) bounds for $ka \ll 1$ are also included for comparison. The results illustrate how the bound depends on the shape and polarization of the electric field. The spheroid simplifies to a sphere for $\ell_1 = \ell_2$ with the bound $D/(Qk^3a^3) = 1$ that is identical to the result in Eqs. 18 and 25 by Thal (2006) using the directivity $D = 3/2$. Several antennas are compared with the bound in Fig. 15 (see also Gustafsson et al. 2009b, 2010; Best 2009; Shahpari et al. 2014). The comparisons show that many antennas perform close to the bounds.

The forward scattering identity in Eq. 20 and bound in Eq. 23 show that the antenna performance is proportional to the polarizability of the antenna structure. Polarizability quantifies the charge-separation properties of a structure, i.e., the induced dipole moment is proportional to the polarizability, $\boldsymbol{p} = \epsilon_0\boldsymbol{\gamma}_{\text{e}} \cdot \boldsymbol{E}$. For example, large metallic regions at the extremities of an antenna increase the maximum achievable D/Q-performance of that antenna, for linear polarization in the direction of the antenna extremities. This is illustrated for a planar capped dipole structure in Fig. 16 (see also Gustafsson et al. 2009a). The polarizability for a planar square metallic rectangle in the xy-plane with side lengths ℓ is $\boldsymbol{\gamma}_\infty \approx 1.04\ell^3(\hat{\boldsymbol{x}}\hat{\boldsymbol{x}} + \hat{\boldsymbol{y}}\hat{\boldsymbol{y}})$. The polarizability for the capped dipole, $\boldsymbol{\gamma}_\infty \approx \ell^3(0.51\hat{\boldsymbol{x}}\hat{\boldsymbol{x}} + 0.93\hat{\boldsymbol{y}}\hat{\boldsymbol{y}})$, is lower as some metal is removed. The polarizability is higher in the $\hat{\boldsymbol{y}}$ -direction than in the $\hat{\boldsymbol{x}}$-direction as the metal strips at the edges allow a large charge separation in the $\hat{\boldsymbol{y}}$-direction. This is further illustrated in the right part of Fig. 16, where the polarizability for planar capped dipoles with area A is depicted. Note that the capped dipoles have zero volume and that the polarizability reduces with approximately a factor of 3, whereas the area decreases a factor of 10^4. The antenna performance for an electric dipole antenna is hence not simply related to the volume or area but to the structures' ability to separate charge.

Closed-form expressions are available for the polarizability dyadics of structures such as spheroids, planar ellipses, half spheres, etc. (Jones 1945; Sohl et al. 2007; Kleinman and Senior 1986; Sihvola et al. 2004). An illustrative example is given by the polarizability of a dielectric sphere with radius a (Jackson 1999; Van Bladel 2007; Cheng 1989; Griffiths 1999), i.e.,

$$\boldsymbol{\gamma}_{\text{e}} = 4\pi a^3\frac{\epsilon_{\text{r}} - 1}{\epsilon_{\text{r}} + 2}\mathbf{I} \leq 4\pi a^3\mathbf{I} = \boldsymbol{\gamma}_\infty. \tag{26}$$

This simple example illustrates two important properties of the polarizability dyadic with implications for antenna design, i.e., the polarizability is increasing in the relative permittivity ϵ_{r} and the size a (Jones 1985; Sjöberg 2009). For other structures the polarizability is calculated numerically using the method of moments (Gustafsson 2011; Helsing and Perfekt 2013; Shahpari et al. 2013) or the finite element method (Sohl et al. 2007). The polarizability and the associated bound on D/Q are approximated by rational functions for cylinders and planar rectangles in Gustafsson (2011); see also the MATLAB code (Gustafsson 2010a).

The forward scattering bounds are generalized to electric and magnetic media by inclusion of the magnetic polarizability $\boldsymbol{\gamma}_{\text{m}}$ in Eqs. 23 and 24 (see Gustafsson et al. 2007a, 2009b). The generalization to

DOI 10.1007/978-981-4560-75-7_18-1

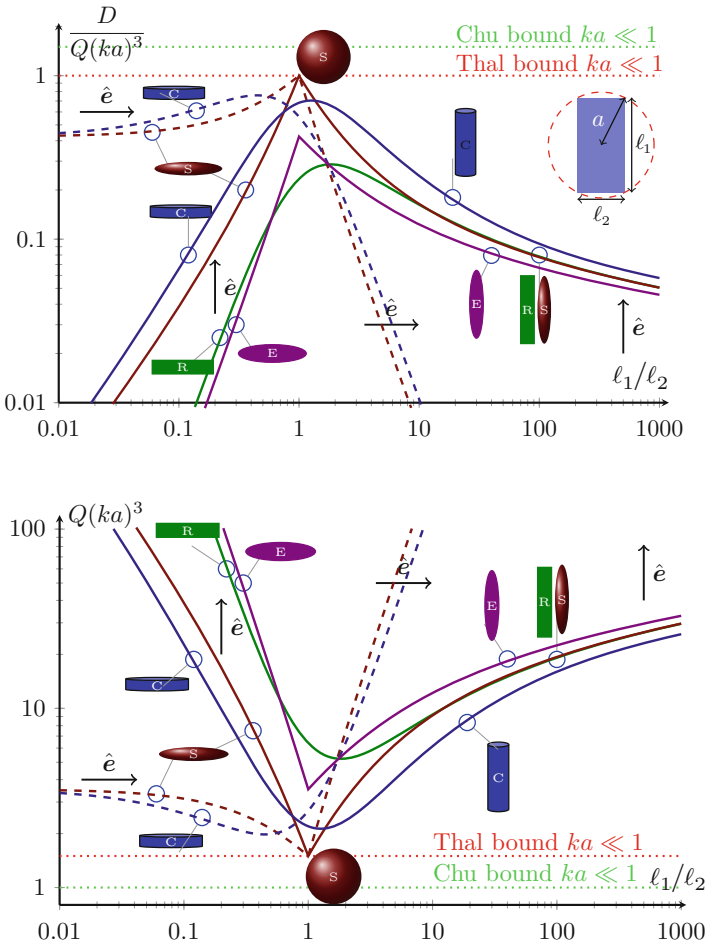


Fig. 14 Forward scattering bound on D/Q (*top*) and Q (*bottom*) normalized with $(ka)^3$ for nonmagnetic spheroids, cylinders, planar rectangles, and planar ellipses with height ℓ_1 and width ℓ_2 (Gustafsson 2011) using $\eta = 1/2$. Vertical polarization in *solid lines* and horizontal polarization in *dashed lines*

elliptic polarization is discussed in Gustafsson and Sohl (2009). Ultra-wideband and multiband antennas are analyzed in Sohl and Gustafsson (2008), and Derneryd et al. (2009). The main drawback with the forward scattering results in Gustafsson et al. (2007a, 2009b) is that they are only useful when the entire volume can be used for the antenna design. They are therefore not useful for antennas integrated (or embedded) into devices such as mobile phones and laptops.

Antenna Current Optimization

Optimization can be used to handle simultaneous requirements on the performance and size of complex radiating structures. Here, a typical wireless device structure composed of an antenna region and a ground plane (see Fig. 17, and Fig. 21) is considered. The entire structure occupies the volume denoted by V and consists of, e.g., screen, battery, electronics, RF circuitry, antenna(s), etc., for a mobile phone. The antenna, which is part of the structure, is restricted to the region occupying the volume V_A. The antenna designer is assumed to be allowed to specify the spatial distribution of metal and dielectrics in the region V_A. The electromagnetic properties of the remaining region $V_G = V - V_A$ is assumed to be fixed. The discussion is restricted to electric current densities and use of the stored energy expression Eqs. 9 and 10.

DOI 10.1007/978-981-4560-75-7_18-1

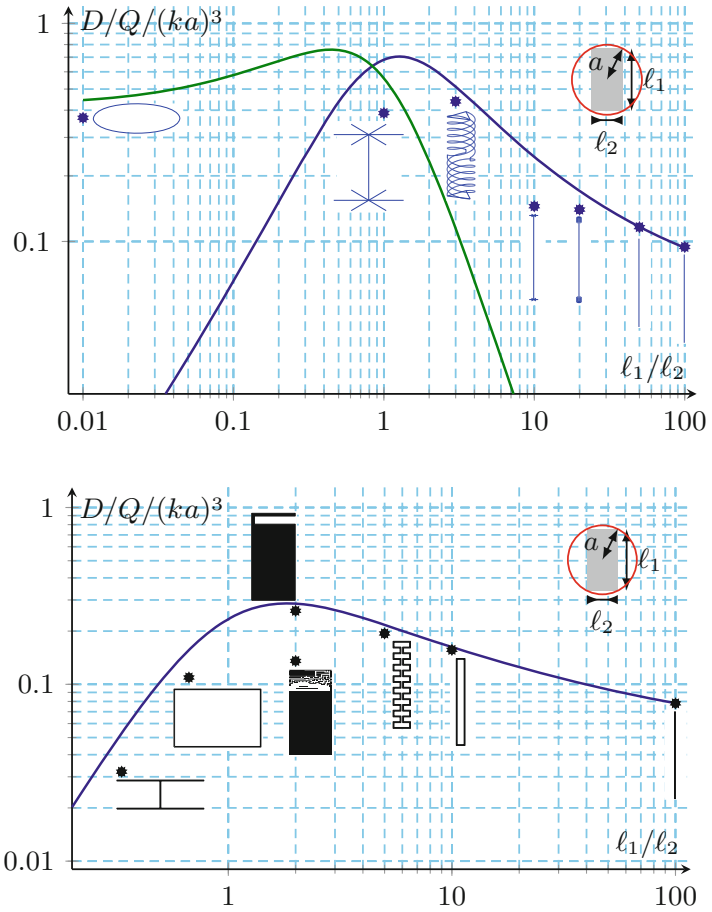


Fig. 15 Forward scattering bound on D/Q using $\eta = 1/2$ normalized with $(ka)^3$ compared with numerical results from Altair Development (2014) for nonmagnetic cylindrical (*top*) (see Table 1) and planar (*bottom*) (see Table 2) structures

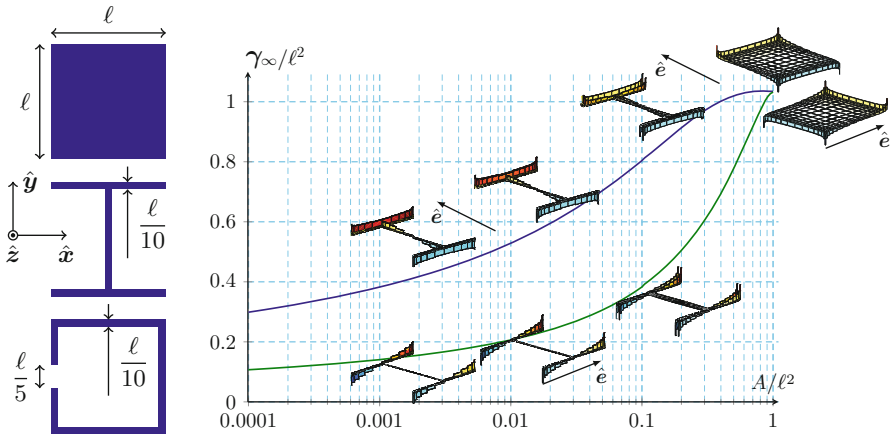


Fig. 16 (*Left*) Planar rectangular structures with polarizability dyadics $\boldsymbol{\gamma}_\infty \approx 1.04\ell^3(\hat{\boldsymbol{x}}\hat{\boldsymbol{x}} + \hat{\boldsymbol{y}}\hat{\boldsymbol{y}})$, $\boldsymbol{\gamma}_\infty \approx \ell^3(0.51\hat{\boldsymbol{x}}\hat{\boldsymbol{x}} + 0.93\hat{\boldsymbol{y}}\hat{\boldsymbol{y}})$, and $\boldsymbol{\gamma}_\infty \approx \ell^3(0.94\hat{\boldsymbol{x}}\hat{\boldsymbol{x}} + 0.96\hat{\boldsymbol{y}}\hat{\boldsymbol{y}})$, for the rectangle, capped dipole, and loop, respectively. (*Right*) Polarizability for planar capped dipoles with area A circumscribed by a rectangle with sides $\ell \times \ell$. Resulting charge distributions are also depicted in the figure (see Gustafsson et al. 2009a)

DOI 10.1007/978-981-4560-75-7_18-1

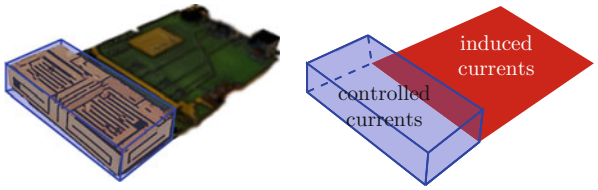


Fig. 17 (*Left*) Circumscribing box for an antenna with a ground plane. Physical bounds are determined for arbitrarily shaped antennas inside of a box in the presence of a fixed structure outside of the box. (*Right*) Current optimization with controllable current in the antenna region (the *box*) and induced current in the remaining structure (here the ground plane)

Magnetic current densities can in many cases lower the bounds on Q and can be analyzed using the stored energy expressions in Jonsson and Gustafsson (2015).

Many antenna design requirements can be formulated as optimization problems. These problems can be further reformulated as optimization problems in terms of antenna current distributions. An example is presented in the following.

Antenna design. From the antenna design perspective, the optimization problem is to design an antenna by proper shaping and choosing of the materials in the antenna region for optimal performance. The current and voltage in one or more feed points determine the matching properties of the antenna. The current distribution in the entire radiating structure determines the radiation properties.

Current distribution. In terms of current density, the antenna optimization problem is to determine the currents in the antenna region that yield optimal performance. These currents are determined without considering the feeding structures such that only radiation properties of the antenna are quantified in the performance of the antenna.

Maximization of G/Q

The partial gain Q-factor quotient is

$$\frac{G(\hat{\boldsymbol{r}},\hat{\boldsymbol{e}})}{Q}=\frac{4\pi P(\hat{\boldsymbol{r}},\hat{\boldsymbol{e}})}{2\omega\max\{W_{\mathrm{e}},W_{\mathrm{m}}\}}=\frac{\pi|\hat{\boldsymbol{e}}^{*}\cdot\boldsymbol{F}(\hat{\boldsymbol{r}})|^{2}}{\omega\eta_{0}\max\{W_{\mathrm{e}},W_{\mathrm{m}}\}}, \tag{27}$$

where $P(\hat{\boldsymbol{r}},\hat{\boldsymbol{e}})$ is the partial radiation intensity for the polarization $\hat{\boldsymbol{e}}$ and direction $\hat{\boldsymbol{r}}$, $\boldsymbol{F}(\hat{\boldsymbol{r}})$ is the far field (Eq. 7), W_{e} is the stored electric energy (Eq. 9), and W_{m} is the stored magnetic energy (Eq. 10). The quantities in Eq. 27 are approximated using a method of moments approach and expanding the current density on the antenna in terms of local basis functions (Eq. 40), as

$$\begin{aligned}
&\hat{\boldsymbol{e}}^{*}\cdot\boldsymbol{F}\approx\mathbf{F}\mathbf{I} && \text{far field in direction } \hat{\boldsymbol{r}} \text{ and polarization } \hat{\boldsymbol{e}},\\
&W_{\mathrm{e}}\approx\frac{1}{4\omega}\mathbf{I}^{\mathrm{H}}\mathbf{X}_{\mathrm{e}}\mathbf{I} && \text{stored E-energy, } \mathbf{X}_{\mathrm{e}} \text{ electric reactance,}\\
&W_{\mathrm{m}}\approx\frac{1}{4\omega}\mathbf{I}^{\mathrm{H}}\mathbf{X}_{\mathrm{m}}\mathbf{I} && \text{stored M-energy, } \mathbf{X}_{\mathrm{m}} \text{ magnetic reactance.}
\end{aligned}$$

The partial gain Q-factor ratio (Eq. 27) becomes

$$\frac{G(\hat{\boldsymbol{r}},\hat{\boldsymbol{e}})}{Q}\approx\frac{4\pi|\mathbf{F}\mathbf{I}|^{2}}{\eta_{0}\max\left\{\mathbf{I}^{\mathrm{H}}\mathbf{X}_{\mathrm{e}}\mathbf{I},\mathbf{I}^{\mathrm{H}}\mathbf{X}_{\mathrm{m}}\mathbf{I}\right\}}. \tag{28}$$

DOI 10.1007/978-981-4560-75-7_18-1

It is assumed that the MoM approximation is sufficiently accurate for the purpose of the analysis.

The optimization problem of maximizing the G/Q ratio in Eq. 28 can be formulated as the problem of minimizing the stored energy for a fixed partial radiation intensity

$$\begin{array}{ll} \text{minimize}_{\mathbf{I}} & \max\left\{\mathbf{I}^{\mathrm{H}}\mathbf{X}_{\mathrm{e}}\mathbf{I}, \mathbf{I}^{\mathrm{H}}\mathbf{X}_{\mathrm{m}}\mathbf{I}\right\} \\ \text{subject to} & |\mathbf{F}\mathbf{I}|^2 = 1, \end{array} \tag{29}$$

where the normalization $|\mathbf{F}\mathbf{I}|^2 = 1$ or equivalently $|\mathbf{F}\mathbf{I}| = 1$ is used. The formulation in Eq. 29 is possible due to the scaling invariance of G/Q in terms of the current matrix $\mathbf{I}$, i.e., G/Q is invariant for the complex scaling $\mathbf{I} \rightarrow \alpha\mathbf{I}$. Moreover, this scaling invariance shows that an arbitrary phase $\mathbf{F}\mathbf{I} = 1$ that removes the absolute value operation (Gustafsson and Nordebo 2013) can be considered. The convex optimization problem of minimizing the stored energy for a fixed partial far field in one direction (Gustafsson and Nordebo 2013), i.e.,

$$\begin{array}{ll} \text{minimize}_{\mathbf{I}} & \max\left\{\mathbf{I}^{\mathrm{H}}\mathbf{X}_{\mathrm{e}}\mathbf{I}, \mathbf{I}^{\mathrm{H}}\mathbf{X}_{\mathrm{m}}\mathbf{I}\right\} \\ \text{subject to} & \mathbf{F}\mathbf{I} = 1 \end{array} \tag{30}$$

is obtained. Let $\mathbf{I}_{\mathrm{o}}$ denote a current matrix that solves Eq. 30. The minimum value of the stored energy in Eq. 30 is unique although the current matrix $\mathbf{I}_{\mathrm{o}}$ is not necessarily unique. This optimum current gives an upper bound on G/Q for the considered direction $\hat{\boldsymbol{r}}$ and polarization $\hat{\boldsymbol{e}}$, i.e.,

$$\frac{G(\hat{\boldsymbol{r}}, \hat{\boldsymbol{e}})}{Q} \leq \left.\frac{G(\hat{\boldsymbol{r}}, \hat{\boldsymbol{e}})}{Q}\right|_{\mathrm{ub}} = \frac{4\pi|\mathbf{F}\mathbf{I}_{\mathrm{o}}|^2}{\eta_0 \max\left\{\mathbf{I}_{\mathrm{o}}^{\mathrm{H}}\mathbf{X}_{\mathrm{e}}\mathbf{I}_{\mathrm{o}}, \mathbf{I}_{\mathrm{o}}^{\mathrm{H}}\mathbf{X}_{\mathrm{m}}\mathbf{I}_{\mathrm{o}}\right\}}. \tag{31}$$

Alternative formulations are available for the convex optimization problem in Eq. 30 (Gustafsson and Nordebo 2013). These formulations can be solved with, e.g., CVX (Grant and Boyd 2011), a dual formulation (Gustafsson et al. 2014b), etc. (see Gustafsson and Nordebo 2013 and Gustafsson et al. 2015 for illustrations).

The optimization problem in Eq. 30 can be solved analytically in the limit of small antennas (Gustafsson et al. 2012a), giving

$$\frac{G(\hat{\boldsymbol{r}},\hat{\boldsymbol{e}})}{Q} \leq \frac{k^3}{4\pi}\hat{\boldsymbol{e}}^* \cdot \boldsymbol{\gamma}_\infty \cdot \hat{\boldsymbol{e}} \quad \text{as} \quad ka \rightarrow 0 \tag{32}$$

for the case of electric, $W_{\mathrm{e}} \geq W_{\mathrm{m}}$, antennas. This result confirms the forward scattering bounds in Eqs. 23 and 24, as also illustrated in Fig. 14. The general case with electric and magnetic current densities is analyzed in Yaghjian et al. (2013) and Jonsson and Gustafsson (2015) and shows that the electric polarizability dyadic in Eq. 32 is replaced with the sum of the electric and magnetic polarizability dyadics. The lower bound on the Q-factor for electric dipoles is (Yaghjian et al. 2013; Jonsson and Gustafsson 2015)

$$Q \geq \frac{6\pi}{k^3 \max \operatorname{eig}\left(\boldsymbol{\gamma}_{\mathrm{e}} + \boldsymbol{\gamma}_{\mathrm{m}}\right)} \quad \text{as} \quad ka \rightarrow 0 \tag{33}$$

(see Yaghjian et al. 2013; Jonsson and Gustafsson 2015 for additional cases and examples).

DOI 10.1007/978-981-4560-75-7_18-1

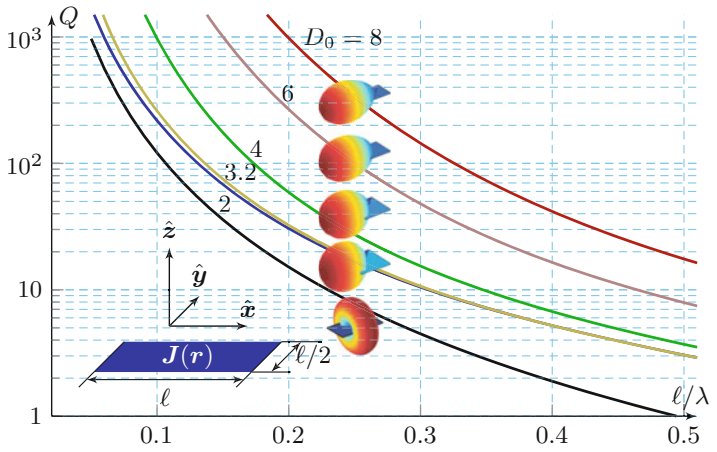


Fig. 18 Illustration of the physical bound on the Q-factor for antennas constrained to a planar rectangle with length ℓ and width $\ell/2$ that radiates with directivity $D(\hat{\boldsymbol{x}}) \geq D_0$ for $D_0 = \{2, 3.2, 4, 6, 8\}$ (see Gustafsson and Nordebo 2013)

Superdirectivity and Prescribed Radiation Patterns

Antennas with a higher directivity than typical antennas of the same size are often referred to as superdirective antennas (Best et al. 2008; Hansen 2006; Kim et al. 2012; Margetis et al. 1998). The increase of the Q-factor for small antennas with high directivity is analyzed by adding the constraint $D \geq D_0$ to Eq. 30. Written in the far field, the partial directivity is at least D_0 if

$$D_0 \leq D = \frac{4\pi|\hat{\boldsymbol{e}}^* \cdot \boldsymbol{F}(\hat{\boldsymbol{r}})|^2}{2\eta_0 P_{\mathrm{r}}} \Rightarrow P_r \leq \frac{2\pi|\hat{\boldsymbol{e}}^* \cdot \boldsymbol{F}(\hat{\boldsymbol{r}})|^2}{\eta_0 D_0}. \tag{34}$$

This is added as the convex constraint $\frac{1}{2}\mathbf{I}^{\mathrm{H}}\mathbf{R}\mathbf{I} \leq 2\pi/(\eta_0 D_0)$ to the optimization problem Eq. 30, giving

$$\begin{array}{ll} \text{minimize}_{\mathbf{I}} & \max\left\{\mathbf{I}^{\mathrm{H}}\mathbf{X}_{\mathrm{e}}\mathbf{I}, \mathbf{I}^{\mathrm{H}}\mathbf{X}_{\mathrm{m}}\mathbf{I}\right\} \\ \text{subject to} & \mathbf{F}\mathbf{I} = 1 \\ & \mathbf{I}^{\mathrm{H}}\mathbf{R}\mathbf{I} \leq \dfrac{4\pi}{\eta_0 D_0}. \end{array} \tag{35}$$

This optimization problem is solved using CVX (Grant and Boyd 2011) for a planar rectangle with side lengths ℓ and $\ell/2$, and $D_0 = \{2, 3.2, 4, 6, 8\}$. The minimum Q-factor, for a radiated field with the partial directivity at least D_0 in the $\hat{\boldsymbol{r}} = \hat{\boldsymbol{x}}$-direction for the polarization $\hat{\boldsymbol{e}} = \hat{\boldsymbol{y}}$, is depicted in Fig. 18. These Q-factors are compared to the Q-factor in Eq. 24 obtained from the forward scattering bound on D/Q in Eq. 23 for the same rectangle. In the forward scattering bound, it is assumed that the antennas radiate as electric dipoles. Under this assumption, the directivity is 1.5 in a direction normal to the plane of the structure (see Fig. 18), and the generalized absorption efficiency is $\eta = 1/2$ (Gustafsson et al. 2009b; Gustafsson 2010a). Note that the constraints $D \geq D_0 = \{4, 6, 8\}$ yield optimum current densities with directivities $D = D_0$. However, the constraint $D \geq 2$ results in an optimum current that contributes both an electric and a magnetic dipole such that the directivity is $D \geq 2.9$, for $\ell/\lambda \geq 0.05$; see the curves labeled 2 and 3.2 in Fig. 18.

Embedded Antennas

The antenna current optimization method can be generalized to derive bounds for antennas embedded in devices, e.g., as that depicted in Fig. 17. This generalization is based on decomposing the current density

DOI 10.1007/978-981-4560-75-7_18-1

into the controllable currents, $\boldsymbol{J}_{\mathrm{A}}$, and induced currents, $\boldsymbol{J}_{\mathrm{G}}$, (Gustafsson and Nordebo 2013; Cismasu and Gustafsson 2014a, b). A region of the considered device, V_{A} (see Fig. 21), is reserved for a structure that can be engineered by, e.g., optimization, manual or computer-aided design, etc. The currents in this region, $\boldsymbol{J}_{\mathrm{A}}$, are considered controllable. The remaining part of the device, V_{G} is considered fixed such that it supports the induced current density $\boldsymbol{J}_{\mathrm{G}}$. Due to the linearity of Maxwell's equations, $\boldsymbol{J}_{\mathrm{G}}$ depends linearly on $\boldsymbol{J}_{\mathrm{A}}$. This dependence can be written $\boldsymbol{J}_{\mathrm{G}} = \mathbf{C}\boldsymbol{J}_{\mathrm{A}}$, where $\mathbf{C}$ is determined from, e.g., the MoM impedance matrix (Peterson et al. 1998; Jin 2011; Johnson and Rahmat-Samii 1999). The bounds on the G/Q of a rectangular, infinitely thin, perfect electric conducting (PEC) device with the dimensions ℓ and $\ell/2$ in which the region with controllable currents may occupy 6 %, 10 %, and 25 % of the device at one end in the ℓ direction are depicted in Fig. 19. These bounds are compared to the bound of the entire structure obtained from the forward scattering sum rule described in section "Forward Scattering Sum Rule" (see also Gustafsson et al. 2007a, 2009b; Gustafsson 2010a). Other cases of embedded devices and details can be found in Gustafsson et al. (2014b), Cismasu and Gustafsson (2014a,b), and Cismasu et al. (2014).

Periodic Array Antennas

Sum rules are used to derive limitations on the bandwidth for array antennas in Doane et al. (2013) and Jonsson et al. (2013). The bandwidth of infinite periodic array antennas is limited by the thickness of the array. Consider a periodic array antenna with thickness d above a PEC ground plane; see Fig. 20. The antenna performance is analyzed in receiving mode by assuming a reciprocal lossless antenna with scattering reflection coefficient, ρ, of equal amplitude as the reflection coefficient in the feed Eq. 1, i.e., $|\Gamma| = |\rho|$. Here, the fundamental mode of the reflected wave is considered. This transforms the antenna to an absorber and the sum rule (Rozanov 2000; Doane et al. 2013; Jonsson et al. 2013)

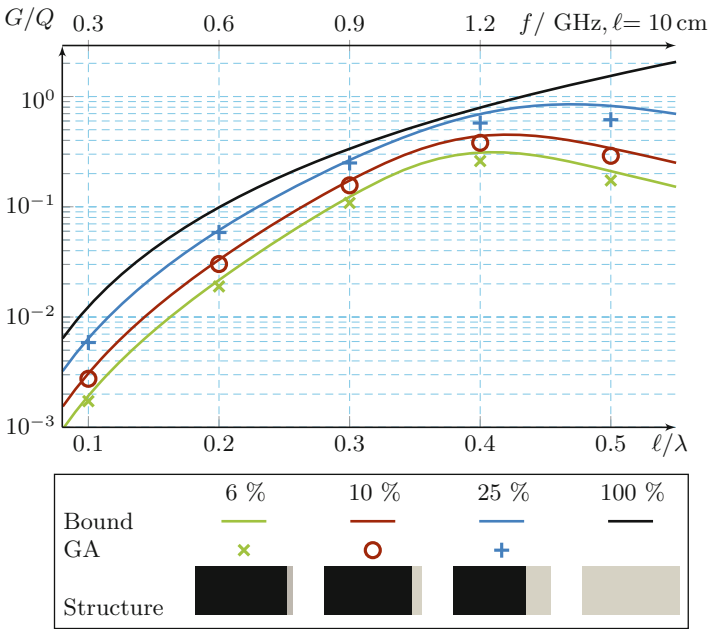


Fig. 19 GA-optimized antenna G/Q ratios (Eq. 28), *marks*, compared with physical bounds on G/Q obtained with a formulation equivalent to Eq. 30 for a rectangular wireless device model in which the antenna may occupy 6 %, 10 %, and 25 % of the device at one end; see Fig. 17. The physical bound on the G/Q ratio of a rectangular, infinitely thin, PEC sheet (Gustafsson et al. 2007a, 2009b; Gustafsson 2010a) is labeled "100 %." *Gray-shaded areas* have controllable currents

DOI 10.1007/978-981-4560-75-7_18-1

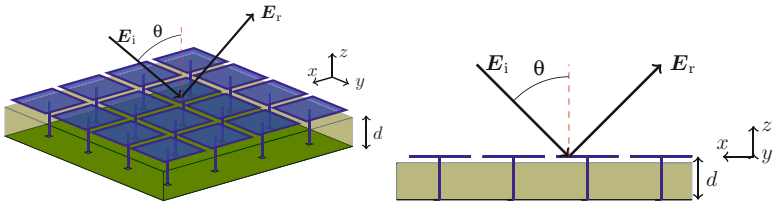


Fig. 20 Periodic array antenna above a PEC ground plane

$$\frac{1}{\pi}\int_{\mathbb{R}} \frac{1}{k^2}\ln\frac{1}{|\rho(k)|}\,\mathrm{d}k \le d\left(1+\frac{\gamma}{2dA}\right) \le d\ \max\{\mu_r\}\begin{cases}\cos\theta & \text{TE}\\ 1/\cos\theta & \text{TM}\end{cases} \tag{36}$$

first derived by Rozanov (2000) for layered structures with $\max\{\mu_r\}$, denoting the maximal static relative permeability of the structure. Here, the low-frequency expansion from Gustafsson and Sjöberg (2011) is used to extend the results to periodic structures.

The sum rule is transformed to the bandwidth bound (Rozanov 2000; Doane et al. 2013; Jonsson et al. 2013)

$$\lambda_2-\lambda_1 \le \frac{2\pi^2 d\max\{\mu_r\}}{\ln|\Gamma_0|^{-1}}\begin{cases}\cos\theta & \text{TE}\\ 1/\cos\theta & \text{TM}\end{cases} \tag{37}$$

for linearly polarized waves, where Γ_0 denotes the threshold level for the reflection coefficient over the wavelength interval $[\lambda_1, \lambda_2]$. Array antennas are analyzed and compared with antenna designs in Doane et al. (2013) and Jonsson et al. (2013). An array figure of merit is also introduced in Jonsson et al. (2013).

Conclusion

Physical bounds on antennas answer questions such as: *how good can an antenna be*? Antenna bounds have evolved from spherical regions (1948) to arbitrary-shape structures (2007) and embedded structures (2013). Here, an overview of methods based on circuit models, mode expansions, sum rules, and optimization and a discussion of some of their pros and cons are presented. This characterization is chosen to emphasize the key characteristics of the majority of the published methods for the study of physical bounds.

All presented bounds are based on assumptions. The forward scattering bounds are, e.g., based on the assumption of linear, reciprocal, time-translational invariant material parameters. These assumptions are valid for most antennas but switches and non-Foster matching can potentially be used to overcome the limitations (Ziolkowski et al. 2013). Moreover, bounds based on the stored energy, such as the Chu bound Eq. 15 and the antenna current optimization in section "Antenna Current Optimization," assume that the stored energy can be determined accurately using Eqs. 8, 9, and 10. There are initial investigations on the stored energy expressions (Hazdra et al. 2011; Gustafsson and Jonsson 2015a, b; Gustafsson et al. 2014a), but much work remains before the stored energy is fully understood.

The D/Q and Q-factor limits for small electric dipole antennas composed of nonmagnetic materials are well understood. There are several independent derivations that provide similar results (Gustafsson et al. 2007a, 2009b, 2010, 2012a; Gustafsson and Nordebo 2013; Yaghjian and Stuart 2010; Yaghjian et al. 2013; Vandenbosch 2011); see also Eqs. 23 and 32 for this case. In addition, many antenna designs are shown to perform close to the bounds (see Gustafsson et al. 2009b, 2010; Best 2009; Shahpari

DOI 10.1007/978-981-4560-75-7_18-1

et al. 2013; Fig. 15). Antennas embedded in finite PEC structures are investigated in Cismasu and Gustafsson (2014a, b) and are also shown to perform close to the bounds. The D/Q results for electric dipole antennas work well up to about half-a-wavelength-sized structures ($ka \approx 1.5$) (Gustafsson et al. 2009b, 2010). The corresponding lower bounds on the Q-factor for larger structures are not well understood, except for the case with prescribed radiation patterns, e.g., dipole patterns (Chu 1948; Thal 2006; Gustafsson and Nordebo 2013).

Although there has been a strong development on the physical bounds in recent years, there are still many open questions. The case with mixed electric and magnetic dipole patterns, magnetic materials, and superdirective antennas Eq. 35 are not as well understood in the sense of realized antenna designs although there are some suggested designs. Efficiency is the most important small antenna parameter besides bandwidth. There are only some initial investigations for lossy structures (Gustafsson 2013), so much work remains on fundamental limitations for efficiency.

Acknowledgment

This work was supported by the Swedish Research Council (VR) and the Swedish Foundation for Strategic Research (SSF) under the program Applied Mathematics and the project Complex analysis and convex optimization for EM design.

Cross-References

- Foster's Reactance Theorem and Non-Foster Impedance Matching in Antenna Systems
- Impedance Matching and Baluns
- Numerical Modeling in Antenna Engineering
- Optimization Methods in Antenna Engineering
- Small Antennas (PIFA PILA Loading Antenna)
- Wheeler-Fano Impedance Matching Principles

Appendix

A Radial Functions

The radial functions in Hansen (1988) are defined as

$$\mathrm{R}_{\tau n}^{(p)}(\kappa) = \begin{cases} z_n^{(p)}(\kappa) & \tau = 1 \\ \dfrac{1}{\kappa} \dfrac{\partial\left(\kappa z_n^{(p)}(\kappa)\right)}{\partial \kappa}, & \tau = 2, \end{cases} \tag{38}$$

where $z_n^{(1)} = \mathrm{j}_n$ are Bessel functions, $z_n^{(2)} = \mathrm{n}_n$ Neumann functions, $z_n^{(3)} = \mathrm{h}_n^{(1)}$ Hankel functions (Hansen 1988), and $\kappa = ka$. The derivatives of $\mathrm{R}_{\tau n}^{(p)}(\kappa)$ are easily expressed in spherical Bessel and Hankel functions as

DOI 10.1007/978-981-4560-75-7_18-1

$$\frac{\partial \mathrm{R}_{\tau n}^{(p)}}{\partial \kappa} = \begin{cases} \dfrac{\partial}{\partial \kappa} z_n^{(p)} & \tau = 1 \\ \dfrac{-\mathrm{R}_{\tau n}^{(p)}}{\kappa} + \dfrac{n(n+n) - \kappa^2}{\kappa^2} z_n^{(p)} & \tau = 2. \end{cases} \tag{39}$$

B High-Contrast Polarizability Dyadics

The high-contrast polarizability dyadic $\boldsymbol{\gamma}_\infty$ is determined from the induced normalized surface charge density, ρ, as

$$\hat{\boldsymbol{e}} \cdot \boldsymbol{\gamma}_\infty \cdot \hat{\boldsymbol{e}} = \frac{1}{E_0} \int_S \hat{\boldsymbol{e}} \cdot \boldsymbol{r} \rho(\boldsymbol{r}) \ \mathrm{dS},$$

where ρ satisfies the integral equation

$$\int_S \frac{\rho(\boldsymbol{r}_1)}{4\pi |\boldsymbol{r} - \boldsymbol{r}_1|} \ \mathrm{dS}_1 = E_0 \boldsymbol{r} \cdot \hat{\boldsymbol{e}} - V$$

with the constraint of zero total charge

$$\int_S \rho(\boldsymbol{r}) \ \mathrm{dS} = 0.$$

used to determine the voltage V.

C Expansion in Terms of Basis Functions

Consider a surface S with current density $\boldsymbol{J} = \boldsymbol{J}(\boldsymbol{r})$; see Fig. 7. Expand the current density in local basis functions

$$\boldsymbol{J}(\boldsymbol{r}) \approx \sum_{n=1}^{N} I_n \boldsymbol{\psi}_n(\boldsymbol{r}) \tag{40}$$

and introduce the $N \times 1$ matrix $\mathbf{I}$ with elements I_n to simplify the notation. The basis functions are assumed to be real valued, divergence conforming, and having vanishing normal components at the boundary (Peterson et al. 1998). For embedded antennas, it is assumed that the currents $\boldsymbol{J}_\mathrm{A}$ in the antenna volume V_A are controllable and induce the currents $\boldsymbol{J}_\mathrm{G}$ in the volume V_G; see Fig. 21.

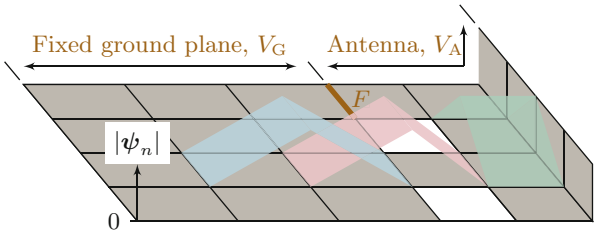


Fig. 21 Illustration of rectangular mesh element discretization and "rooftop" basis function amplitude for a three-dimensional radiating structure. Metal areas are depicted in gray shading. The amplitudes of three of the total $6 \times 3 + 5 \times 4 - 4 - 3$ basis functions are depicted in *blue*, *pink*, and *green shading*. The feeding edge is marked F

DOI 10.1007/978-981-4560-75-7_18-1

A standard MoM implementation of the EFIE using the Galerkin procedure computes the impedance matrix $\mathbf{Z} = \mathbf{R} + \mathrm{j}\mathbf{X}$ with the elements

$$\frac{Z_{mn}}{\eta_0} = \mathrm{j}\int_S\int_S \left(k^2\boldsymbol{\psi}_{m1}\cdot\boldsymbol{\psi}_{n2} - \nabla_1\cdot\boldsymbol{\psi}_{m1}\nabla_2\cdot\boldsymbol{\psi}_{n2}\right)\frac{e^{-\mathrm{j}kR_{12}}}{4\pi kR_{12}}\ \mathrm{dS}_1\,\mathrm{dS}_2,$$

where $\psi_{n1} = \psi_n(\boldsymbol{r}_1), \psi_{n2} = \psi_n(\boldsymbol{r}_2), n = 1,\ldots,N$, and $R_{12} = |\boldsymbol{r}_1 - \boldsymbol{r}_2|$. The differentiated MoM impedance matrix

$$\begin{aligned}\frac{k\partial Z_{mn}}{\eta_0\partial k} = \int_S\int_S &\mathrm{j}\left(k^2\boldsymbol{\psi}_{m1}\cdot\boldsymbol{\psi}_{n2} + \nabla_1\cdot\boldsymbol{\psi}_{m1}\nabla_2\cdot\boldsymbol{\psi}_{n2}\right)\frac{\mathrm{e}^{-\mathrm{j}kR_{12}}}{4\pi kR_{12}}\\ &+\left(k^2\boldsymbol{\psi}_{m1}\cdot\boldsymbol{\psi}_{n2} - \nabla_1\cdot\boldsymbol{\psi}_{m1}\nabla_2\cdot\boldsymbol{\psi}_{n2}\right)\frac{\mathrm{e}^{-\mathrm{j}kR_{12}}}{4\pi}\ \mathrm{dS}_1\mathrm{dS}_2\end{aligned}$$

is used to estimate the stored energy. The differentiated reactance matrix, $\mathbf{X}' = \mathrm{Im}\mathbf{Z}'$, gives the stored magnetic and electric energies

$$W_\mathrm{m} \approx \frac{1}{8}\mathbf{I}^\mathrm{H}\left(\frac{\partial\mathbf{X}}{\partial\omega} + \frac{\mathbf{X}}{\omega}\right)\mathbf{I} \quad \text{and} \quad W_\mathrm{e} \approx \frac{1}{8}\mathbf{I}^\mathrm{H}\left(\frac{\partial\mathbf{X}}{\partial\omega} - \frac{\mathbf{X}}{\omega}\right)\mathbf{I},$$

respectively. These expressions are identical to the stored energy expressions introduced by Vandenbosch (Vandenbosch 2010) using a MoM approximation (see also Harrington 1972; 1975). Introduce the electric $\mathbf{X}_\mathrm{e}$, and magnetic $\mathbf{X}_\mathrm{m}$, reactance matrices

$$\mathbf{X}_\mathrm{e} = \frac{1}{2}\left(\omega\frac{\partial\mathbf{X}}{\partial\omega} - \mathbf{X}\right) \quad \text{and} \quad \mathbf{X}_\mathrm{m} = \frac{1}{2}\left(\omega\frac{\partial\mathbf{X}}{\partial\omega} + \mathbf{X}\right).$$

The total radiated power is determined from $\mathbf{R} = \mathrm{Re}\{\mathbf{Z}\}$.

The far field Eq. 7 in the direction $\hat{\boldsymbol{r}}$ projected on $\hat{\boldsymbol{e}}$ is approximated by the $N \times 1$ matrix $\mathbf{FI} \approx \hat{\boldsymbol{e}}^* \cdot \boldsymbol{F}(\hat{\boldsymbol{r}})$ defined as

$$\mathbf{FI} \approx -\mathrm{j}k\eta_0\sum_{n=1}^{N} I_n\int_S \hat{\boldsymbol{e}}^*\cdot\boldsymbol{\psi}_n(\boldsymbol{r})\frac{\mathrm{e}^{\mathrm{j}k\hat{\boldsymbol{r}}\cdot\boldsymbol{r}}}{4\pi}\,\mathrm{dS}. \tag{41}$$

It is assumed that the numerical approximation is sufficiently accurate so the approximate equal to ($\approx$) above can be replaced with equalities.

D Numerical Results for the Presented Antennas

Numerical simulation data for the antennas presented in this chapter are collected in Table 1 for wire antennas and Table 2 for planar antennas.

DOI 10.1007/978-981-4560-75-7_18-1

Table 1 Numerical results of the wire antennas simulated in Altair Development (2014); see also Fig. 15. The forward scattering bounding boxes are either cylindrical or spheroidal depending on the geometry of the structure

	ℓ_1/ℓ_2	ka	D	Q	$\frac{D/Q}{(ka)^3}$	$\frac{4\pi D/Q}{\gamma k^3}$
	100	1.481	1.63	5	0.094	1.00
	50	1.463	1.63	5	0.116	1.00
	20	1.188	1.61	7	0.139	0.82
	10	1.184	1.61	7	0.145	0.59
	1	0.587	1.50	19	0.388	0.69
	0.01	1.075	2.34	5	0.369	0.83
	2	0.421	1.50	49	0.411	0.86
	1	0.137	1.50	675	0.867	0.87
	3	0.497	1.49	29	0.425	0.85
	1.5	0.266	1.50	142	0.557	0.83

DOI 10.1007/978-981-4560-75-7_18-1

Table 2 Numerical results of the planar antennas simulated in Altair Development (2014); see also Fig. 15. Results are for a rectangular forward scattering bounding box

	ℓ_1/ℓ_2	ka	D	Q	$\frac{D/Q}{(ka)^3}$	$\frac{4\pi D/Q}{\gamma k^3}$
		1.486	1.64	6	0.078	0.99
	10	1.319	1.62	7	0.105	0.64
	5	1.178	1.61	7	0.132	0.61
	0.33	1.072	1.47	37	0.032	0.61
	10	1.411	1.64	4	0.157	0.96
	0.67	1.246	2.48	12	0.109	0.70
	5	0.953	1.59	12	0.148	0.68
	5	1.050	1.57	7	0.194	0.89
	2	0.708	1.55	18	0.245	0.86
	2	1.359	1.67	3	0.260	0.91
	2	0.953	1.53	19	0.095	0.33
	2	0.351	1.51	257	0.136	0.47

DOI 10.1007/978-981-4560-75-7_18-1

References

Adams JJ, Duoss EB, Malkowski TF, Motala MJ, Ahn BY, Nuzzo RG, Bernhard JT, Lewis JA (2011) Conformal printing of electrically small antennas on three-dimensional surfaces. Adv Mater 23(11):1335–1340

Altair Development S.A. (Pty) Ltd, Stellenbosch. FEKO, Field computations involving bodies of arbitrary shape, Suite 7.0. 2014

Bernland A (2012) Bandwidth limitations for scattering of higher order electromagnetic spherical waves with implications for the antenna scattering matrix. IEEE Trans Antennas Propag 60(9):4345–4353

Bernland A, Gustafsson M, Nordebo S (2011a) Physical limitations on the scattering of electromagnetic vector spherical waves. J Phys A: Math Theor 44(14):145401

Bernland A, Luger A, Gustafsson M (2011b) Sum rules and constraints on passive systems. J Phys A: Math Theor 44(14):145205. doi:10.1088/1751-8113/44/14/145205

Best SR (2004) The radiation properties of electrically small folded spherical helix antennas. IEEE Trans Antennas Propag 52(4):953–960

Best SR (2009) A comparison of the cylindrical folded helix *Q* to the Gustafsson limit. In: Antennas and propagation, 2009. EuCAP 2009. 3rd European conference on. IEEE, Berlin, pp 2554–2557

Best SR, Hanna DL (2010) A performance comparison of fundamental small-antenna designs. IEEE Antennas Propag Mag 52(1):47–70

Best SR, Altshuler EE, Yaghjian AD, McGinthy JM, O'Donnell TH (2008) An impedance-matched 2-element superdirective array. IEEE Antennas Wirel Propag Lett 7:302–305

Carpenter CJ (1989) Electromagnetic energy and power in terms of charges and potentials instead of fields. IEE Proc A 136(2):55–65

Cheng DK (1989) Field and wave electromagnetics. Addison-Wesley, Reading

Chu LJ (1948) Physical limitations of omnidirectional antennas. J Appl Phys 19:1163–1175

Cismasu M, Gustafsson M (2014a) Antenna bandwidth optimization with single frequency simulation. IEEE Trans Antennas Propag 62(3):1304–1311

Cismasu M, Gustafsson M (2014b) Multiband antenna *Q* optimization using stored energy expressions. IEEE Antennas Wirel Propag Lett 13(2014):646–649

Cismasu M, Tayli D, Gustafsson M (2014) Stored energy based 3D antenna analysis and design. Technical Report LUTEDX/(TEAT-7231)/1-18/(2014). Lund University, Department of Electrical and Information Technology, Lund

Collin RE, Rothschild S (1964) Evaluation of antenna *Q*. IEEE Trans Antennas Propag 12:23–27

Davis W, Yang T, Caswell E, Stutzman W (2011) Fundamental limits on antenna size: a new limit. IET Microwaves Antennas Propag 5(11):1297–1302

Derneryd A, Gustafsson M, Kristensson G, Sohl C (2009) Application of gain-bandwidth bounds on loaded dipoles. IET Microwaves Antennas Propag 3(6):959–966

Doane JP, Sertel K, Volakis JL (2013) Matching bandwidth limits for arrays backed by a conducting ground plane. IEEE Trans Antennas Propag 61(5):2511–2518

Doane JP, Sertel K, Volakis JL (2014) Bandwidth limits for lossless, reciprocal PEC-backed arrays of arbitrary polarization. IEEE Trans Antennas Propag 62(5):2531–2542

Fano RM (1950) Theoretical limitations on the broadband matching of arbitrary impedances. J Franklin Inst 249(1, 2):57–83, 139–154

Fante RL (1969) Quality factor of general antennas. IEEE Trans Antennas Propag 17(2):151–155

Foltz HD, McLean JS (1999) Limits on the radiation *Q* of electrically small antennas restricted to oblong bounding regions. In: IEEE Antennas and Propagation Society international symposium, vol 4. IEEE, Orlando, pp 2702–2705

Geyi W (2003a) Physical limitations of antenna. IEEE Trans Antennas Propag 51(8):2116–2123
Geyi W (2003b) A method for the evaluation of small antenna *Q*. IEEE Trans Antennas Propag 51(8):2124–2129
Geyi W (2011) Foundations of applied electrodynamics. Wiley, Hoboken
Glazunov AA, Gustafsson M, Molisch A (2011) On the physical limitations of the interaction of a spherical aperture and a random field. IEEE Trans Antennas Propag 59(1):119–128
Grant M, Boyd S (2011) CVX: Matlab software for disciplined convex programming, version 1.21. cvxr.com/cvx
Griffiths DJ (1999) Introduction to electrodynamics, 3rd edn. Prentice-Hall, Englewood Cliffs
Gustafsson M (2010a) AntennaQ – MATLAB script that computes physical bounds on Q and D/Q for antennas. http://www.mathworks.se/matlabcentral/fileexchange/26806-antennaq
Gustafsson M (2010b) Time-domain approach to the forward scattering sum rule. Proc R Soc A 466:3579–3592
Gustafsson M (2010c) Sum rules for lossless antennas. IET Microwaves Antennas Propag 4(4):501–511
Gustafsson M (2011) Physical bounds on antennas of arbitrary shape. In: Antennas and propagation conference (LAPC), 2011, Loughborough. IEEE, pp 1–5
Gustafsson M (2013) Efficiency and Q for small antennas using Pareto optimality. In: 2013 IEEE international symposium on antennas and propagation, Orlando, pp 1–2
Gustafsson M, Jonsson BLG (2015a) Antenna Q and stored energy expressed in the fields, currents, and input impedance. IEEE Trans Antennas Propag 63(1):240–249
Gustafsson M, Jonsson BLG (2015b) Stored electromagnetic energy and antenna Q. Prog Electromagn Res (PIER) 150:13–27
Gustafsson M, Nordebo S (2006) Bandwidth, Q factor, and resonance models of antennas. Prog Electromagn Res 62:1–20
Gustafsson M, Nordebo S (2007) On the spectral efficiency of a sphere. Prog Electromagn Res 67:275–296
Gustafsson M, Nordebo S (2013) Optimal antenna currents for Q, superdirectivity, and radiation patterns using convex optimization. IEEE Trans Antennas Propag 61(3):1109–1118
Gustafsson M, Sjöberg D (2010) Sum rules and physical bounds on passive metamaterials. New J Phys 12:043046
Gustafsson M, Sjöberg D (2011) Physical bounds and sum rules for high-impedance surfaces. IEEE Trans Antennas Propag 59(6):2196–2204
Gustafsson M, Sohl C (2009) New physical bounds on elliptically polarized antennas. In: Proceedings of the third European conference on antennas and propagation, Berlin, 23–27 Mar 2009. The Institution of Engineering and Technology, pp 400–402
Gustafsson M, Sohl C, Kristensson G (2007a) Physical limitations on antennas of arbitrary shape. Proc R Soc A 463:2589–2607
Gustafsson M, Kristensson G, Nordebo S, Larsson C, Bernland A, Sjöberg D (2009a) Physical bounds and sum rules in scattering and antenna theory. In: International conference on electromagnetics in advanced applications (ICEAA), Turin, 14–18 Sept, pp 600–603
Gustafsson M, Sohl C, Kristensson G (2009b) Illustrations of new physical bounds on linearly polarized antennas. IEEE Trans Antennas Propag 57(5):1319–1327. doi:10.1109/TAP.2009.2016683
Gustafsson M, Cismasu M, Nordebo S (2010) Absorption efficiency and physical bounds on antennas. Int J Antennas Propag 2010(Article ID 946746):1–7
Gustafsson M, Cismasu M, Jonsson BLG (2012a) Physical bounds and optimal currents on antennas. IEEE Trans Antennas Propag 60(6):2672–2681

Gustafsson M, Vakili I, Keskin SEB, Sjöberg D, Larsson C (2012b) Optical theorem and forward scattering sum rule for periodic structures. IEEE Trans Antennas Propag 60(8):3818–3826
Gustafsson M, Fridén J, Colombi D (2015) Antenna current optimization for lossy media with near field constraints. IEEE Antennas Wirel Propag Lett 14:1538–1541
Gustafsson M, Tayli D, Cismasu M (2014a) Q factors for antennas in dispersive media. Technical report LUTEDX/(TEAT-7232)/1−24/(2014). Lund University, Department of Electrical and Information Technology, Lund. http://www.eit.lth.se
Gustafsson M, Tayli D, Ehrenborg C, Cismasu M, Nordebo S (2014b) Tutorial on antenna current optimization using MATLAB and CVX. Prepared for submission to www.e-fermat.org
Hansen RC (1981) Fundamental limitations in antennas. Proc IEEE 69(2):170–182
Hansen JE (ed) (1988) Spherical near-field antenna measurements. IEE electromagnetic waves series, vol 26. Peter Peregrinus, Stevenage
Hansen RC (2006) Electrically small, superdirective, and superconductive antennas. Wiley, Hoboken
Hansen RC, Collin RE (2009) A new Chu formula for Q. IEEE Antennas Propag Mag 51(5):38–41
Hansen RC, Collin RE (2011) Small antenna handbook. Wiley, Hoboken
Hansen TV, Kim OS, Breinbjerg O (2012) Stored energy and quality factor of spherical wave functions – in relation to spherical antennas with material cores. IEEE Trans Antennas Propag 60(3):1281–1290
Harrington RF (1961) Time harmonic electromagnetic fields. McGraw-Hill, New York
Harrington RF, Mautz JR (1972) Control of radar scattering by reactive loading. IEEE Trans Antennas Propag 446–454
Harrington R (1975) Characteristic modes for antennas and scatterers. In: Mittra R (ed) Numerical and asymptotic techniques in electromagnetics. Topics in applied physics, vol 3. Springer, Berlin/Heidelberg, pp 51–87
Hazdra P, Capek M, Eichler J (2011) Radiation Q-factors of thin-wire dipole arrangements. IEEE Antennas Wirel Propag Lett 10:556–560
Helsing J, Perfekt K-M (2013) On the polarizability and capacitance of the cube. Appl Comput Harmon Anal 34(3):445–468
IEEE145-1993 (1993) IEEE standard definition of terms for antennas. Antenna Standards Committee of the IEEE Antennas and Propagation Society
Jackson JD (1999) Classical electrodynamics, 3rd edn. Wiley, New York
Jin JM (2011) Theory and computation of electromagnetic fields. Wiley, Hoboken. ISBN 9781118088111
Johnson JM, Rahmat-Samii Y (1999) Genetic algorithms and method of moments GA/MOM for the design of integrated antennas. IEEE Trans Antennas Propag 47(10):1606–1614
Jones RC (1945) A generalization of the dielectric ellipsoid problem. Phys Rev 68(3–4):93–96
Jones DS (1985) Scattering by inhomogeneous dielectric particles. Q J Mech Appl Math 38:135–155
Jonsson BLG, Gustafsson M (2015) Stored energies in electric and magnetic current densities for small antennas. Proc R Soc Lond A 471(2176):1–23
Jonsson BLG, Kolitsidas CI, Hussain N (2013) Array antenna limitations. IEEE Antennas Wirel Propag Lett 12:1539–1542
Kajfez D Jr, Wheless WP (1986) Invariant definitions of the unloaded Q factor. IEEE Trans Microwave Theory Tech 34(7):840–841
Karlsson A (2004) Physical limitations of antennas in a lossy medium. IEEE Trans Antennas Propag 52:2027–2033
Kildal PS, Best SR (2008) Further investigations of fundamental directivity limitations of small antennas with and without ground planes. In: Antennas and Propagation Society international symposium. AP-S 2008. IEEE, San Diego, pp 1–4
Kim O (2012) Minimum Q electrically small antennas. IEEE Trans Antennas Propag 60(8):3551–3558

Kim O, Breinbjerg O, Yaghjian A (2010) Electrically small magnetic dipole antennas with quality factors approaching the Chu lower bound. IEEE Trans Antennas Propag 58(6):1898–1906

Kim O, Pivnenko S, Breinbjerg O (2012) Superdirective magnetic dipole array as a first-order probe for spherical near-field antenna measurements. IEEE Trans Antennas Propag 60(10):4670–4676

Kleinman RE, Senior TBA (1986) Rayleigh scattering, chapter 1. In: Varadan VV, Varadan VK (eds) Low and high frequency asymptotics. Handbook on acoustic, electromagnetic and elastic wave scattering, vol 2. Elsevier Science, Amsterdam, pp 1–70

Kwon D-H (2005) On the radiation Q and the gain of crossed electric and magnetic dipole moments. IEEE Trans Antennas Propag 53(5):1681–1687

Margetis D, Fikioris G, Myers JM, Wu TT (1998) Highly directive current distributions: general theory. Phys Rev E 58(2):2531

McLean JS (1996) A re-examination of the fundamental limits on the radiation Q of electrically small antennas. IEEE Trans Antennas Propag 44(5):672–676

Nordebo S, Gustafsson M, Kristensson G (2006) On the capacity of the free space antenna channel. In: IEEE Antennas and Propagation Society international symposium. IEEE Press, Albuquerque, pp 3105–3108

Nussenzveig HM (1972) Causality and dispersion relations. Academic, London

Nyberg D, Kildal P-S, Carlsson J (2010) Effects of intrinsic radiation Q on mismatch factor of three types of small antennas: single-resonance, gradual-transition and cascaded-resonance types. IET Microwaves Antennas Propag 4:83 – 90(7)

Peterson AF, Ray SL, Mittra R (1998) Computational methods for electromagnetics. IEEE Press, New York

Pozar DM (1998) Microwave engineering. Wiley, New York

Pozar DM (2009) New results for minimum Q, maximum gain, and polarization properties of electrically small arbitrary antennas. In: Antennas and propagation, 2009. EuCAP 2009. 3rd European conference on, pp 1993–1996

Purcell EM (1969) On the absorption and emission of light by interstellar grains. J Astrophys 158:433–440

Rahola J (2009) Bandwidth potential and electromagnetic isolation: tools for analysing the impedance behaviour of antenna systems. In: Antennas and propagation, 2009. EuCAP 2009. 3rd European conference on, pp 944–948

Rozanov KN (2000) Ultimate thickness to bandwidth ratio of radar absorbers. IEEE Trans Antennas Propag 48(8):1230–1234

Shahpari M, Thiel DV, Lewis A (2013) Polarizability of 2D and 3D conducting objects using method of moments. ANZIAM J 54:C446–C458

Shahpari M, Thiel D, Lewis A (2014) An investigation into the Gustafsson limit for small planar antennas using optimization. IEEE Trans Antennas Propag 62(2):950–955

Sievenpiper DF, Dawson DC, Jacob MM, Kanar T, Kim S, Long J, Quarfoth RG (2012) Experimental validation of performance limits and design guidelines for small antennas. IEEE Trans Antennas Propag 60(1):8–19

Sihvola A, Yla-Oijala P, Jarvenpaa S, Avelin J (2004) Polarizabilities of platonic solids. IEEE Trans Antennas Propag 52(9):2226–2233

Sjöberg D (2009) Variational principles for the static electric and magnetic polarizabilities of anisotropic media with perfect electric conductor inclusions. J Phys A: Math Theor 42:335403

Skrivervik AK, Gustafsson M (2012) Fundamental limitations. In: Joffre L, Martinez-Vasquez M, Serrano R, Roquetta G (eds) Handbook on small antennas, EurAAP seventh framework programme. Universitat Politecnica de Catalunya, Barcelona, pp 5–59

DOI 10.1007/978-981-4560-75-7_18-1

Skrivervik AK, Zürcher J-F, Staub O, Mosig JR (2001) PCS antenna design: the challenge of miniaturization. IEEE Antennas Propag Mag 43(4):12–27

Sohl C, Gustafsson M (2008) A priori estimates on the partial realized gain of Ultra-Wideband (UWB) antennas. Q J Mech Appl Math 61(3):415–430

Sohl C, Gustafsson M, Kristensson G (2007) Physical limitations on broadband scattering by heterogeneous obstacles. J Phys A: Math Theor 40:11165–11182

Sohl C, Larsson C, Gustafsson M, Kristensson G (2008) A scattering and absorption identity for metamaterials: experimental results and comparison with theory. J Appl Phys 103(5):054906

Sten JC-E, Koivisto PK, Hujanen A (2001) Limitations for the radiation Q of a small antenna enclosed in a spheroidal volume: axial polarisation. AEÜ Int J Electron Commun 55(3):198–204

Stuart H, Best S, Yaghjian A (2007) Limitations in relating quality factor to bandwidth in a double resonance small antenna. IEEE Antennas Wirel Propag Lett 6:460–463

Taylor TT (1955) Design of line-source antennas for narrow beamwidth and low side lobes. Antennas Propag, Trans IRE Prof Group 3(1):16–28

Thal HL (1978) Exact circuit analysis of spherical waves. IEEE Trans Antennas Propag 26(2):282–287

Thal HL (2006) New radiation Q limits for spherical wire antennas. IEEE Trans Antennas Propag 54(10):2757–2763

Thal HL (2009) Gain and Q bounds for coupled TM-TE modes. IEEE Trans Antennas Propag 57(7):1879–1885

Thal HL (2012) Q bounds for arbitrary small antennas: a circuit approach. IEEE Trans Antennas Propag 60(7):3120–3128

Thiele G, Detweiler P, Penno R (2003) On the lower bound of the radiation Q for electrically small antennas. IEEE Trans Antennas Propag 51(6):1263–1269

Van Bladel JG (2007) Electromagnetic fields, 2nd edn. IEEE Press, Piscataway

Vandenbosch GAE (2010) Reactive energies, impedance, and Q factor of radiating structures. IEEE Trans Antennas Propag 58(4):1112–1127

Vandenbosch GAE (2011) Simple procedure to derive lower bounds for radiation Q of electrically small devices of arbitrary topology. IEEE Trans Antennas Propag 59(6):2217–2225

Volakis J, Chen CC, Fujimoto K (2010) Small antennas: miniaturization techniques & applications. McGraw-Hill, New York

Wheeler HA (1947) Fundamental limitations of small antennas. Proc IRE 35(12):1479–1484

Yaghjian AD, Best SR (2005) Impedance, bandwidth, and Q of antennas. IEEE Trans Antennas Propag 53(4):1298–1324

Yaghjian AD, Stuart HR (2010) Lower bounds on the Q of electrically small dipole antennas. IEEE Trans Antennas Propag 58(10):3114–3121

Yaghjian AD, Gustafsson M, Jonsson BLG (2013) Minimum Q for lossy and lossless electrically small dipole antennas. Prog Electromagn Res 143:641–673

Zemanian AH (1987) Distribution theory and transform analysis: an introduction to generalized functions, with applications. Dover Publications, New York

Ziolkowski RW, Tang M-C, Zhu N (2013) An efficient, broad bandwidth, high directivity, electrically small antenna. Microwave Opt Technol Lett 55(6):1430–1434

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_21-1

Transmission-Line Based Metamaterials in Antenna Engineering

Marco A. Antoniades[a]*, Hassan Mirzaei[b] and George V. Eleftheriades[b]
[a]Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus
[b]The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada

Abstract

In this chapter, transmission-line-based metamaterials are presented, and their application to the design of passive and active antennas is outlined. Transmission-line metamaterials, also termed negative-refractive-index transmission-line (NRI-TL) metamaterials, are formed by periodically loading a transmission line with lumped-element series capacitors and shunt inductors, and it is shown that they can support both forward and backward waves, as well as standing waves with a zero propagation constant. These rich propagation characteristics form the underlying basis for their use in many antenna applications, including leaky-wave antennas, compact resonant antennas, and multiband antennas. The resonant characteristics of the NRI-TL metamaterial structures reveal how these structures can be designed to offer multiband responses whose resonant frequencies are not harmonically related while offering large degrees of miniaturization. Design equations for rapid prototyping are presented, enabling the simple design of metamaterial antennas to a given specification using standard microwave substrates and loading elements in either fully printed form or surface-mount chip components. A number of passive metamaterial antenna applications are presented, including examples of zeroth-order resonant antennas, negative-order resonant antennas, epsilon-negative antennas, mu-negative antennas, metamaterial dipole antennas, and metamaterial-inspired antennas. Active non-Foster matching networks for small antennas are also presented using negative impedance converters (NICs) and negative impedance inverters (NIIs), and it is demonstrated how these can be applied to metamaterial-inspired antennas. Finally, a new method of implementing reactive non-Foster elements using loss-compensated negative-group-delay (NGD) networks is presented that exhibits improved stability, dispersion, and achievable bandwidth.

Keywords

Negative-refractive-index transmission line (NRI-TL); Metamaterials; Electrically small antennas; Compact antennas; Resonant antennas; Planar antennas; Multiband antennas; Dispersion engineering; Composite right-/left-handed (CRLH) materials; Active antennas; Active non-Foster matching networks; Negative impedance converters; Negative impedance inverters

Introduction

The emergence of a new class of engineered electromagnetic materials, known as negative-refractive-index (NRI) metamaterials, realized using either split-ring resonators and wires or reactively loaded transmission lines, has generated great interest within the electromagnetics community, for their potential to create new devices that exhibit altogether new phenomena or improved performance characteristics compared to their conventional counterparts.

*Email: mantonia@ucy.ac.cy

DOI 10.1007/978-981-4560-75-7_21-1

The term NRI derives from the fact that these materials can have simultaneously negative material parameters (permittivity ϵ and permeability μ), and therefore a negative-refractive index, and as such they exhibit phenomena that are not readily encountered in nature, hence the prefix "meta." More generally, the material parameters can be engineered to have positive, negative, and zero values by what is known as "dispersion engineering." Furthermore, by controlling the material parameters as a function of space, this allows the electric and magnetic fields, and therefore the power flow, to be arbitrarily directed within a structure, which has opened up an entirely new field of "transformation optics" (Pendry et al. 2006). Some examples of unusual phenomena that have been obtained using metamaterials are the creation of a completely flat perfect lens (Pendry 2000), sub-wavelength resolution imaging beyond the diffraction limit (Grbic and Eleftheriades 2004), electromagnetic cloaking (Pendry et al. 2006; Zedler and Eleftheriades 2011), and backward-wave and broadside radiation from a planar leaky-wave structure (Grbic and Eleftheriades 2002; Iyer and Eleftheriades 2004).

The recent growth in metamaterial research has also resulted in an analogous increase in the amount of publications in this area. Notable among these are several books that have been published recently on metamaterials, each with a different focus (Eleftheriades and Balmain 2005; Engheta and Ziolkowski 2006; Caloz and Itoh 2006; Marques et al. 2007; Capolino 2009; Cui et al. 2010), and several antenna-related textbooks and handbooks that contain dedicated sections on metamaterials (Balanis 2008, 2012; Volakis 2007; Volakis et al. 2010). In general, metamaterials can be designed for either guided-wave applications (e.g., phase shifters, couplers, power dividers, resonators, etc.) or radiating applications (leaky-wave antennas, small resonant antennas, and active variants of these). The focus of this contribution is on *transmission-line*-based metamaterials, as these pertain to the design of passive and active resonant antennas.

The *transmission-line* approach to synthesizing NRI metamaterials relies on periodically loading conventional microwave transmission lines with lumped-element series capacitors and shunt inductors. This approach was introduced in Eleftheriades et al. (2002), Antoniades and Eleftheriades (2003), Caloz and Itoh (2003), and Sanada et al. (2004), and subsequently the materials were termed negative-refractive-index transmission-line (NRI-TL) metamaterials to reflect the fact that a negative-refractive index can be achieved by reactively loading a host transmission line. An example of such a two-dimensional NRI-TL metamaterial unit cell is shown in Fig. 1.

NRI-TL metamaterials have since proven to be very practical due to their planar nature and the large negative-index bandwidth and low transmission losses that can be achieved. Furthermore, as will be subsequently shown in this chapter, NRI-TL metamaterials exhibit propagation characteristics with

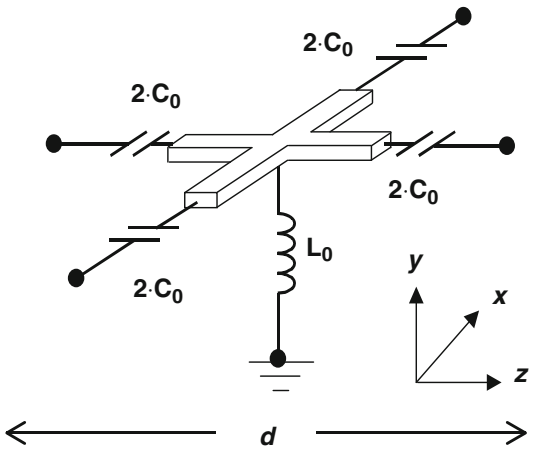


Fig. 1 Two-dimensional NRI-TL metamaterial unit cell (Iyer et al. 2003). © 2003 OSA

DOI 10.1007/978-981-4560-75-7_21-1

alternating stopbands and passbands, supporting backward waves in the left-handed band (NRI region), forward waves in the right-handed band (positive-refractive-index (PRI) region), and standing waves with a zero propagation constant between the left-handed and right-handed bands. These versatile propagation characteristics are a key element for the use of NRI-TL metamaterials in the design of multiband antennas, enabling the creation of tailored frequency responses. They also enable a large degree of miniaturization by exploiting the fact that the obtained zero propagation constant is independent of the length of the structure and the fact that in the backward-wave region, the propagation constant is inversely proportional to the frequency. The transmission-line approach to synthesizing NRI-TL metamaterials has also proven to be very useful for the integration of NRI-TL devices and antennas with other circuits on the same substrate, since they have a completely planar form factor.

Here, it should be noted that due to the relatively recent development of metamaterials, there are different terminologies that are used in the literature to denote these materials. Transmission-line based metamaterials have also been termed composite right-/left-handed (CRLH) materials (Lai et al. 2004), a terminology that reflects both their left-handed and right-handed propagation characteristics. Other designations for NRI media in which the effective permittivity and permeability are simultaneously negative include double-negative (DNG) media (Engheta and Ziolkowski 2006) and left-handed media (LHM) (Veselago 1968). Furthermore, single-negative (SNG) media denote media where one of the material parameters is negative and include two variants: epsilon-negative (ENG) media, where the permittivity ϵ is negative, and mu-negative (MNG) media where the permeability μ is negative. Finally, for epsilon-near-zero (ENZ) and epsilon-zero (EZR) media, the permittivity ϵ is near zero and zero, respectively, while for mu-near-zero (MNZ) and mu-zero (MZR) media, the permeability is near zero and zero, respectively.

Negative-Refractive-Index Transmission-Line Theory

NRI-TL Metamaterial Structure

In order to understand the operation of resonant antennas based on transmission-line metamaterials, the propagation and impedance properties of one-dimensional NRI-TL metamaterials will first be outlined. This allows the visualization of the different antenna modes that these metamaterials have to offer, while providing simple design equations for the design of single-band and multiband resonant antennas.

The one-dimensional NRI-TL metamaterial phase-shifting line shown in Fig. 2 has been constructed by periodically loading a host transmission line with a characteristic impedance of Z_0 with series capacitors C_0 and shunt inductors L_0. Here, $Z_0 = \sqrt{L/C}$, where L and C are the per unit length inductance and capacitance of the transmission line. It can be designed to occupy an arbitrary length and to provide an arbitrary phase shift, and has an inherent phase-compensating nature. This is because it incorporates both a transmission-line component with an equivalent low-pass topology that contributes a negative phase shift and a backward-wave component with a high-pass topology that contributes a positive phase shift.

The phase-shifting line consists of identical repeating symmetrical unit cells, where each of these unit cells can be considered to be a small NRI-TL metamaterial phase-compensating structure. Figure 3 shows two possible implementations of the elementary phase-compensating NRI-TL metamaterial unit cell; the T configuration was developed in Eleftheriades et al. (2002) and Antoniades and Eleftheriades (2003) and the Π configuration was subsequently developed in Elek and Eleftheriades (2005). It can be recognized that the T and Π unit cells are simply related by a shift of the reference planes defining each unit cell within the larger periodic structure shown in Fig. 2. As such, it is expected that their propagation characteristics will be identical, something which will be subsequently verified in the following section. The T unit cell has a host transmission line with characteristic impedance Z_0 and length d that is loaded with two series

DOI 10.1007/978-981-4560-75-7_21-1

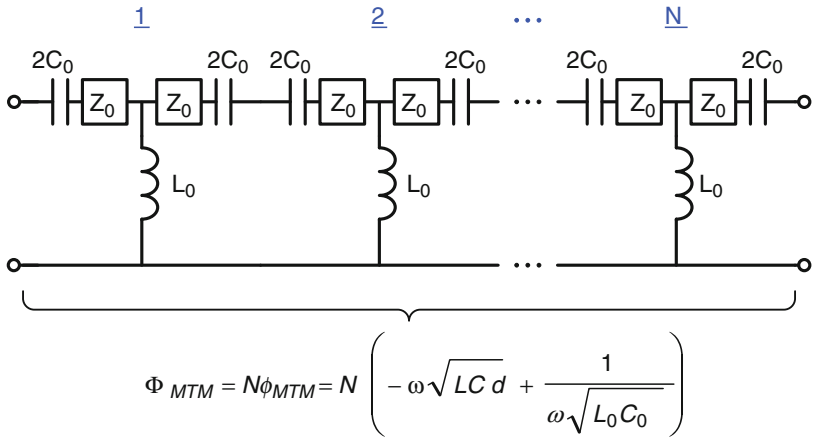


Fig. 2 *N*-stage NRI-TL metamaterial phase-shifting line

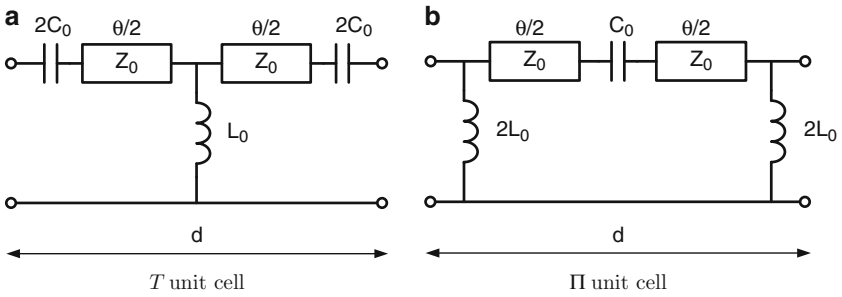


Fig. 3 NRI-TL metamaterial unit cells

capacitors $2C_0$ and a shunt inductor L_0, while the Π unit cell has a host transmission line also with characteristic impedance Z_0 and length d that is loaded with one series capacitor C_0 and two shunt inductors $2L_0$. The choice of which topology to use for a particular design mainly depends on the intended application and the technology in which the circuits are realized. Implementations of both unit cells will be shown in the section "Antenna Design Using TL-Metamaterials" for both microstrip and coplanar waveguide designs.

Propagation and Impedance Characteristics of the *T* Unit Cell

The propagation characteristics of a metamaterial line that consists of identical symmetric unit cells can be determined by conducting a periodic Bloch-Floquet analysis (Collin 1992) on each unit cell in Fig. 3. The details of the periodic analysis can be found in Antoniades (2004, 2009), and the key results are summarized here as these relate to the design of phase-shifting lines for antenna applications.

The dispersion relation for a periodic structure comprising an infinite number of metamaterial T unit cells as shown in Fig. 3a is

$$\cos(\beta_{BL}d) = \left(1 - \frac{1}{4\omega^2 L_0 C_0}\right)\cos(\theta) + \left(\frac{1}{2\omega C_0 Z_0} + \frac{Z_0}{2\omega L_0}\right)\sin(\theta) - \frac{1}{4\omega^2 L_0 C_0}. \quad (1)$$

Here, β_{BL} is the Bloch propagation constant for the periodic structure with periodicity d, as shown in Fig. 3. The Bloch propagation constant therefore provides a complete picture of the propagation characteristics of the periodic structure, which are typically displayed as a function of frequency on a dispersion diagram.

DOI 10.1007/978-981-4560-75-7_21-1

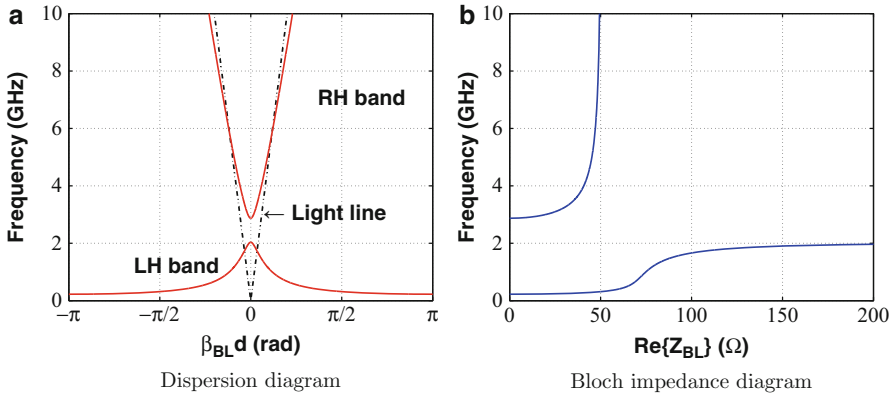


Fig. 4 Characteristics of a representative NRI-TL metamaterial T unit cell with parameters $C_0 = 5$ pF, $L_0 = 25$ nH, $Z_0 = 50\ \Omega$, $d = 3$ mm, and $\theta = 8.8°$ at 2 GHz. Note that in this case, $Z_{0,\mathrm{BW}} = \sqrt{L_0/C_0} > Z_0$

A representative dispersion diagram obtained using Eq. 1 is shown in Fig. 4a. It can be observed that the propagation characteristics of the structure exhibit alternating passbands and stopbands. The two passbands that are of interest are the lower left-handed (LH) band which supports backward waves and the upper right-handed (RH) band which supports forward waves. The light line is also shown in Fig. 4a, which demarcates the transition between slow-wave and fast-wave propagation. Therefore, the metamaterial structure supports backward and forward waves that can be either slow or fast. This indicates that the metamaterial can be used for either guided-wave applications in the slow-wave regions outside of the light cone (Islam and Eleftheriades 2007, 2012; Eleftheriades 2007; Lai et al. 2004) or leaky-wave antenna applications in the fast-wave region within the light cone (Antoniades and Eleftheriades 2008a; Mehdipour and Eleftheriades 2014; Hashemi and Itoh 2011). Furthermore, at the two frequency points where $\beta_{\mathrm{BL}} d = 0$, the metamaterial can also be used for resonant antenna applications. These, and additional resonant modes that can be excited on this structure, will be outlined in the section "Antenna Design Using TL-Metamaterials."

The characteristic impedance of the periodic structure, known as the Bloch impedance, is defined at the terminals of each unit cell and has positive and negative solutions, which correspond to the forward and reflected traveling waves, respectively. It should be noted that the Bloch impedance of a periodic structure is not unique and will depend on the location of the reference planes for each unit cell. Therefore it is expected that the two metamaterial unit cells of Fig. 3 will have different Bloch impedances. The Bloch impedance for the periodic structure comprising an infinite number of T metamaterial unit cells can be written as

$$Z_{\mathrm{BL,T}} = \pm \sqrt{\frac{\left(Z + \frac{Z^2 Y}{8} + \frac{Y Z_0^2}{2}\right)\cos(\theta) + \frac{j}{2}\left(\frac{Z^2 Y_0}{2} + Y Z Z_0 + 2Z_0\right)\sin(\theta) + \frac{Z^2 Y}{8} - \frac{Y Z_0^2}{2}}{\frac{Y}{2}\cos(\theta) + jY_0 \sin(\theta) + \frac{Y}{2}}}, \tag{2}$$

where

$$Z = \frac{1}{j\omega C_0} \quad \& \quad Y = \frac{1}{j\omega L_0}. \tag{3}$$

DOI 10.1007/978-981-4560-75-7_21-1

A representative Bloch impedance diagram for the metamaterial T unit cell obtained using Eq. 2 is shown in Fig. 4b. It can be observed that a real Bloch impedance exists only within the passbands of the periodic structure, while within the stopbands, the Bloch impedance is imaginary. It can also be observed that the Bloch impedance exhibits a large variation throughout the left-handed band and remains around 50 Ω for a very small frequency range. As such, the metamaterial structure will exhibit the undesirable feature of a very narrow impedance bandwidth when it is matched to a 50 Ω feed line, as is frequently done for antenna applications. On the contrary, in the right-handed band, the Bloch impedance converges to a constant value of 50 Ω as the frequency increases. This is a very desirable feature that enables broadband matching of the periodically loaded metamaterial line to a feed line or a terminating load.

In the design of the metamaterial unit cells, it is interesting to note that depending on the values of the elements loading the host transmission line, the Bloch impedance characteristics can be significantly different. In order to further understand the Bloch impedance behavior, the characteristic impedance of the host transmission line Z_0 must be compared with the characteristic impedance of the backward-wave line that loads the host TL, $Z_{0,\mathrm{BW}} = \sqrt{L_0/C_0}$. For the example shown in Fig. 4 with loading-element parameters of $C_0 = 5$ pF and $L_0 = 25$ nH, the characteristic impedance of the backward-wave line is $Z_{0,\mathrm{BW}} = 70.7$ Ω, which is greater than the characteristic impedance of the host transmission line, $Z_0 = 50$ Ω. Considering now the case where $Z_{0,\mathrm{BW}}$ is less than Z_0, the values of $C_0 = 15$ pF and $L_0 = 25$ nH are chosen, resulting in $Z_{0\ \mathrm{BW}} = 40.8$ Ω.

The dispersion and Bloch impedance diagrams for this second case are shown in Fig. 5. Here, it can be observed that even though the general dispersion characteristics are very similar to the ones shown in Fig. 4a, the Bloch impedance characteristics have changed significantly. Within the lower left-handed band, the Bloch impedance no longer passes through the 50 Ω point, but rather attains a maximum value of only 39 Ω, indicating that it is impossible to match the metamaterial line to 50 Ω within the lower left-handed band. This is an important consideration that should be taken into account when designing metamaterial lines using T unit cells with $Z_{0,\mathrm{BW}} < Z_0$. If, however, it is desired to match the line to 50 Ω within the upper right-handed band, then the T unit cell with $Z_{0,\mathrm{BW}} < Z_0$ is still an attractive option.

It can therefore be concluded that in order to obtain a Bloch impedance diagram of the form shown in Fig. 4b, where the metamaterial line consisting of T unit cells can be matched to a specific impedance (in this case $Z_0 = 50$ Ω) in both the lower left-handed and the upper right-handed bands, then the following condition must be satisfied:

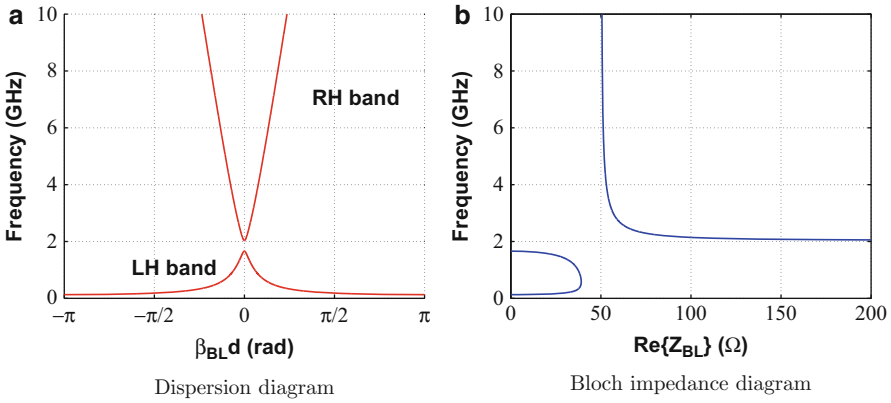


Fig. 5 Characteristics of a representative NRI-TL metamaterial T unit cell with parameters $C_0 = 15$ pF, $L_0 = 25$ nH, $Z_0 = 50$ Ω, $d = 3$ mm, and $\theta = 8.8°$ at 2 GHz. Note that in this case, $Z_{0,\mathrm{BW}} = \sqrt{L_0/C_0} < Z_0$

DOI 10.1007/978-981-4560-75-7_21-1

$$Z_{0,\mathrm{BW}} > Z_0 \quad \text{or} \quad \sqrt{\frac{L_0}{C_0}} > \sqrt{\frac{L}{C}}. \tag{4}$$

Relation to Compact Resonant Antennas

One might wonder how the analysis of an infinitely long periodic structure relates to the design of compact, resonant antennas that typically would consist of only a few unit cells. The answer lies in the fact that even though the above analysis was carried out assuming an infinitely long periodic structure, the propagation and impedance characteristics can be retained even for a finite structure, by simply terminating it on both ends in its Bloch impedance. This eliminates any reflections at the source and load and allows the propagating waves along the line to effectively see an infinite periodic medium. The periodic structure can therefore be made arbitrarily small, and in the extreme case it can consist of a single unit cell, without affecting its propagation and impedance characteristics, simply by ensuring that it is excited and terminated in its Bloch impedance. For compact antennas, which are inherently one-port devices, the terminating impedance condition can be achieved by designing the structures such that the radiation resistance is as close as possible to the Bloch impedance, which for typical antennas is chosen to be 50 Ω. As will be seen in the section "Antenna Applications of TL-Metamaterials," antenna designs based on transmission-line metamaterials enable a high radiation resistance close to 50 Ω to be achieved even for very-low-profile antennas, which also translates into the additional benefit of providing a high radiation efficiency.

Propagation Characteristics of the *Π* Unit Cell

A similar procedure can be carried out for the analysis of the metamaterial *Π* unit cell of Fig. 3b. It has been found that the dispersion characteristics of the *Π* unit cell are identical to those of the *T* unit cell and are also given by Eq. 1 (Antoniades 2009). The Bloch impedance of the metamaterial *Π* unit cell is given by

$$Z_{\mathrm{BL},\Pi} = \pm \sqrt{\frac{\frac{Z}{2}\cos(\theta) + jZ_0\sin(\theta) + \frac{Z}{2}}{\left(Y + \frac{ZY^2}{8} + \frac{ZY_0^2}{2}\right)\cos(\theta) + \frac{j}{2}\left(\frac{Y^2Z_0}{2} + YZY_0 + 2Y_0\right)\sin(\theta) + \frac{ZY^2}{8} - \frac{ZY_0^2}{2}}}, \tag{5}$$

with the same expressions for Z and Y from Eq. 3.

Representative dispersion and Bloch impedance diagrams for the metamaterial *Π* unit cell are shown in Fig. 6 for the case where $Z_{0,\mathrm{BW}} > Z_0$ and in Fig. 7 for the case where $Z_{0,\mathrm{BW}} < Z_0$. As expected, the dispersion characteristics of the *Π* unit cell for the case where $Z_{0,\mathrm{BW}} < Z_0$ shown in Fig. 7a are identical to the dispersion characteristics of the *T* unit cell shown in Fig. 5a. The Bloch impedance of the two unit cells also has very similar characteristics in the upper right-handed band, as can be verified from Figs. 5b and 7b. In the lower left-handed band, however, the Bloch impedance for *Π* unit cell with $Z_{0,\mathrm{BW}} < Z_0$ has the additional advantage that it passes through 50 Ω, thus also enabling the matching of the metamaterial line within this band. For the *Π* unit cell operating under the condition that $Z_{0,\mathrm{BW}} > Z_0$ shown in Fig. 6, the Bloch impedance does not pass through the 50 Ω point in the left-handed band; therefore, it is not possible to match the line to 50 Ω within this band. Thus, in order to match the line to 50 Ω in both the left-handed and right-handed bands, the following condition for the *Π* unit cells must be satisfied:

$$Z_{0,\mathrm{BW}} < Z_0 \quad \text{or} \quad \sqrt{\frac{L_0}{C_0}} < \sqrt{\frac{L}{C}}. \tag{6}$$

DOI 10.1007/978-981-4560-75-7_21-1

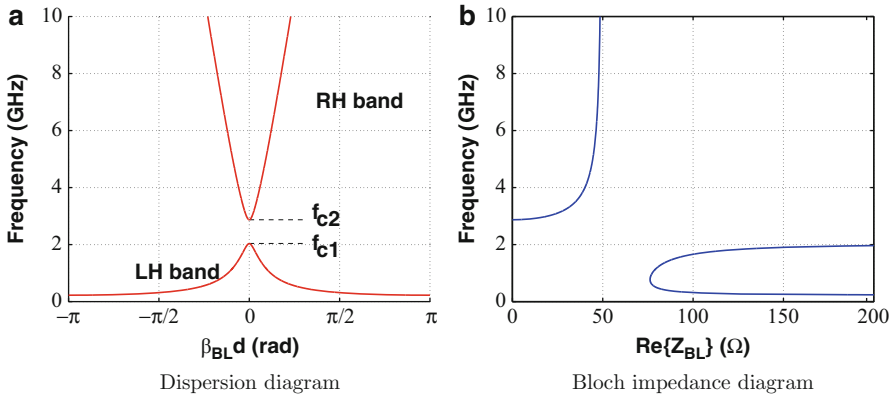


Fig. 6 Characteristics of a representative NRI-TL metamaterial Π unit cell with parameters $C_0 = 5$ pF, $L_0 = 25$ nH, $Z_0 = 50\ \Omega$, $d = 3$ mm, and $\theta = 8.8°$ at 2 GHz. Note that in this case, $Z_{0,\mathrm{BW}} = \sqrt{L_0/C_0} > Z_0$

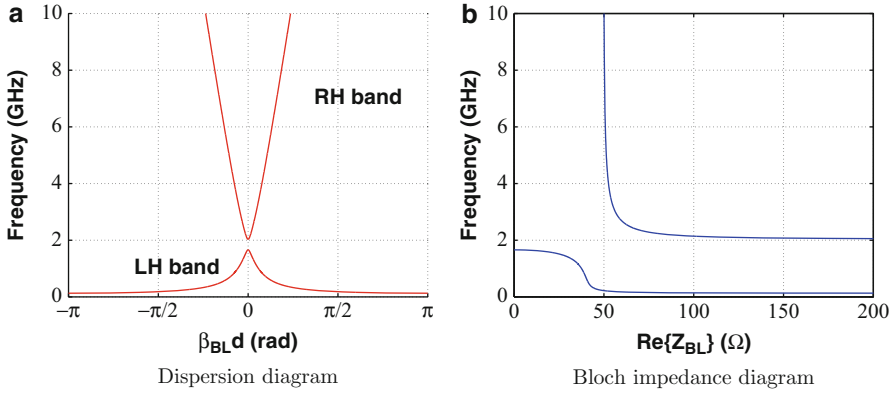


Fig. 7 Characteristics of a representative NRI-TL metamaterial Π unit cell with parameters $C_0 = 15$ pF, $L_0 = 25$ nH, $Z_0 = 50\ \Omega$, $d = 3$ mm, and $\theta = 8.8°$ at 2 GHz. Note that in this case, $Z_{0,\mathrm{BW}} = \sqrt{L_0/C_0} < Z_0$

Furthermore, the useful cases of the T unit cell with $Z_{0,\mathrm{BW}} > Z_0$ shown in Fig. 4 and the Π unit cell with $Z_{0,\mathrm{BW}} < Z_0$ shown in Fig. 7 can be compared. It can be observed that the dispersion characteristics of the Π unit cell are similar to the dispersion characteristics of the T unit cell; however, the Bloch impedances of the two unit cells exhibit complementary characteristics, as can be verified from Figs. 4b and 7b. These complementary characteristics can be attributed to the complementary forms of the Bloch impedances given by Eqs. 2 and 5.

By inspection of both Figs. 4b and 7b, it can also be observed that if the stopband were closed, then a metamaterial line that consists of either T or Π unit cells would not suffer from the rapid Bloch impedance changes around the vicinity of the stopband and would remain close to 50 Ω over a much wider bandwidth extending down to low frequencies. The method by which the stopband can be closed will be outlined in the following section, where it will be demonstrated that indeed when the stopband is closed, the Bloch impedance exhibits a significantly wider bandwidth.

Effective Medium Propagation Characteristics

The dispersion relation of Eq. 1 is sufficient to fully characterize the propagation characteristics of the NRI-TL metamaterial structure. For practical applications, it is useful to consider the case where the

DOI 10.1007/978-981-4560-75-7_21-1

periodically loaded metamaterial line can be considered an effective medium, thus providing further insight into its operation.

In order to consider a series of cascaded unit cells as an effective periodic medium, the physical length of the unit cell must be much smaller than a wavelength, translating into a small electrical length θ, i.e., $\theta \ll 1$. In addition, a small phase shift per unit cell is required in order to avoid the large phase shifts associated with the lower Bragg cutoff frequency, i.e., $\beta_{\mathrm{BL}} d \ll 1$. In order to reflect the effective nature of the periodic metamaterial medium, the Bloch propagation constant β_{BL} will be written as β_{MTM} henceforth. Thus, under the aforementioned assumptions that $\theta \ll 1$ and $\beta_{\mathrm{BL}} d \ll 1$ and considering a periodicity of d for each unit cell, the effective metamaterial propagation constant can be written as

$$\beta_{\mathrm{MTM}} = \pm\omega\sqrt{L_{\mathrm{eff}} C_{\mathrm{eff}}} = \pm\omega\sqrt{\left[L - \frac{1}{\omega^2 C_0 d}\right]\left[C - \frac{1}{\omega^2 L_0 d}\right]}. \tag{7}$$

Equation 7 demonstrates that the effective propagation constant of the metamaterial medium has a similar form to the propagation constant of a conventional transmission line but with effective inductance and capacitance terms, L_{eff} and C_{eff} given by the expressions in the square brackets of Eq. 7. Examining this equation, it can be observed how this medium can exhibit a positive, a negative, and a zero propagation constant, simply by changing the values of the loading parameters L_0 and C_0. When the loading reactance is negligible, the host transmission-line parameters L and C dominate and a forward wave is present on the structure. On the other hand, when the loading reactance is dominant and greater than the individual L and C values, L_{eff} and C_{eff} become negative, and thus the structure supports a backward wave. Finally, when the loading reactances are equal to the host transmission-line parameters L and C, then the effective propagation constant is zero, and there is no propagation along the structure. In this case, a standing wave is established along the structure that has a constant amplitude (i.e., no variation along the z-direction) that varies with time. The lumped-element equivalent circuit of the NRI-TL unit cell under the effective medium conditions $\theta \ll 1$ and $\beta_{\mathrm{BL}} d \ll 1$ is shown in Fig. 8, where L_{eff} can be considered the effective inductance of the series branch, while C_{eff} can be considered the effective capacitance of the shunt branch.

Setting each of the effective material parameters L_{eff} and C_{eff} equal to zero in Eq. 7 results in expressions for the cutoff frequencies of the lower and the upper edges of the stopband, f_{c1} and f_{c2}, respectively, as shown in Fig. 6a:

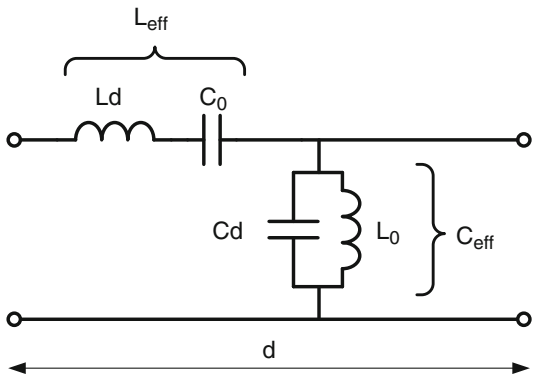


Fig. 8 Lumped-element equivalent circuit of the NRI-TL unit cell under the effective medium conditions $\theta \ll 1$ and $\beta_{\mathrm{BL}} d \ll 1$

DOI 10.1007/978-981-4560-75-7_21-1

$$f_{c1} = \min\left\{ \frac{1}{2\pi\sqrt{LC_0 d}}, \quad \frac{1}{2\pi\sqrt{L_0 Cd}} \right\} = f_{-0}, \tag{8}$$

$$f_{c2} = \max\left\{ \frac{1}{2\pi\sqrt{LC_0 d}}, \quad \frac{1}{2\pi\sqrt{L_0 Cd}} \right\} = f_{+0}. \tag{9}$$

Note that f_{c1} and f_{c2} also correspond to the zero-degree resonance frequencies f_{-0} and f_{+0}, respectively, discussed in the section "Resonant Characteristics of NRI-TL Metamaterials." Considering the scenario where $1/2\pi\sqrt{LC_0 d} < 1/2\pi\sqrt{L_0 Cd}$, the lower edge of the stopband, f_{c1}, is given by the series resonance between the total series inductance of the transmission line, Ld, and the loading capacitor C_0, while the upper edge of the stopband, f_{c2}, is given by the shunt resonance between the total shunt capacitance of the transmission-line section, Cd, and the loading inductance L_0.

By equating f_{c1} and f_{c2}, the stopband between the two cutoff frequencies can be closed, and a continuous band is formed between the lower left-handed and the upper right-handed bands. The impedance matching condition for closing the stopband in a NRI-TL metamaterial structure can therefore be written as (Eleftheriades et al. 2002)

$$Z_0 = \sqrt{\frac{L_0}{C_0}}. \tag{10}$$

The condition of Eq. 10 can also be written as

$$Z_{0,\mathrm{BW}} = Z_0 \quad \text{or} \quad \sqrt{\frac{L_0}{C_0}} = \sqrt{\frac{L}{C}}. \tag{11}$$

Thus, according to the above impedance matching condition, for a uniform transition between the left-handed band and the right-handed band, the characteristic impedance of the NRI backward-wave line components that load the line must be the same as the characteristic impedance of the PRI host transmission line.

Furthermore, under the effective medium conditions $\theta \ll 1$ and $\beta_{\mathrm{BL}} d \ll 1$, the Bloch impedances for both the T and the Π unit cells reduce to the characteristic impedance of the backward-wave line. Therefore, all three characteristic impedances are equal, resulting in a perfectly matched NRI-TL metamaterial structure:

$$Z_{\mathrm{BL}} = Z_{0,\mathrm{BW}} = Z_0. \tag{12}$$

Figure 9a shows a representative dispersion diagram obtained using Eq. 1 for either a T or a Π metamaterial unit cell when the impedance matching condition of Eq. 10 is imposed. It can be observed that compared to the open dispersion diagram of Fig. 6a, there is now a single passband comprising the lower left-handed band that smoothly transitions into the upper right-handed band at the design frequency of $f_0 = 2$ GHz. Thus, depending on the region of operation, Fig. 9a highlights the fact that the NRI-TL metamaterial structure can support forward and backward waves as well as the peculiar case of standing waves with a zero-degree phase shift ($\beta_{\mathrm{BL}} d = 0$) at the closed stopband point of $f_0 = 2$ GHz. Additionally, around the design frequency, the phase response is linear and broadband with frequency.

Figure 9b shows the Bloch impedance diagrams for both the T and π unit cells under the impedance matching condition of Eq. 10, which exhibit as expected complementary characteristics. From Fig. 9b, it

DOI 10.1007/978-981-4560-75-7_21-1

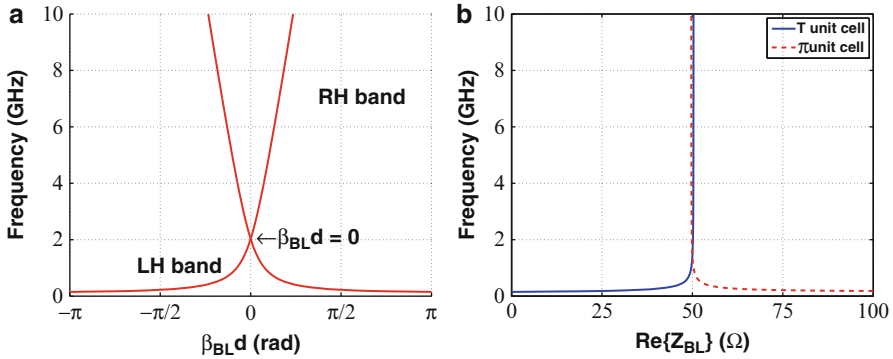


Fig. 9 Characteristics of representative NRI-TL metamaterial T and Π unit cells with parameters $C_0 = 10$ pF, $L_0 = 25$ nH, $Z_0 = 50\ \Omega$, $d = 3$ mm, and $\theta = 8.8°$ at 2 GHz. Note that in this case, the impedance matching condition of $Z_0 = \sqrt{L_0/C_0}$ is imposed, and the stopband is closed

can be observed that the Bloch impedance remains remarkably constant around 50 Ω for a much wider bandwidth compared to the open stopband cases of Figs. 4b and 7b, which suffer from rapid changes in the Bloch impedance around the vicinity of the stopbands.

The impedance matching condition of Eq. 10 can subsequently be used in Eq. 7 to derive an approximate expression for the effective propagation constant of the line, given as

$$\beta_{\text{MTM}} = \omega\sqrt{LC} + \frac{-1}{\omega\sqrt{L_0C_0}d} \tag{13}$$

$$\beta_{\text{MTM}} = \beta_{\text{H-TL}} + \beta_{\text{BW}}. \tag{14}$$

This expression can be interpreted as the sum of the propagation constants of the host transmission line, $\beta_{\text{H-TL}}$, and a uniform backward-wave line, β_{BW}, formed by the loading elements L_0 and C_0. The analogous effective phase shift per unit cell can be written as

$$\phi_{\text{MTM}} = -\beta_{\text{MTM}}d. \tag{15}$$

Inserting Eq. 13 into Eq. 15, the effective phase shift per NRI-TL metamaterial unit cell under the impedance matching condition of Eq. 10 can be written as (Antoniades and Eleftheriades 2003)

$$\phi_{\text{MTM}} = -\omega\sqrt{LC}d + \frac{1}{\omega\sqrt{L_0C_0}} \tag{16}$$

$$\phi_{\text{MTM}} = \phi_{\text{H-TL}} + \phi_{\text{BW}}. \tag{17}$$

Equation 16 can be recognized as the sum of the phases incurred by the host transmission line, $\phi_{\text{H-TL}}$, and the backward-wave line, ϕ_{BW}, which was described initially in Fig. 2. This equation succinctly describes the inherent phase-compensating nature of each NRI-TL metamaterial unit cell. By adjusting the values of the loading elements L_0 and C_0 while maintaining the impedance matching condition, the effective phase shift across each unit cell can be tailored to produce a net positive, negative, or even a 0° phase shift at a given frequency. It should be emphasized that the necessary conditions for the phase relationship presented in Eq. 16 to be valid are that the impedance matching condition of Eq. 10 must be

DOI 10.1007/978-981-4560-75-7_21-1

satisfied, the physical length of the unit cell must be small compared to the wavelength, i.e., $|\phi_{\text{H-TL}}| \ll 1$, and the phase shift per unit cell must also be small, i.e., $|\phi_{\text{MTM}}| \ll 1$.

Multistage NRI-TL Metamaterial Phase-Shifting Lines

The propagation characteristics of a single metamaterial unit cell presented in the sections "Propagation and Impedance Characteristics of the T Unit Cell" and "Propagation Characteristics of the Π Unit Cell" can be used to design multistage metamaterial lines, provided that the metamaterial line is terminated in its Bloch impedance. In this manner, each unit cell within the finite metamaterial line will effectively see an infinite periodic medium and will thus retain its propagation characteristics. Consequently, the phase expressions derived assuming an infinite periodic medium can be used in order to characterize the phase response of a finite metamaterial structure.

If the source and load terminations are equal to the characteristic impedance of the metamaterial line in the matched case (Eq. 12), then from Eq. 16, the total phase incurred by an N-stage metamaterial line as shown in Fig. 2 can be written as the sum of the phase incurred by each constituent unit cell:

$$\Phi_{\text{MTM}} = N\phi_{\text{MTM}} = N\left(-\omega\sqrt{LC}d + \frac{1}{\omega\sqrt{L_0C_0}}\right). \tag{18}$$

In order to verify the validity of the above arguments, the phase response of a representative four-stage metamaterial line was obtained using the effective medium phase shift of Eq. 18, periodic analysis of a cascade of four identical T unit cells, and the Agilent-ADS circuit simulator, and the results are shown in Fig. 10.

It can be observed that the phase responses obtained from the effective medium analysis, from the periodic analysis, and from Agilent-ADS are identical within the passband of the structure. As expected, the effective medium response of Eq. 18 does not exhibit a low-frequency cutoff. At low frequencies below 0.15 GHz, the structure has a stopband, and therefore the periodic analysis is not able to predict the phase. Nevertheless, these results confirm that the propagation characteristics derived for the infinitely periodic case can also be used to design finite-length metamaterial phase-shifting lines, provided that the lines are terminated in their Bloch impedance.

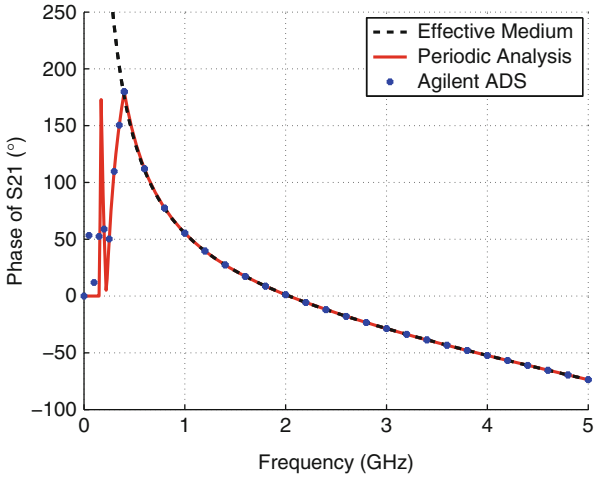


Fig. 10 Phase responses of a four-stage NRI-TL metamaterial line obtained using the effective medium phase shift of Eq. 18, periodic analysis, and the simulator Agilent-ADS. Parameters used: $Z_{\text{term}} = 50\ \Omega$, $C_0 = 10$ pF, $L_0 = 25$ nH, $Z_0 = 50\ \Omega$, $d = 3$ mm, and $\theta = 8.8°$ at 2 GHz

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_21-1

Antenna Design Using TL Metamaterials

Resonant Characteristics of NRI-TL Metamaterials

Multiband Operation

The rich dispersion characteristics of metamaterials form the underlying basis for their use in many antenna applications. In this section, the resonant characteristics of NRI-TL metamaterial structures are analyzed by considering their dispersion diagrams. These can be used to visualize how these structures can be designed to offer multiband responses whose resonant frequencies are not harmonically related while offering large degrees of miniaturization.

Figure 11 shows a representative dispersion diagram obtained using Eq. 1, where the propagation characteristics of the metamaterial can be observed. When investigating the resonant properties of this structure, of particular interest are the frequencies where the total electrical length is equal to zero and integer multiples of π (Eleftheriades 2009), i.e.,

$$\beta_{\mathrm{BL}}\ell_{\mathrm{TOT}} = n\pi \quad ; \quad n = \pm 0, \quad \pm 1, \quad \pm 2 \ldots \tag{19}$$

Thus, for resonant antenna applications, effective radiation can be achieved from the metamaterial structure when its total length ℓ_{TOT} is equal to integer multiples of half a wavelength (Schussler et al. 2004a; Caloz and Itoh 2006):

$$\ell_{\mathrm{TOT}} = n\frac{\lambda}{2} \quad ; \quad n = \pm 0, \quad \pm 1, \quad \pm 2 \ldots \tag{20}$$

In the general case where N unit cells are used in order to implement the metamaterial structure with a total length of $\ell_{\mathrm{TOT}} = Nd$, then the analogous electrical length per unit cell becomes

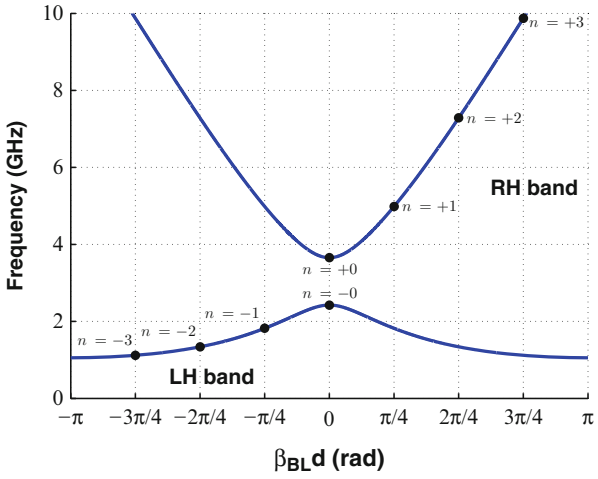


Fig. 11 Typical dispersion diagram for a NRI-TL metamaterial unit cell with parameters $C_0 = 0.14$ pF, $L_0 = 30$ nH, $Z_0 = 300\ \Omega$, $d = 12.5$ mm, and $\theta = 45°$ at 3 GHz. Superimposed are the locations of the nth order resonances for a metamaterial structure that consists of a cascade of $N = 4$ unit cells. Note the contraction of the resonant frequencies in the left-handed (LH) band compared to the right-handed (RH) band and the fact that the resonant frequencies are not harmonically related

DOI 10.1007/978-981-4560-75-7_21-1

$$\beta_{\mathrm{BL}} d = \frac{n\pi}{N} \quad ; \quad n = \pm 0, \quad \pm 1, \quad \pm 2 \ldots \pm (N-1) \tag{21}$$

An N-stage metamaterial structure can therefore support negative-order ($n < 0$) resonances in the left-handed band, zeroth-order ($n = 0$) resonances at the $\beta_{\mathrm{BL}} d = 0$ point, and positive-order ($n > 0$) resonances in the right-handed band. This is in contrast to conventional right-handed structures that can support only positive-order resonances within the right-handed region of their respective dispersion diagram.

The resonances that are achievable with a metamaterial structure are depicted graphically on the dispersion diagram of Fig. 11 for the representative case of $N = 4$ unit cells. Thus, within the irreducible Brillouin zone of $-\pi < \beta_{\mathrm{BL}} d < \pi$ on the dispersion diagram, there are potentially $2N$ resonances that can be excited. For every nth resonance that occurs at each of the $\beta_{\mathrm{BL}} d = \frac{n\pi}{N}$ points along the horizontal axis, there is an analogous resonance frequency, f_n, along the vertical axis. In general, the total number of resonances that are excited within a metamaterial structure will be less than $2N$ and will depend on the excitation mechanism and the terminal impedances. Since each nth resonance corresponds to a particular mode, i.e., a particular field distribution on the structure, the excitation mechanism will determine which modes are coupled to and therefore which modes appear in the resulting frequency response. In practical antenna applications, it is not in general possible to couple to all of the modes simultaneously using a single feed; therefore, the number of resonances that will typically appear will be less than $2N$. This can be observed in many of the examples subsequently shown in the section "Antenna Applications of TL-Metamaterials."

In order to keep the antenna size small, typically only the $n = \pm 0$ and $n = \pm 1$ resonances are employed in the design of metamaterial antennas, i.e., when $\beta_{\mathrm{BL}} \ell_{\mathrm{TOT}} = \pm 0$ and $\beta_{\mathrm{BL}} \ell_{\mathrm{TOT}} = \pm \pi$. At these frequencies, resonant standing waves are excited on the structure. Two conventional applications which employ a $+\pi$ resonance are a $\lambda/2$ dipole antenna whose current distribution has a phase shift of π along the length of the dipole and a $\lambda/2$ patch antenna which also establishes a π phase shift between its two radiating edges. As has been outlined above, structures that employ NRI-TL metamaterial loading also have the ability to excite the $+\pi$ resonance but have the additional benefit and flexibility of exciting the $\pm$zero and the $-\pi$ resonances at nonharmonic frequencies.

Compact Designs with a Small Number of Unit Cells, *N*

The $n = \pm 0$ zeroth-order resonances (ZOR) are independent of the total number of unit cells N in the structure (Antoniades and Eleftheriades 2003), since each constituent unit cell incurs in and of itself 0°. This can be seen by setting $n = 0$ in Eq. 21. Thus, any number of unit cells can be chosen in order to implement the $n = \pm 0$ resonances, including the limiting case of $N = 1$. The zeroth-order resonances are therefore well suited for antenna miniaturization purposes, as will be outlined in the subsequent section. For the general case of a dispersion relation with an open stopband, as shown in Fig. 11, the zero-degree resonance will occur at two resonant frequencies, f_{-0} and f_{+0}. Note that these two frequencies are the same as the cutoff frequencies f_{c1} and f_{c2} in Fig. 6a. Note also that in terms of nomenclature, the $n = \pm 0$ zeroth-order resonance (ZOR) frequencies are also known as zero-degree frequencies, zero-index frequencies, infinite-wavelength frequencies, epsilon-zero (EZR) frequencies, and mu-zero (MZR) frequencies.

The $\pm\pi$ resonances will depend on the total number of unit cells comprising the metamaterial structure. Thus, for a structure consisting of N unit cells, the $\pm\pi$ resonances will be excited when the electrical length per unit cell is equal to $\beta_{\mathrm{BL}} d = \pm \pi/\mathrm{N}$, from Eq. 21. It should be noted that the number of unit cells used to implement the metamaterial structure is in theory arbitrary; however, in practice it is limited by the space that is available to implement the host transmission line and the loading elements L_0 and C_0 for each metamaterial unit cell.

DOI 10.1007/978-981-4560-75-7_21-1

As will be shown in the section "Design Equations for Rapid Prototyping," a larger number of unit cells results in larger loading-element values L_0 and C_0 required in order to implement the desired phase shift per unit cell. From Eqs. 24 and 25, if N unit cells are used in order to implement the structure, then the values of the required loading elements will be N times larger than the implementation with a single unit cell. Thus, N should be kept small in order to facilitate the physical realization of the loading elements L_0 and C_0, especially if these are implemented in fully printed form. Additionally, if the loading elements L_0 and C_0 are implemented using off-the-shelf surface-mount chip components, then it is advantageous to keep their values low since these generally come in smaller packages and have higher self-resonant frequencies (see for example the Coilcraft Inc. chip inductor datasheet (Coilcraft Inc 2015)). Finally, a single unit cell realization of the $\pm\pi$ resonances should be avoided, since this corresponds to operation at the onset of the lower and upper stopband regions, which are associated with high insertion losses.

Antenna Miniaturization

Negative-Order Resonant Modes, $n < 0$

For the purpose of antenna miniaturization, the resonant frequencies that lie within the lower left-handed band on the dispersion diagram of Fig. 11 are of most interest. This is because these negative-order resonant modes with $n < 0$ are associated with large propagation constants at low frequencies and a contraction of the guided wavelength as the frequency decreases (Schussler et al. 2004a; Lee et al. 2006; Iizuka and Hall 2007; Eleftheriades 2009).

Within the lower left-handed band, the backward-wave propagation constant $\beta_{\mathrm{BW}} = -1/\omega\sqrt{L_0C_0}d$ is dominant (see Eq. 13), which is inversely proportional to the frequency. Therefore, the propagation constant increases as the frequency decreases, allowing large propagation constants to be achieved at low frequencies, facilitating the realization of the $-\pi$ resonance. Furthermore, the propagation constant is also inversely proportional to the loading-element values L_0 and C_0, therefore enabling large propagation constants to be achieved with small loading-element values. This enables the design of compact resonant antennas with loading elements that can be easily realized in printed form, while maintaining the overall antenna size much smaller than $\lambda/2$.

The above scenario is in contrast to conventional right-handed lines, where the propagation constant is directly proportional to the frequency, therefore necessitating antenna sizes that are on the order of $\lambda/2$ in order to create resonant antennas. The size of resonant right-handed antennas can be reduced by employing slow-wave low-pass loading; however, since the propagation constant is proportional to the loading-element values, the required loading-element values are significantly larger than the ones required for the metamaterial-loaded antennas with the same size reduction.

With reference to Fig. 11, note that in the left-handed band, the resonant frequencies are spaced much closer together compared to the resonances in the right-handed band, and they are *not* harmonically related. This can again be attributed to the fact that the dominant backward-wave propagation constant in this band is inversely proportional to the frequency. In the right-handed band, the propagation constant of the host transmission line $\beta_{\mathrm{H\text{-}TL}} = \omega\sqrt{LC}$ is dominant; therefore, the propagation constant is directly proportional to the frequency, and the resonant frequencies *are* harmonically related. Thus, the metamaterial structure exhibits a contraction of the resonant frequencies in the left-handed (LH) band compared to the right-handed (RH) band. Furthermore, the locations of the resonant frequencies can be directly controlled by adjusting the loading-element values L_0 and C_0, through a process called "dispersion engineering."

DOI 10.1007/978-981-4560-75-7_21-1

Zeroth-Order Resonant (ZOR) Modes, $\mathbf{n = \pm 0}$

Antenna miniaturization can also be achieved at the two points on the dispersion diagram where the propagation constant is zero, i.e., for the $n = \pm 0$ ZOR frequencies of f_{-0} and f_{+0}. In the closed stopband case shown in Fig. 9, these two frequencies merge into one ZOR frequency, $f_{\pm 0}$. Note that the two ZOR frequencies are not harmonically related and can be easily adjusted by changing the metamaterial structure's transmission line and loading parameters (see Eqs. 8 and 9). This provides great flexibility when designing ZOR antennas, since it allows the ZOR frequencies to be placed arbitrarily.

At the two ZOR frequencies, the phase shift that each metamaterial unit cell incurs is 0° and is independent of the unit cell's physical length. Therefore, any number of unit cells, N, can be cascaded together, and the resulting overall phase shift that the metamaterial structure incurs will still be 0°. Thus, in theory the zero-degree metamaterial structures can be made infinitesimally small, enabling large degrees of antenna miniaturization. In practice, the size of the zero-degree unit cells is limited by the available space required to implement the host transmission line and the loading elements L_0 and C_0, especially if these are realized in printed form. In order to achieve the equivalent -2π phase shift from a conventional transmission line, its length must be one guided wavelength, λ_g. Consequently, the physical size of the zero-degree metamaterial structures can be made significantly smaller than that of a λ_g transmission line and in the limiting case can be simply a single zero-degree metamaterial unit cell.

Note that for antenna applications, a single unit cell realization of the zero-degree resonances is not problematic, as was the case for the $\pm\pi$ resonances described in the section "Compact Designs with a Small Number of Unit Cells, N." This is because for the case of a closed stopband, the $\beta_{\mathrm{BL}}d = \pm 0$ point is not at the edge of a stopband region, and in the open stopband case, even though the two $\beta_{\mathrm{BL}}d = \pm 0$ points are at the edge of the central stopband, the impedance seen at the terminals of the antenna primarily consists of the radiation resistance of the structure. For the latter case, if the radiation resistance of a single unit cell is not high enough, this could necessitate the use of additional unit cells in order to increase the overall antenna radiation resistance.

Design Equations for Rapid Prototyping

When designing compact antennas, a designer will typically try to achieve a targeted performance within a specified antenna volume or the targeted performance could be sought while attempting to minimize the antenna volume. Alternatively, the maximum antenna performance could be sought for a specified volume. Performance metrics could include input impedance bandwidth, multiband operation, polarization, directive properties, and radiation efficiency.

For the design of compact antennas based on transmission-line metamaterials, once the mode of operation is chosen based on the analysis of the section "Resonant Characteristics of NRI-TL Metamaterials," e.g., $n = 0$ or $n = -1$, the primary concern is to obtain the loading-element values L_0 and C_0 required to realize the metamaterial. Thus, for a given antenna volume, the designer must choose the number of metamaterial unit cells N that will be used in the realization of the antenna, as well as the technology that the transmission lines will be implemented in (e.g., microstrip, coplanar waveguide, etc.). Typically, based on the volume constraints and the system impedance, the designer will also choose the parameters of the host transmission line, namely, the characteristic impedance Z_0 and the propagation constant $\beta_{\text{H-TL}}$, and the length of each unit cell d. This will in turn define the electrical length of each transmission-line section $\theta = \beta_{\text{H-TL}}d$ and the corresponding phase that each transmission-line section incurs $\phi_{\text{H-TL}} = -\theta = -\beta_{\text{H-TL}}d$. These parameters will subsequently allow the designer to calculate the loading-element values required to achieve a specific phase shift for each metamaterial unit cell. Below are outlined two methods for obtaining these loading-element values: the first method provides simple and intuitive approximate design equations derived from the effective medium case of the section "Effective Medium Propagation Characteristics," and the second method provides exact design equations derived

DOI 10.1007/978-981-4560-75-7_21-1

for the periodic medium case of the sections "Propagation and Impedance Characteristics of the T Unit Cell" and "Propagation Characteristics of the Π Unit Cell."

Approximate Expressions for NRI-TL Metamaterials

For the effective medium outlined in the section "Effective Medium Propagation Characteristics," closed-form expressions can be derived for the values of the loading elements L_0 and C_0 required to produce a desired phase shift per unit cell ϕ_{MTM} at a given frequency ω_0, given a section of host transmission line with parameters Z_0 and $\phi_{\text{H-TL}}$. Approximate expressions for the loading elements can be obtained by substituting the impedance matching condition of Eq. 10 into Eq. 16 and solving for L_0 and C_0

$$L_0 = \frac{Z_0}{\omega_0(\phi_{\text{MTM}} - \phi_{\text{H-TL}})}, \tag{22}$$

$$C_0 = \frac{1}{\omega_0 Z_0(\phi_{\text{MTM}} - \phi_{\text{H-TL}})}. \tag{23}$$

It should be noted that since the impedance matching condition of Eq. 10 was used in the derivation of Eqs. 22 and 23, these are only valid when the stopband is closed and in the effective medium limit of $|\phi_{\text{MTM}}| \ll 1$ and $|\phi_{\text{H-TL}}| \ll 1$. Nonetheless, these expressions are very simple and intuitive and clearly highlight the factors that affect the values of the loading elements.

The loading-element values can also be expressed in terms of the total phase shift that the metamaterial structure incurs. If the metamaterial structure consists of N unit cells, and the total phase shift that it incurs is $\Phi_{\text{MTM}} = N\phi_{\text{MTM}}$ given by Eq. 18, while the total phase shift that the host transmission line incurs is $\Phi_{\text{H-TL}} = N\phi_{\text{H-TL}}$, then from Eqs. 22 and 23 the loading-element values can be written as

$$L_0 = N\left(\frac{Z_0}{\omega_0(\Phi_{\text{MTM}} - \Phi_{\text{H-TL}})}\right), \tag{24}$$

$$C_0 = N\left(\frac{1}{\omega_0 Z_0(\Phi_{\text{MTM}} - \Phi_{\text{H-TL}})}\right). \tag{25}$$

These expressions reveal that from a design point of view, it is advantageous to keep N small, i.e., use a small number of unit cells, resulting in smaller values for the required loading-element values.

Exact Expressions for NRI-TL Metamaterials

For the periodic medium of the sections "Propagation and Impedance Characteristics of the T Unit Cell" and "Propagation Characteristics of the Π Unit Cell," the exact expressions for the loading elements can be obtained by employing the full dispersion relation of Eq. 1 for the general case where the stopband can be either open or closed. In both cases, it is desired to calculate the loading elements for a given phase shift per unit cell, $\phi_{\text{MTM}} = -\beta_{\text{BL}}d$, and Bloch impedance, Z_{BL}.

Closed Stopband

Considering first the case when the stopband is closed, recall from Fig. 9 that the Bloch impedance remains very close to the value of Z_0 over a large frequency range. Thus, a very good approximation for the Bloch impedance for a large range of frequencies is actually the value of Z_0. Therefore, the impedance matching condition of Eq. 10 is rearranged to obtain an expression of the loading inductance L_0 in terms of C_0 and Z_0:

DOI 10.1007/978-981-4560-75-7_21-1

$$L_0 = C_0 Z_0^2. \tag{26}$$

By substituting Eq. 26 into Eq. 1, a quadratic equation in terms of C_0 is obtained:

$$(\cos(\beta_{BL}d) - \cos(\theta))C_0^2 - \left(\frac{\sin(\theta)}{\omega_0 Z_0}\right)C_0 + \frac{1+\cos(\theta)}{4\omega_0^2 Z_0^2} = 0. \tag{27}$$

The solution to the above quadratic equation can be written as

$$C_0 = \frac{\left(\frac{\sin(\theta)}{\omega_0 Z_0}\right) \pm \sqrt{\left(\frac{\sin(\theta)}{\omega_0 Z_0}\right)^2 - 4(\cos(\beta_{BL}d) - \cos(\theta))\left(\frac{1+\cos(\theta)}{4\omega_0^2 Z_0^2}\right)}}{2(\cos(\beta_{BL}d) - \cos(\theta))}. \tag{28}$$

The positive solution of the quadratic equation is sought, therefore resulting in unique values for L_0 and Co from Eqs. 26 and 28. A particular case of Eq. 27 arises when $\beta_{BL}d$ is equal to θ (or equivalently $|\phi_{MTM}| = |\phi_{H\text{-}TL}|$). In this case, the first term of Eq. 27 is equal to zero, and the loading capacitance C_0 can be simply expressed as

$$C_0 = \frac{1+\cos(\theta)}{4\omega_0 Z_0 \sin(\theta)}. \tag{29}$$

Open Stopband

The open stopband case is the most general case and proceeds in a similar, yet slightly more complicated manner. The two design parameters in this case are the phase shift per unit cell, given by the dispersion relation of Eq. 1, and the Bloch impedance given by Eq. 2 or 5. Both the dispersion and Bloch impedance relations are functions of L_0 and C_0; however, an attempt to find a closed-form expression for these reveals that the expression quickly becomes very complicated. A simple way to overcome this problem is to solve for L_0 and C_0 graphically by expressing C_0 in terms of L_0 through the general form of the dispersion relation of Eq. 1, given in terms of Z and Y. Therefore, Eq. 1 can be written as

$$Z = \frac{\cos(\beta_{BL}d) - \cos(\theta) - j\frac{YZ_0}{2}\sin(\theta)}{\frac{Y}{4}\cos(\theta) + j\frac{Y_0}{2}\sin(\theta) + \frac{Y}{4}} = \frac{1}{j\omega C_0}. \tag{30}$$

Recall from Eq. 3 that $Z = 1/j\omega C_0$ and $Y = 1/j\omega L_0$. Substituting Eq. 30 into the Bloch impedance expressions of Eq. 2 or 5 and by sweeping the value of L_0, a range of values for the Bloch impedance is obtained. The final value of L_0 is obtained by selecting the value that corresponds to the desired Bloch impedance. The loading capacitance C_0 is subsequently found by substituting the value of L_0 into $Y = 1/j\omega L_0$ and then into Eq. 30.

In summary, NRI-TL metamaterial structures have the ability to support negative-order resonances, positive-order resonances, and zeroth-order resonances, meaning that they can incur a positive, negative, or zero phase shift depending on the values of the loading elements while maintaining a small physical size. In addition, only a few unit cells are required in order to achieve the resonant characteristics necessary to implement compact resonant antennas. Furthermore, antenna miniaturization can be achieved for structures that support negative-order resonances and/or zeroth-order resonances. With

DOI 10.1007/978-981-4560-75-7_21-1

these in mind, the following section focuses on various antenna applications that have exploited the beneficial properties of metamaterials to create multiband, compact, and efficient antennas.

Antenna Applications of TL Metamaterials

Zeroth-Order Resonant Antennas, $n = \pm 0$

A Zero-Index NRI-TL Metamaterial Folded-Monopole Antenna

The first antenna described herein that employs the $n = 0$ zeroth-order resonance of NRI-TL metamaterials is an electrically small folded-monopole antenna, as shown in Fig. 12 (Antoniades and Eleftheriades 2008b). This is also referred to as a zero-index metamaterial antenna, since at the $\beta_{\mathrm{BL}} d = 0$ point, the refractive index is equal to zero. This antenna was first presented in Eleftheriades et al. (2004) and subsequently in Eleftheriades et al. (2007) using fully printed loading elements, and variations using surface-mount lumped-element components were presented in Qureshi et al. (2005) and Eleftheriades (2007).

The metamaterial antenna consists of $N = 4$ microstrip NRI-TL metamaterial unit cells that are arranged in a 2×2 configuration over a ground plane. Each metamaterial unit cell was designed to excite the $n = 0$ zeroth-order resonant mode under a closed stopband condition at a design frequency of $f_0 = 3$ GHz. The antenna was designed based on the symmetric NRI-TL metamaterial Π unit cell shown in Fig. 3b. The Π topology was chosen because it reveals the natural monopole folding effect that the circuit provides, which is achieved by feeding the antenna from the base of one of the inductors.

In order to feed each of the vertical vias of the antenna in phase, it is required that each metamaterial unit cell incurs 0° at the design frequency, f_0. Therefore, for a chosen unit cell size, Eqs. 22 and 23 were used to design metamaterial unit cells that incurred a phase shift of $\phi_{\mathrm{MTM}} = 0°$ at 3 GHz, under the closed stopband condition of Eq. 10.

Even-Odd Mode Analysis

The proposed zero-index metamaterial antenna of Fig. 12 can be analyzed by decomposing the antenna current into a superposition of an even mode (I_e) and an odd mode (I_o). Since the antenna is symmetrical, analysis of a single unit cell that contains two vias will reveal the general characteristics of the entire antenna, namely, that the even-mode current is dominant on the structure. Even though the analysis of a single unit cell only models two folded arms (vias), nevertheless this provides a qualitative measure of the antenna performance. Furthermore, Goubau has analyzed the case of an N-element folded monopole where only one of the monopoles is excited and has demonstrated that the results can be generalized from the two-element case (Goubau 1976; Vaughan and Bach-Andersen 2003, pp. 515–518).

Beginning with the even-mode excitation, two equal voltages (V/2) are applied to each end of the metamaterial unit cell of Fig. 3b as shown in Fig. 13a. Due to the symmetry of the unit cell, this effectively places an open circuit (OC) at the center of the structure, thus producing two identical decoupled circuits. Note that in the circuit of Fig. 13a, the vertical vias connecting the inductors to ground have been explicitly included in order to highlight their importance as the main radiating elements in the structure.

Each via can then be replaced with its corresponding radiation resistance R_r and its associated inductance L_{via} can be added to the lumped inductor value L_0', resulting in a total inductance of L_0. By replacing the short transmission-line sections with their equivalent series inductance $L' = Ld$ and capacitance $C' = Cd$, the circuit of Fig. 13b is obtained. No current will pass through the series resonator formed by $L'/2$ and $2C_0$ because it is open circuited on one end; therefore, these components can be removed from the circuit. At resonance (when $\phi_{\mathrm{MTM}} = 0°$), the series resonator formed by $2L_0$ and $C'/2$

DOI 10.1007/978-981-4560-75-7_21-1

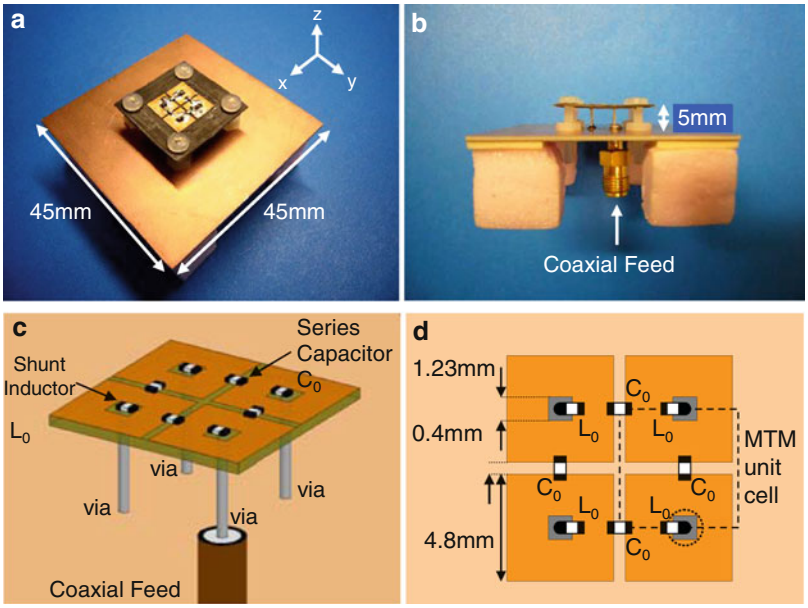


Fig. 12 Electrically small zero-index metamaterial folded-monopole antenna: (**a**) perspective view, (**b**) side view, (**c**) three-dimensional diagram, and (**d**) top view diagram. Ideal metamaterial parameters used: $N = 4, f_0 = 3$ GHz, $\phi_{\text{MTM}} = 0°$, $\phi_{\text{H-TL}} = -17.69°$, $Z_0 = 126.15\ \Omega$, $C_0 = 1.36$ pF, $L_0 = 21.67$ nH (Antoniades and Eleftheriades 2008b) © 2008 IEEE

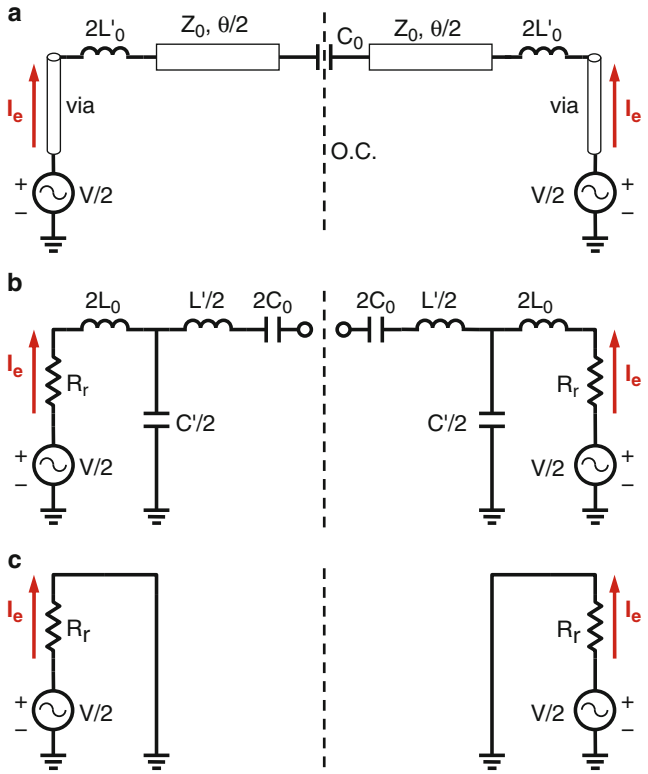


Fig. 13 Even-mode equivalent circuits for a single unit cell of the zero-index metamaterial antenna: (**a**) TL-based metamaterial unit cell, (**b**) lumped-element equivalent circuit, and (**c**) simplified lumped-element circuit at resonance (Antoniades and Eleftheriades 2008b) © 2008 IEEE

DOI 10.1007/978-981-4560-75-7_21-1

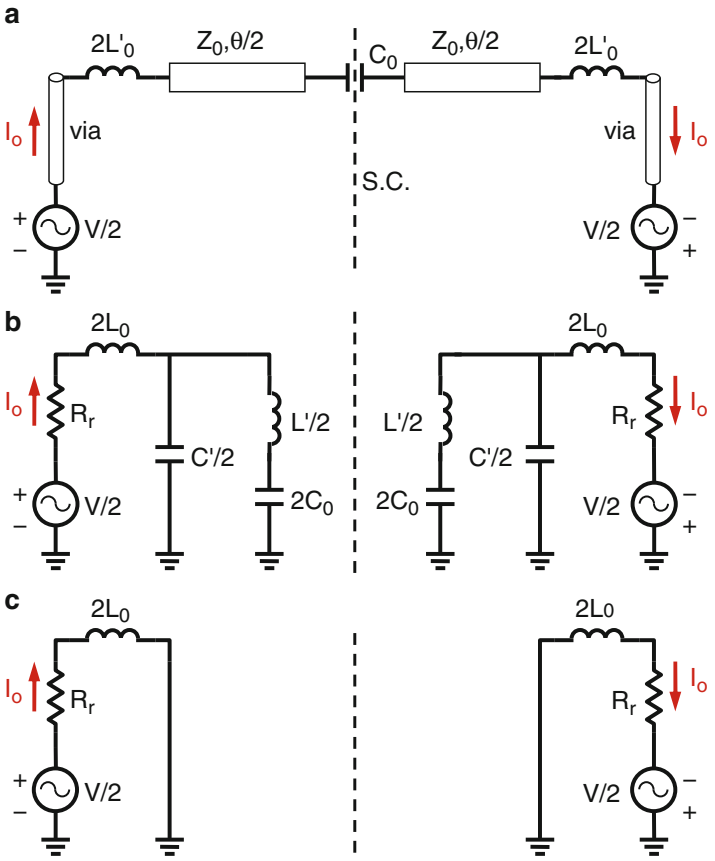


Fig. 14 Odd-mode equivalent circuits for a single unit cell of the zero-index metamaterial antenna: (**a**) TL-based metamaterial unit cell, (**b**) lumped-element equivalent circuit, and (**c**) simplified lumped-element circuit at resonance (Antoniades and Eleftheriades 2008b) © 2008 IEEE

will become a short circuit, therefore resulting in the final simplified circuit of Fig. 13c. Since no other impedances are present in the circuit other than R_r, maximum current will be delivered to the radiation resistance, given by $I_e = (V/2)/R_r$.

Considering now the odd-mode excitation, an equal and opposite voltage is applied to each end of the metamaterial unit cell as shown in Fig. 14a. Due to the symmetry of the unit cell, this effectively places a short circuit (SC) at the center of the structure, thus producing two identical decoupled circuits with equal and opposite excitations. If the vias are then replaced with their corresponding radiation resistance R_r and inductance L_{via} and the short transmission-line sections with their equivalent series inductance $L' = Ld$ and capacitance $C' = Cd$, the circuit of Fig. 14b is obtained. At resonance, the series resonator formed by $L'/2$ and $2C_0$ will become a short circuit, thus shorting out $C'/2$ and therefore resulting in the final simplified circuit of Fig. 14c. It can now be observed that the odd-mode current will be given by $I_o = (V/2)/(R_r + j\omega 2L_0)$. Therefore, the $2L_0$ inductor plays a pivotal role in adjusting the odd-mode current, and for any value of the loading inductor $2L_0$, I_e will always be greater than I_o.

Considering some representative values, a short monopole with a uniform current distribution will have a radiation resistance of (Schelkunoff and Friis 1952)

$$R_r = 160\pi^2 (h/\lambda)^2. \tag{31}$$

DOI 10.1007/978-981-4560-75-7_21-1

A close approximation to a uniform current distribution can be achieved by top loading the monopole with a metal plate, something that is inherently achieved in the microstrip design shown in Fig. 12. Therefore, for a monopole with $h/\lambda = 1/20$, the radiation resistance from Eq. 31 will be $R_r = 4\ \Omega$. Assuming that the excitation voltage is V = 1 volt, for a typical value of $L_0 = 20$ nH, at 3 GHz the magnitude of the odd-mode current is $|I_o| = 0.66$ mA. The corresponding even-mode current is $|I_e| = 127$ mA, which is approximately 190 times greater than $|I_o|$. Therefore, since the majority of the current flowing on the vertical vias lies in the even mode, this enables the metamaterial structure to act as a good radiator.

From the above discussion, it can be seen that a short monopole has a radiation resistance in the range of a few ohms, which makes it hard to match to 50 Ω. A simple way to increase the input impedance of a resonant antenna is to use a multiple folding technique (Best 2005), which relies on the fact that the currents in each arm of the antenna are in phase. As has been demonstrated with the even-odd mode analysis, this is naturally achieved with the NRI-TL metamaterial unit cell. For an antenna with N folded arms, the input impedance is given by (Best 2005)

$$R_{\text{in}} \approx N^2 R_r. \tag{32}$$

Thus, an antenna with $N = 4$ folded arms that are each $\lambda/20$ long and have $R_r = 4\ \Omega$ would have an input impedance of $R_{\text{in}} = 64\ \Omega$. This is sufficiently close to 50 Ω to provide a good match; therefore, the height of the antenna was chosen to be $h = \lambda/20 = 5$ mm.

Having established through the even-odd mode analysis that at the design frequency the zero-index metamaterial antenna maximizes the even-mode current, this enables it to be modeled as a folded monopole, thus exhibiting a substantial increase in the radiation resistance from Eq. 32. Consequently, all four vias of the antenna are excited with equal currents in terms of their magnitude and phase. Figure 15a, b shows the geometry and the simulated current distribution of the metamaterial antenna obtained using a full-wave simulator, where it can be observed that indeed all four vias are excited with equal currents. The four-unit-cell design was further extended to a design with eight unit cells as shown Fig. 15c, d, where it can again be observed that all eight vias are excited equally.

Figure 15e, f compares the input impedance of the two antennas, which verifies that indeed the radiation resistance increases according to N^2. For the four-via antenna of Fig. 15a, the full-wave simulated input resistance at resonance is $R_{\text{in}} = 70.6\ \Omega$, which is very close to the theoretical value of 64 Ω that was calculated above using Eq. 32. For the eight-via antenna of Fig. 15c, the full-wave simulated input resistance at resonance is as expected approximately four times that of the four-via antenna, i.e., $R_{\text{in}} \approx 8^2(70.6/4^2) = 282\ \Omega$. It can be noted that this result is different than the one presented in Best (2014), where the driven via is out of phase with the remaining vias. Here, all the vias are excited in phase and the folding effect of Eq. 32 is fully realized.

The final antenna of Fig. 12 has dimensions of $\lambda_0/10 \times \lambda_0/10 \times \lambda_0/20$ over a $0.45\lambda_0 \times 0.45\lambda_0$ ground plane, and it exhibits a 53 MHz −10 dB measured $|S_{11}|$ bandwidth around 3.1 GHz, as shown in Fig. 16a, and a relatively high radiation efficiency of 70 %. The radiation patterns shown in Fig. 16b, c verify that this antenna radiates a vertically polarized linear electric field vector, consistent with a vertically oriented folded-monopole geometry.

A Planar Dual-Mode NRI-TL Metamaterial-Loaded Monopole Antenna

In order to increase the bandwidth of the metamaterial antenna shown in the section "A Zero-Index NRI-TL Metamaterial Folded-Monopole Antenna," a natural extension is to create the aforementioned zero-index folded-monopole topology within the widely popular printed monopole structure. This suggests that the corresponding vertical monopole geometry of Fig. 12 is flattened into a single horizontal plane in line with the feed, which reduces significantly the height of the antenna. Figure 17 shows such a

DOI 10.1007/978-981-4560-75-7_21-1

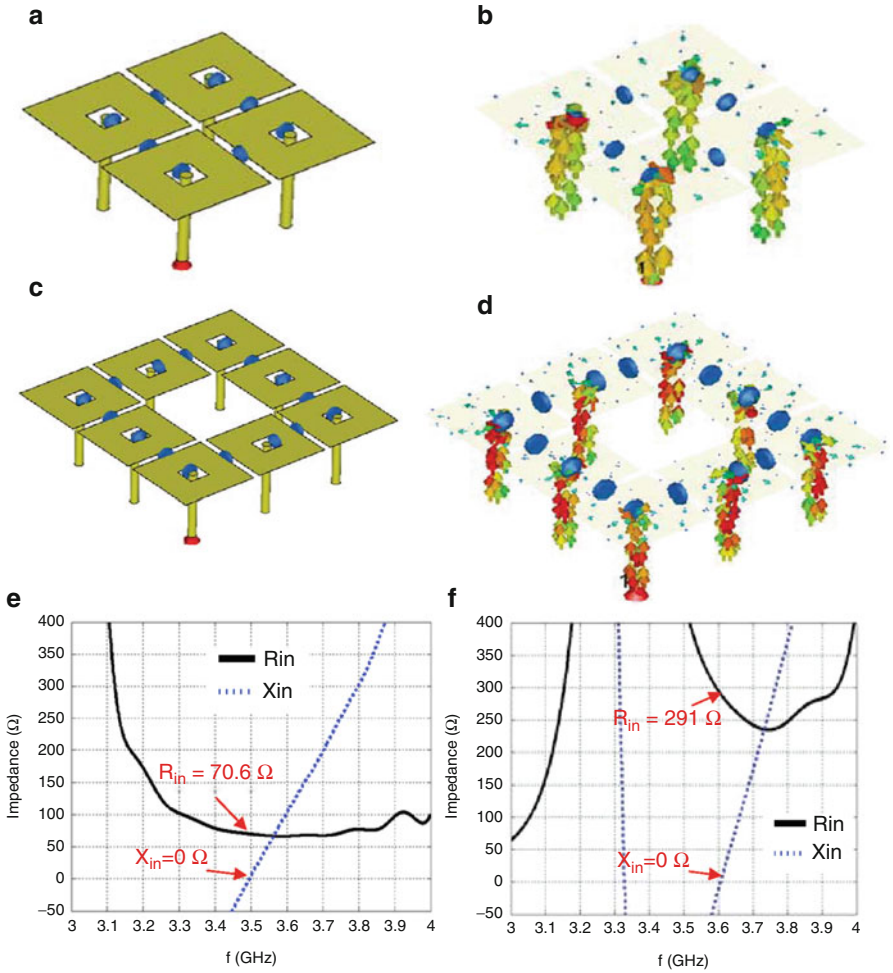


Fig. 15 (**a**) Zero-index metamaterial antenna geometry from Antoniades and Eleftheriades (2008b) with N = 4 unit cells, (**b**) current distribution at the resonant frequency of the antenna with N = 4 unit cells (all vias are in-phase), (**c**) antenna with N = 8 unit cells, (**d**) current distribution at the resonant frequency of the antenna with N = 8 unit cells (all vias are in-phase), (**e**) input impedance of the antenna with N = 4 unit cells, and (**f**) input impedance of the antenna with N = 8 unit cells. Note that the radiation resistance increases according to $R_{\text{in}} = N^2 R_r$ (Simulations courtesy of Dr. Aidin Mehdipour)

metamaterial-loaded antenna, where a single NRI-TL metamaterial unit cell has been integrated directly onto a printed monopole (Antoniades and Eleftheriades 2009).

The metamaterial unit cell was designed to operate at the $n = 0$ zeroth-order resonance under a closed stopband condition at a design frequency of $f_0 = 5.5$ GHz. Thus, at 5.5 GHz, the antenna acts as a folded monopole in the x-direction, while at 3.55 GHz, the metamaterial loading enables the top edge of the ground plane to act as a half-wavelength dipole in the orthogonal y-direction, effectively creating an additional $n = -1$ resonant mode. By virtue of the orthogonality between the two radiating modes, the metamaterial antenna exhibits an $|S_{11}|$ response with a dual resonance, as seen in Fig. 17b, and therefore a very wide measured −10 dB impedance bandwidth of 4.06 GHz.

The dual-mode operation of the antenna can be explained by considering the current distribution on the metamaterial-loaded antenna at each of the resonant frequencies, as shown in Figs. 18a and 19a. At 5.5 GHz, the metamaterial-loaded monopole was designed to achieve a zero phase shift, $\phi_{\text{MTM}} = 0°$; therefore, the currents along the monopole and along the bottom thin inductive strip were in phase. Thus, at this frequency, the metamaterial loading was used to create a two-arm folded monopole, similar to the four-arm folded monopole of the section "A Zero-Index NRI-TL Metamaterial Folded-Monopole

DOI 10.1007/978-981-4560-75-7_21-1

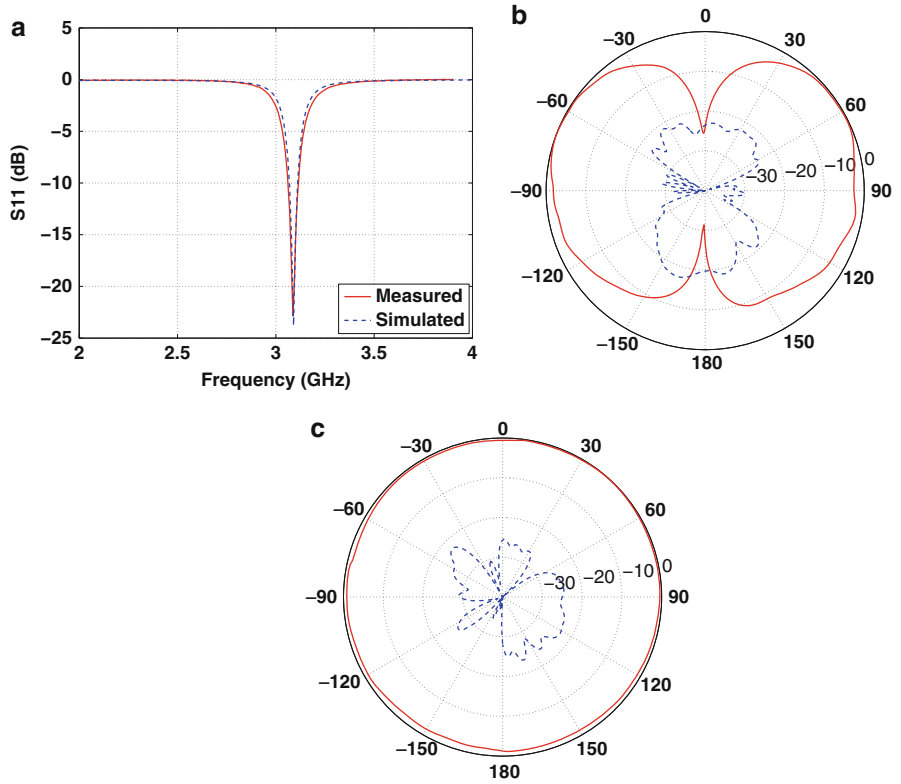


Fig. 16 (**a**) $|S_{11}|$ responses for the zero-index metamaterial antenna of Fig. 12. (**b**) E-plane (*xz*-plane) and (c) H-plane (*xy*-plane) measured radiation patterns at 3.1 GHz. *Solid line*: co-polarization, *dashed line*: cross-polarization (Antoniades and Eleftheriades 2008b) © 2008 IEEE

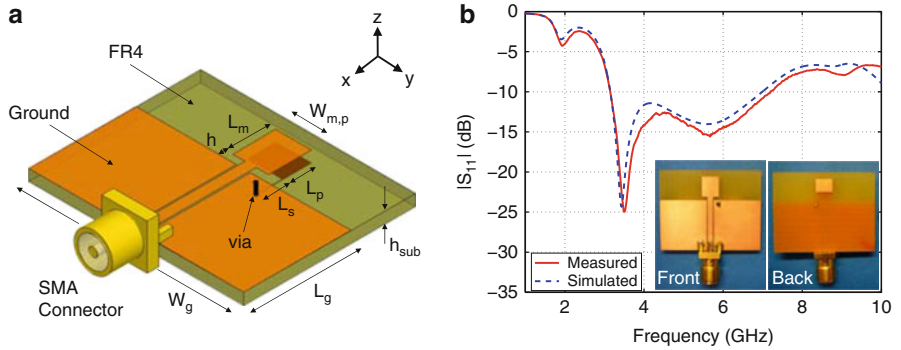


Fig. 17 (**a**) Planar dual-mode NRI-TL metamaterial-loaded monopole antenna. Dimensions (mm): $L_m = 6$, $W_m = 5$, $L_g = 15$, $W_g = 30$, h $= 1$, $L_p = 4$, $W_p = 5$, $L_s = 3.45$, $W_s = 0.1$, $h_{sub} = 1.59$, $S_{cpw} = 0.2$, $W_{cpw} = 1.55$, via diameter $= 0.5$. (**b**) $|S_{11}|$ responses for the metamaterial-loaded monopole antenna shown in the inset photographs (Antoniades and Eleftheriades 2009), © 2009 IEEE

Antenna." As was outlined above, by adjusting the value of the loading inductance L_0, it is possible to effectively eliminate the odd-mode current on the monopole, enabling the even-mode current along the *x*-direction to radiate, as shown in Fig. 18a. Additionally, at 5.5 GHz, the currents on the two ground planes are out of phase and the balanced CPW mode is preserved; therefore, these currents do not contribute to any radiation.

DOI 10.1007/978-981-4560-75-7_21-1

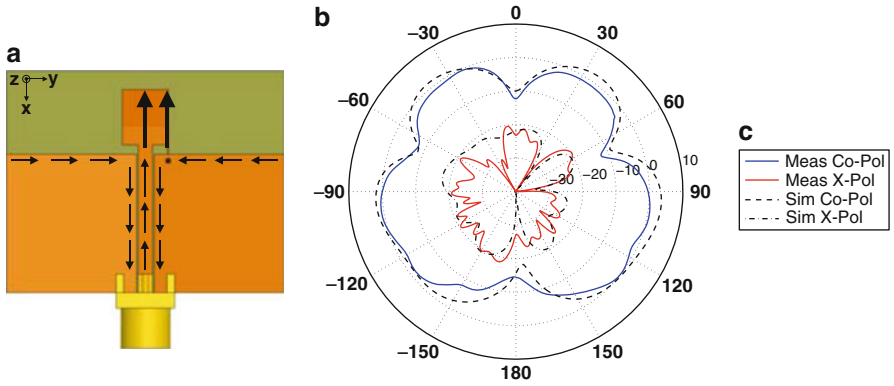


Fig. 18 (**a**) Simulated current distributions on the conductors of the metamaterial-loaded monopole antenna. (**b**) Radiation patterns in the *xy*-plane at 5.5 GHz. (**c**) Legend (Antoniades and Eleftheriades 2009), © 2009 IEEE

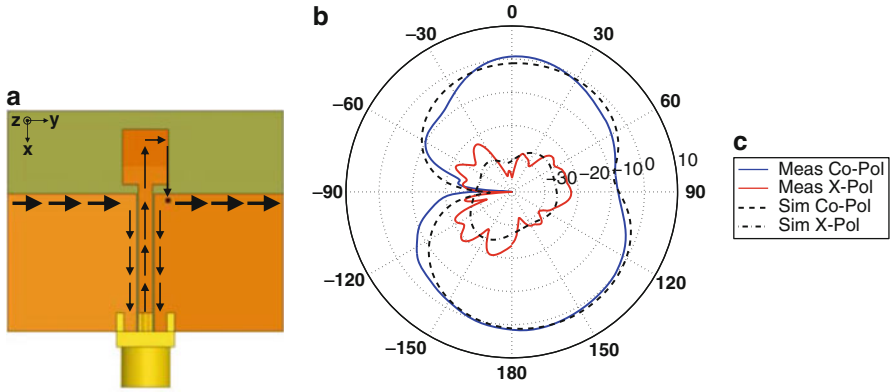


Fig. 19 (**a**) Simulated current distributions on the conductors of the metamaterial-loaded monopole antenna. (**b**) Radiation patterns in the *xy*-plane at 3.55 GHz. (**c**) Legend (Antoniades and Eleftheriades 2009), © 2009 IEEE

The radiation patterns in the *xy*-plane for the metamaterial-loaded antenna at 5.5 GHz are shown in Fig. 18b. Here, it can be observed that the antenna exhibits radiation patterns with a horizontal *x*-directed linear electric field polarization, consistent with *x*-directed currents along the monopole and the bottom thin inductive strip, shown in Fig. 18a. Thus, the radiation patterns verify that at 5.5 GHz the metamaterial loading of the monopole antenna enables it to operate as a short folded monopole.

At 3.55 GHz, the antenna no longer acts as a folded monopole along the *x*-axis, but rather as a dipole oriented along the *y*-axis. At this frequency, the metamaterial-loaded monopole acts as a balun for the ground plane currents, resulting in co-phased currents along the top edges of both the left and right ground plane sections, as shown in Fig. 19a. This renders the ground plane as the main radiating element at this frequency.

The radiation patterns in the *xy*-plane for the metamaterial-loaded antenna at 3.55 GHz are shown in Fig. 19b. At this frequency, the antenna exhibits radiation patterns with a horizontal *y*-directed linear electric field polarization, consistent with a *y*-directed current along the ground plane of the structure, as shown in Fig. 19a. Thus, the radiation patterns verify that at 3.55 GHz, the ground plane acts as the main radiating element for the antenna, providing an orthogonal polarization to the one observed at 5.5 GHz.

Thus, the metamaterial-loaded monopole antenna achieves a very-wide-band $|S_{11}|$ response by merging the resonances of the $n = 0$ zeroth-order resonant mode with the $n = -1$ resonant mode, which each provide orthogonal radiation patterns. The total size of the antenna is only $22 \times 30 \times 1.59$ mm, and the

DOI 10.1007/978-981-4560-75-7_21-1

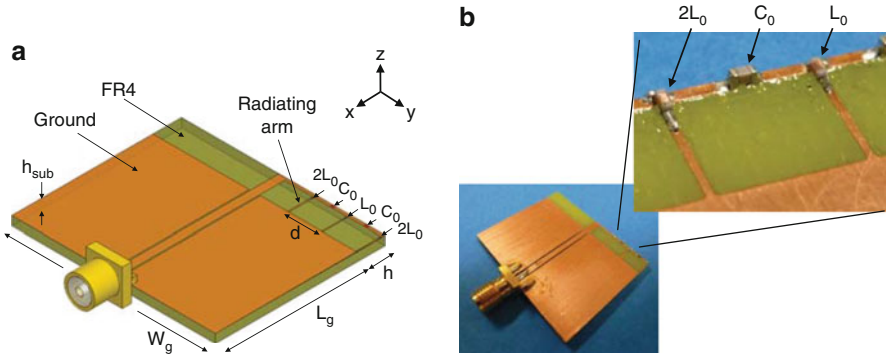


Fig. 20 Planar multi-mode NRI-TL metamaterial folded-monopole antenna. Dimensions (mm): $W_g = 36$, $L_g = 25$, $h = 5$, $d = 6$, $W_{cpw} = 1.55$, $S_{cpw} = 0.2$, $h_{sub} = 1.59$. Loading: $C_0 = 0.4$ pF, $L_0 = 12$ nH (Antoniades et al. 2013), © 2013 IEEE

measured radiation efficiency is on the order of 90 % at both 3.55 GHz and 5.5 GHz. Furthermore, its compact and low-profile design is fully printed and does not require the use of any chip lumped-element components or an external matching network.

A Planar Multi-mode NRI-TL Metamaterial Folded-Monopole Antenna

Another related antenna that is based on a fully planar topology, but which is designed to excite multiple modes, is the folded-monopole antenna shown in Fig. 20 (Antoniades et al. 2013). Here, a coplanar waveguide (CPW) transmission line has been used to feed two NRI-TL metamaterial Π unit cells that are placed asymmetrically at the top right-hand side of the structure and which are implemented using surface-mount chip components. The shunt connections to the CPW ground plane form the thin radiating arms when the antenna is excited at its zero-index folded-monopole frequency.

The metamaterial unit cells were designed to operate under an open stopband condition with a $n = 0$ zeroth-order resonance frequency around 4 GHz. At this frequency, the currents are in phase on the thin radiating arms of each unit cell, and this allows the antenna to be treated as a short multi-arm folded monopole, enabling it to be matched to 50 Ω, even though the height of each arm is only $\lambda_0/14$. Similar to the antenna presented in the section "A Planar Dual-Mode NRI-TL Metamaterial-Loaded Monopole Antenna," Fig. 21a (top) shows the current distribution at 4.18 GHz, where it can be observed that indeed at this zero-index frequency, the currents on all of the radiating arms of the metamaterial unit cells, as well as the CPW feed line, are in phase and are along the x-direction. Figure 21a (bottom) shows the current distribution at 1.92 GHz, where it can be observed that at this frequency an orthogonal dipolar mode is excited along the y-direction, equivalent to the $n = -1$ resonant mode. These two resonances can be observed in the $|S_{11}|$ response shown in Fig. 21b, in addition to another two resonances at 0.98 GHz and 1.43 GHz. The resonance at 1.43 GHz is equivalent to the $n = -2$ mode, while the resonance at 0.98 GHz is a mode created by the CPW ground plane.

The planar metamaterial folded-monopole antenna of Fig. 20 has a total size of $30 \times 36 \times 1.59$ mm^3, while the height of each radiating arm is only 5 mm. It exhibits three resonances that are matched below −10 dB, at 1.43 GHz, 1.95 GHz, and 3.81 GHz, with associated bandwidths of 25 MHz, 110 MHz, and 405 MHz, respectively, with measured radiation efficiencies in the range of 70–92 %.

Other Zeroth-Order Resonant Antennas

There are numerous other examples of antennas reported in the literature that have used the $n = 0$ zeroth-order resonance in their design to achieve beneficial characteristics such as size reduction, monopolar radiation, and multiband performance. Notable among these are the designs outlined below.

DOI 10.1007/978-981-4560-75-7_21-1

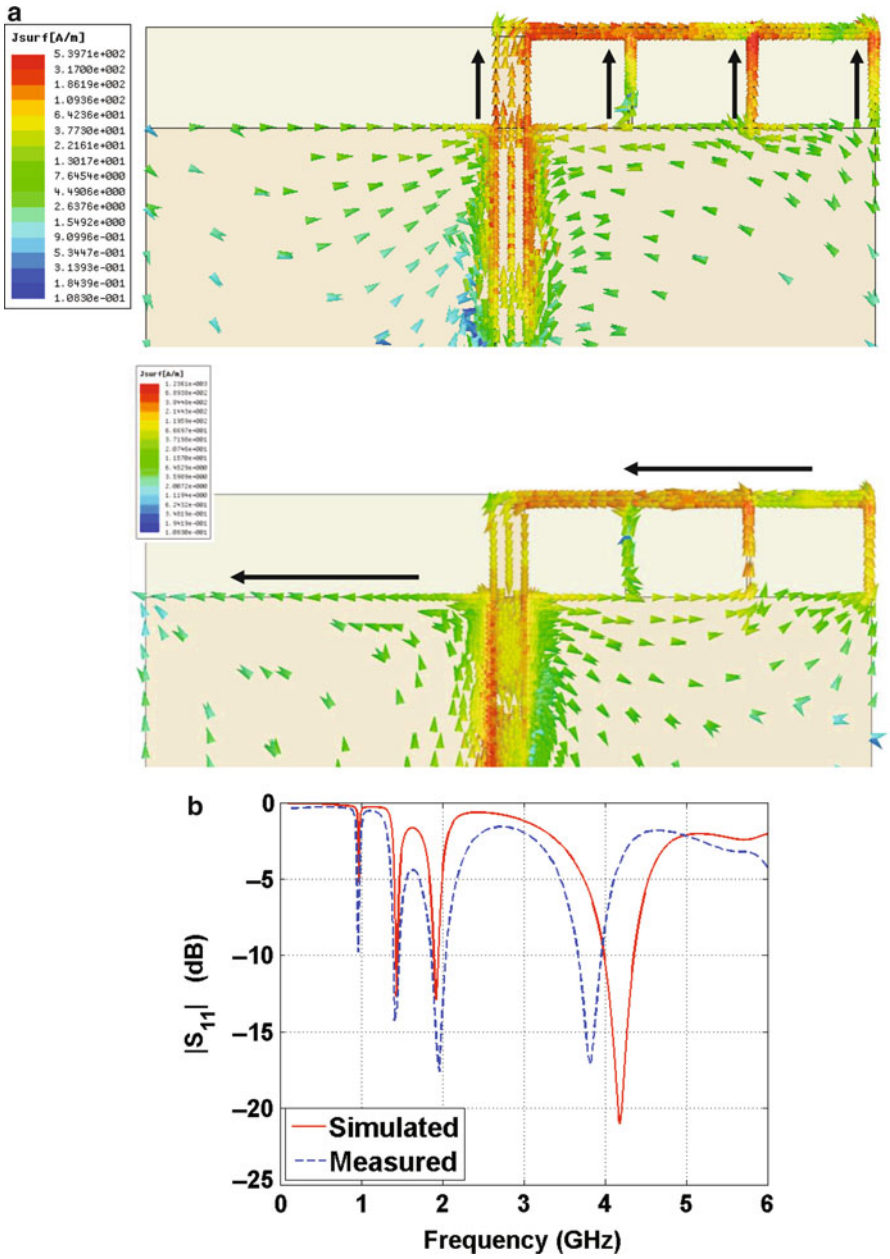


Fig. 21 (**a**) Simulated current distributions: (*top*) at 4.18 GHz, (*bottom*) at 1.92 GHz. (**b**) $|S_{11}|$ responses for the multi-mode folded-monopole antenna of Fig. 20 (Antoniades et al. 2013), © 2013 IEEE

Among the fully printed designs, a compact monopole antenna with single-cell metamaterial loading is described in Zhu et al. (2010), in which a planar monopole antenna is loaded with a via-less NRI-TL metamaterial unit cell, similar to the antenna described in the section "A Planar Dual-Mode NRI-TL Metamaterial-Loaded Monopole Antenna," and a defected ground plane is used in order to achieve a tri-band response with high radiation efficiencies. A zeroth-order resonant antenna is presented in Lee (2011) that comprises two fully printed metamaterial unit cells implemented in an asymmetric coplanar waveguide configuration. The antenna achieves in-phase horizontally directed currents on both of its shunt inductive strips at the ZOR frequency, which results in radiation patterns with a horizontal linear electric field polarization analogous to that of a short dipole directed along the axis of the inductive strips. A compact zeroth-order resonant antenna based on a fully printed dual-arm spiral configuration is

DOI 10.1007/978-981-4560-75-7_21-1

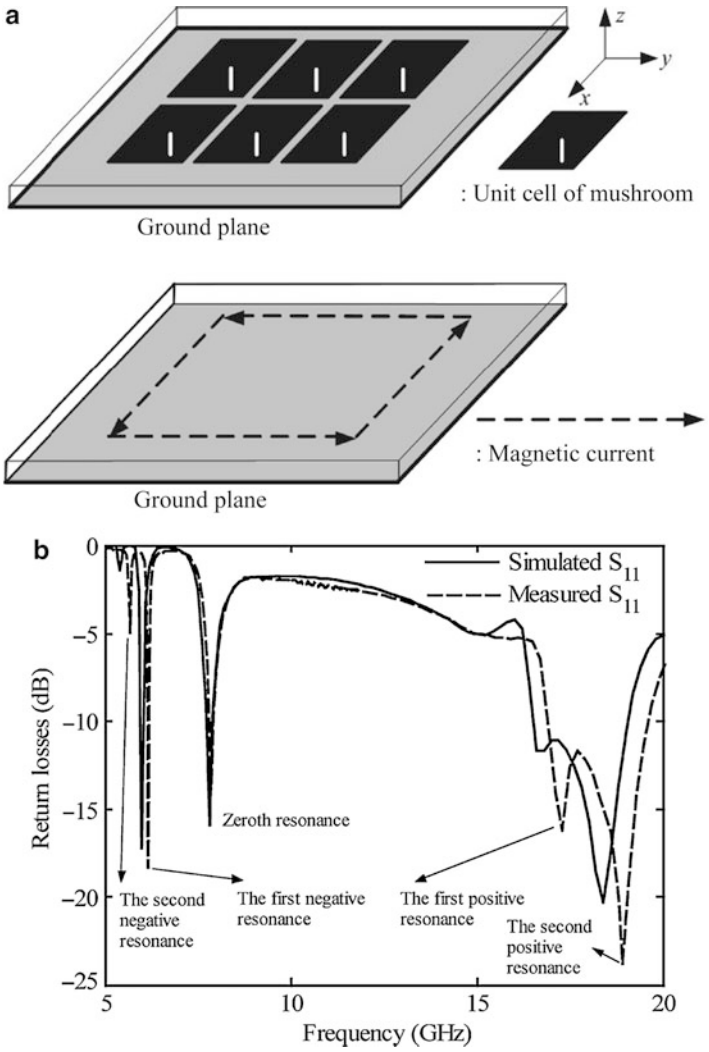


Fig. 22 (**a**) Proposed 3 × 2 mushroom ZOR antenna: (*top*) layout, (*bottom*) equivalent magnetic loop current. (**b**) $|S_{11}|$ responses (Lee and Lee 2007) © 2007 IEEE

presented in Liu et al. (2012), which is very compact and also exhibits dipole-like radiation patterns, albeit with a narrow bandwidth. In Bertin et al. (2012), three versions of fully printed metamaterial-loaded monopoles are presented, which exhibit monopolar radiation patterns and multiband responses that cover the DCS-1800, UMTS, Wi-Fi, and part of the LTE frequency bands.

Several designs have also been reported that are based on the Sievenpiper mushroom structure (Sievenpiper et al. 1999). In Lai et al. (2007), a microstrip mushroom structure is used to implement one-dimensional infinite-wavelength resonant antennas with monopolar radiation patterns, which exhibit resonances at both the $n = -1$ and $n = 0$ ZOR frequencies. The mushroom structure of Lai et al. (2007) was further miniaturized in Baek and Lim (2009) by etching a spiral slot in the ground plane, thus increasing the value of the shunt inductive loading at the expense of degraded radiation patterns. A two-dimensional version of the mushroom structure is presented in Lee and Lee (2007), seen in Fig. 22a, and it is shown that this structure emulates a horizontal magnetic loop current over a perfect electric conductor, thus producing monopolar radiation patterns at its ZOR frequency. The $|S_{11}|$ response of the antenna, shown in Fig. 22b, exhibits a $n = 0$ zeroth-order resonance at around 7.8 GHz, in addition to $n = -2$, $n = -1$, $n = 1$, and $n = 2$ resonances.

DOI 10.1007/978-981-4560-75-7_21-1

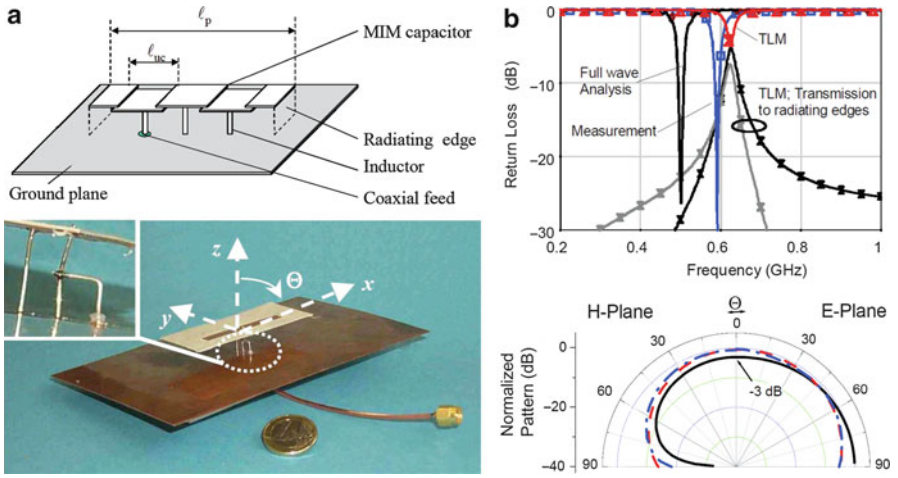


Fig. 23 (**a**) Compact metamaterial-loaded patch antenna. (**b**) (*Top*) $|S_{11}|$ response (*bottom*) normalized radiation patterns. *Black solid line*, simulated on an infinite ground plane; *red dotted line*, simulated on a finite ground plane; *blue dash-dot line*, measured on a finite ground plane (Schussler et al. 2004a) © 2004 IEEE

Negative-Order Resonant Antennas, $n < 0$

A Compact Metamaterial-Loaded Patch Antenna

As outlined in the section "Resonant Characteristics of NRI-TL Metamaterials," when operated in the left-handed band, i.e., with negative-order resonances, NRI-TL metamaterials can be used to miniaturize the size of conventional antennas that would normally require a resonant length of $\lambda/2$ to achieve a π phase shift across their length. Thus, the size of a conventional patch antenna that has a π phase shift between its two radiating edges can be significantly reduced by using NRI-TL metamaterials to achieve the equivalent $-\pi$ phase shift between its two radiating edges or equivalently the $n = -1$ resonant mode.

The first antenna described herein that employs the $n = -1$ resonant mode of NRI-TL metamaterials to shrink the physical size of a microstrip patch antenna is presented in Schussler et al. (2004a, b) and is shown in Fig. 23a. The fabricated prototype shown in Fig. 23a consists of two microstrip unit cells that are fed using a tapped-inductor feeding technique. Here, the shunt loading inductor L_0 has been realized using vertical wires and the series loading capacitor C_0 has been realized using a metal-insulator-metal (MIM) capacitor structure. The configuration of Fig. 23a can also be recognized as the one-dimensional version of the Sievenpiper mushroom structure, albeit with different loading compared to the $n = 0$ designs outlined in the section "Other Zeroth-Order Resonant Antennas."

The fabricated prototype of the metamaterial patch antenna has dimensions of $\lambda_0/40 \times \lambda_0/15 \times \lambda_0/40$ over a 0.6×0.6 mm^2 ground plane, and the $|S_{11}|$ response shown in Fig. 23b shows that the metamaterial patch antenna exhibits a simulated $n = -1$ resonance at 0.5 GHz with a -10 dB bandwidth of 1.5 %, while the measured resonance occurs at approximately 0.6 GHz. The measured radiation patterns in the E-plane (*xz*-plane) and the H-plane (*yz*-plane), also shown in Fig. 23b, indicate that the antenna radiates in the same manner as a conventional patch antenna, but with reduced efficiency and gain.

A Metamaterial-Based Low-Profile Broadband Mushroom Antenna

A two-dimensional antenna that uses a negative-order resonance in order to achieve a π phase shift between its radiating edges is presented in Liu et al. (2014). The antenna, shown in Fig. 24a, is a Sievenpiper mushroom structure that consists of a 4×4 array of unit cells over a ground plane that is fed by a microstrip line through a slot cut into the ground plane. The distinguishing feature of this antenna is that by placing the feeding slot directly beneath the center gap between the mushroom cells, both the

DOI 10.1007/978-981-4560-75-7_21-1

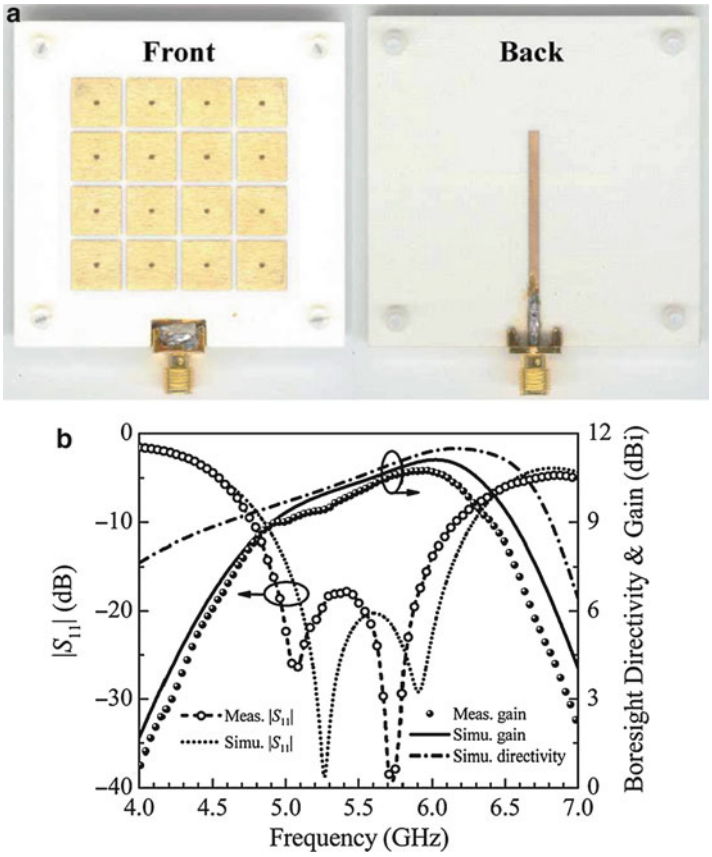


Fig. 24 (**a**) Photograph of the slot-fed metamaterial mushroom antenna. (**b**) $|S_{11}|$, directivity, and gain responses (Liu et al. 2014) © 2014 IEEE

TM_{10} and the TM_{20} modes can be excited simultaneously at nonharmonic adjacent frequencies, thus providing a broadband $|S_{11}|$ response, as shown in Fig. 24b, while maintaining radiation at broadside throughout the operating band.

Figure 25a shows the simulated electric field distribution for the mushroom antenna at 4.97 GHz, where it can be observed that the field distribution resembles the TM_{10} mode of a conventional patch antenna. A notable difference, however, is that the electric fields in the gaps between the mushroom cells and at the radiating edges at the two open ends of the antenna are in phase, therefore producing radiation. Thus, the quality factor of the mushroom antenna decreases compared to the conventional patch antenna, leading to a bandwidth enhancement.

Figure 25b shows the simulated electric field distribution for the mushroom antenna at 5.98 GHz, where it can be observed that the field distribution resembles the antiphase TM_{20} mode of a conventional patch antenna. The slot excitation on the ground plane beneath the central mushroom gap enforces opposite electric fields across the central slot region, thus exciting the antiphase TM_{20} mode with out-of-phase E_z components in the central region. This, in turn, results in co-phased electric fields in the gaps between the mushroom cells and at the radiating edges at the two open ends of the antenna, therefore also producing radiation.

The metamaterial mushroom antenna of Fig. 24a has an overall size of $60 \times 60 \times 4.1$ mm^3 and attains a measured −10 dB bandwidth of 25 %, from 4.77 to 6.16 GHz, with an average gain of 9.9 dBi. Across the bandwidth, the measured antenna efficiency is greater than 76 %, and the cross-polarization levels are less than 20 dB.

DOI 10.1007/978-981-4560-75-7_21-1

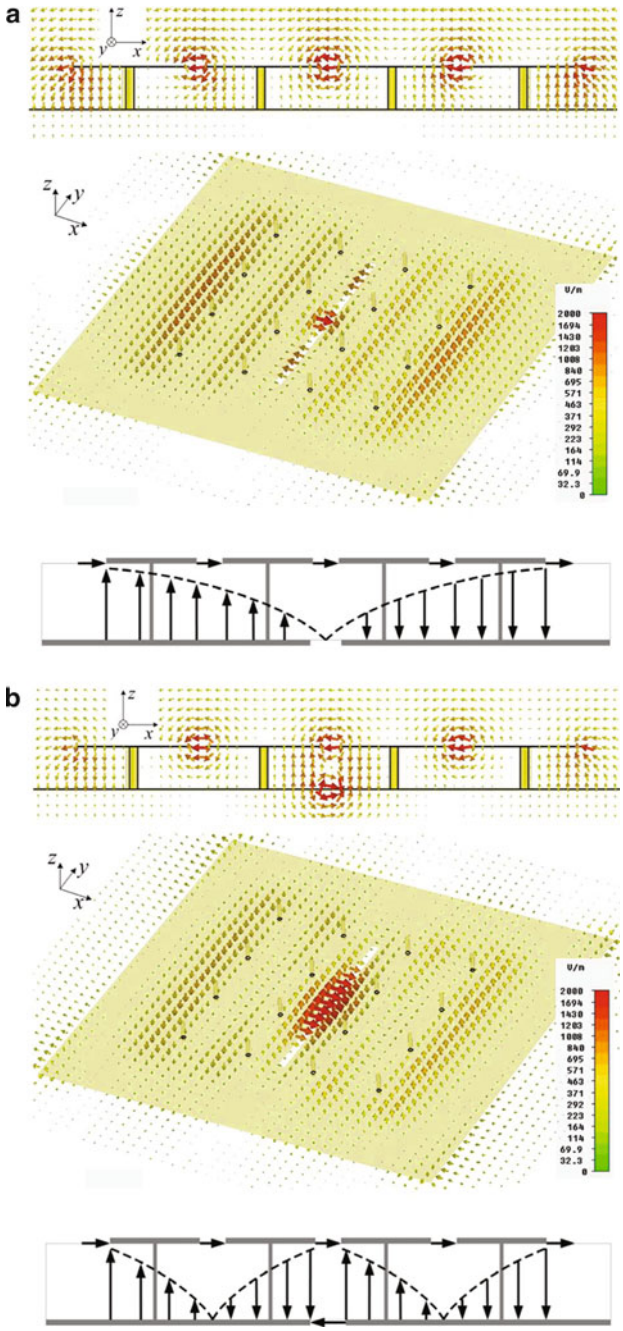


Fig. 25 (**a**) TM_{10} mode and (**b**) antiphase TM_{20} mode: (*top*) simulated E-field distribution, (*bottom*) sketch of the operating mechanism (Liu et al. 2014) © 2014 IEEE

Metamaterial-Loaded Substrate Integrated Waveguide Slot Antennas

Another type of antenna that uses negative-order resonances in order to achieve size miniaturization is shown in Fig. 26a (Dong and Itoh 2010). The miniaturized waveguide slot antenna consists of a substrate integrated waveguide (SIW) that has a slot etched on its surface. The slot acts as a series loading capacitor C_0 and forms the main radiating element, while the shunt inductive posts of the SIW provide the shunt inductive loading L_0, thus enabling this type of antenna to exhibit NRI-TL metamaterial characteristics. By operating the antenna at its negative-order resonant frequencies, which are well below the initial waveguide cutoff frequency, this leads to a significant degree of miniaturization. Two different versions of

DOI 10.1007/978-981-4560-75-7_21-1

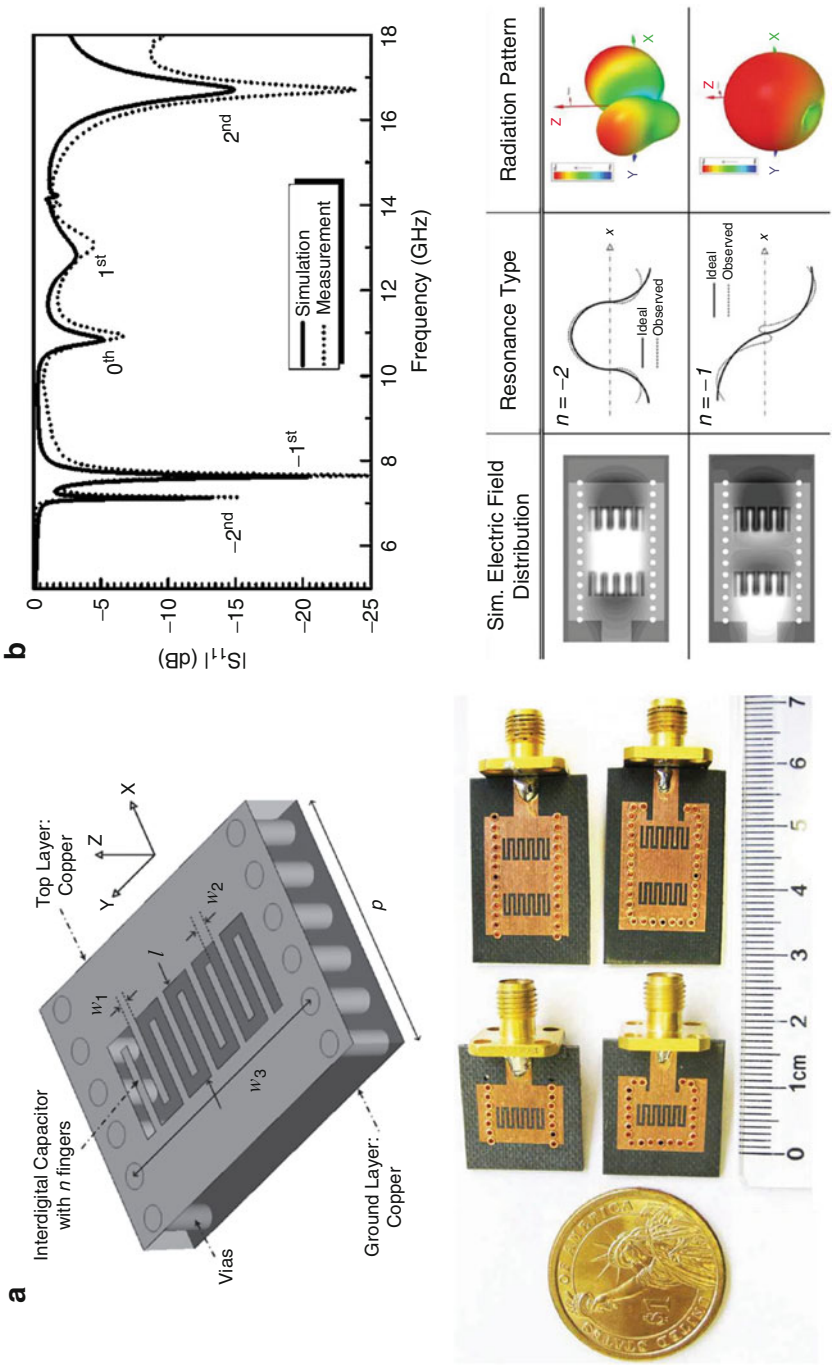


Fig. 26 (**a**) Metamaterial-loaded substrate integrated waveguide slot antennas, (**b**) $|S_{11}|$ responses, simulated electric field distributions, and radiation patterns at the $n = -2$ and $n = -1$ resonances (Dong and Itoh 2010) © 2010 IEEE

DOI 10.1007/978-981-4560-75-7_21-1

the antenna are presented, one which is open ended and one which is short ended, as shown in the prototypes on the bottom of Fig. 26a. The short-ended antenna, which represents a quasi-quarter-wavelength resonator, is essentially a cavity-backed slot antenna and therefore exhibits a high gain, while the open-ended antenna has a small size owing to its quasi-half-wavelength operation in the left-handed region.

The characteristics of a two-stage open-ended slot antenna are shown in Fig. 26b. From the $|S_{11}|$ responses, the multiband nature of the antenna can be observed, where the five resonances $n = -2$, $n = -1$, $n = 0$, $n = 1$, and $n = 2$ can be identified. The field distributions and radiation patterns for the $n = -2$ and $n = -1$ modes shown on the bottom of Fig. 26b indicate that the antenna can exhibit both dipolar and patch-type radiation characteristics at closely spaced resonant frequencies. The measured bandwidths of the antennas range from 1.5 % to 2.6 %, the measured gain from 3.2 to 6.8 dBi, and the measured efficiencies from 77 % to 91 %. The overall sizes of the antennas range from $0.265\lambda_0 \times 0.318\lambda_0 \times 0.03\lambda_0$ to $0.506\lambda_0 \times 0.343\lambda_0 \times 0.03\lambda_0$.

Other Negative-Order Resonant Patch Antennas

Various other metamaterial patch antennas have been reported in the literature that exploit the advantages of negative-order resonances that NRI-TL metamaterials have to offer (Lee et al. 2006; Tretyakov and Ermutlu 2005; Alu et al. 2007; Herraiz-Martinez et al. 2008a; Wang et al. 2010; Dong et al. 2011). One- and two-dimensional versions of metamaterial patch antenna designs that have used negative-order resonances to achieve size reduction and multiband responses have been shown in Lee et al. (2006). Additionally, in Tretyakov and Ermutlu (2005), a theoretical analysis has been presented on the benefits of using transmission-line-based metamaterials to create patch antennas with enhanced bandwidth characteristics and reduced size. Finally, in Alu et al. (2007), the possibility of designing sub-wavelength resonant patch antennas using metamaterials has been theoretically analyzed, and it is demonstrated that these designs may exhibit in principle an arbitrarily low resonant frequency for a fixed dimension, but they may not necessarily radiate efficiently when their size is electrically small. By employing a circular patch geometry, it has been shown that specific modes can be selected that enable the patch antenna to be compact, but yet exhibit comparable radiation performance to that of a conventional patch antenna of standard dimensions.

Epsilon-Negative (ENG) Antennas

Another transmission-line-based metamaterial structure that supports a zeroth-order resonant mode is the so-called epsilon-negative (ENG) structure shown in Fig. 27a in its distributed form and in Fig. 27b in its lumped-element form. It can be observed that this topology is simply an NRI-TL metamaterial unit cell

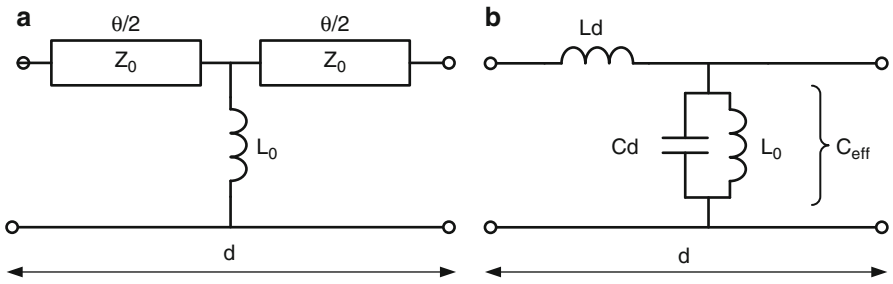


Fig. 27 Epsilon-negative (ENG) metamaterial unit cell: (**a**) symmetric distributed equivalent circuit, (**b**) lumped-element equivalent circuit

DOI 10.1007/978-981-4560-75-7_21-1

from Fig. 3 with the series loading capacitor C_0 removed, resulting in a shunt inductor-loaded transmission line.

Considering the dispersion diagram of Fig. 11, it has been outlined in the section "Resonant Characteristics of NRI-TL Metamaterials" that the NRI-TL metamaterial structure exhibits two $n = \pm 0$ zeroth-order resonant frequencies f_{-0} and f_{+0}, which can be adjusted by changing the values of the loading elements L_0 and C_0, according to Eqs. 8 and 9. Removing the series loading capacitor C_0 is equivalent to letting its value go to infinity, $C_0 = \infty$; therefore, the zeroth-order resonant frequencies become

$$f_{-0} = 0, \tag{33}$$

$$f_{+0} = \frac{1}{2\pi\sqrt{L_0(Cd)}}. \tag{34}$$

These results indicate that the ENG metamaterial structure has only one zeroth-order resonant frequency, which is determined by the values of the shunt elements in the unit cell of Fig. 27b. Therefore, for a fixed length of host transmission line, the zeroth-order resonant frequency f_{+0} can be easily tuned by simply adjusting the value of the loading inductance L_0. Additionally, like the NRI-TL metamaterial structure, the zeroth-order resonant frequency of the ENG structure is also independent of its physical length; however, since the loading capacitor C_0 has been eliminated, this allows larger degrees of antenna miniaturization to be achieved.

The dispersion relation for the ENG metamaterial structure of Fig. 27a can be obtained by setting $C_0 = \infty$ in the dispersion relation of the NRI-TL metamaterial structure given by Eq. 1 and can be written as

$$\cos(\beta_{\mathrm{BL}} d) = \cos(\theta) + \left(\frac{Z_0}{2\omega L_0}\right)\sin(\theta). \tag{35}$$

Under the effective medium conditions $\theta \ll 1$ and $\beta_{\mathrm{BL}} d \ll 1$, Eq. 35 becomes

$$\cos(\beta_{\mathrm{BL}} d) \approx 1 - \frac{1}{2}\left(\omega^2 LCd^2 - \frac{Ld}{L_0}\right). \tag{36}$$

Using Eq. 7, the effective propagation constant of the ENG metamaterial structure can be written as

$$\beta_{\mathrm{ENG}} = \omega\sqrt{L_{\mathrm{eff}} C_{\mathrm{eff}}} = \omega\sqrt{L\left[C - \frac{1}{\omega^2 L_0 d}\right]}, \tag{37}$$

$$\beta_{\mathrm{ENG}} = \sqrt{\omega^2 LC - \frac{L}{L_0 d}}. \tag{38}$$

Equation 37 demonstrates that with the appropriate value of the shunt inductor L_0, the effective shunt capacitance C_{eff} of the ENG structure (see Fig. 27b) can be made equal to zero, resulting in a zero propagation constant and therefore a zeroth-order resonance. Exactly at the zeroth-order resonance frequency f_{+0}, the effective permittivity of the structure is equal to zero and is therefore termed the epsilon-zero (EZR) point.

DOI 10.1007/978-981-4560-75-7_21-1

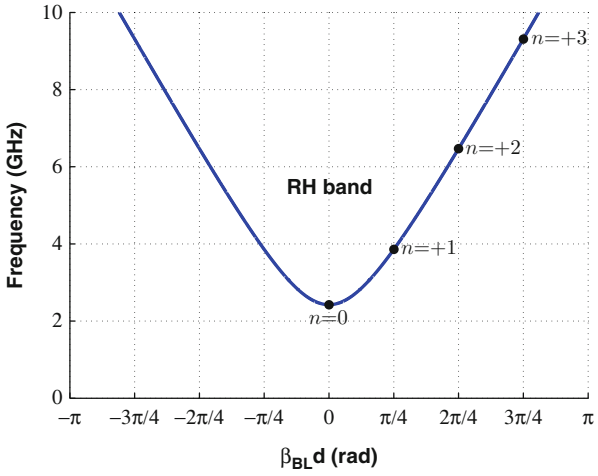


Fig. 28 Typical dispersion diagram for an ENG metamaterial unit cell with parameters $L_0 = 30$ nH, $Z_0 = 300\ \Omega$, $d = 12.5$ mm, and $\theta = 45°$ at 3 GHz. Superimposed are the locations of the nth order resonances for an ENG metamaterial structure that consists of a cascade of $N = 4$ unit cells

A representative ENG dispersion diagram is shown in Fig. 28, which was obtained using Eq. 35 and the same parameters as the NRI-TL metamaterial structure in Fig. 11. It can be observed that the ENG structure exhibits a lower stopband and a forward right-handed propagation band, and the backward left-handed band has been eliminated. Thus, unlike the NRI-TL structure which supports both forward and backward waves, the ENG structure supports only forward propagating waves, similar to a conventional transmission line. However, unlike a transmission line, propagation on the ENG structure begins above the zeroth-order resonance frequency of f_{+0} given approximately by Eq. 34; therefore, it exhibits a high-pass behavior with a DC offset.

In the general case where N unit cells are used in order to implement the ENG metamaterial structure with a total length of $\ell_{\text{TOT}} = Nd$, then the electrical length per unit cell, analogous to Eq. 21, becomes

$$\beta_{\text{ENG}} d = \frac{n\pi}{N} \quad ; \quad n = 0, 1, 2 \ldots (N-1) \tag{39}$$

An N-stage ENG metamaterial structure can therefore support a zeroth-order ($n = 0$) resonance at the $\beta_{\text{BL}} d = 0$ point, as well as positive-order ($n > 0$) resonances in the right-handed band, as shown in Fig. 28.

A Double-Resonant ENG Metamaterial Folded-Monopole Antenna

An example of an ENG metamaterial-based compact antenna is shown in Fig. 29 (Zhu and Eleftheriades 2009b). It uses the same folding technique and has a similar form to the NRI-TL metamaterial folded-monopole antenna described in the section "A Zero-Index NRI-TL Metamaterial Folded-Monopole Antenna"; however, it is implemented using ENG metamaterial unit cells. Recall that for the NRI-TL folded monopole, the shunt loading inductor L_0 was responsible for maximizing the even-mode current, and the series loading capacitor C_0 did not play a pivotal role in this process. Therefore, the elimination of the series loading capacitor C_0 in the ENG structure does not affect the ability of the antenna to act as a folded monopole, and the even-mode current is still maximized. Moreover, the ENG folded-monopole antenna of Fig. 29 employs a doubly resonant metamaterial structure in order to increase its bandwidth. The antenna consists of two ENG metamaterial arms, and each arm is designed to exhibit zeroth-order resonances at two closely spaced frequencies. Each arm comprises a microstrip transmission line loaded

DOI 10.1007/978-981-4560-75-7_21-1

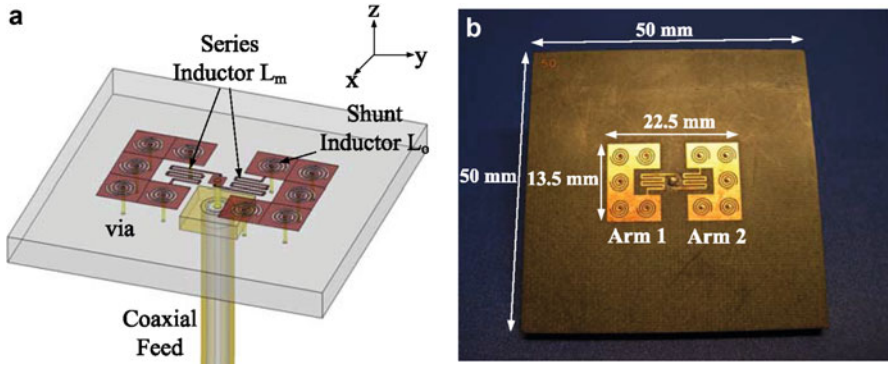


Fig. 29 Double-resonant ENG metamaterial folded-monopole antenna: (**a**) three-dimensional diagram, (**b**) photograph of the fabricated prototype (Zhu and Eleftheriades 2009b) © 2009 IEEE

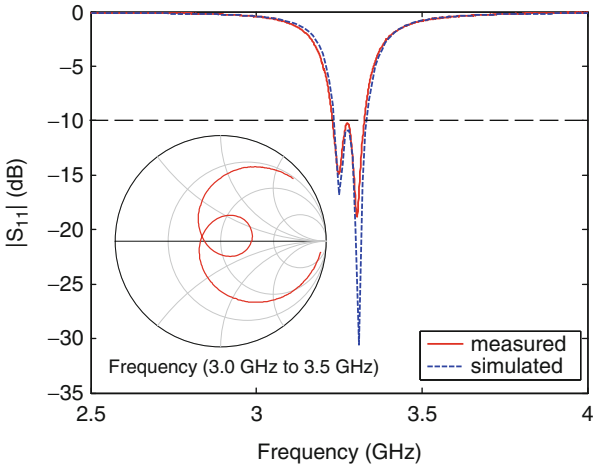


Fig. 30 $|S_{11}|$ responses for the ENG metamaterial antenna of Fig. 29 (Zhu and Eleftheriades 2009b) © 2009 IEEE

with five shunt spiral inductors, whose zero-index frequency can be tuned by adjusting the value of the loading inductors. A wideband performance is achieved by merging the corresponding two resonances together into a single passband.

The antenna has dimensions of $\lambda_0/4 \times \lambda_0/7 \times \lambda_0/29$ over a $0.55\lambda_0 \times 0.55\lambda_0$ ground plane. A series meander-line inductor is employed at the feed point to compensate the capacitive input impedance, thus achieving a good impedance match to 50 Ω over a 100 MHz −10 dB measured $|S_{11}|$ bandwidth around 3.3 GHz, as shown in Fig. 30. The antenna also exhibits a vertical linear electric field polarization, similar to that of a short monopole over a small ground plane, a measured gain of 0.79 dBi, and a measured radiation efficiency of 66 %.

Other Epsilon-Negative (ENG) Antennas

Various other ENG metamaterial antennas have been reported in the literature (Lai et al. 2007; Park et al. 2007; Park and Lee 2011; Niu et al. 2013; Niu and Feng 2013; Kim et al. 2009). In Lai et al. (2007), an inductor-loaded transmission line, essentially a one-dimensional microstrip mushroom structure without the series gaps between each of the unit cells, is used to implement a one-dimensional infinite-wavelength resonant antenna with monopolar radiation patterns at the $n = 0$ ZOR frequency. This antenna was then extended to a two-dimensional version in order to increase the gain and to reduce asymmetries in the radiation patterns. In Park et al. (2007), also a one-dimensional mushroom structure

DOI 10.1007/978-981-4560-75-7_21-1

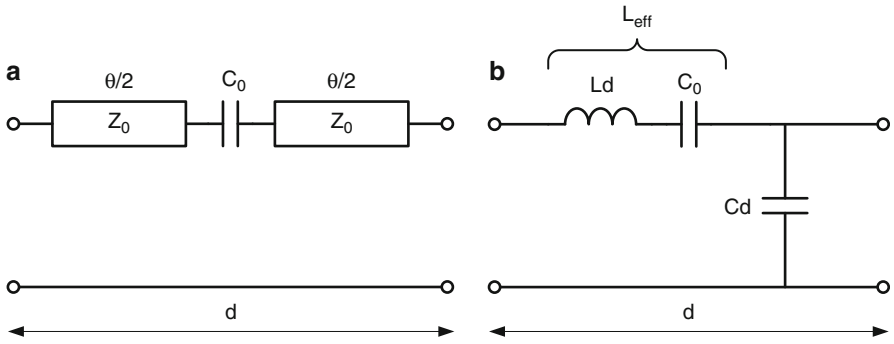


Fig. 31 Mu-negative (MNG) metamaterial unit cell: (**a**) symmetric distributed equivalent circuit, (**b**) lumped-element equivalent circuit

without the series gaps is investigated for different unit cell sizes with good gain performance. Finally, in Niu et al. (2013), a fully planar ENG antenna is presented that is implemented in asymmetric coplanar waveguide technology and which achieves dual-band and wideband performance by exciting both the $n = 0$ and the $n = 1$ resonant modes.

Mu-Negative (MNG) Antennas

Another transmission-line-based metamaterial structure that supports a zeroth-order resonant mode, which is also a variation of the NRI-TL metamaterial structure, is the so-called mu-negative (MNG) structure shown in Fig. 31a in its distributed form and in Fig. 31b in its lumped-element form. It can be observed that this topology is simply an NRI-TL metamaterial unit cell from Fig. 3 with the shunt loading inductor L_0 removed, resulting in a series capacitor-loaded transmission line.

Removing the shunt loading inductor L_0 from the NRI-TL metamaterial structure is equivalent to letting its value go to infinity, $L_0 = \infty$; therefore, from Eqs. 8 and 9, the zeroth-order resonant frequencies for the MNG metamaterial become

$$f_{-0} = 0, \tag{40}$$

$$f_{+0} = \frac{1}{2\pi\sqrt{(Ld)C_0}}. \tag{41}$$

These results indicate that the MNG metamaterial structure, like the ENG structure, has only one zeroth-order resonant frequency, which is determined by the values of the series elements in the unit cell of Fig. 31b. Therefore, for a fixed length of host transmission line, the zeroth-order resonant frequency f_{+0} can be easily tuned by simply adjusting the value of the loading capacitance C_0. Additionally, like the NRI-TL metamaterial structure, the $n = 0$ zeroth-order resonant frequency of the MNG structure is also independent of its physical length; however, since the loading inductor L_0 has been eliminated, this allows larger degrees of antenna miniaturization to be achieved.

The dispersion relation for the MNG metamaterial structure of Fig. 31a can be obtained by setting $L_0 = \infty$ in the dispersion relation of the NRI-TL metamaterial structure given by Eq. 1 and can be written as

$$\cos(\beta_{\mathrm{BL}} d) = \cos(\theta) + \left(\frac{1}{2\omega C_0 Z_0}\right)\sin(\theta). \tag{42}$$

DOI 10.1007/978-981-4560-75-7_21-1

Under the effective medium conditions $\theta \ll 1$ and $\beta_{\mathrm{BL}} d \ll 1$, Eq. 42 becomes

$$\cos\left(\beta_{\mathrm{BL}} d\right) \approx 1 - \frac{1}{2}\left(\omega^2 LCd^2 - \frac{Ld}{C_0}\right). \tag{43}$$

Using Eq. 7, the effective propagation constant of the MNG metamaterial structure can be written as

$$\beta_{\mathrm{MNG}} = \omega\sqrt{L_{\mathrm{eff}} C_{\mathrm{eff}}} = \omega\sqrt{\left[L - \frac{1}{\omega^2 C_0 d}\right] C}, \tag{44}$$

$$\beta_{\mathrm{MNG}} = \sqrt{\omega^2 LC - \frac{C}{C_0 d}}. \tag{45}$$

Equation 44 demonstrates that with the appropriate value of the series capacitor C_0, the effective series inductance L_{eff} of the MNG structure (see Fig. 31b) can be made equal to zero, resulting in a zero propagation constant and therefore a zeroth-order resonance. Exactly at the zeroth-order resonance frequency f_{+0}, the effective permeability of the structure is equal to zero and is therefore termed the mu-zero (MZR) point.

A representative MNG dispersion diagram is shown in Fig. 32, which was obtained using Eq. 42 and the same parameters as the NRI-TL metamaterial structure in Fig. 11. It can be observed that the MNG structure exhibits a very similar response to the ENG structure, with a lower stopband and a forward right-handed propagation band, and the backward left-handed band has been eliminated. Thus, unlike the NRI-TL structure which supports both forward and backward waves, and similar to the ENG structure, the MNG structure supports only forward propagating waves, similar to a conventional transmission line. However, unlike a transmission line, propagation on the MNG structure begins above the zeroth-order resonance frequency of f_{+0} given approximately by Eq. 41; therefore, it exhibits a high-pass behavior with a DC offset. Note also that by comparing the three dispersion diagrams shown in Figs. 11, 28, and 32, the location of the $n = -0$ resonance for the NRI-TL structure corresponds to the $n = 0$ resonance of the

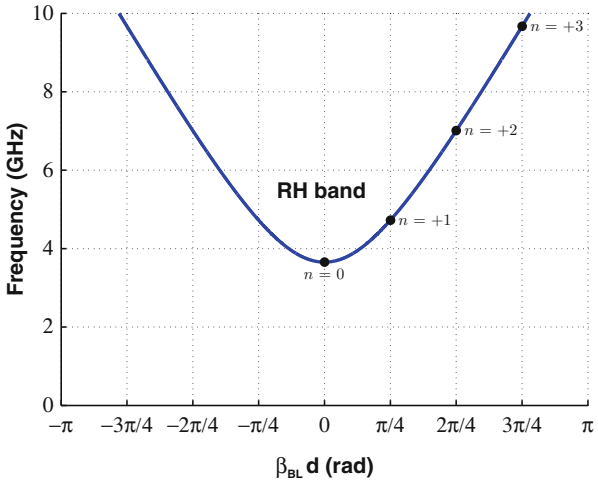


Fig. 32 Typical dispersion diagram for an MNG metamaterial unit cell with parameters $C_0 = 0.14$ pF, $Z_0 = 300\ \Omega$, $d = 12.5$ mm, and $\theta = 45°$ at 3 GHz. Superimposed are the locations of the nth order resonances for an MNG metamaterial structure that consists of a cascade of $N = 4$ unit cells

DOI 10.1007/978-981-4560-75-7_21-1

ENG structure, and the location of the $n = +0$ resonance for the NRI-TL structure corresponds to the $n = 0$ resonance of the MNG structure.

In the general case where N unit cells are used in order to implement the MNG metamaterial structure with a total length of $\ell_{TOT} = Nd$, then the electrical length per unit cell, analogous to Eq. 21, becomes

$$\beta_{MNG} d = \frac{n\pi}{N} \quad ; \quad n = 0, 1, 2, \ldots (N-1) \tag{46}$$

An N-stage MNG metamaterial structure can therefore support a zeroth-order ($n = 0$) resonance at the $\beta_{BL} d = 0$ point, as well as positive-order ($n > 0$) resonances in the right-handed band, as shown in Fig. 32.

A Dual-Band MNG Metamaterial Loop Antenna

An example of a dual-band MNG metamaterial loop antenna is shown in Fig. 33a (Park et al. 2010). It uses a microstrip transmission line that is loaded with series interdigital capacitors in order to implement the MNG metamaterial unit cells. Interdigital capacitors with different finger numbers were investigated, and it was found that ones with a higher number of fingers result, as expected, in a larger series loading capacitance, C_0, and therefore a lower zeroth-order resonance frequency. However, the trade-off of implementing a larger loading capacitance is a larger size.

It is noted that because the zeroth-order resonance for the MNG structure is obtained at the resonant frequency of the series branch of the MNG structure of Fig. 31b, short-circuit boundary conditions are necessary at the ends of an MNG resonant structure in order to support a resonance. This is in contrast to ENG resonant structures, which require open-circuit boundary conditions in order to support a zeroth-order resonance, which is due to the shunt branch of the ENG structure, as seen in Fig. 27b. The MNG antenna is therefore formed by short-circuiting a capacitively loaded microstrip transmission line at its

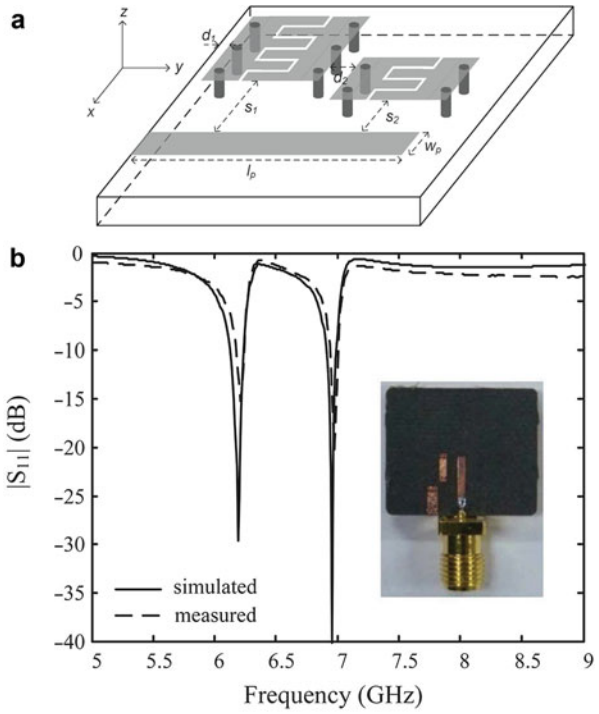


Fig. 33 (**a**) Dual-band MNG metamaterial loop antenna, (**b**) $|S_{11}|$ responses (Park et al. 2010) © 2010 IEEE

DOI 10.1007/978-981-4560-75-7_21-1

two ends with metallic vias, and it is fed using a microstrip line that magnetically couples into the resulting loop. The microstrip feed line is open ended and is designed to have a length equal to a quarter wavelength at the antenna's zeroth-order resonant frequency. The use of the microstrip feed line therefore limits the minimum achievable size for the MNG antenna to a quarter wavelength.

A dual-band antenna is created by adding a second MNG structure with a different size, and therefore a different zeroth-order resonance, to the same feed line, as shown in Fig. 33a. In this design, the first MNG structure consists of a single unit cell that has an interdigital capacitor with six fingers that resonate at a frequency of 6.2 GHz, and the second MNG structure also consists of a single unit cell that has an interdigital capacitor with four fingers that resonate at a frequency of 7 GHz. The distance between each of the MNG structures and the feed line was optimized in each case in order to maximize the coupling to each of the resonators.

The $|S_{11}|$ response of the antenna is shown in Fig. 33b, where the dual-band performance can be observed. The MNG antenna achieves 1.03 % and 0.95 % fractional bandwidths at 6.2 GHz and 7 GHz, respectively, with associated measured gains of 2.3 dBi and 3.3 dBi and efficiencies of 83 % and 84 %. The size of the antenna is $0.108\lambda_0 \times 0.175\lambda_0$ at 6.2 GHz and $0.121\lambda_0 \times 0.197\lambda_0$ at 7 GHz.

Other Mu-Negative (MNG) Antennas

Various other MNG metamaterial antennas have been reported in the literature (Bilotti et al. 2008; Wei et al. 2012a, b). In Bilotti et al. (2008), the size reduction of a circular patch antenna that is printed on a substrate that is partially with an MNG metamaterial is theoretically investigated. It is shown that a sub-wavelength metamaterial patch antenna can be achieved with satisfactory matching and radiating performance by employing split-ring resonators as the magnetic inclusions beneath the patch antenna. In Wei et al. (2012a, b), two versions of MNG loop antennas are presented, which consist of circular loop antennas that are periodically loaded with capacitors. This configuration enables uniform, in-phase current along the loop, therefore achieving a horizontally polarized omnidirectional radiation pattern, similar to a magnetic dipole antenna, even though the perimeter of the loop is comparable to the operating wavelength. Additionally, the capacitive loading enables a wide impedance bandwidth to be achieved.

NRI-TL Metamaterial Dipole Antennas

Most of the metamaterial antenna applications that have been described so far have been based on folded monopoles, printed monopoles, and patch antennas. Another type of antenna that can be designed based on metamaterial techniques, and which exhibits negative-order, zeroth-order, and positive-order resonances, is the differentially fed dipole antenna.

Some examples of metamaterial-loaded dipole antennas that have been reported in the literature are Ziolkowski and Erentok (2006), Jin and Ziolkowski (2010), Herraiz-Martinez et al. (2008b, 2011), Iizuka and Hall (2007), Liu et al. (2009), and Antoniades and Eleftheriades (2011a, b, 2012). In Ziolkowski and Erentok (2006), an electrically small electric dipole antenna was surrounded by a negative-permittivity shell, and it was demonstrated that the distributed inductance of the shell could be resonantly matched to the capacitive electrically small dipole. Thus, a resonant system was achieved which was matched to the source impedance, therefore leading to high overall efficiencies. Metamaterial-engineered planar dipole antennas have also been presented in Jin and Ziolkowski (2010), which have linear or circular polarizations. Another idea presented in Herraiz-Martinez et al. (2008b) demonstrated that a printed dipole antenna that is loaded with split-ring resonators can achieve a dual-band $|S_{11}|$ response.

The authors in Iizuka and Hall (2007) and Liu et al. (2009) have demonstrated that creating a dipole antenna from an asymmetrically loaded left-handed transmission line leads to a difference in the amplitudes of the out-of-phase currents on each of the transmission-line conductors. Therefore, the two currents do not completely cancel each other out in the far field, enabling the structure to radiate. Even

DOI 10.1007/978-981-4560-75-7_21-1

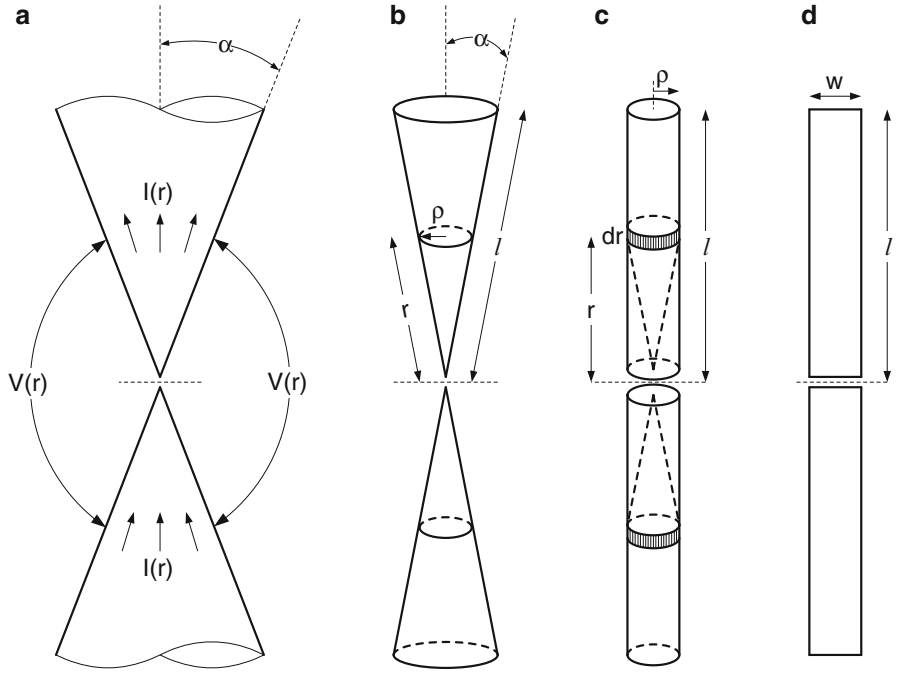


Fig. 34 The evolution of Schelkunoff's infinite biconical antenna into a planar dipole antenna. (**a**) Infinite biconical antenna, (**b**) finite biconical antenna, (**c**) cylindrical dipole antenna, and (**d**) planar dipole antenna (Antoniades and Eleftheriades 2012) © 2012 IEEE

though a multiband $|S_{11}|$ response is achieved, the reported efficiency and gain of the antennas in Iizuka and Hall (2007) are quite low, due to the partial cancelation of the current, combined with the presence of material losses in the printed components. In Liu et al. (2009), the authors have optimized the design of the left-handed dipoles in order to improve the efficiency of their antennas.

Herein, a technique is adopted for the design of a metamaterial-loaded dipole antenna that does not rely on an asymmetrically loaded transmission line in order to create an unbalanced radiating current. Instead, in an approach initially proposed by Schelkunoff, the two arms of a dipole antenna are modeled as the two conductors of a biconical transmission line. The biconical transmission line can subsequently be transformed into its equivalent dipole counterpart through the process shown in Fig. 34. This allows the dipole antenna to be treated as the host transmission line in a conventional NRI-TL metamaterial structure and provides a suitable medium for the placement of the loading elements. As such, the NRI-TL properties of the structure are maintained, while still creating an efficient radiator.

The proposed NRI-TL metamaterial-loaded dipole antenna is shown in Fig. 35 (Antoniades and Eleftheriades 2012). It consists of a planar dipole antenna on an FR4 substrate in the shape of a bow tie with a small flare angle, two series capacitive gaps forming the loading capacitance C_0, and two shunt inductive strips forming the loading inductance L_0, resulting a two-unit-cell metamaterial-loaded dipole antenna.

Referring to Fig. 35, the first NRI-TL unit cell is formed by the two central dipole conductors connected to the feed, which are joined at their ends through two vias to a thin inductive strip on the bottom of the substrate. Thus, the two central dipole conductors effectively form a transmission line with a characteristic impedance Z_{01} and length l_1 which is inductively loaded at its end with a thin inductive strip with an inductance L_{01}. Since the dipole arms have a finite length, Schelkunoff showed that this can be modeled as a terminal impedance, Z_{t1}, connected to the end of the transmission line. Therefore, this terminal impedance must be added in parallel to the impedance of the thin inductive strip. The transmission line

DOI 10.1007/978-981-4560-75-7_21-1

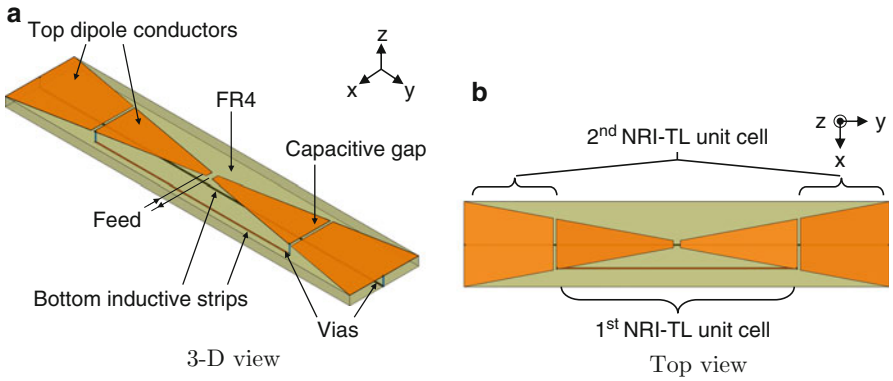


Fig. 35 NRI-TL metamaterial-loaded dipole antenna, consisting of two NRI-TL unit cells loading a printed dipole antenna (Antoniades and Eleftheriades 2012) © 2012 IEEE

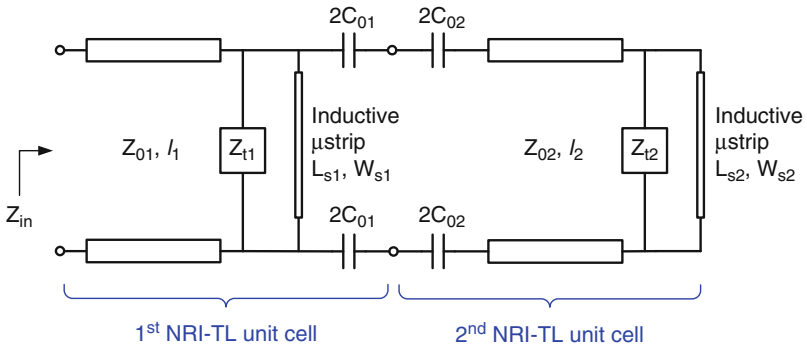


Fig. 36 Equivalent circuit of the NRI-TL dipole antenna of Fig. 35 (Antoniades and Eleftheriades 2012) © 2012 IEEE

is also connected to a series capacitance $2C_{01}$, formed in part due to the capacitive gaps in each of the dipole arms.

The second NRI-TL unit cell is formed by the two outer dipole conductors, which are also joined at their ends through two vias to a longer thin inductive strip, which was centered on the bottom of the board. Thus, the two outer dipole conductors effectively form another transmission line with a characteristic impedance Z_{02} and length l_2, which is inductively loaded at its end with a longer inductive strip with an inductance L_{02}. To the impedance of the thin inductive strip, the terminal impedance, Z_{t2}, of the finite dipole arms must be added as with the first unit cell. The transmission line is also connected at its input to a series capacitance $2C_{02}$, formed in part due to the capacitive gaps in each of the dipole arms.

In this manner, the equivalent circuit of the NRI-TL dipole antenna shown in Fig. 36 is formed, which consists of two cascaded asymmetric NRI-TL metamaterial unit cells. Figure 37a shows the $|S_{11}|$ responses obtained from the Agilent-ADS circuit simulator for the equivalent circuit of the dipole antenna of Fig. 36 for different values of the inductive strip length L_{s2}. Note that while L_{s2} was varied, the total length of the antenna was kept at $L_2 = 50$ mm, and all of the other geometrical parameters were also kept constant. Figure 37b shows the full-wave simulated $|S_{11}|$ responses obtained from Ansoft-HFSS for the NRI-TL dipole antenna of Fig. 35 for different values of the bottom inductive strip L_{s2}. It can be observed that the general performance of the antenna equivalent circuit matches well with the performance obtained from the full-wave HFSS results, thus validating the antenna equivalent circuit. It can also be observed that all of the antenna responses exhibit multiband characteristics, with four distinct resonances in the frequency range shown.

DOI 10.1007/978-981-4560-75-7_21-1

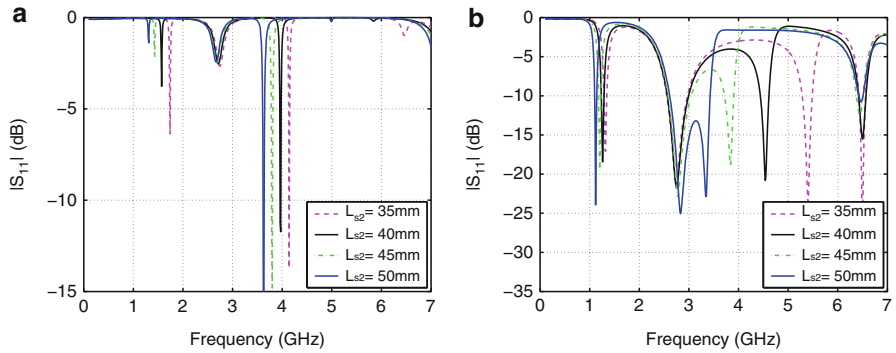


Fig. 37 (**a**) ADS simulated $|S_{11}|$ responses of the equivalent circuit of the NRI-TL dipole antenna of Fig. 36 for different values of the inductive strip length L_{s2}, using the equivalent circuit parameters extracted from Fig. 35. Parameters: $C_{01} = C_{02} = 0.2$ pF, $L_{s1} = 28$ mm, $W_{s1} = 0.1$ mm, $L_{s2} = 35 \rightarrow 50$ mm, $W_{s2} = 0.1$ mm, $Z_{01} = 312.6\ \Omega$, $Z_{02} = 355.2\ \Omega$, $l_1 = 13.6$ mm, $l_2 = 10.9$ mm, and $\beta l_1 = 49°$ and $\beta l_2 = 39.2°$ at 3 GHz. (**b**) HFSS simulated $|S_{11}|$ responses of the multiband NRI-TL metamaterial-loaded dipole antenna of Fig. 35 for different values of the bottom inductive strip length, $L_{s2} = 35 \rightarrow 50$ mm (Antoniades and Eleftheriades 2012) © 2012 IEEE

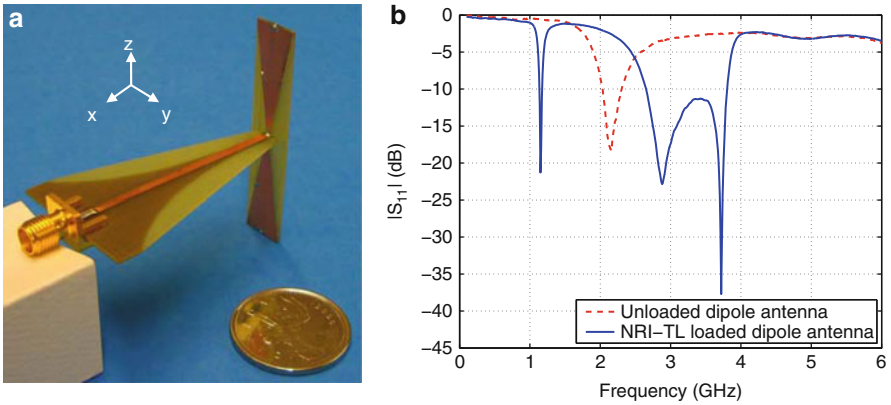


Fig. 38 Photograph of the NRI-TL metamaterial-loaded dipole antenna of Fig. 35 with $L_{s2} = 50$ mm, fed using a broadband transition. (**b**) Measured $|S_{11}|$ response of the antenna in (**a**), compared with the measured $|S_{11}|$ response of a reference unloaded dipole antenna fed using the same transition and with the same overall dimensions (Antoniades and Eleftheriades 2012) © 2012 IEEE

The fabricated prototype of the NRI-TL dipole antenna is shown in Fig. 38a. The measured $|S_{11}|$ response of the antenna is shown in Fig. 38b, compared to the measured $|S_{11}|$ response of a reference unloaded dipole antenna. It can be observed that the NRI-TL dipole antenna exhibits three distinct resonances corresponding to the $n = -1$, $n = -0$, and $n = +0$ resonances at 1.15 GHz, 2.88 GHz, and 3.72 GHz, with associated -10 dB bandwidths of 37 MHz and 1150 MHz, while the unloaded antenna exhibits only a single resonance at 2.15 GHz with a bandwidth of 275 MHz. Thus, the NRI-TL dipole antenna exhibits a 47 % reduction of the lowest resonant frequency compared to the reference unloaded dipole antenna. Furthermore, at 1.15 GHz, the length of the antenna, $L_2 = 50$ mm, is $0.19\lambda_0$, which represents a miniaturization factor of approximately two compared to the unloaded dipole antenna which has a length of $0.36\lambda_0$ at its resonant frequency of 2.15 GHz. Throughout the bands of operation, the antenna maintains a linear electric field polarization, and the measured gain and radiation efficiency vary from 0.11 dBi to 3.26 dBi and 49.5–95.6 %, respectively.

DOI 10.1007/978-981-4560-75-7_21-1

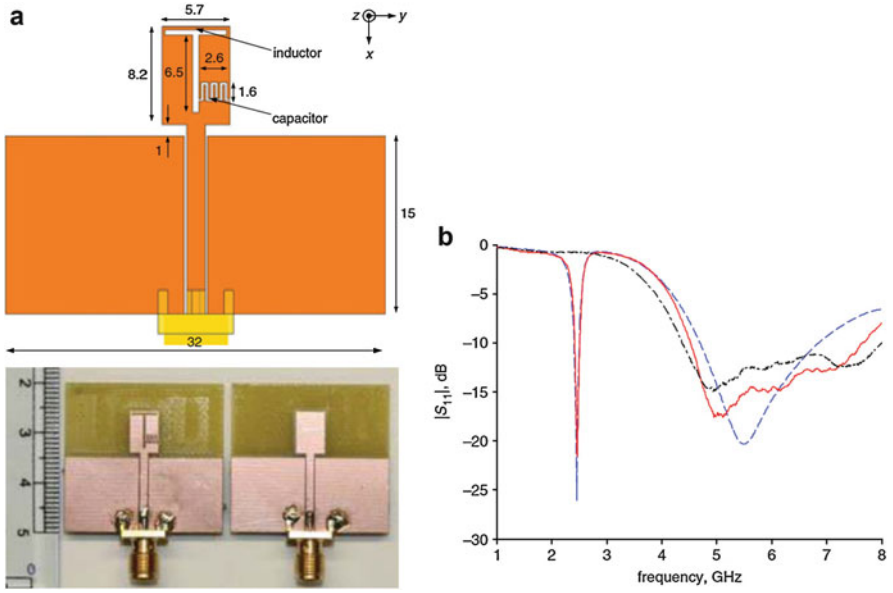


Fig. 39 (**a**) Dual-band metamaterial-inspired small monopole antenna for Wi-Fi applications, (**b**) $|S_{11}|$ responses: *red solid line*, measured results for the metamaterial-inspired antenna; *blue dashed line*, simulated results for the metamaterial-inspired antenna; and *black dash-dot line*, measured results for the unloaded monopole antenna (Zhu and Eleftheriades 2009a) © 2009 IET

Metamaterial-Inspired Antennas

The numerous antenna designs that have been published on transmission-line-based metamaterials, as well as volumetric metamaterials based on split-ring resonators and wires, have also led to the creation of many antenna designs that are not metamaterial-based per se but that have been inspired by metamaterial concepts. Thus, metamaterials have provided a conceptual route for implementing many small resonant antennas (Zhu and Eleftheriades 2009a, 2010; He and Eleftheriades 2012; Ryan and Eleftheriades 2012).

A dual-band monopole antenna that employs metamaterial-inspired reactive loading is shown in Fig. 39a (Zhu and Eleftheriades 2009a). It comprises a two-arm fork-like CPW-fed monopole with a thin-strip inductor loaded on top of the monopole and an interdigital capacitor loaded on the right-side arm. Alternatively, it can be seen as a T-shaped slot cut out of the rectangular patch with a capacitor loaded on its right. This metamaterial-inspired loading enables the antenna to operate in two modes, covering the lower Wi-Fi band of 2.40–2.48 GHz and the upper Wi-Fi band of 5.15–5.80 GHz. The first mode is a monopole mode that operates at the upper Wi-Fi band, where the capacitor becomes a short circuit and the inductor becomes an open circuit. In addition to the monopole resonance at the upper Wi-Fi band, the metamaterial-inspired reactive loading introduces a second resonant mode the lower Wi-Fi band. At this frequency, the antenna no longer acts as a monopole along its axis, but rather as a slot along this axis.

Figure 39b shows the $|S_{11}|$ response of the metamaterial-inspired antenna, compared to the $|S_{11}|$ response of a reference unloaded antenna, where a dual-band performance of the metamaterial-inspired antenna can be clearly seen compared to the unloaded CPW-fed monopole antenna. The antenna exhibits a measured −10 dB bandwidth of 90 MHz at the lower Wi-Fi band from 2.42 to 2.51 GHz and a bandwidth of 3.2 GHz from 4.52 to 7.72 GHz for the upper Wi-Fi band, with measured efficiencies of 89.2 % at 5.50 GHz and 64.0 % at 2.46 GHz and measured gains of 1.53 at 5.50 GHz and 0.71 at 2.46 GHz. The measured radiation patterns confirm the operation of the antenna as a conventional printed monopole in the upper band and a slot in the lower band.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_21-1

Active Non-Foster Matching Networks for Small Antennas

Having described numerous passive antenna applications of transmission-line metamaterials in the previous section, the focus of this section shifts to active devices and specifically how these can be used in the design of compact and broadband antennas.

Antenna matching for receivers improves the received signal-to-noise ratio (SNR) and, for transmitters, enhances the radiated power efficiency. However, for electrically small antennas, because of their high-quality factor (high-Q), matching becomes a challenging task. In fact, a fundamental limit on the quality factor of a linearly polarized antenna with a radiation efficiency of η and an electrical size of ka, due to Wheeler (1947), Chu (1948), and Harrington (1960), is given by

$$Q = \eta\left(\frac{1}{k^3a^3} + \frac{1}{ka}\right). \tag{47}$$

The Chu limit above inversely relates the minimum achievable Q of an antenna to its electrical size; therefore, a small antenna assumes a high-Q. In turn, to match a high-Q electrically small antenna with a lossless passive matching network, the achievable matching bandwidth is constrained by another fundamental limit – namely, the Bode-Fano limit (Bode 1947; Fano 1950). This limit inversely relates the matching bandwidth to the Q of the load to be matched (here the antenna). All previously discussed small metamaterial antennas in this chapter were passive and thus subject to these matching limitations. The employment of active components, including reactive non-Foster elements, in the antenna or in the matching network, is sought to overcome either of these fundamental limits, and this will be the subject of this section. But, what are reactive non-Foster elements?

Zobel and Foster established that passive, non-dissipative, two-terminal reactances exhibit a positive slope with frequency (Zobel 1923; Foster 1924) – that is, the reactance X and the susceptance B of such networks for all frequencies satisfy:

$$\frac{\partial X}{\partial \omega} > 0 \quad \text{and} \quad \frac{\partial B}{\partial \omega} > 0\,. \tag{48}$$

Nonetheless, one can think of reactive elements with an anomalous behavior which, in contrast to Eq. 48, exhibit a negative slope with frequency; hence, they are called non-Foster reactive elements. For the reactances X_{NF} and susceptances B_{NF} of such networks and at least in part of the frequency spectrum:

$$\frac{\partial X_{NF}}{\partial \omega} < 0 \quad \text{and} \quad \frac{\partial B_{NF}}{\partial \omega} < 0\,. \tag{49}$$

Simple representative examples of non-Foster elements include negative capacitors and negative inductors.

Implementation of Reactive Non-Foster Elements

The networks employed for synthesizing non-Foster reactive elements must inevitably lack the condition of Foster's reactance theorem in Eq. 48. Therefore, these elements need to be designed using lossy (Mirzaei and Eleftheriades 2013c) or non-passive – i.e., active – networks (Linvill 1953). Traditionally, non-Foster reactive elements are implemented by two groups of circuits called negative impedance converters (NICs) and negative impedance inverters (NIIs). The operation of these circuits is schematically shown in Fig. 40.

DOI 10.1007/978-981-4560-75-7_21-1

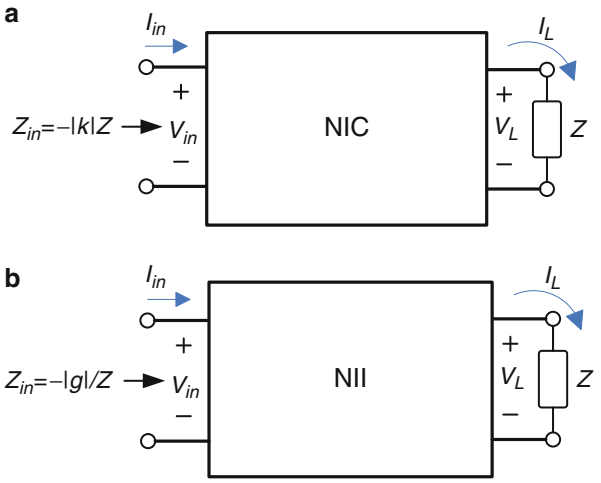


Fig. 40 The operation of (**a**) a negative impedance converter (NIC) and (**b**) a negative impedance inverters (NII) as two-port networks

According to Fig. 40a, NICs are two-port networks terminated at one port to an impedance, and looking into the other port, the negative of the termination impedance, with a scaling factor, is seen. For example, terminating such a network to a capacitor results in a negative capacitor. This property remains invariant by interchanging the input and output ports. Moreover, NICs change either the direction of the load current or invert the polarity of the load voltage, as seen from the input terminals. These two types are, respectively, named current-inversion NICs (INICs) and voltage-inversion NICs (VNICs). Furthermore, NIC circuits are usually open-circuit stable (OCS) at one port and short-circuit stable (SCS) at the other port (Brownlie 1966). A port is said to be OCS (SCS), if for any arbitrary passive impedance Z connected to the other port, the network obtained by open-circuiting (short-circuiting) the OSC (SCS) port remains stable. However, the knowledge of OCS and SCS ports does not provide enough information about the stability of a given NIC circuit terminated to arbitrary loads except for an open or short circuit (Stearns 2011).

On the other hand, according to Fig. 40b, for terminated NIIs, the driving-point impedance at the input port is the inverse of the termination impedance, with a negative scaling factor. The operation of NIIs is similar to gyrators, but with a difference that NIIs have a negative gyration resistance (or conductance). As such, terminating an NII to a capacitor results in a negative inductor, as seen from the input port.

An inventory of some of the NIC circuits has been gathered in Sussman-Fort (1998) and Stearns (2011). These circuits include Linvill's single-ended and balanced VNICs (Linvill 1953), Larky's INIC (Larky 1956, 1957), Yanagisawa's INIC (Yanagisawa 1957), Sandberg and Nagata's INIC (Sandberg 1960; Nagata 1965), Hakim's VNICs and INIC (Hakim 1965), and Myers' VNICs and INICs (Myers 1965). In addition to these NIC circuits, a popular NII realization that uses two FET devices can be found in Brucher et al. (1995) and Kolev et al. (2001).

It should be noted that different NIC and NII circuits perform differently in terms of stability margin, noise, nonlinear behavior, and sensitivity to the transistor parameters. More importantly, the stability issues of a selected NIC or NII need to be carefully treated for a successful implementation of non-Foster reactive elements. The difficulty of this treatment becomes more obvious by noticing that all numerical port-based stability tests, including Rollet's k factor and μ test, which work sufficiently well in many microwave amplifier design tasks, widely fail in predicting the stability of non-Foster circuits (Stearns 2011, 2012, 2013).

DOI 10.1007/978-981-4560-75-7_21-1

A complete assessment of the network poles, by calculating the zeros of the network determinant (or its normalized determinant), can be a comprehensive way of determining the stability of non-Foster circuits, where a stable circuit must have no right-hand plane (RHP) poles (Bode 1947). In this regard, by building up on the concept of the normalized determinant function (NDF) and a simple way of calculating the NDF using "return ratios" (Bode 1947), Struble and Platzker implemented a method to evaluate the NDF using CAD tools (Struble and Platzker 1993; Platzker and Struble 1994). These days, the NDF method can be found integrated into commercial microwave simulation tools. The NDF analysis takes into account the effects of all loops in the network and is a reliable method. Nevertheless, it requires access and control over the internal linear-dependent sources in the models of the active devices used in the design. Therefore, when having access only to the S-parameters or the compiled linear models with no control over the active dependent sources inside the model, this method cannot be employed. In these situations, the Nyquist test applied to an accurate estimation of the loop gain in the feedback network can be a compelling choice (Middlebrook 1975; Tian et al. 2001).

More recently, an alternative approach for implementing reactive non-Foster elements using loss-compensated negative-group-delay (NGD) networks has been presented (Mirzaei and Eleftheriades 2013c). This approach naturally leads to stable implementations. It is developed by observing that non-Foster reactive elements and loss-compensated NGD networks influence propagating waves in a similar fashion.

Antennas with External Non-Foster Matching Networks

A brief review of the early applications of non-Foster reactive elements in small antenna matching networks can be found in Sussman-Fort and Rudish (2009). In this regard, the early works of Harris et al. (Harris and Myers 1968) and Perry (Albert 1973), which are briefly explained below, remarkably stand out. Other related works have been reported in Albee (1976), Bahr (1977), Sussman-Fort and Rudish (2009), Stearns (2011, 2013), White et al. (2012) and Xu et al. (2012). These applications can be explained with the help of a representative equivalent-circuit model.

An electrically small dipole or monopole antenna can be modeled with a simple series-*RC* equivalent circuit, where the radiation resistance is proportional to the square of frequency – i.e., $R_r = R_0(f/f_0)^2$. In this relation, R_0 represents the radiation resistance at a given frequency f_0.

Assuming that one has access to non-Foster inductances and capacitances, the antenna can be perfectly matched to the characteristic impedance of the system Z_0 as shown in Fig. 41a, where a negative capacitor

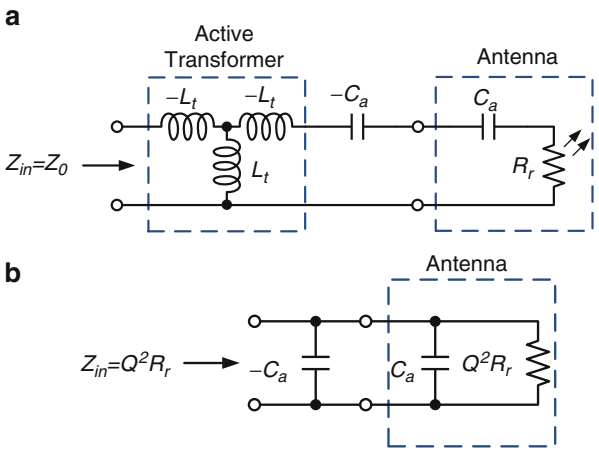


Fig. 41 (**a**) Perfect matching and (**b**) a more practical approach for matching an electrically small monopole using reactive non-Foster elements

DOI 10.1007/978-981-4560-75-7_21-1

tunes out the capacitance of the antenna and a step-up non-Foster T-transformer is used to cancel out the quadratic frequency dependence of R_r and match it to Z_0 (Skahill et al. 1998).

This scheme is too complicated, and considering all practical trade-offs, including difficulty of implementation, biasing, and noise, usually a much simpler approach involving only one series or parallel non-Foster capacitor at the terminals of the antenna is adopted. The scheme using a series floating negative capacitor is the same as the one in Fig. 41a without the non-Foster T-transformer. Using this approach, the antenna reactance is effectively eliminated, but the input resistance remains small and highly frequency dependent. Thereby, the antenna is not matched to the characteristic impedance of the system. Another scheme using one parallel capacitor, due to Harris et al. (Harris and Myers 1968), can be explained by converting the equivalent circuit of the small monopole antenna from a series *RC* to parallel *RC* as shown in Fig. 41b. This equivalent circuit is valid because of the large Q of an electrically small antenna. Then, by placing a single-negative capacitor parallel to the terminals of the antenna, the reactive part of the input admittance can be effectively eliminated over a broad frequency range, but the remaining resistive part is highly frequency dependent, assumes very large values, and is not matched to the characteristic impedance of the system. Nevertheless, overall improvements in gain (Harris and Myers 1968; Albert 1973) and SNR (Sussman-Fort and Rudish 2009) in a wide frequency range have been reported using such simple schemes involving one series or parallel non-Foster capacitor.

These ideas of matching have attracted strong interest over the past few decades; however, the actual implementations have remained challenging due to the difficulty of realizing high-Q reactive non-Foster elements – mainly due to the stability issues which were previously discussed. In addition, in a practical implementation, other relevant issues, including the total power consumption, nonlinearity, and noise, need to be addressed.

Antennas with Embedded Non-Foster Matching Networks

Small antennas internally augmented by non-Foster reactive elements can provide a wider matching bandwidth at the input terminals compared to their passive counterparts. In fact, augmentation with reactive non-Foster elements is an extension to the antenna loading with passive Foster reactive elements which is widely employed for controlling the antenna parameters. Non-Foster reactive elements can potentially replace these Foster elements, leading to broadband operation. In particular, for broadband matching applications, the starting point is an antenna whose resonant frequency can be configured with a passive reactive Foster element, and herein a procedure is presented for replacing it with reactive non-Foster elements to make it broadband at the antenna terminals.

To explain the idea, it is assumed that a certain frequency-reconfigurable antenna around the resonant frequency can be modeled with a simple resonant equivalent circuit – say, a *series-RLC* circuit – and the resonant frequency of the antenna can be tuned using a variable capacitor C_t, as shown in Fig. 42a. It is also assumed that the antenna has a large radiation resistance close to the characteristic impedance of the system. This implies that, if electrically small, the antenna provides a very narrow bandwidth due to its large Q.

Subsequently, by assuming access to ideal non-Foster reactive elements, C_t can be replaced with a parallel combination of $-C_a$ and $-L_a$ to cancel out the antenna reactances, as shown in Fig. 42b, and the antenna will be matched over the entire frequency range where the model is valid. Note that such a simple model cannot represent the general complicated model of most antennas over a broad bandwidth, but it simply and sufficiently motivates the idea. Besides, for the actual design, as explained next in this section, the antenna tuning data, rather than its exact model, is required.

DOI 10.1007/978-981-4560-75-7_21-1

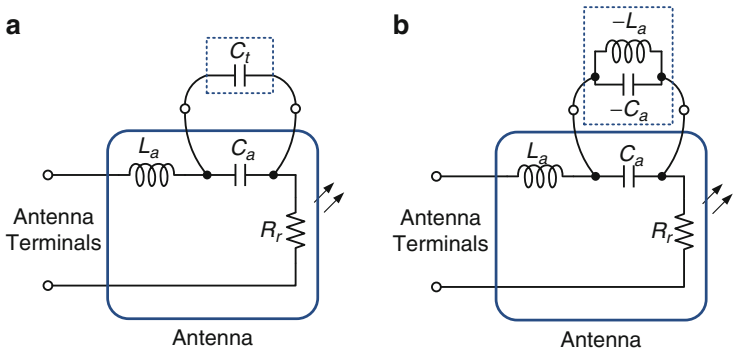


Fig. 42 (**a**) A simple series resonant model for an antenna for which the resonant frequency can be tuned using a tuning capacitor C_t. (**b**) C_t can ideally be replaced with a combination of non-Foster reactive elements to satisfy the resonance condition over a broad frequency range and to obtain a broadband matching at the antenna input terminals

Practical Design of Embedded Non-Foster Matching Networks

The design of an embedded non-Foster matching network can be performed totally based on the frequency reconfigurability of the underlying passive antenna. The design process leads to the synthesis of a two-terminal network which can replace the tuning capacitor or inductor, C_t or L_t, in the passive antenna to achieve a broadband operation. In this section, the design procedure is explained in three steps:

Step I: Selecting a Suitable Passive Antenna

Suitable antennas for the purpose of embedding non-Foster matching networks include frequency-reconfigurable (or frequency-agile) antennas for which the resonant frequency can be swept over a broad range using a variable capacitor C_t or an inductor L_t.

An example of such an antenna is shown in Fig. 43, which is a metamaterial-inspired monopole patch loaded with a varactor and fed by a coplanar waveguide (CPW), as shown in the section "Metamaterial-Inspired Antennas" (Mirzaei and Eleftheriades 2011b; Zhu and Eleftheriades 2009a). Other examples include a PIFA antenna (Di Nallo et al. 2007; Bit-Babik et al. 2007), as well as antennas presented in Zhu and Ziolkowski (2012a, b) and Ziolkowski et al. (2013). In fact, there is no limitation on the size of a suitable antenna as long as the non-Foster circuit can be fitted into the space provided. Therefore, this method can be attractive for matching electrically small antennas.

Step II: Obtaining the Frequency Reconfigurability and Tuning Behavior

The resonant frequency of the selected antenna can be tuned by the tuning element loading the antenna (C_t or L_t). The tuning behavior is a graph of $B_t = \omega C_t$ or $X_t = \omega L_t$ versus resonant frequencies ω, which can be measured or simulated over the desired bandwidth in some discrete frequencies for the passive antenna.

For example, a scaled version of the antenna from Fig. 43 and its tuning behavior are demonstrated in Fig. 44. This figure shows a large tuning bandwidth in the lower UHF band. The results in this graph are translated to a $B_t = \omega C_t$ versus ω graph in Fig. 44c, which exhibits a negative slope in accordance with the expected non-Foster behavior.

Step III: Fitting a Synthesizable Reactance Function to the Tuning Data and Calculating the Parameters of the Embedded Matching Network

In this step, an appropriate function has to be fitted to the graph representing the antenna tuning behavior. This function should represent a synthesizable reactance (susceptance function $B_t(\omega)$ or inductance function $X_t(\omega)$) using a combination of Foster and non-Foster elements. The idea here is to replace the

DOI 10.1007/978-981-4560-75-7_21-1

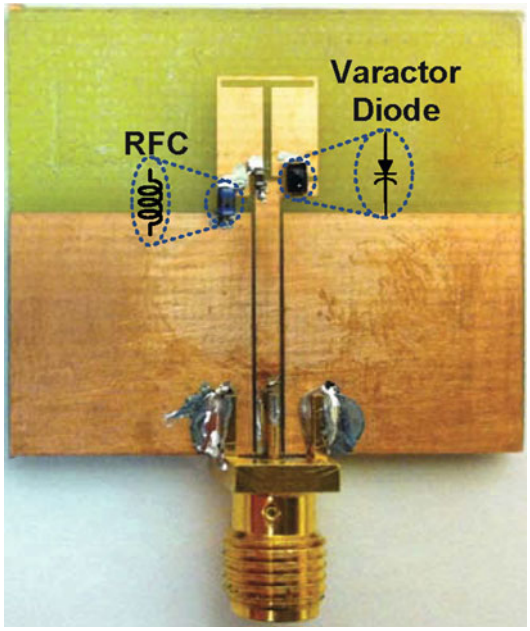


Fig. 43 Example of a metamaterial-inspired frequency-reconfigurable antenna, suited for embedding a non-Foster matching network (Mirzaei and Eleftheriades 2011b) © 2011 IEEE

tuning element C_t or L_t with a network that satisfies the resonance condition in a broad frequency range. As the sample graph in Fig. 44c illustrates, such a network synthesizes a non-Foster reactance at its terminals; hence, it can be implemented using a combination of non-Foster $-C$ and $-L$ and Foster C and L elements. The following criteria can be used to select the most appropriate non-Foster network among all possible combinations:

- Error function: that is, how closely the synthesized reactance follows the tuning data.
- Complexity: that is, how difficult the implementation of the network is. In particular, implementations with fewer non-Foster components are preferred.
- Component values: that is, very large or very small values for the capacitors and inductors are not desirable.
- The sensitivity of the fitted curve to the value of the network elements: that is, how the antenna response changes if there exist some variation, dispersion, and error in the implementation of the network elements. This variation is more pronounced for the non-Foster elements, because of the more difficult implementation and limited quality factor.

After finalizing the non-Foster network selection, such a network can be implemented using appropriate NIC or NII circuits. Two implementation examples are shown in Fig. 45, where the antenna in Fig. 45a is the non-Foster version of the passive antennas shown in Figs. 43 and 44a. For the implementations in Fig. 45a, b, a 10 dB fractional bandwidth equal to 8.1 % and 8.2 % for return losses, respectively, around the center frequencies of 465 MHz and 306.7 MHz, has been obtained. For these two antennas, ka is equal to 0.486 and 0.506 at these center frequencies, where k is the free-space wave number and a is the radius of the smallest sphere circumscribing the antennas. According to a widely accepted criterion, $ka \approx 0.5$ marks the value below which the antennas can be regarded as electrically small; therefore, these two implementations approximately satisfy the condition for being electrically small. The bandwidth of 8.1 % of the antenna in Fig. 45a should be compared to the original 2.9 % bandwidth of the passive antenna

DOI 10.1007/978-981-4560-75-7_21-1

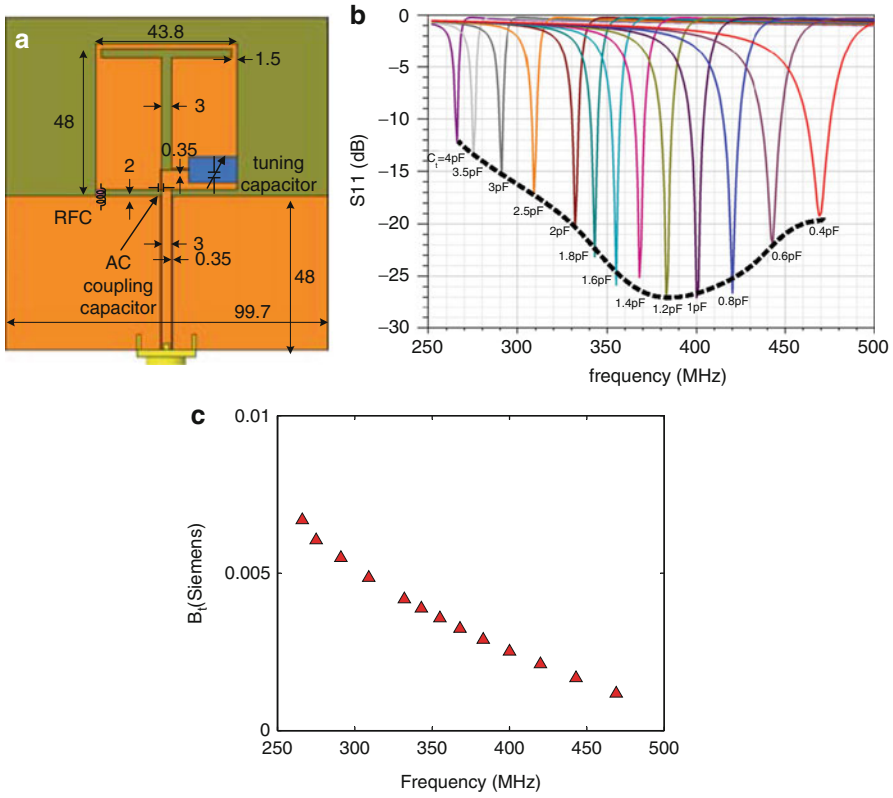


Fig. 44 (**a**) Dimensions of the frequency-reconfigurable antenna suitable for embedding a non-Foster matching network. (**b**) Simulation data using a finite-element field solver demonstrates that the resonant frequency of the antenna can be tuned by tuning the loading capacitor C_t. (**c**) Graph of the tuning susceptance $B_t = \omega C_t$ vs. frequency (Mirzaei and Eleftheriades 2013b) © 2013 IEEE

shown in Fig. 44b. However, it should be noted that due to practical limitations (discussed below), a very broad bandwidth, similar to the dashed-line graph in Fig. 44b, cannot be achieved. Nonetheless, the actual implementations, exemplified above, show a significant improvement in the antenna matching bandwidth (Mirzaei and Eleftheriades 2011a; Mirzaei and Eleftheriades 2013b; Zhu and Ziolkowski 2012a, b; Ziolkowski et al. 2013).

Promises and Challenges of Non-Foster Matching Networks for Antennas

The ideas of antennas with non-Foster matching networks have attracted strong interest over the past few decades; however, the actual implementations have remained challenging due to the difficulty of realizing high-Q reactive non-Foster elements – mainly due to the stability issues which were previously discussed. In practice, an unavoidable dispersion in the implementation of the reactive non-Foster elements, their limited quality factors, and sensitivity of the matching bandwidth to these variations tend to significantly limit the achievable bandwidth (Mirzaei and Eleftheriades 2013b). In addition, in a practical implementation, other relevant issues, including the total power consumption, nonlinearity, and noise, need to be addressed. For example, although a non-Foster matching network for a receiving antenna improves the received signal level by improving the matching bandwidth, it also increases the noise level. For such an application, the achievable signal-to-noise ratio at the antenna terminals, rather than the actual matching bandwidth, will utterly matter.

DOI 10.1007/978-981-4560-75-7_21-1

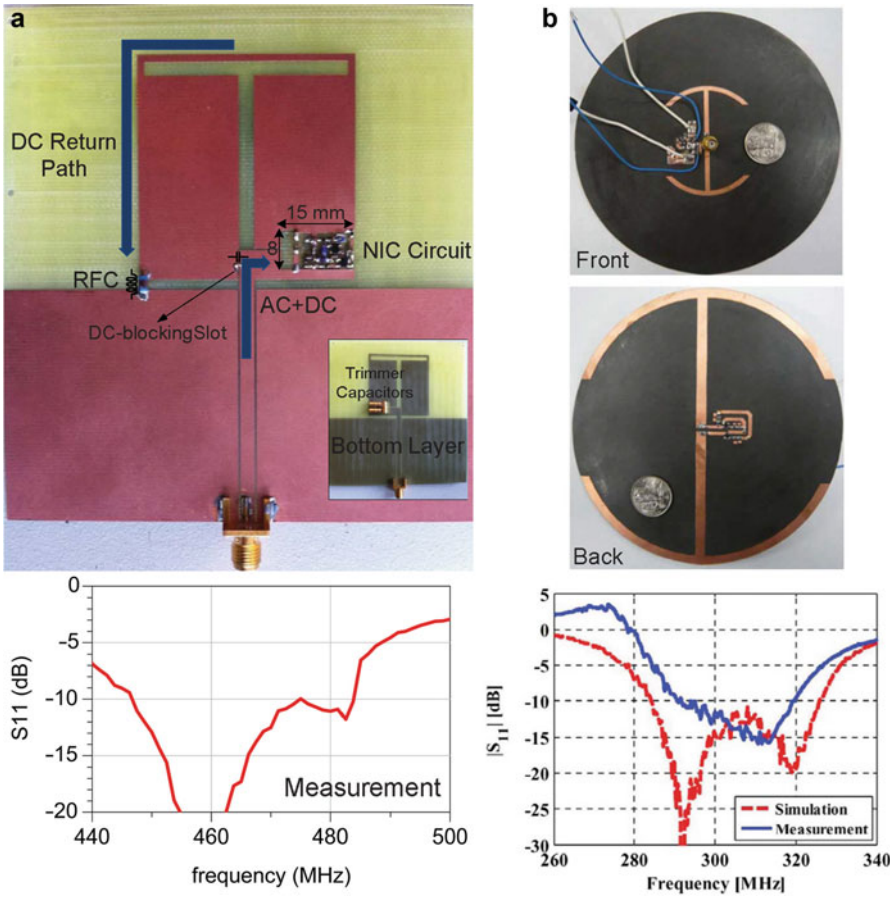


Fig. 45 (**a**) Metamaterial-inspired monopole antenna with an embedded non-Foster matching network (Mirzaei and Eleftheriades 2011a, 2013b), © 2011, 2013 IEEE. (**b**) Near-field resonant parasitic "Egyptian axe dipole" antenna with an embedded non-Foster matching network (Zhu and Ziolkowski 2012a) © 2012 IEEE

Some of these challenges, including the stability and dispersion issues, can be addressed by a recently proposed method for implementing reactive non-Foster elements using loss-compensated negative-group-delay (NGD) networks (Mirzaei and Eleftheriades 2013c). In this kind of non-Foster elements, the negative slope of the reactance is synthesized in a lossy NGD network, and the losses are compensated separately by amplification. This is in contrast to the traditional NIC and NII networks where the active and passive parts are intertwined in a circuit with positive feedback loops that tend to destabilize the circuit. The full application of this kind of non-Foster elements is due to further development of bilateral amplification blocks. Nevertheless, when using regular unilateral amplifiers, a special unilateral version of non-Foster reactances can be obtained which can prove useful for specific applications (Mirzaei and Eleftheriades 2013a, c, 2014). For example, a unilateral floating negative capacitor can be useful for matching applications, for which a −2.4 pF implementation is shown in Fig. 46 in the frequency range of 1–1.5 GHz. The experimental results show that a low-dispersion unilateral capacitor with a good quality factor has been obtained in this bandwidth.

DOI 10.1007/978-981-4560-75-7_21-1

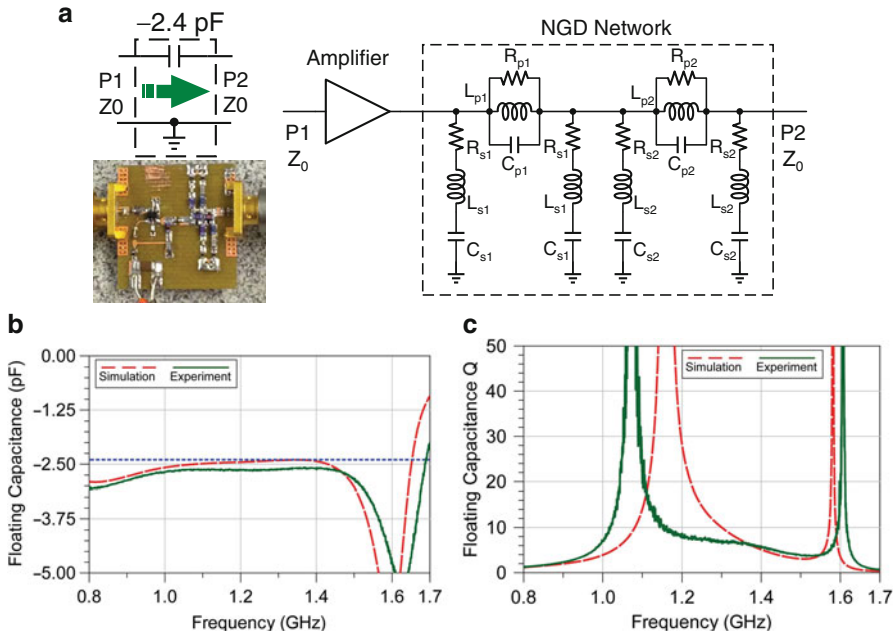


Fig. 46 (**a**) Photograph and schematic of a fabricated unilateral floating capacitor made by cascading an NGD network and an amplifier. (**b**) Floating capacitance value. (**c**) Q, extracted from the S-parameters, shows a low-dispersion capacitance with a good quality factor (Mirzaei and Eleftheriades 2013c) © 2013 IEEE

Conclusion

In this chapter, the theory and operation of transmission-line-based metamaterials has been presented, and it has been shown how this can be used to design various types of passive and active antennas. The transmission-line-based metamaterial structure was formed by periodically loading a conventional microwave transmission line with lumped-element series capacitors and shunt inductors. The resulting structure was termed a negative-refractive-index transmission-line (NRI-TL) metamaterial, and it was shown through a dispersion analysis that it can support both backward waves in the left-handed NRI region and forward waves in the right-handed positive-refractive-index (PRI) region. Furthermore, it can support standing waves with a zero propagation constant at the transition point between the NRI and PRI regions.

The rich propagation characteristics of NRI-TL metamaterials form the underlying basis for their use in many antenna applications, including leaky-wave antennas, compact resonant antennas, and multiband antennas. The resonant characteristics of the NRI-TL metamaterial structures were analyzed with the aid of a dispersion diagram, enabling one to clearly visualize how these structures can be designed to offer multiband responses whose resonant frequencies are not harmonically related while offering large degrees of miniaturization.

Specifically, it was outlined how effective radiation can be achieved from a metamaterial structure when its total electrical length is equal to integer multiples of π, including zero and negative integer values, referred to as the zeroth-order and negative-order resonant modes, respectively. The individual resonant frequencies can be adjusted by changing the values of the reactive loading of the transmission line through a "dispersion engineering" process, therefore allowing the resonances to be placed at multiple nonharmonic frequencies. Subsequently, it was demonstrated how NRI-TL metamaterial antennas can be designed that are independent of their overall physical size, which is in contrast to conventional

DOI 10.1007/978-981-4560-75-7_21-1

resonant antennas whose size is on the order of $\lambda/2$. This useful property allows compact metamaterial antenna designs to be achieved at a fraction of the size of conventional antennas by employing the zeroth-order and negative-order modes of the NRI-TL metamaterials.

Design equations for rapid prototyping have been presented, allowing an antenna designer to easily determine the loading-element values required for a metamaterial antenna with certain volume and system impedance constraints. Guidelines have also been provided for the physical realization of the loading elements, either in fully printed form or using off-the-shelf surface-mount chip components.

In order to highlight the advantages that transmission-line metamaterials have to offer in the design of antennas, a number of passive metamaterial antenna applications have been presented. These include examples of zeroth-order resonant antennas, negative-order resonant antennas, epsilon-negative antennas, mu-negative antennas, metamaterial dipole antennas, and metamaterial-inspired antennas.

In the concluding section, active non-Foster matching networks for small antennas have been presented, and it has been demonstrated how these can be applied to metamaterial-inspired antennas. The implementation of reactive non-Foster elements was presented using two networks: negative impedance converters (NICs) and negative impedance inverters (NIIs). The limitations of antennas that employ external non-Foster matching networks were described, followed by the benefits of using antennas with embedded non-Foster matching networks. Subsequently, a practical design procedure for antennas with embedded non-Foster matching networks was presented, and it was shown how this can be applied to a specific metamaterial antenna design.

Finally, in order to overcome some of the limitations faced by conventional reactive non-Foster elements, including stability, dispersion, and achievable bandwidth, a new method of implementing reactive non-Foster elements using loss-compensated negative-group-delay (NGD) networks was presented. Using this method, it was demonstrated that a low-dispersion unilateral negative capacitor with a good quality factor can be achieved over a wide bandwidth.

Cross-References

See also the following chapters in the Handbook of Antenna Technology:

- ▶ Fosters Reactance Theorem and Non-Foster Impedance Matching in Antenna Systems
- ▶ Metamaterials and Antennas
- ▶ Small Antennas (PIFA/PILA/Loading Antenna/etc.)
- ▶ Wheeler-Fano Impedance Matching Principles

References

Albee TK (1976) Broadband VLF loop antenna system. US Patent 3,953,799

Albert KP (1973) Broadband antennas systems realized by active circuit conjugate impedance matching. Master's thesis, Naval Postgraduate School, Monterey. Acc. No. AD769800

Alu A, Bilotti F, Engheta N, Vegni L (2007) Subwavelength, compact, resonant patch antennas loaded with metamaterials. IEEE Trans Antennas Propag 55(1):13–25

Antoniades MA (2004) Compact linear metamaterial phase shifters for broadband applications. Master's thesis, University of Toronto, Toronto

Antoniades MA (2009) Microwave devices and antennas based on negative-refractive-index transmission-line metamaterials. Ph D thesis, University of Toronto, Toronto

Antoniades MA, Eleftheriades GV (2003) Compact linear lead/lag metamaterial phase shifters for broadband applications. IEEE Antennas Wirel Propag Lett 2(1):103–106
Antoniades MA, Eleftheriades GV (2008a) A CPS leaky-wave antenna with reduced beam squinting using NRI-TL metamaterials. IEEE Trans Antennas Propag 56(3):708–721
Antoniades MA, Eleftheriades GV (2008b) A folded-monopole model for electrically small NRI-TL metamaterial antennas. IEEE Antennas Wirel Propag Lett 7:425–428
Antoniades MA, Eleftheriades GV (2009) A broadband dual-mode monopole antenna using NRI-TL metamaterial loading. IEEE Antennas Wirel Propag Lett 8:258–261
Antoniades MA, Eleftheriades GV (2011a) A multi-band NRI-TL metamaterial-loaded bow-tie antenna. In: Proceedings IEEE AP-S international symposium on antennas and propagation, Spokane, pp 1−4
Antoniades MA, Eleftheriades GV (2011b) A NRI-TL metamaterial-loaded bow-tie antenna. In: Proceedings fifth European conference on antennas and propagation, Rome, pp 1−4
Antoniades MA, Eleftheriades GV (2012) Multiband compact printed dipole antennas using NRI-TL metamaterial loading. IEEE Trans Antennas Propag 60(12):5613–5626
Antoniades MA, Abbosh A, Razali AR (2013) A compact multiband NRI-TL metamaterial-loaded planar antenna for heart failure monitoring. In: Proceedings IEEE AP-S international symposium on antennas and propagation, Orlando, pp 1372−1373
Baek S, Lim S (2009) Miniaturised zeroth-order antenna on spiral slotted ground plane. Electron Lett 45(20):1012–1014
Bahr A (1977) On the use of active coupling networks with electrically small receiving antennas. IEEE Trans Antennas Propag 25(6):841–845
Balanis CA (ed) (2008) Modern antenna handbook. Wiley, Hoboken
Balanis CA (2012) Advanced engineering electromagnetics, 2nd edn. Wiley, New York
Bertin G, Bilotti F, Piovano B, Vallauri R, Vegni L (2012) Switched beam antenna employing metamaterial-inspired radiators. IEEE Trans Antennas Propag 60(8):3583–3593
Best SR (2005) The performance properties of electrically small resonant multiple-arm folded wire antennas. IEEE Antennas Propag Mag 47(4):13–27
Best SR (2014) The significance of composite right/left-handed (CRLH) transmission-line theory and reactive loading in the design of small antennas. IEEE Antennas Propag Mag 56(4):15–33
Bilotti F, Alu A, Vegni L (2008) Design of miniaturized metamaterial patch antennas with μ-negative loading. IEEE Trans Antennas Propag 56(6):1640–1647
Bit-Babik G, Di Nallo C, Svigelj J, Faraone A (2007) Small wideband antenna with non-Foster loading elements. In: Proceedings International conference on electromagnetics in advanced applications (ICEAA), Torino, Italy, pp 105−107
Bode HW (1947) Network analysis and feedback amplifier design. D. Van Nostrand, New York
Brownlie J (1966) On the stability properties of a negative impedance converter. IEEE Trans Circuit Theory 13(1):98–99
Brucher A, Meunier PH, Jarry B, Guilion P, Sussman-Fort SE (1995) Negative resistance monolithic circuits for microwave planar active filter losses compensation. In: Proceedings 25th European microwave conference (EuMC), vol 2, Bologna, Italy, pp 910−915
Caloz C, Itoh T (2003) Novel microwave devices and structures based on the transmission line approach of meta-materials. In: Proceedings IEEE MTT-S international microwave symposium, vol 1, Philadelphia, pp 195−198
Caloz C, Itoh T (2006) Electromagnetic metamaterials: transmission line theory and microwave applications. Wiley, Hoboken
Capolino F (ed) (2009) Metamaterials handbook: applications of metamaterials. CRC Press, Boca Raton
Chu LJ (1948) Physical limitations of omni-directional antennas. J Appl Phys 19(12):1163–1175

Coilcraft Inc (2015) 0402CS (1005) Ceramic chip inductors. http://www.coilcraft.com/0402cs.cfm. Document 198-1. Accessed 1 Feb 2015
Collin RE (1992) Foundations for microwave engineering, 2nd edn. McGraw-Hill, New York
Cui TJ, Smith DR, Liu R (eds) (2010) Metamaterials: theory, design, and applications. Springer, New York
Di Nallo C, Bit-Babik G, Faraone A (2007) Wideband antenna using non-Foster loading elements. In: Proceedings IEEE AP-S international symposium antennas on propagation, Honolulu, HI, USA, pp 4501–4504
Dong Y, Itoh T (2010) Miniaturized substrate integrated waveguide slot antennas based on negative order resonance. IEEE Trans Antennas Propag 58(12):3856–3864
Dong Y, Toyao H, Itoh T (2011) Compact circularly-polarized patch antenna loaded with metamaterial structures. IEEE Trans Antennas Propag 59(11):4329–4333
Eleftheriades GV (2007) Enabling RF/microwave devices using negative-refractive-index transmission-line (NRI-TL) metamaterials. IEEE Antennas Propag Mag 49(2):34–51
Eleftheriades GV (2009) EM transmission-line metamaterials. Mater Today 12:30–41
Eleftheriades GV, Balmain KG (eds) (2005) Negative-refraction metamaterials: fundamental principles and applications. Wiley, Hoboken
Eleftheriades GV, Iyer AK, Kremer PC (2002) Planar negative refractive index media using periodically L-C loaded transmission lines. IEEE Trans Microw Theory Tech 50(12):2702–2712
Eleftheriades GV, Grbic A, Antoniades MA (2004) Negative-refractive-index transmission-line metamaterials and enabling electromagnetic applications. In: Proceedings IEEE AP-S international symposium antennas on propagation, vol 2, Monterey, pp 1399−1402
Eleftheriades GV, Antoniades MA, Qureshi F (2007) Antenna applications of negative-refractive-index transmission-line structures. IET Microw Antennas Propag 1(1):12–22
Elek F, Eleftheriades GV (2005) A two-dimensional uniplanar transmission-line metamaterial with a negative index of refraction. New J Phys 7(163):1–18
Engheta N, Ziolkowski RW (eds) (2006) Metamaterials: physics and engineering explorations. Wiley, Hoboken
Fano RM (1950) Theoretical limitations on the broadband matching of arbitrary impedances. J Franklin Inst 249(1):57–83
Foster RM (1924) A reactance theorem. Bell Syst Tech J 3:259–267
Goubau G (1976) Multi-element monopole antennas. In: Proceedings ECOM-ARO workshop on electrically small antennas, Ft. Monmouth, pp 63−67
Grbic A, Eleftheriades GV (2002) A backward-wave antenna based on negative refractive index L-C networks. In: Proceedings IEEE AP-S international symposium antennas on propagation, vol 4, San Antonio, pp 340−343
Grbic A, Eleftheriades GV (2004) Overcoming the diffraction limit with a planar left-handed transmission-line lens. Phys Rev Lett 92(11):117403
Hakim SS (1965) Some new negative-impedance convertors. Electron Lett 1(1):9–10
Harrington RF (1960) Effect of antenna size on gain, bandwidth and efficiency. J Res Natl Bur Stand 64D(1):1–12
Harris AD, Myers GA (1968) An investigation of broadband miniature antennas. Technical report AD0677320, Naval Postgraduate School, Monterey
Hashemi MRM, Itoh T (2011) Evolution of composite right/left-handed leaky-wave antennas. Proc IEEE 99(10):1746–1754
He Y, Eleftheriades GV (2012) Metamaterial-inspired wideband circular monopole antenna. In: Proceedings IEEE AP-S international symposium antennas on propagation, Chicago, pp 1−2

Herraiz-Martinez FJ, Gonzalez-Posadas V, Garcia-Munoz LE, Segovia-Vargas D (2008a) Multifrequency and dual-mode patch antennas partially filled with left-handed structures. IEEE Trans Antennas Propag 56(8):2527–2539

Herraiz-Martinez FJ, Segovia-Vargas D, Garcia-Munoz LE, Gonzalez-Posadas V (2008b) Dual-frequency printed dipole loaded with meta-material particles. In: Proceedings IEEE AP-S international symposium antennas on propagation, San Diego, pp 1−4

Herraiz-Martinez FJ, Hall PS, Liu Q, Segovia-Vargas D (2011) Left-handed wire antennas over ground plane with wideband tuning. IEEE Trans Antennas Propag 59(5):1460–1471

Iizuka H, Hall PS (2007) Left-handed dipole antennas and their implementations. IEEE Trans Antennas Propag 55(55):1246–1253

Islam R, Eleftheriades GV (2007) Miniaturized microwave components and antennas using negative-refractive-index transmission-line (NRI-TL) metamaterials. Metamaterials (Elsevier) 1:53–61

Islam R, Eleftheriades GV (2012) A review of the microstrip/negative-refractive-index transmission-line coupled-line couplers. IET Microw Antennas Propag 6(1):31–45

Iyer AK, Eleftheriades GV (2004) Leaky-wave radiation from planar negative-refractive-index transmission-line metamaterials. In: Proceedings IEEE MTT-S international microwave symposium, vol 2, Forth Worth, pp 1411−1414

Iyer AK, Kremer PC, Eleftheriades GV (2003) Experimental and theoretical verification of focusing in a large, periodically loaded transmission line negative refractive index metamaterial. Opt Express 11(7):696–708

Jin P, Ziolkowski RW (2010) Linearly and circularly polarized, planar, electrically small, metamaterial-engineered dipole antennas. In: Proceedings IEEE AP-S international symposium antennas on propagation, Toronto, pp 1−4

Kim J, Kim G, Seong W, Choi J (2009) A tunable internal antenna with an epsilon negative zeroth order resonator for DVB-H service. IEEE Trans Antennas Propag 57(12):4014–4017

Kolev S, Delacressonniere B, Gautier J-L (2001) Using a negative capacitance to increase the tuning range of a varactor diode in MMIC technology. IEEE Trans Microw Theory Tech 49(12):2425–2430

Lai A, Itoh T, Caloz C (2004) Composite right/left-handed transmission line metamaterials. IEEE Microw Mag 5(3):34–50

Lai A, Leong KMKH, Itoh T (2007) Infinite wavelength resonant antennas with monopolar radiation pattern based on periodic structures. IEEE Trans Antennas Propag 55(3):868–876

Larky AI (1956) Negative-impedance converter design. Ph D thesis, Stanford University

Larky AI (1957) Negative-impedance converters. IRE Trans Circuit Theory 4(3):124–131

Lee H-M (2011) A compact zeroth-order resonant antenna employing novel composite right/left-handed transmission-line unit-cells structure. IEEE Antennas Wirel Propag Lett 10:1377–1380

Lee J-G, Lee J-H (2007) Zeroth order resonance loop antenna. IEEE Trans Antennas Propag 55(3):994–997

Lee C-J, Leong KMKH, Itoh T (2006) Composite right/left-handed transmission line based compact resonant antennas for RF module integration. IEEE Trans Antennas Propag 54(8):2283–2291

Linvill JG (1953) Transistor negative-impedance converters. Proc IRE 41(6):725–729

Liu Q, Hall PS, Borja AL (2009) Efficiency of electrically small dipole antennas loaded with left-handed transmission lines. IEEE Trans Antennas Propag 57(10):3009–3017

Liu C-C, Chi P-L, Lin Y-D (2012) Compact zeroth-order resonant antenna based on dual-arm spiral configuration. IEEE Antennas Wirel Propag Lett 11:318–321

Liu W, Chen ZN, Qing X (2014) Metamaterial-based low-profile broadband mushroom antenna. IEEE Trans Antennas Propag 62(3):1165–1172

Marques R, Martin F, Sorolla M (2007) Metamaterials with negative parameters: theory, design and microwave applications. Wiley, Hoboken

Mehdipour A, Eleftheriades GV (2014) Leaky-wave antennas using negative-refractive-index transmission-line metamaterial supercells. IEEE Trans Antennas Propag 62(8):3929–3942

Middlebrook RD (1975) Measurement of loop gain in feedback systems. Int J Electron 38(4):485–512

Mirzaei H, Eleftheriades GV (2011a) A wideband metamaterial-inspired compact antenna using embedded non-Foster matching. In: Proceedings IEEE AP-S international symposium antennas on propagation, Spokane, WA, USA, pp 1950–1953

Mirzaei H, Eleftheriades GV (2011b) A compact frequency-reconfigurable metamaterial-inspired antenna. IEEE Antennas Wirel Propag Lett 10:1154–1157

Mirzaei H, Eleftheriades GV (2013a) Unilateral non-Foster elements using loss-compensated negative-group-delay networks for guided-wave applications. In: Proceedings IEEE MTT-S international microwave symposium, Seattle, WA, USA, pp 1–4

Mirzaei H, Eleftheriades GV (2013b) A resonant printed monopole antenna with an embedded non-Foster matching network. IEEE Trans Antennas Propag 61(11):5363–5371

Mirzaei H, Eleftheriades GV (2013c) Realizing non-Foster reactive elements using negative-group-delay networks. IEEE Trans Microw Theory Tech 61(12):4322–4332

Mirzaei H, Eleftheriades GV (2014) Realizing non-Foster reactances using negative-group-delay networks and applications to antennas. In: Proceedings IEEE radio wireless symposium (RWS), Newport Beach, CA, USA, pp 58−60

Myers BR (1965) New subclass of negative-impedance convertors with improved gain-product sensitivities. Electron Lett 1(3):68–70

Nagata M (1965) A simple negative impedance circuit with no internal bias supplies and good linearity. IEEE Trans Circuit Theory 12(3):433–434

Niu B-J, Feng Q-Y (2013) Bandwidth enhancement of CPW-fed antenna based on epsilon negative zeroth- and first-order resonators. IEEE Antennas Wirel Propag Lett 12:1125–1128

Niu B-J, Feng Q-Y, Shu P-L (2013) Epsilon negative zeroth- and first-order resonant antennas with extended bandwidth and high efficiency. IEEE Trans Antennas Propag 61(12):5878–5884

Park B-C, Lee J-H (2011) Omnidirectional circularly polarized antenna utilizing zeroth-order resonance of epsilon negative transmission line. IEEE Trans Antennas Propag 59(7):2717–2721

Park J-H, Ryu Y-H, Lee J-G, Lee J-H (2007) Epsilon negative zeroth-order resonator antenna. IEEE Trans Antennas Propag 55(12):3710–3712

Park JH, Ryu Y-H, Lee J-H (2010) Mu-zero resonance antenna. IEEE Trans Antennas Propag 58(6):1865–1875

Pendry JB (2000) Negative refraction makes a perfect lens. Phys Rev Lett 85(18):3966–3969

Pendry JB, Schurig D, Smith DR (2006) Controlling electromagnetic fields. Science 312:1780–1782

Platzker A, Struble W (1994) Rigorous determination of the stability of linear n-node circuits from network determinants and the appropriate role of the stability factor K of their reduced two-ports. In: Proceedings 3rd international workshop on integrated nonlinear microwave and millimeterwave circuits, Duisburg, Germany, pp 93−107

Qureshi F, Antoniades MA, Eleftheriades GV (2005) A compact and low-profile metamaterial ring antenna with vertical polarization. IEEE Antennas Wirel Propag Lett 4:333–336

Ryan CGM, Eleftheriades GV (2012) Two compact, wideband, and decoupled meander-line antennas based on metamaterial concepts. IEEE Antennas Wirel Propag Lett 11:1277–1280

Sanada A, Caloz C, Itoh T (2004) Planar distributed structures with negative refractive index. IEEE Trans Microw Theory Tech 52(4):1252–1263

Sandberg IW (1960) Synthesis of driving-point impedances with active RC networks. Bell Syst Tech J 39(4):947–962

Schelkunoff SA, Friis HT (1952) Antennas: theory and practice. Wiley, New York, p 309

Schussler M, Freese J, Jakoby R (2004a) Design of compact planar antennas using LH-transmission lines. In: Proceedings IEEE MTT-S international microwave symposium, vol 1, Forth Worth, pp 209−212

Schussler M, Oertel M, Fritsche C, Freese J, Jakoby R (2004b) Design of periodically L-C loaded patch antennas. In: Proceedings 27th ESA antenna technology workshop on innovative periodic antennas, Santiago de Compostela

Sievenpiper D, Lijun Z, Broas RFJ, Alexopoulos NG, Yablonovitch E (1999) High-impedance electromagnetic surfaces with a forbidden frequency band. IEEE Trans Microw Theory Tech 47(11):2059–2074

Skahill G, Rudish RM, Piero JA (1998) Electrically small, efficient, wideband, low-noise antenna elements. In: Proceedings antenna application symposium, Monticello, IL, USA, pp 214−231

Stearns SD (2011) Non-Foster circuits and stability theory. In: Proceedings IEEE AP-S international symposium antennas on propagation, Spokane, WA, USA, pp 1942–1945

Stearns SD (2012) Incorrect stability criteria for non-Foster circuits. In: Proceedings IEEE AP-S international symposium antennas on propagation, Chicago, IL, USA, pp 1–4

Stearns SD (2013) Circuit stability theory for non-Foster circuits. In: Proceedings IEEE MTT-S international microwave symposium, Seattle, WA, USA, pp 1–4

Struble W, Platzker A (1993) A rigorous yet simple method for determining stability of linear N-port networks [and MMIC application]. In: Proceedings GaAs IC symposium digest, San Jose, CA, USA, pp 251−254

Sussman-Fort SE (1998) Gyrator-based biquad filters and negative impedance converters for microwaves. Int J RF Microw Comput Aided Eng 8(2):86–101

Sussman-Fort SE, Rudish RM (2009) Non-Foster impedance matching of electrically-small antennas. IEEE Trans Antennas Propag 57(8):2230–2241

Tian M, Visvanathan V, Hantgan J, Kundert K (2001) Striving for small-signal stability. IEEE Circuits Devices Mag 17(1):31–41

Tretyakov SA, Ermutlu M (2005) Modeling of patch antennas partially loaded with dispersive backward-wave materials. IEEE Antennas Wirel Propag Lett 4:266–269

Vaughan R, Bach-Andersen J (2003) Channels, propagation and antennas for mobile communications. IEE, London

Veselago VG (1968) The electrodynamics of substances with simultaneously negative values of $\in$ and μ. Soviet Phys Uspekhi 10(4):509–514

Volakis JL (2007) Antenna engineering handbook, 4th edn. McGraw-Hill Professional, New York

Volakis JL, Chen C-C, Fujimoto K (2010) Small antennas: miniaturization techniques & applications. McGraw-Hill Professional, New York

Wang C, Hu B-J, Zhang X-Y (2010) Compact triband patch antenna with large scale of frequency ratio using CRLH-TL structures. IEEE Antennas Wirel Propag Lett 9:744–747

Wei K, Zhang Z, Feng Z (2012a) Design of a wideband horizontally polarized omnidirectional printed loop antenna. IEEE Antennas Wirel Propag Lett 11:49–52

Wei K, Zhang Z, Feng Z, Iskander MF (2012b) A MNG-TL loop antenna array with horizontally polarized omnidirectional patterns. IEEE Trans Antennas Propag 60(6):2702–2710

Wheeler HA (1947) Fundamental limitations of small antennas. Proc IRE 35(12):1479–1484

White CR, Colburn JS, Nagele RG (2012) A non-Foster VHF monopole antenna. IEEE Antennas Wirel Propag Lett 11:584–587

DOI 10.1007/978-981-4560-75-7_21-1

Xu ZA, White CR, Yung MW, Yoon YJ, Hitko DA, Colburn JS (2012) Non-Foster circuit adaptation for stable broadband operation. IEEE Microw Wirel Compon Lett 22(11):571–573

Yanagisawa T (1957) RC active networks using current inversion type negative impedance converters. IRE Trans Circuit Theory 4(3):140–144

Zedler M, Eleftheriades GV (2011) Anisotropic transmission-line metamaterials for 2-D transformation optics applications. Proc IEEE 99(10):1634–1645

Zhu J, Eleftheriades GV (2009a) Dual-band metamaterial-inspired small monopole antenna for WiFi applications. Electron Lett 45(22):1104–1106

Zhu J, Eleftheriades GV (2009b) A compact transmission-line metamaterial antenna with extended bandwidth. IEEE Antennas Wireless Propag Lett 8:295–298

Zhu J, Eleftheriades GV (2010) A simple approach for reducing mutual coupling in two closely spaced metamaterial-inspired monopole antennas. IEEE Antennas Wireless Propag Lett 9:379–382

Zhu N, Ziolkowski RW (2012a) Broad-bandwidth, electrically small antenna augmented with an internal non-Foster element. IEEE Antennas Wireless Propag Lett 11:1116–1120

Zhu N, Ziolkowski RW (2012b) Design and measurements of an electrically small, broad bandwidth, non-Foster circuit-augmented protractor antenna. Appl Phys Lett 101(2):024107

Zhu J, Antoniades MA, Eleftheriades GV (2010) A compact tri-band monopole antenna with single-cell metamaterial loading. IEEE Trans Antennas Propag 58(4):1031–1038

Ziolkowski RW, Erentok A (2006) Metamaterial-based efficient electrically small antennas. IEEE Trans Antennas Propag 54(7):2113–2130

Ziolkowski RW, Tang M-C, Zhu N (2013) An efficient, broad bandwidth, high directivity, electrically small antenna. Microw Opt Tech Lett 55(6):1430–1434

Zobel OJ (1923) Theory and design of uniform and composite electric wave-filters. Bell Syst Tech J 2(1):1–46

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_22-1

Theory of Transformation Optics in Antenna Design

Di Bao and Tie Jun Cui*
State Key Laboratory of Millimeter Waves, Department of Radio Engineering, Southeast University, Nanjing, P. R. China

Abstract

Transformation optics provides a bridge between the electromagnetic functionality of the device and the material properties of the custom-engineered media. This chapter includes an overview of transformation optics theory and their application in antenna engineering. The basic theory of transformation optics is analyzed, including the general transformation and quasi-conformal mapping. Reviews are focused on the planar lens antenna, the multibeam antenna, the Luneburg lens antenna, and the metasurface Luneburg lens.

Keywords

Transformation optics; conformal mapping; quasi-conformal transformation optics; metamaterials; lens antenna; luneburg lens; metasurface

Introduction

Einstein's general theory of relativity (Crelinsten 2006) indicates that the matter-energy densities result in motion of matter and light propagating in a curved space-time domain (Wald 1984). In 1961, Dolin applied the idea of curved space to electromagnetic system with inhomogeneous fillings (Dolin 1961) for the first time. He proved that Maxwell's equations are form invariant under a space-deforming transformation and proposed an early form of invisibility cloak. In the 1990s, Pendry's group's insightful work reestablished the field of transformation optics (TO), or transformation electromagnetics (Ward and Pendry 1996, 1998). However, limited by naturally available materials to physically realize the design, the concept did not attract much attention.

Metamaterials have been a hot research field since the beginning of this century (Pendry et al. 1999; Shelby et al. 2001). In 2006, its impressive progress and capability to realize gradient index (GRIN) structure with spatially varying permittivity and permeability values inspired scientists that it is possible to implement the inhomogeneous and even anisotropic constitutive parameters of the TO structure with metamaterials (Pendry et al. 2006; Leonhardt 2006). TO has finally attracted intensive attentions, which leads to the development of many novel functional devices, such as the invisible free-space cloak (Pendry et al. 2006; Schurig et al. 2006), carpet cloak (Li and Pendry 2008; Liu et al. 2009; Ma et al. 2009), omnidirectional absorber (Narimanov and Kildishev 2009; Cheng et al. 2010), EM concentrator (Rahm et al. 2008; Yaghjian and Maci 2009), etc. (Mei and Cui 2012; Werner and Kwon 2014).

Transformation optics has also offered great flexibility to design novel antenna, such as a planar focusing antenna (Kong et al. 2007; Kwon and Werner 2009), highly directive multibeam antenna (Jiang et al. 2008b), small aperture antenna (Luo et al. 2009; Lu et al. 2009), layered lens antenna (Jiang et al. 2008a; Tichit et al. 2009), and transformed Luneburg lens (Demetriadou and Hao 2011). Most of

*Email: tjcui@seu.edu.cn

DOI 10.1007/978-981-4560-75-7_22-1

these early works contain inhomogeneous and anisotropic materials with both tensor ε and μ values, which are difficult to realize with only a few physical verification examples (Tichit et al. 2011; Jiang et al. 2011, 2012) and suffer from high loss and narrow bandwidth. In 2008, the quasi-conformal transformation optics (QCTO) was proposed, which is capable of designing all-dielectric transformation devices while still maintaining the performance. Following this theory, many kinds of transformation antenna have been reported (Mei et al. 2010; Tang et al. 2010; Yang et al. 2011a, b; Kwon 2012; Wu et al. 2013; Oliveri et al. 2014) and experimental demonstrations have been carried out in both 2D and 3D cases (kundtz and Smith 2010; Ma and Cui 2010; Mei et al. 2011; Liang and Hum 2014; Mateo-Segura et al. 2014) and surface wave (Wan et al. 2014).

In this chapter, the general transformation optics is introduced first, and its applications in multibeam antenna and layered lens antenna are reviewed. Then, the quasi-conformal mapping is presented along with its utilization in flattened Luneburg lens, surface flattened Luneburg lens, and planar lens design. In both parts, the first device is explained in detail to show how the transformation works.

General Transformation Optics

Theory of General Transformation Optics

Based on coordinate transformation, transformation optics can map a free space (virtual system) into a distorted space (physical system) and link the spatial distortions to changes in the distributions of the material's constitutive parameter tensors. The physical implementation of the materials in the transformed space is the designed transformation optics device, where the route of the wave is decided by how the space is distorted or, equivalently, how the material properties vary across the geometry (Pendry et al. 2006; Leonhardt 2006).

Considering a coordinate transformation which maps a virtual system (x_1, x_2, x_3) to a physical system (x_1', x_2', x_3'), the form invariance of Maxwell's equations can be employed to derive the constitutive parameters of the medium in the transformed space as (Pendry et al. 2006)

$$\overline{\varepsilon'} = \frac{\overline{\Lambda}\overline{\varepsilon}\overline{\Lambda}^T}{det\left(\overline{\Lambda}\right)}, \quad \overline{\mu'} = \frac{\overline{\Lambda}\overline{\mu}\overline{\Lambda}^T}{det\left(\overline{\Lambda}\right)} \tag{1}$$

where $\overline{\Lambda}$ is the Jacobian matrix and can be expressed as

$$\overline{\Lambda} = \begin{pmatrix} \frac{\partial x_1'}{\partial x_1} & \frac{\partial x_1'}{\partial x_2} & \frac{\partial x_1'}{\partial x_3} \\ \frac{\partial x_2'}{\partial x_1} & \frac{\partial x_2'}{\partial x_2} & \frac{\partial x_2'}{\partial x_3} \\ \frac{\partial x_3'}{\partial x_1} & \frac{\partial x_3'}{\partial x_2} & \frac{\partial x_3'}{\partial x_3} \end{pmatrix} \tag{2}$$

In the design of most transformation lens antenna, 2D cases with TE wave incidence are usually considered in the first step. 3D design can be obtained by rotating the 2D results, and TM mode can be calculated with the principle of duality.

DOI 10.1007/978-981-4560-75-7_22-1

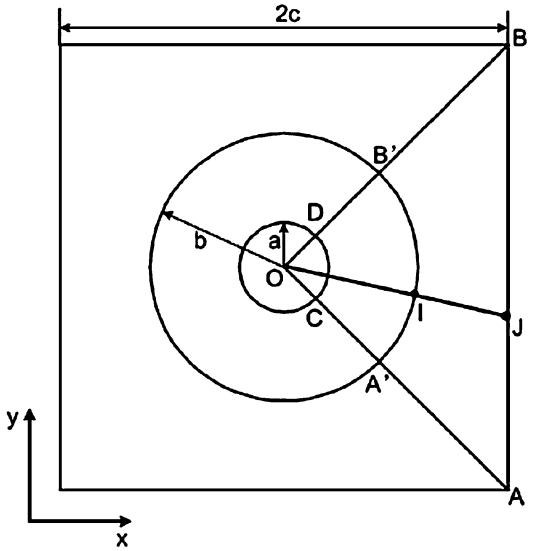


Fig. 1 Configuration of cylindrical-to-plane-wave conversion in the Cartesian coordinate (Jiang et al. 2008b)

In the Cartesian coordinate system, the TE wave consists of only H_x', H_y', and E_z' components; then only μ_{xx}', μ_{yy}', μ_{xy}', μ_{yx}', and ε_{zz}' contribute. So the Jacobian matrix $\overline{\Lambda}$ can be simplified as

$$\overline{\Lambda} = \begin{pmatrix} \dfrac{\partial x'}{\partial x} & \dfrac{\partial x'}{\partial y} & 0 \\ \dfrac{\partial y'}{\partial x} & \dfrac{\partial y'}{\partial y} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3}$$

Even in 2D cases, the permittivity and permeability values generated from Eq. 1 are generally anisotropic and inhomogeneous diagonal tensors with extremely high or lower than unity values. Metamaterials are capable of accessing these extreme medium properties in the design with the dispersive resonant unit cell, and the anisotropic properties can be implemented by rotating the directions of the metamaterial unit cell (Schurig et al. 2006; Jiang et al. 2011, 2012). Such resonant structure will inevitably increase the losses and reduce the bandwidth of operation.

Multibeam Antenna via Continuous Transformation Optics

Based on coordinate transformation, the conversion from cylindrical waves to plane waves in a short range has been proposed (Jiang et al. 2008b) with potential applications in a four-beam antenna or a compact range for near-field measurement of plane waves.

As shown in Fig. 1, the square domain is divided into four identical triangles, and one of the triangles, OAB, is discussed first for simplicity. If the fan-shaped area $\widehat{OCD}$ is transformed to a bigger space in sector $\widehat{OA'B'}$ and the arc $CA'B'D$ is transformed to a trapezia-like region $A'ABB'$, then cylindrical waves excited at the origin would propagate through the square $A'ABB'$ with the equiphase surface changing from circle to plane. The transformation in triangles OAB could be written as

DOI 10.1007/978-981-4560-75-7_22-1

$$r' = \begin{cases} br/a, & 0 \le r \le a, \\ \dfrac{cr - bx}{(b-a)x}(r-a) + b, & a \le r \le b, \end{cases} \tag{4}$$

where (x, y) is an arbitrary point in the virtual (original) space, (x', y') is the corresponding point in the physical (transformed) space, $r = \sqrt{x^2 + y^2}$, and $r' = \sqrt{x'^2 + y'^2}$.

The coordinate transformation can be mathematically described as

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{cases} \begin{pmatrix} bx/a \\ by/a \\ z \end{pmatrix}, & 0 \le r \le a, \\ \begin{pmatrix} \dfrac{(c - bx/r)\ (r-a)}{b-a} + \dfrac{bx}{r} \\ \dfrac{(c - bx/r)\ (r-a)y}{(b-a)x} + \dfrac{by}{r} \end{pmatrix}, & a \le r \le b. \end{cases} \tag{5}$$

With Eqs. 3 and 5, the Jacobian transformation can be obtained as

$$\overline{\Lambda} = \begin{cases} \begin{pmatrix} b/a & 0 & 0 \\ 0 & b/a & 0 \\ 0 & 0 & 1 \end{pmatrix}, & 0 \le r \le a, \\ \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix}, & a \le r \le b, \end{cases} \tag{6}$$

in which

$$\begin{aligned} a_{11} &= \frac{cr^2x - br^3 + b^2y^2}{(b-a)r^3}, \\ a_{12} &= \frac{\left(cr^2 - b^2x\right)y}{(b-a)r^3}, \\ a_{21} &= \frac{\left(acr^3 - b^2x^3 - cr^2y^2\right)y}{x^2(b-a)r^3}, \\ a_{22} &= \frac{b^2x^3 + cr^4 - acr^3 + cr^2y^2 - br^3x}{(b-a)r^3x}. \end{aligned}$$

The determinant of the above Jacobian matrix can be calculated as

$$\det \overline{\Lambda} = \begin{cases} b^2/a^2, & 0 \le r \le a \\ a_{11}a_{22} - a_{12}a_{21}, & a \le r \le b \end{cases}. \tag{7}$$

According to Eq. 1, the permittivity and permeability tensors of the medium in the physical space can be obtained as

DOI 10.1007/978-981-4560-75-7_22-1

$$\overline{\varepsilon_r'} = \overline{\mu_r'} = \begin{cases} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & a^2/b^2 \end{pmatrix} & 0 \le r' \le b, \\ \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & 0 \\ \varepsilon_{yx} & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{pmatrix} & b \le r', \quad x' \le c, \end{cases} \tag{8}$$

in which

$$\varepsilon_{xx} = \frac{a_{11}^2 + a_{12}^2}{\det(\overline{\Lambda})}, \tag{9a}$$

$$\varepsilon_{xy} = \frac{a_{11}a_{21} + a_{12}a_{22}}{\det(\overline{\Lambda})} = \varepsilon_{yx}, \tag{9b}$$

$$\varepsilon_{yy} = \frac{a_{21}^2 + a_{22}^2}{\det(\overline{\Lambda})}, \tag{9c}$$

$$\varepsilon_{zz} = \frac{1}{\det(\overline{\Lambda})}. \tag{9d}$$

The constitutive parameter tensors for the triangle *OAB* are now fully provided in Eq. 9 in the transformed coordinates (x', y'). Rotating these equations with angles of $\pi/2$, π, and $3\pi/2$ around the z-axis, the corresponding relative permittivity and permeability tensors could be obtained in the whole square domain. To verify the effectiveness of the design, full-wave simulations have been carried out based on finite-element method. The cylindrical waves are excited by a line-current source of infinite length oriented in the z direction in the origin of the coordinate system at 8 GHz. The electric field distribution inside and outside the conversion materials is shown in Fig. 2. It can be observed that when the cylindrical

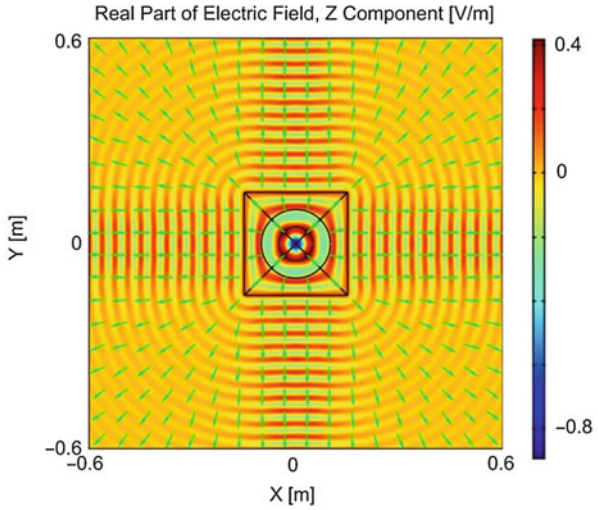


Fig. 2 The distributions of electric fields and power-flow lines for the conversion from cylindrical waves to plane waves via the layered optical transformation (Jiang et al. 2008b)

DOI 10.1007/978-981-4560-75-7_22-1

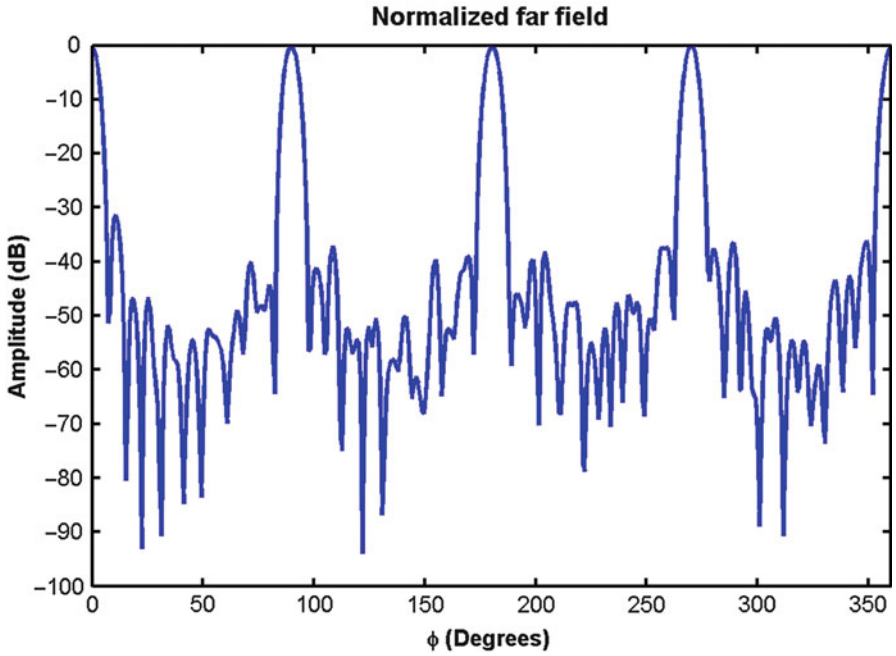


Fig. 3 The far-field radiation pattern of the multibeam antenna (Jiang et al. 2008b)

wave propagates through the short conversion metamaterial layer, four beams of plane waves can be clearly observed in the neighboring area. The far-field radiation pattern of the multibeam antenna is also calculated and plotted in Fig. 3, which exhibits little scattering effect at the four corners of the square and high-directivity radiation to the four directions with the antenna aperture as small as eight wavelengths in free space.

A similar highly directive four-beam antenna has then been designed and experimentally demonstrated by D. H. Werner's group in 2011 (Jiang et al. 2011) by mapping a long and narrow isosceles triangles into another isosceles triangles with big vertex angle. The lens is physically realized with split-ring resonators (SRRs) in the G-band with an embedded monopole antenna source. The anisotropic material parameters of the transformed medium are implemented by rotating the directions of the SRR unit cell. Experimental results indicate that the lens obtains a quadbeam radiation pattern over 20 % bandwidth with a gain enhancement of about 6.5 dB.

Lens Antennas Based on Layered Optical Transformation

The optical transformation has been applied to design a 2D layered homogeneous lens with uniaxially anisotropic material (Jiang et al. 2008a). As shown in Fig. 4, the virtual space is a rectangle $ABCD$, while the physical space is a trapezia $ABC'D'$. Both spaces are divided into n layers as the shaded region indicates. For the kth $(1 \leq k \leq n)$ layer, an elongating coordinate transformation is defined as

$$x' = x + \frac{(k - 0.5)\ (b - a)x}{na}, \quad (1 \leq k \leq n), \tag{10a}$$

$$y' = y, \tag{10b}$$

$$z' = z. \tag{10c}$$

DOI 10.1007/978-981-4560-75-7_22-1

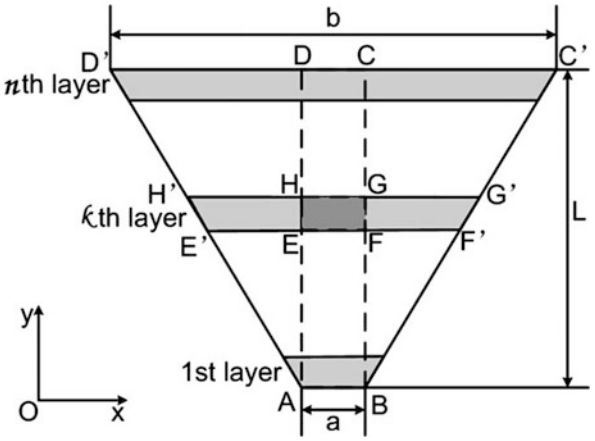


Fig. 4 The 2D layered lens in the Cartesian coordinate system (Jiang et al. 2008a)

If the incident signal is emitted from port *AB*, then the local properties of the EM fields in the small region *ABCD* will be mapped to a large region $ABC'D'$ while still maintaining the equiphase front at port $C'D'$ as port *CD*, which promises a high-gain performance for the lens.

The full expressions of the constitutive parameter tensors in each layer of the physical space can then be derived from Eq. 1 as

$$\varepsilon_{xx}^{k} = \mu_{xx}^{k} = \alpha_k, \tag{11a}$$

$$\varepsilon_{yy}^{k} = \mu_{yy}^{k} = 1/\alpha_k, \tag{11b}$$

$$\varepsilon_{zz}^{k} = \mu_{zz}^{k} = 1/\alpha_k. \tag{11c}$$

where all the off-diagonal parameters are zero and

$$\alpha_k = 1 + \frac{(k - 0.5)\ (b - a)}{na}. \tag{12}$$

Full expressions of the constitutive parameter tensors are provided in Eq. 11. In the practical case of TE wave illumination with E field polarized along the *z*-axis, only ε_{zz}^{k}, μ_{xx}^{k}, and μ_{yy}^{k} are required in Eq. 11. In these equations, the dispersion relations of the physical space can be kept unchanged if the products of $\mu_{xx}^{k}\varepsilon_{zz}^{k}$ and $\mu_{yy}^{k}\varepsilon_{zz}^{k}$ are identical as those in the above equations. One of the convenient sets for physical realization is

$$\mu_{xx}^{k} = 1, \quad \varepsilon_{zz}^{k} = 1, \quad \mu_{yy}^{k} = 1/\alpha_k^2, \quad 1 \le k \le n, \tag{13}$$

where only μ_{yy}^{k} has different value in each layer of the transformed lens antenna, so the lens is easy to realize physically with metamaterials. There would be no reflection for normal incidence because of the impedance matching.

For comparison, the lens antenna is also designed based on continuous optical transformation in (Jiang et al. 2008a), and full-wave simulations of these lens antennas placed inside a PEC horn antenna are performed at 6 GHz. The electric field distributions of the lens antenna composed of layered homogeneous material defined as Eq. 11 are first studied as shown in Fig. 5a. The lens in Fig. 5b is composed of

DOI 10.1007/978-981-4560-75-7_22-1

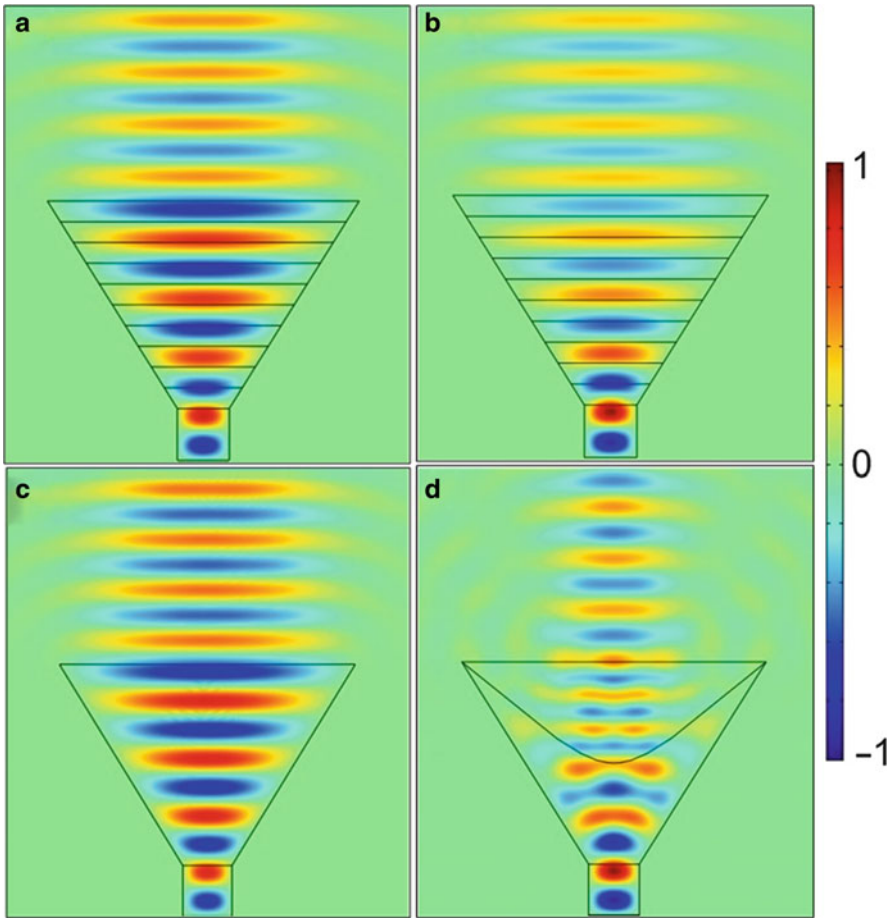


Fig. 5 The electric field distribution of the lens antennas located in the same PEC horn antenna excited with a line-current source. The lens is composed of (**a**) layered homogeneous medium following Eq. 11, (**b**) layered homogeneous medium following Eq. 13, (**c**) inhomogeneous and anisotropic materials defined by continuous transformation, and (**d**) regular dielectric (Jiang et al. 2008a)

layered homogeneous material defined according to the simplified parameters in Eq. 13. It can be observed that the reflections in Fig. 5b are less than those in Fig. 5a. Figure 5c demonstrates the field distribution of the lens designed by continuous optical transformation, and it reveals strong standing waves and reflection inside the lens.

As a comparison with the conventional lens, a horn antenna filled with collimating lens composed of regular dielectric in accordance with the method in (Liberti and Rappaport 1999) is shown in Fig. 5d. The dimension of the PEC horn is the same in all four subfigures, so it is obvious that the three OT lens antennas can concentrate almost all EM powers as beam of plane waves in front of the antennas and hence much better performance than the conventional lens. Such transformation can be applied to design multibeam antennas as well.

Quasi-Conformal Transformation Optics

Mathematically, there are infinite kinds of mapping from one virtual domain to one physical domain with different analytical or numerical methods (Chen et al. 2009; Chang et al. 2010). Researchers can optimize the transformation to minimize the anisotropy of the designed medium to quasi-isotropic material for

DOI 10.1007/978-981-4560-75-7_22-1

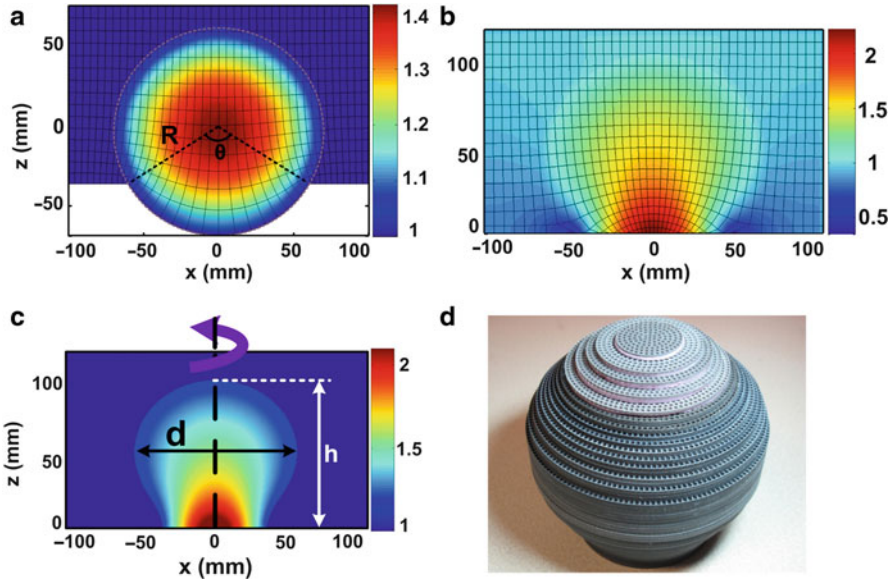


Fig. 6 (**a**) The refraction index distribution of 2D Luneburg lens embedded in the virtual space. (**b**) The refractive index distribution of 2D flattened Luneburg lens in the physical space. (**c**) The final refract index distribution in the *xoz* plane. (**d**) The photo of the fabricated 3D lens (Ma and Cui 2010)

engineering application (Li and Pendry 2008; Werner and Kwon 2014). The detailed theory and operating mechanism of QCTO are presented in the following subsection with the design examples.

Three-Dimensional Luneburg Lens Based on Quasi-Conformal Transformation Optics

A Luneburg lens is a spherically symmetric gradient index lens which could guide the incoming collimated rays from infinity to a focal point on the opposite surface of the lens without any aberrations. The application of Luneburg lens is restricted from the gradient index distribution of refractive index, which is difficult to implement, and the spherical surface, which leads to mismatch to the planar feeding source or detector array. In 2010, Kundtz and Smith proposed a solution by flattening the spherical surface into a planar surface with the QCTO carried out a 2D experimental demonstration with metamaterials (Kundtz and Smith 2010). A series of Luneburg lens has been designed in Yang's group (Demetriadou and Hao 2011; Quevedo-Teruel et al. 2012; Quevedo-Teruel and Hao 2013) with an all-dielectric ceramic flat lens fabricated (Mateo-Segura et al. 2014). Cui's group designed a 3D flattened Luneburg lens based on the QCTO and fabricated the lens with all-dielectric metamaterials in three dimensions (3D) at Ku-band (Ma and Cui 2010). The designed 3D lens is more advantageous than conventional lens with no aberration, a zero focal distance, a flattened focal surface, and the ability to construct radiation beams at extremely large angles.

For simplicity, the 2D case is considered first with TE wave incidence. The Luneburg-based lens and its background area as shown in Fig. 6a are mapped to a new lens in a rectangular area with a part of flat surface in Fig. 6b. The Luneburg surface in the virtual space with a radius R has the refractive index distribution in free space as

$$n = \sqrt{2 - \frac{r^2}{R^2}}. \tag{14}$$

DOI 10.1007/978-981-4560-75-7_22-1

In the virtual space, the very fine grid is generated with near-orthogonal cells such that

$$\frac{\partial x'}{\partial y} \simeq 0, \tag{15a}$$

$$\frac{\partial y'}{\partial x} \simeq 0. \tag{15b}$$

Since x' and y' are functions of both x and y, Eq. 15 can be also written according to the chain rule as

$$\frac{\partial x'}{\partial y} = \frac{\partial x'}{\partial x}\,\frac{\partial x}{\partial y} \simeq 0, \tag{16a}$$

$$\frac{\partial y'}{\partial x} = \frac{\partial y'}{\partial y}\,\frac{\partial y}{\partial x} \simeq 0. \tag{16b}$$

So the Jacobian matrix in Eq. 3 can be reduced to

$$\overline{\Lambda} = \begin{pmatrix} \dfrac{\partial x'}{\partial x} & 0 & 0 \\ 0 & \dfrac{\partial y'}{\partial y} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{17}$$

The permittivity and permeability values can then be calculated by Eq. 1 as

$$\epsilon'_z = \frac{\epsilon_z}{det\Lambda} \simeq \epsilon_z \frac{\Delta x \Delta y}{\Delta x' \Delta y'}, \tag{18a}$$

$$\mu' = \mu_0 \begin{pmatrix} \dfrac{\left(\frac{\partial x'}{\partial x}\right)^2}{det\Lambda} & 0 \\ 0 & \dfrac{\left(\frac{\partial y'}{\partial y}\right)^2}{det\Lambda} \end{pmatrix} \tag{18b}$$

where Δx, Δy, $\Delta x'$, and $\Delta y'$ are the dimensions of each cell in the two coordinate systems in Fig. 6.

According to the theoretical study in (Li and Pendry 2008), the suitable coordinate transform has been chosen in the design so that the anisotropic factor can be dismissed with unit permeability value. The physical medium can then be described by Eq. 18a alone. The refractive index value in the transformed lens can be expressed as

$$n' \simeq n\sqrt{\frac{\Delta x \Delta y}{\Delta x' \Delta y'}}. \tag{19}$$

The refractive index distribution is drawn in Fig. 6b. The background cells with refractive index value close to unity are all set to be air, and the final 2D refractive index profile is plotted in Fig. 6c. The 3D refractive index distribution can now be generated by rotating the 2D surface along the z-axis in Fig. 6c.

DOI 10.1007/978-981-4560-75-7_22-1

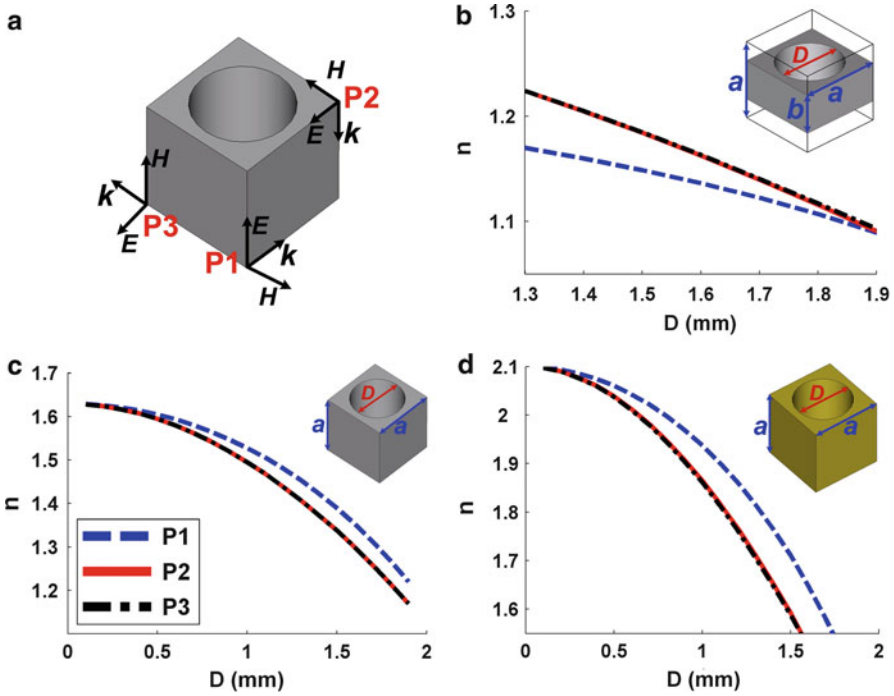


Fig. 7 The relations between the refractive index and the hole's diameter D at 15 GHz under different polarizations. (**a**) Three polarizations P1, P2, and P3 for a unit cell, (**b**) the 2 × 2 × 1 mm^3 F4B unit cell, (**c**) the 2 × 2 × 2 mm^3 F4B unit cell, (**d**) the 2 × 2 × 2 mm^3 R4B unit cell (Ma and Cui 2010)

The transformed 3D Luneburg lens is then fabricated with multilayered nonresonant metameterials, which is realized by drilling inhomogeneous holes in two kinds of dielectric plates, FR4 ($\varepsilon = 4.4$, $\delta = 0.025$) and F4B($\varepsilon = 2.65$, $\delta = 0.001$). Three kinds of unit cells have been studied to fulfill the designed refractive index distribution, and the effective refractive index values of these structures are calculated with the S-parameter retrieval method (Smith et al. 2002) and plotted in Fig. 7. All three orthogonal polarizations of incident waves are considered and quasi-isotropic properties have been observed.

Simulated and measured near electric fields of the lens have been carried out with excellent agreement, and the measured radiation pattern of the 3D lens in Ku-band is plotted in Fig. 8 (Ma and Cui 2010). A coax-to-waveguide device with the aperture size of 16 × 8 mm^2 was applied as the feeding source at three typical source positions on the flattened focal surface. As shown in Fig. 8, the patterns in the left column illustrate far-field radiation diagrams in E-plane (horizontally polarization) with different feeding locations at 12.5, 15, and 18 GHz, and those in the right column are the patterns in H-plane (vertically polarization). Highly directive beam is observed, and the radiation direction can be effectively controlled by the feeding position in a range of 50° with respect to the normal direction (*z*-axis).

Flattened Metasurface Luneburg Lens

Quasi-conformal mapping has also been successfully utilized in metasurface antenna design, such as the flattened metasurface Luneburg lens (Wan et al. 2014). The transformation of the metasurface lens is similar to the 2D part of the previous 3D Luneburg lens design, as shown in Fig. 6a–c. In the physical realization, the surface refractive indexes are related to the surface impedances, which can be readily obtained with U-shaped metallic unit cell as shown in Fig. 9a, and the surface refractive index values can be controlled by tuning the parameter *h*. The designed flattened Luneburg metasurface lens is fabricated as shown in Fig. 9b, with three antipodal Vivaldi antennas function as the planar source array to excite the

DOI 10.1007/978-981-4560-75-7_22-1

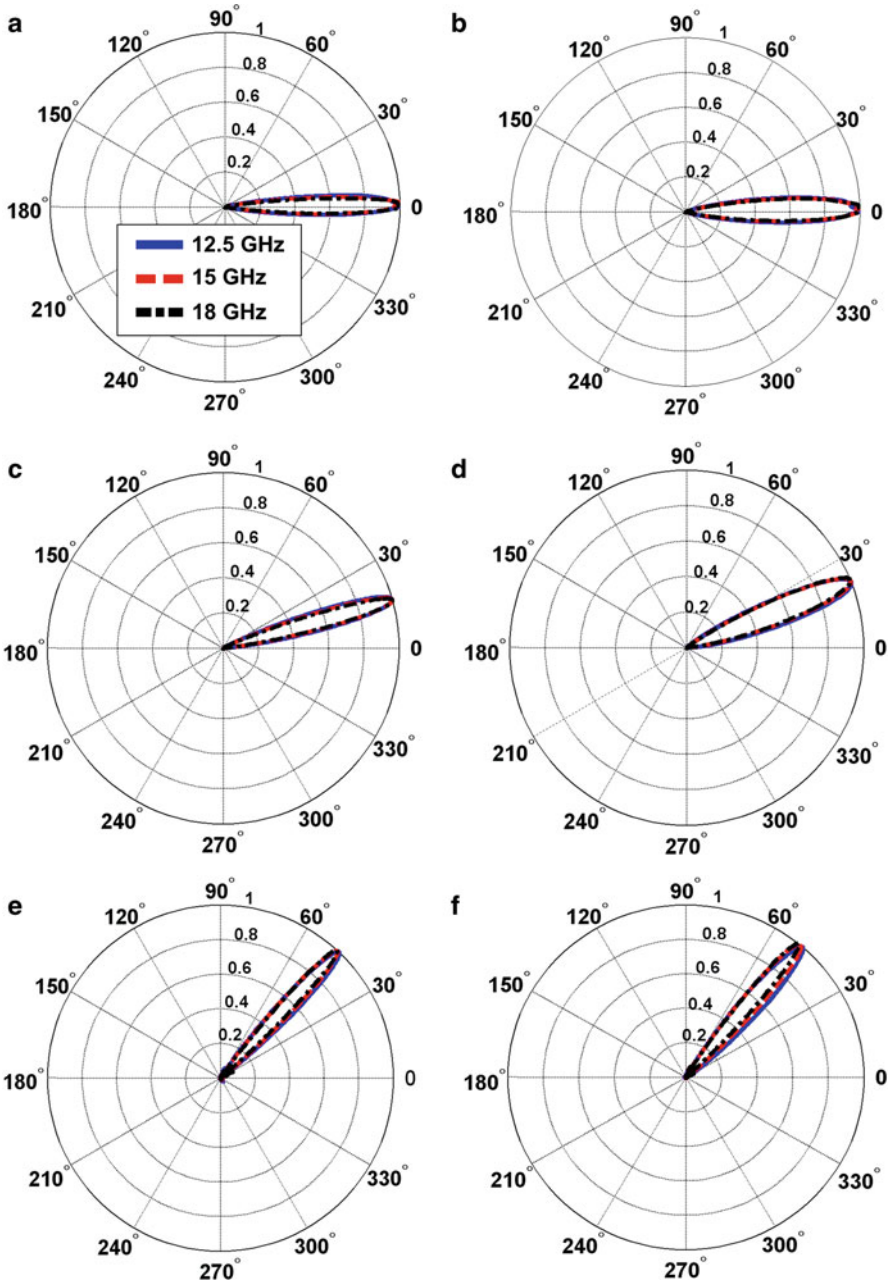


Fig. 8 The measured far-field pattern of the 3D lens in E-plane (**a**, **c**, **e**) and H-plane (**b**, **d**, **f**) at 12.5, 15, and 18 GHz. The location of the feeding source is (**a**, **b**): ($x = 0$ and $y = 0$), (**c**, **d**): ($x = -10mm$ and $y = 0$), and (**e**, **f**): ($x = -30mm$ and $y = 0$), respectively (Ma and Cui 2010)

TE-mode surface wave on the surface. The measured near-electric field distribution at 9 GHz is shown in Fig. 9c. It can be observed that the Vivaldi antennas at different locations can excite surface waves propagating along different directions from the Luneburg lens. More simulation and experimental results indicate that the lens is also working at 8 and 10 GHz (Wan et al. 2014).

Planar Focusing Reflector Lens Antenna

Parabolic antenna can produce the narrow beamwidths and high directivity, and it has been widely used in point-to-point communications, such as microwave relay links to carry telephone and television signals

DOI 10.1007/978-981-4560-75-7_22-1

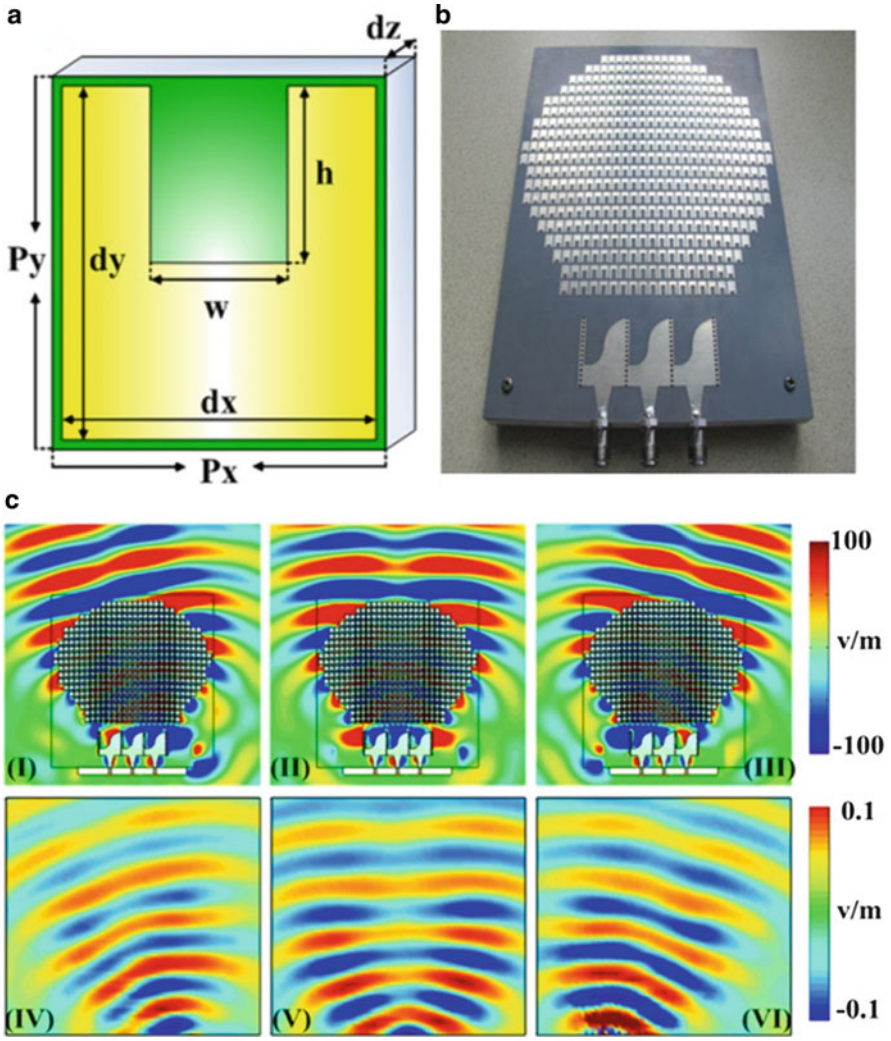


Fig. 9 (**a**) The U-shaped particle. (**b**) The photo of the designed flattened Luneburg lens. (**c**) The simulated (*top three subfigures*) and measured (*bottom three subfigures*) near-field distribution when exciting the right antenna (*I* and *IV*), the central antenna (*II* and *V*), and the left antenna (*III* and *VI*), respectively (Wan et al. 2014)

between neighboring cities, satellite and spacecraft communication antennas, radio telescopes, and so on. In 2007, J. A. Kong's group suggested the design of a low-profile planar focusing antenna based on the general transformation of a parabolic antenna (Kong et al. 2007) and verified the performance of the designed anisotropic and inhomogeneous antenna with simulation. Based on quasi-conformal mapping, the novel planar focusing antenna has been designed and verified through numerical simulation with commercial software (Mei et al. 2010) and the finite-difference time-domain method (Tang et al. 2010) independently in 2010 and was experimentally demonstrated in 2011 (Mei et al. 2011).

The virtual and physical spaces in the experimental verification work (Mei et al. 2011) are illustrated in Fig. 10a, b. The curved boundary at the bottom of the virtual space is a 2D parabolic reflector. Based on quasi-conformal mapping, the quasi-rectangular region in the virtual space as Fig. 10a is transformed into a standard rectangle in the physical space as Fig. 10b in TE mode. The refractive index distribution in the designed physical space is shown in Fig. 10c. The less than one material parameters are all set to one to avoid dispersive resonant structure in the physical implementation. The refractive index value in most part of region I is close to one, so it is excluded from the final design and the domain of the flat lens antenna is then reduced to region II.

DOI 10.1007/978-981-4560-75-7_22-1

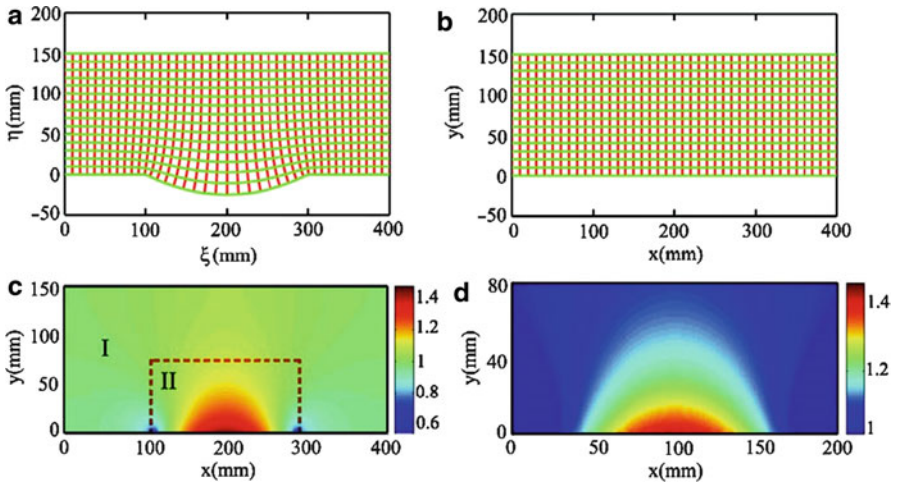


Fig. 10 (**a**) The virtual space of the parabolic reflector. (**b**) The physical space of the planar antenna with quasi-conformal mapping. (**c**) The refractive index distribution in (**b**). (**d**) The refractive index distribution of the final lens antenna with the n less than one set to one (Mei et al. 2011)

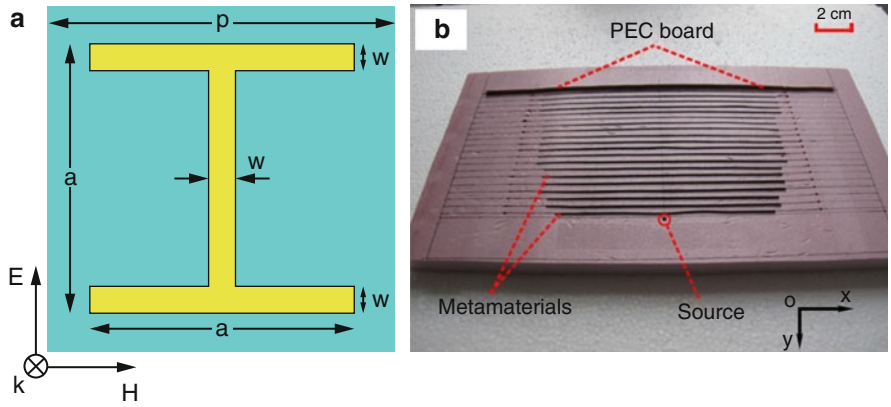


Fig. 11 (**a**) The geometry of the I-shaped unit cell. (**b**) The schematic of the fabricated planar antenna (Mei et al. 2011)

The planar antenna is designed in X-band with the broadband and low-loss I-shaped metamaterials as depicted in Fig. 11a. A different refractive index value in the transformed device can be obtained by changing the height of the I-shaped structure, *a*. PCB strips supported by refractive index value close to one foam structure are constructed to build the planar lens as illustrated in Fig. 11b. The measured near-field distributions of the parabolic reflector and planar lens antenna at 10 GHz agree well with each other as shown in Fig. 12. Further simulation and experimental results indicate that the planar antennas have a broadband performance from 8.5 to 11.5 GHz.

Conclusion

In this chapter, the transformation optics under two main types has been reviewed: the general transformation and the quasi-conformal mapping. Based on these theories, application examples including the multibeam antenna, layered lens antenna, 3D flattened Luneburg lens, flattened metasurface Luneburg lens, and planar reflector lens are introduced to demonstrate the flexibility of TO provided to manipulate EM wave and design novel antennas.

DOI 10.1007/978-981-4560-75-7_22-1

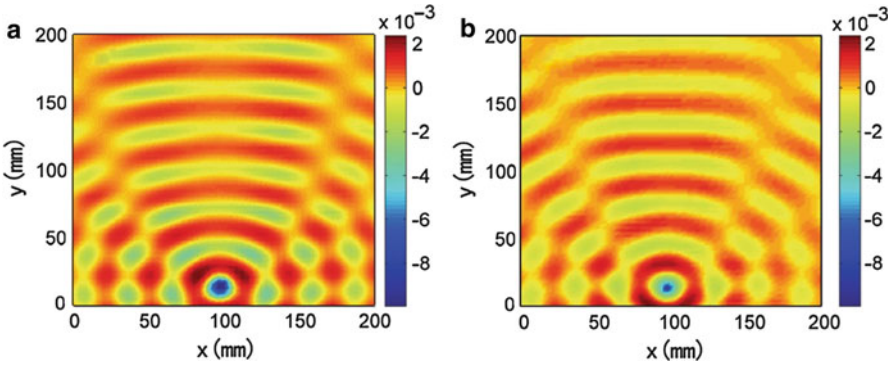


Fig. 12 Measured near-field distribution at 10 GHz of (**a**) the parabolic reflector and (**b**) the transformed planar antenna (Mei et al. 2011)

Cross-References

- ▶ Broadband and Multiband Planar Antennas
- ▶ Lens Antennas
- ▶ Metamaterials and Antennas
- ▶ Near-Field Antenna Measurement Techniques
- ▶ Radio Frequency Beamforming for Scanned and Multibeam Antenna Systems

References

Chang Z, Zhou X, Hu J, Hu G (2010) Design method for quasi-isotropic transformation materials based on inverse laplace's equation with sliding boundaries. Opt Express 18(6):6089–6096

Chen X, Fu Y, Yuan N (2009) Invisible cloak design with controlled constitutive parameters and arbitrary shaped boundaries through helmholtz's equation. Opt Express 17(5):3581–3586

Cheng Q, Cui TJ, Jiang WX, Cai BG (2010) An omnidirectional electromagnetic absorber made of metamaterials. New J Phys 12(6):063006

Crelinsten J (2006) Einstein's jury: the race to test relativity. Princeton University Press, Princeton

Demetriadou A, Hao Y (2011) Slim Luneburg lens for antenna applications. Opt Express 19(21):19925–19934

Dolin L (1961) On a possibility of comparing three-dimensional electromagnetic systems with inhomogeneous filling. Izv Vyssh Uchebn Zaved Radiofiz 4:964–967

Jiang WX, Cui TJ, Ma HF, Yang XM, Cheng Q (2008a) Layered high-gain lens antennas via discrete optical transformation. Appl Phys Lett 93(22):221906

Jiang WX, Cui TJ, Ma HF, Zhou XY, Cheng Q (2008b) Cylindrical-to-plane-wave conversion via embedded optical transformation. Appl Phys Lett 92(26):261903

Jiang ZH, Gregory MD, Werner DH (2011) Experimental demonstration of a broadband transformation optics lens for highly directive multibeam emission. Phys Rev B 84(16):165111

Jiang ZH, Gregory MD, Werner DH (2012) Broadband high directivity multibeam emission through transformation optics-enabled metamaterial lenses. IEEE Trans Antennas Propag 60(11):5063–5074

Kong F, Wu B-I, Kong JA, Huangfu J, Xi S, Chen H (2007) Planar focusing antenna design by using coordinate transformation technology. Appl Phys Lett 91(25):253509

Kundtz N, Smith DR (2010) Extreme-angle broadband metamaterial lens. Nat Mater 9(2):129–132

DOI 10.1007/978-981-4560-75-7_22-1

Kwon D-H (2012) Quasi-conformal transformation optics lenses for conformal arrays. IEEE Antennas Wirel Propag Lett 11:1125–1128

Kwon D-H, Werner DH (2009) Flat focusing lens designs having minimized reflection based on coordinate transformation techniques. Opt Express 17(10):7807–7817

Leonhardt U (2006) Optical conformal mapping. Science 312:17771780

Li J, Pendry J (2008) Hiding under the carpet: a new strategy for cloaking. Phys Rev Lett 101(20):203901

Liang L, Hum SV (2014) Realizing a flat UWB 2-D reflector designed using transformation optics. IEEE Trans Antennas Propag 62(5):2481–2487

Liberti JC, Rappaport TS (1999) Smart antennas for wireless communications: IS-95 and third generation CDMA applications. Prentice Hall PTR, Upper Saddle River

Liu R, Ji C, Mock JJ, Chin JY, Cui TJ, Smith DR (2009) Broadband ground-plane cloak. Science 323(5912):366–369

Lu W, Lin Z, Chen H, Chan CT (2009) Transformation media based super focusing antenna. J Phys D Appl Phys 42(21):212002

Luo Y, Zhang J, Chen H, Huangfu J, Ran L (2009) High-directivity antenna with small antenna aperture. Appl Phys Lett 95(19):193506

Ma HF, Cui TJ (2010) Three-dimensional broadband and broad-angle transformation-optics lens. Nat Commun 1:124

Ma H, Jiang W, Yang X, Zhou X, Cui T (2009) Compact-sized and broadband carpet cloak and free-space cloak. Opt Express 17(22):19947–19959

Mateo-Segura C, Dyke A, Dyke H, Haq S, Hao Y (2014) Flat luneburg lens via transformation optics for directive antenna applications. IEEE Trans Antennas Propag 62(4, 2):1945–1953

Mei ZL, Cui TJ (2012) Transformation electromagnetics and its applications. Int J RF Microwave Comput Aided Eng 22(4, SI):496–511

Mei Z-L, Bai J, Niu TM, Cui T-J (2010) A planar focusing antenna design with the quasi-conformal mapping. Prog Electromagn Res M 13:261–273

Mei ZL, Bai J, Cui TJ (2011) Experimental verification of a broadband planar focusing antenna based on transformation optics. New J Phys 13:063028

Narimanov E, Kildishev A (2009) Optical black hole: broadband omnidirectional light absorber. Appl Phys Lett 95:041106

Oliveri G, Bekele ET, Werner DH, Turpin JP, Massa A (2014) Generalized QCTO for metamaterial-lens-coated conformal arrays. IEEE Trans Antennas Propag 62(8):4089–4095

Pendry J, Holden A, Robbins D, Stewart W (1999) Magnetism from conductors and enhanced nonlinear phenomena. IEEE Trans Microwave Theory Tech 47(11):2075–2084

Pendry JB, Schurig D, Smith DR (2006) Controlling electromagnetic fields. Science 312:1780–1782

Quevedo-Teruel O, Hao Y (2013) Directive radiation from a diffuse Luneburg lens. Opt Lett 38(4):392–394

Quevedo-Teruel O, Tang W, Hao Y (2012) Isotropic and nondispersive planar fed Luneburg lens from Hamiltonian transformation optics. Opt Lett 37(23):4850–4852

Rahm M, Schurig D, Roberts DA, Cummer SA, Smith DR, Pendry JB (2008) Design of electromagnetic cloaks and concentrators using form-invariant coordinate transformations of maxwell's equations. Photonics Nanostruct Fundam Appl 6(1):87–95

Schurig D, Mock JJ, Justice BJ, Cummer SA, Pendry JB, Starr AF, Smith DR (2006) Metamaterial electromagnetic cloak at microwave frequencies. Science 314:977980

Shelby RA, Smith DR, Schultz S (2001) Experimental verification of a negative index of refraction. Science 292:7779

DOI 10.1007/978-981-4560-75-7_22-1

Smith D, Schultz S, Markoš P, Soukoulis C (2002) Determination of effective permittivity and permeability of metamaterials from reflection and transmission coefficients. Phys Rev B 65(19):195104

Tang W, Argyropoulos C, Kallos E, Song W, Hao Y (2010) Discrete coordinate transformation for designing all-dielectric flat antennas. IEEE Trans Antennas Propag 58(12):3795–3804

Tichit PH, Burokur SN, de Lustrac A (2009) Ultradirective antenna via transformation optics. J Appl Phys 105(10):104912

Tichit PH, Burokur SN, Germain D, de Lustrac A (2011) Design and experimental demonstration of a high-directive emission with transformation optics. Phys Rev B 83(15):155108

Wald R (1984) General relativity. University of Chicago Press, Chicago

Wan X, Jiang WX, Ma HF, Cui TJ (2014) A broadband transformation-optics metasurface lens. Appl Phys Lett 104(15):151601

Ward A, Pendry J (1996) Refraction and geometry in maxwell's equations. J Mod Opt 43(4):773–793

Ward A, Pendry J (1998) Calculating photonic greens functions using a nonorthogonal finite-difference time-domain method. Phys Rev B 58(11):7252

Werner DH, Kwon D-H (2014) Transformation electromagnetics and metamaterials. Springer, London

Wu Q, Jiang ZH, Quevedo-Teruel O, Turpin JP, Tang W, Hao Y, Werner DH (2013) Transformation optics inspired multibeam lens antennas for broadband directive radiation. IEEE Trans Antennas Propag 61(12):5910–5922

Yaghjian AD, Maci S (2009) Alternative derivation of electromagnetic cloaks and concentrators. New J Phys 10(11):115022

Yang R, Tang W, Hao Y (2011a) Wideband beam-steerable flat reflectors via transformation optics. IEEE Antennas Wirel Propag Lett 10:1290–1294

Yang R, Tang W, Hao Y, Youngs I (2011b) A coordinate transformation-based broadband flat lens via microstrip array. IEEE Antennas Wirel Propag Lett 10:99–102

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_23-1

Frequency Selective Surfaces

De Song Wang[a]*, Shi-Wei Qu[b] and Chi Hou Chan[a]
[a]State Key Laboratory of Millimeter Waves, Partner Laboratory in City University of Hong Kong, Hong Kong, SAR, China
[b]School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, China

Abstract

Traditionally, frequency selective surfaces (FSSs) comprising structures with periodicity in two dimensions have important applications as spatial filters in microwave and optics. Due to the manufacturing process, they are usually in the form of printed patches on a dielectric substrate or apertures in a conducting screen. Multiple FSS screens and dielectric layers can be stacked together to meet desirable spectral filter responses. For these structures, a surface discretization of the unknown currents and/or electric fields is more convenient, and efficient integral equation solutions are sought thereafter. Recent developments in manufacturing processes, material properties, and new wave phenomena call for periodic structures with three-dimensional (3D) unit-cell elements. There are situations that a volumetric discretization of the unit cell is more appropriate, and differential equation solutions are preferred. The advent of powerful commercial simulation tools allows effective FSS designs with more flexibility. Compounded with the popularity of 3D printings, some previously unimaginable FSS structures can now be cost-effectively realized. Exploitation of transmission and reflection information obtained from FSS also paves the way for better antenna designs.

Keywords

Planar frequency selective surface; Frequency selective surface with 3D unit cells; Lens antennas; Reflectarrays; 3D printing

Introduction

Frequency selective surfaces (FSSs) comprising structures with periodicity in two dimensions have important applications as filters in microwave and optics. One household application is the microwave oven door screen which allows us to observe the food cooking inside the oven but prevents the leakage of microwave radiation to ensure user safety. In the US patent, *Microwave Oven Door Screen* (US4, 051, 341 approved Sept. 27, 1977), it described that the door screen generally comprises a pair of transparent plates and a microwave shielding material made with punched holes on aluminum plate of thickness 0.1–0.35 mm. The aperture has a diameter of 1.2 mm or less, and the center-to-center spacing between two apertures is from 1.4 to 1.8 mm for effective microwave shielding. Figure 1a shows a typical modern microwave oven door whose design is not much different from what was described in the patent more than 37 years ago. On a larger scale, Fig. 1b shows a typical honeycomb panel used for RFI/EMI shielded enclosures which provides electromagnetic shielding and air ventilation.

*Email: dswang2@cityu.edu.hk

DOI 10.1007/978-981-4560-75-7_23-1

FSSs with periodic apertures are inductive whose spectral response is typically a high-pass/band-pass filter. In contrast, FSSs with periodic array of patches are capacitive, behaving as a low-pass/band-stop filter. Various spectral responses can be achieved through a combination of the inductive and capacitive screens, dielectric loading, unit-cell geometry, and spacing between unit cells.

The scattering mechanism of periodic structures was reported by D. Rittenhouse in 1786 (Stroke 1967). He became interested in the diffraction effects observed through a silk handkerchief at a distant lamp. He constructed one-dimensional (1D) periodic structure in the form of equally spaced parallel hairs. J. von Fraunhofer independently invented the first ruled optical gratings in 1821. From then on, the investigations of FSSs were basically through experimental means and physical interpretations, including the Wood's anomaly observed experimentally with ruled diffraction gratings (Wood 1902).

Spectral response of FSSs was predicted by numerical modeling as early as 1960s. Scattering by a doubly periodic array of square apertures on a conducting plane was modeled with variational method by Kieburtz and Ishimaru (1961) in which the accuracy of the results strongly depends on the chosen trial functions. With the advent of electronic computing and numerical methods, more accurate characterizations of FSSs were reported, and point matching (Ott et al. 1967), waveguide modes (Chen 1970), and plane-wave expansion (Lee 1971) were employed for the analysis of rectangular patches or apertures.

The shape of the unit-cell element also plays a key role to spectral response of the FSS. For thin wire structure, Pelton and Munk (1979) approximated the induced currents on the cross dipole element by two cosinusoidal components. For currents in ring elements, Parker and Hamdy (1981) employed a series of cosine and sine functions which are periodic in the angular position. In the analysis of square loops, Hamdy and Parker (1982) treated the currents as a superposition of those on four coupled rectangular strips expanded with sinusoidal functions. Spectral-domain formulation for the analysis of FSSs with cross dipole and Jerusalem cross elements was introduced by Tsao and Mittra (1984) in which the convolution in the integral equation is conveniently replaced by a product in the spectral domain. A junction basis function was also introduced to correctly account for the current discontinuity at the junction of the cross. The same formulation was extended to analyze circular patch FSSs (Mittra et al. 1984). In the spectral domain formulation, basis functions are chosen such that they are analytically Fourier transformable. For the circular patch case, Chebyshev functions of the first and second kinds were employed. Their Fourier transforms are combinations of different orders of Bessel function of the first kind. In all these approaches, known functions representing the unknown currents are regarded as entire-domain basis functions, and a few of them are sufficient. Therefore, the unknown weighting coefficients can be obtained by a simple inversion of the obtained impedance matrix. In contrast, to provide modeling flexibility, a larger number of subdomain basis functions are employed instead.

Rubin and Bertoni (1983) employed subsectional basis functions to approximate current distributions on a periodically perforated plane. The impedance matrix elements involve double infinite summations that need to be judicially truncated. These double summations were rapidly summed using fast Fourier transforms (FFTs) by Cwik and Mittra (1987) by enforcing uniform, rectangular discretizations. However, the number of summation terms was limited to the size of the FFT. This problem was alleviated by Chan and Mittra (1990) by showing that due to the periodicity of the Floquet modes, terms with the same exponentials could be summed prior to the use of FFTs. The double summations converge rapidly, and it also demonstrates that the relative convergence phenomenon discussed in Lee (1971) does not exist if sufficient number of terms is included in the summation. This efficient method has also been employed to investigate FSS periodically loaded with nonlinear devices (Epp et al. 1992). A review paper summarizing the analysis methods for planar FSSs was published by Mittra et al. (1988), and the efficient approach presented in Chan and Mittra (1990) was validated in infrared (Schimert et al. 1990) and millimeter waves (Schimert et al. 1991).

DOI 10.1007/978-981-4560-75-7_23-1

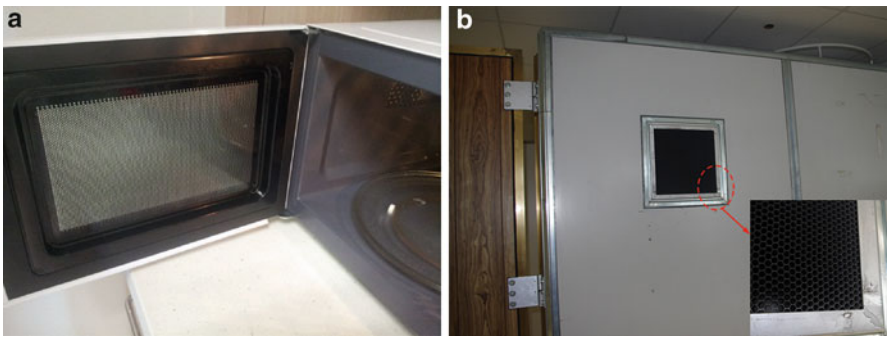


Fig. 1 (**a**) A typical microwave oven door and (**b**) honeycomb panel used for shield room

The use of uniform rectangular discretization, however, limits the flexibility in modeling the unit-cell geometry. A hybrid spectral/spatial approach for a free-standing FSS based on the Poisson summation formula was demonstrated by Jorgenson and Mittra (1990) in which no assumptions about the discretization of the unit cell were made. The spatial domain formulation converges rapidly when the Green's function between the source point and field point is subtracted by another Green's function with the field point slightly lifted off from the plane of the FSS. The subtracted contribution is compensated by the spectral domain formulation which also converges rapidly due to an exponential decay when evaluated off the FSS. This approach allows evaluation of the honeycomb-type panel shown in Fig. 1b (Jorgenson and Mittra 1991). For FSS embedded in layered dielectrics, Kipp and Chan (1994) adopted the same hybrid approach with the spatial domain Green's function evaluated by the complex image method (Chow et al. 1991). The spatial contribution is further accelerated by using Shank's transformation (Singh and Singh 1990). In the complex image method, the spatial Green's function is broken into three parts: the quasi-dynamic image, the complex image, and the surface wave terms. Double summation of the first two terms can be further accelerated by Ewald's method (Jordan et al. 1986) and the last term by the lattice-sum method (Chin et al. 1994). These techniques are then integrated into the hybrid method (Yu and Chan 2000).

When the FSS is not planar, numerical methods presented above would not be applicable and one has to resort to other techniques such as the finite-element method (FEM) and the finite-difference time-domain (FDTD) method in which volumetric discretization of the 3D unit cell is often adopted. A switchable, tunable grating was presented by Bushbeck and Chan (1993). The grating is in the form of a plastic slab with periodic channels filled with fluid. The surface of the slab bulges up periodically when pressure is applied to the fluid. Spectral response of this grating was analyzed with two-dimensional (2D) FE-M. In general, 3D FEM method allows for flexible modeling of doubly periodic structures (Eibert et al. 1999). Another partial-differential equation approach is the FDTD method. Veysoglu et al. (1993) proposed a Floquet field mapping to account for the periodic boundary conditions for the leading and trailing ends of the unit cell at oblique incidence. The performance of the method deteriorates with the increase of incidence angle. This problem can be alleviated with a combination of sine and cosine excitations (Harms et al. 1994). In addition to Floquet field mapping for the periodic FDTD equations, Roden et al. (1998) employed a split-field update method which led to a less stringent stability criterion and more accurate geometrical modeling with nonorthogonal FDTD grids. On the other hand, the advantage of obtaining a range of frequency response in a single time-domain simulation disappears due to the use of Floquet field mapping. 3D FSSs with double periodicity can also be solved using volumetric/surface integral equation approach with the periodic Green's function obtained by Ewald's transformation (Shi and Chan 2010).

DOI 10.1007/978-981-4560-75-7_23-1

Recent Development

In addition to a large volume of journal papers on FSSs and periodic structures, there are also many books published over the years, ranging from theory, analysis, optimization, and design (Wu 1995; Vardaxoglou 1977; Rahmat-Samii and Michelssen 1999; Munk 2000; Capolino 2009). Except for Capolino (2009) which focuses on new electromagnetic properties of metamaterials in the form of periodic 2D/3D structures, these books have been published for over a decade. Therefore, it is also beneficial to provide a brief review on some selected recent work on FSSs, and obviously many other works are left out here given the limited space.

Shen and his research group proposed 3D FSS structures that consist of a 2D periodic array of 3D unit cells formed by vertically placed microstrip structures (Rashid and Shen 2010, 2011; Rashid et al. 2012, 2014). The 3D FSSs offered arbitrarily pseudo-elliptical performance by exciting a suitable number of propagating modes and controlling their coupling with air. Compared to conventional 2D FSSs, 3D FSS structures provide more design flexibility although they require more physical assembly. The use of resonant metallic elements in FSSs could be problematic at very high power. J. H. Barton et al. (2012, 2014) investigated all-dielectric FSSs for high-power microwaves. Because of lack of metals, arcing at field concentration points and heating in the conductors were successfully avoided. This technology could find a variety of applications in high-power microwave systems, including radomes, beam shaping, and more.

Ohira et al. (2004, 2005) and Bossard et al. (2005, 2006) employed genetic algorithm (GA) to design FSSs with desired performances. This methodology was very direct and flexible but could be time consuming as there is no one-to-one correspondence between the filter response and the unit-cell geometry. Novel waveguide filters are designed using GA to determine which pixels of a bitmap are printed with conductor (Ohira et al. 2005) on each FSS screen and cascading three screens to form a waveguide band-pass filter with a steep roll-off. Switches are incorporated into the FSS in Bossard et al. (2005) which makes the surface reconfigurable for either single-, dual-, or tri-band operation. The surface can also be polarization dependent or independent, subject to different on and off patterns of the switches. Other active FSS structures were reported by Sanz-Izquierdo et al. (2010, 2011) and Sanz-Izquierdo and Parker (2014) on which active components were loaded to tune the filtering characteristics. The designs could be used for various applications including modification of the electromagnetic architecture of buildings and the control of electromagnetic wave propagation to improve the efficiency of radio spectrum use.

Practical applications of traditional FSSs are restricted in limited space, so it is difficult to arrange larger number of elements with large physical sizes to make a finite FSS with better performance. In addition, large element sizes and large periodicities lead to difficulties, especially in conforming to curved surfaces. In Liu et al. (2009), Yan et al. (2014), and Yu et al. (2014), miniaturized FSSs were investigated by respectively loading lumped reactive components and meandering wires. The proposed FSS structures exhibited superior miniaturization and provided favorable resonant stability at various incidence angles. Substrate integrated waveguide (SIW) technology was applied to design high-performance FSS structures by Luo et al. (2005, 2006, 2007). The high-Q property of SIW cavities greatly improved the frequency selectivity of the FSSs, which was significant for the applications in antenna radome, object stealth, etc. Resulting from the coupling between cavity mode and aperture mode, SIW FSSs have additional transmission zeros nearby the passbands, which improved the frequency selectivity.

Metallic wire grid can be viewed as a nonresonant inductive component, and metallic patches separated by thin air gaps can be viewed as a nonresonant capacitive component. Based on this theory, Sarabandi and Behdad (2007) proposed a variety of low-profile FSSs with sub-wavelength unit cells by cascading these nonresonant components. Further studies on low-profile FSSs designed by combining resonant and

DOI 10.1007/978-981-4560-75-7_23-1

nonresonant elements were presented by Bayatpur and Sarabandi (2008), Behdad et al. (2009), and Al-Joumayly and Behdad (2009). Actually, an FSS with any arbitrary multipole or non-commensurate multiband filtering response can be obtained using the approach described above. In addition, the proposed FSS structures have extremely small overall profiles and stable resonant frequency with respect to the incidence angle.

Recently, research on terahertz (THz) devices has been rapidly expanded due to a wide range of applications, including chemical and biological sensing, remote sensing, imaging, and submillimeter-wave applications. Some fabrication approaches with high manufacturing accuracy, such as wafer micro-fabrication technology and electron-beam evaporation technology, were used to fabricate THz FSS structures. Dickie et al. (2005, 2009, 2014) proposed a variety of FSSs for THz applications. Based on a precision micromachining technique, the fabricated metallic FSSs have low insertion loss due to lack of substrate loss. Active tuning of FSS at 0.86 THz was reported by Chang et al. (2013). Fish-net FSS was printed on polymer-dispersed liquid crystal, and the resonant frequency can be tuned by varying the voltage applied to the FSS.

FSSs can be applied to the design of antennas to reduce planar antenna radar cross section (RCS), enhance gain and directivity, and improve bandwidth. Many researchers have designed antennas based on FSS structures to obtain desired performances (Jazi and Denidni 2010, 2013; Foroozesh and Shafai 2010; Genovesi et al. 2012). Jazi and Denidni (2010, 2013) applied different active FSSs to design sweeping-beam radiation-pattern antenna and reconfigurable-agile radiation-pattern antenna. These antennas provided superior radiation performances in terms of bandwidth, radiation pattern, and gain. Additionally, there was more freedom to control the antenna radiation characteristics. Foroozesh and Shafai (2010) used a highly reflective patch-type FSS as the superstrate of a cavity resonant antenna to obtain high peak gains and offer wide input impedance and gain bandwidths. Genovesi et al. (2012) investigated how to reduce the RCS of a common patch array antenna at the out-of-band frequencies by using FSSs. Finally, replacing the solid ground plane of the array antenna with a hybrid FSS was proven to be an effective means, which reduced the RCS but also preserved the radiation performance of the antenna.

Due to the advent in both computer hardware and computational methods, the emphases of FSS research are shifted toward more complex unit-cell geometry to satisfy new design specifications and higher operating frequencies. Based on the review of more recent works, FSS design is first modeled with commercially available software such as HFSS (Ansoft HFSS, 3D EM Simulator, version 14.0.0). 3D unit-cell elements are used for providing more design flexibility. Conventional multilayered printed-circuit technologies may not be sufficient to meet new demands, and alternative fabrication processes are needed. Some of these issues will be addressed in this chapter.

Multilayered FSS Filter

FSSs have been widely used as spatial filters; however, most FSS designers only focus their efforts in getting the right passband and stopband, multiband response, or stability against different angles of incidence. In reality, design specifications of an FSS filter are not different from those of a filter in microwave circuitry. For example, an FSS band-pass filter should have a low insertion loss in the passband and a steep roll-off at the band skirts which are often overlooked (Chiu and Chang 2009; Sarabandi and Behdad 2007). The sharp roll-off can be achieved by high-order filters by stacking several FSS screens, but this may increase the insertion loss. One approach to improve stopband rejection is to introduce transmission zeros in the filter design (Ma et al. 2006; Chu and Wang 2008). Ma et al. (2006) presented how the electric coupling, magnetic coupling, and their combination between two resonators can affect the presence and locations of the transmission zero (TZ) using two separate coupling paths.

DOI 10.1007/978-981-4560-75-7_23-1

Similar approach will be adopted here to illustrate how FSS designers can improve the spectral responses of their designs.

Recently, because of the availability of unlicensed 60 GHz spectrum, researches on 60 GHz devices are being rapidly expanded, but little research work has been reported on the 60 GHz FSSs. In general, there are two approaches to obtain high filter selectivity, including increasing resonance orders and adding TZs. Higher-order structures would be complicated and large, so in this work TZs are added to improve frequency selectivity. Certainly, higher-order characteristics are also investigated in section "Stability and Higher-Order Characteristics." In Luo et al. (2007) and Rashid et al. (2012), cross coupling has been introduced to obtain multiple TZs around the passband. In Luo et al. (2008), two FSSs with one TZ each have been combined to realize an FSS structure with multiple TZs. However, these high-selectivity FSSs are difficult to design at 60 GHz or THz because of their 3D structures.

In this section, a dual-polarized FSS structure at 60 GHz based on aperture-coupled resonators is proposed. It provides multiple TZs resulting from the mixed electric and magnetic couplings, therefore achieving high-frequency selectivity. In addition, the FSS is easy to design and fabricate due to the simple and planar structure. In contrast, an SIW FSS structure in section "SIW FSS with Multiple TZs" will be presented, similarly providing two or three TZs. However, as mentioned above, this SIW structure is difficult to fabricate due to the presence of many small metallic vias.

Tightly Coupled FSS Structure and Analysis

The objective of this work is to design a 60 GHz FSS with high frequency selectivity. In Pous and Pozar (1991), a coupling FSS structure based on the aperture-coupled microstrip patch has been designed. However, there is no TZ around the passband; the same is true in the patch-aperture-patch or aperture-patch-aperture FSS (Chan et al. 1992). In this work, based on the coupling resonator theory, an FSS with mixed electric and magnetic couplings is designed, producing two TZs at finite frequencies.

Tightly Coupled FSS Structure

The 60 GHz FSS is designed using two cross resonators coupled by rectangular apertures. It consists of three metal layers separated by two substrates. In this work, Rogers Duroid 5880 is applied as the substrate to support the structure because of its stable dielectric constant of 2.2 and low loss tangent of 0.004 at

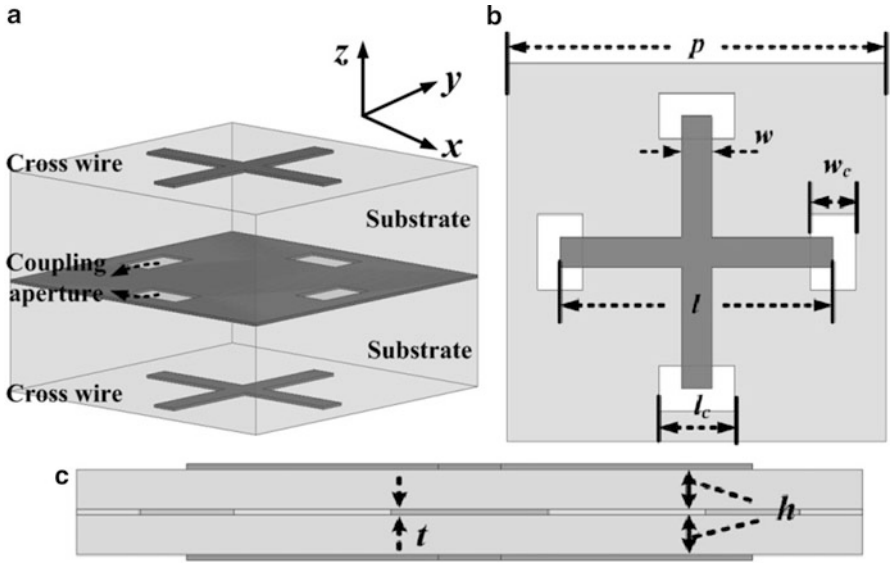


Fig. 2 Second-order FSS unit cell. (**a**) Perspective view, (**b**) top view, and (**c**) side view

DOI 10.1007/978-981-4560-75-7_23-1

Table 1 Dimensions of the proposed tightly coupled FSS structure

Physical parameters	p	l	w	l_c	w_c	t	h
Values (mm)	2.5	1.8	0.2	0.5	0.3	0.017	0.127

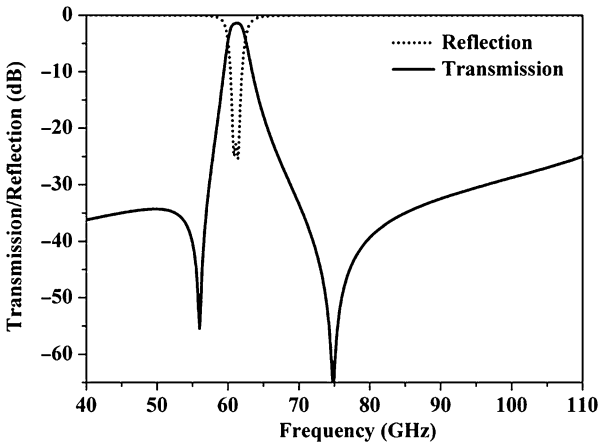


Fig. 3 Frequency responses of the proposed FSS

60 GHz (Li and Luk 2014). Figure 2a–c shows the configuration of the proposed FSS unit cell. In the unit cell, the cross resonators on the top and bottom layers are identical and viewed as two half-wavelength resonators. Four identical rectangular apertures etched on the middle ground are directly located between each end of the top and bottom crosses. The final FSS is realized by periodical arrangement of the unit cells.

The proposed FSS is modeled conveniently using the integral equation as described in Mittra et al. (1988). As the method has been experimentally verified at various frequencies (Schimert et al. 1990, 1991), only the numerical results are focused for this structure. Table 1 lists the dimensions of the proposed FSS. The simulated frequency responses of the proposed FSS for an x- or y-polarized normal incidence are plotted in Fig. 3. The FSS produces two transmission poles (TPs) in the passband. It has insertion and return losses of 1.4 and 23 dB at the operating frequency of 61 GHz, respectively. This second-order structure has symmetric response with a fractional 3-dB bandwidth of 4.1 % (2.5 GHz). The two produced TZs are located, respectively, at 56 and 75 GHz both with an attenuation level of around 60 dB, which are helpful to achieve superior frequency selectivity and suppress the lower and upper sidebands.

Analysis of the FSS

Figure 4 shows the coupling diagram for the proposed FSS in Fig. 2, in which the resonators on the top and bottom layers are represented by the squares. Through the coupling apertures on the middle layer, mixed electric and magnetic couplings exist between the two resonators. From the theory of coupling filter, the FSS structure obtains a narrow bandpass response with two TPs in the passband. One of TPs is due to the individual resonators, and the other one appears as a combined effect of the resonators plus the effect of the coupling apertures. Due to the mixed couplings, two TZs can be produced at real finite frequencies and their frequencies can be adjusted by the location and dimensions of the coupling apertures. As described in Ma et al. (2006), when the electric coupling is dominant, the two produced TZs would be located at lower and upper sides of the passband, respectively. While the magnetic coupling is dominant, two TZs would be produced and both located at upper side of the passband.

DOI 10.1007/978-981-4560-75-7_23-1

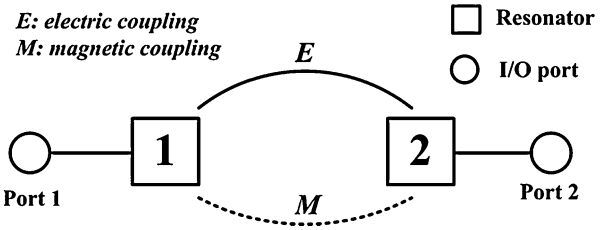


Fig. 4 Coupling diagram for the proposed FSS

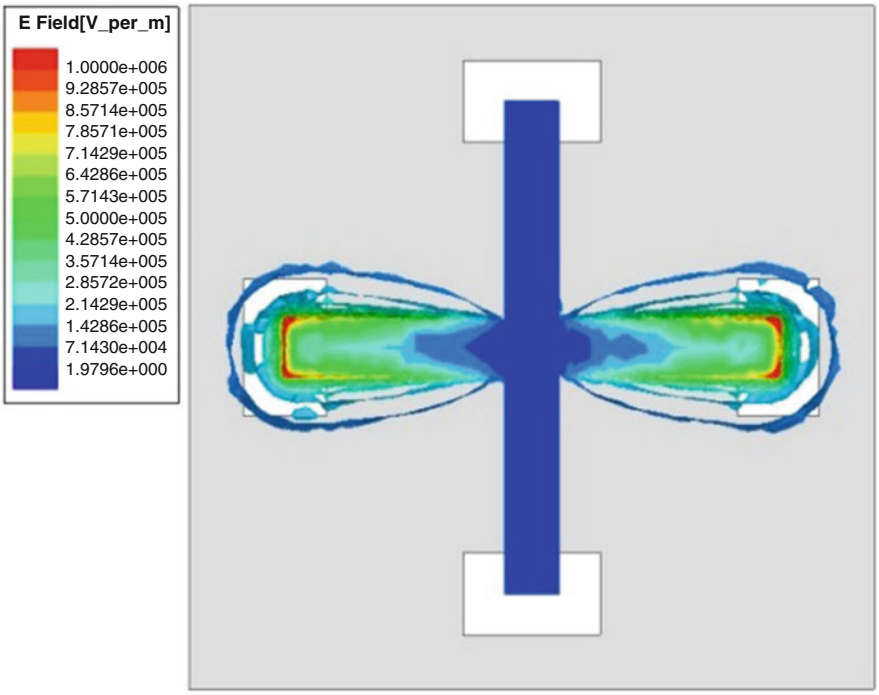


Fig. 5 E-field distributions of the FSS unit cell at the operating frequency for an x-polarized normal incidence

Figure 5 shows the electric field (E-field) distributions of the FSS unit cell at 61 GHz in the passband for an x-polarized normal incidence. As observed, the electric field is mainly concentrated around the ends of the cross resonators. This fact demonstrates that the coupling between the resonators is mainly electric coupling. Therefore, there are two TZs, respectively, located at the lower and upper sidebands, as shown in Fig. 4. Adjusting appropriately the locations of the coupling apertures can affect the strength of the electric and magnetic couplings and then change the frequencies of the TZs. Figure 6 shows the transmission responses with various locations of the coupling apertures. It is noted that all the coupling apertures are relocated symmetrically, as shown in the inset of Fig. 6. As found, one of the TPs changes slightly, while the other shifts to lower frequencies with relocation of the coupling apertures. Additionally, when the coupling apertures are located around the ends of the crosses ($l_m = 0.1$ mm), with dominant electric coupling, the produced TZs are separately located at lower and upper sidebands. When the coupling apertures are close to the centers of the crosses ($l_m = 0.2$ mm or 0.3 mm), with dominant magnetic coupling, all the produced TZs are located at upper sidebands. At last, as the apertures are moved to the centers of the crosses ($l_m = 0.4$ mm), there are the strongest magnetic coupling and the weakest electric coupling that no TZs appear at finite frequencies. Actually, the FSS structure investigated in Pous and Pozar (1991) is a special case of the design presented here.

DOI 10.1007/978-981-4560-75-7_23-1

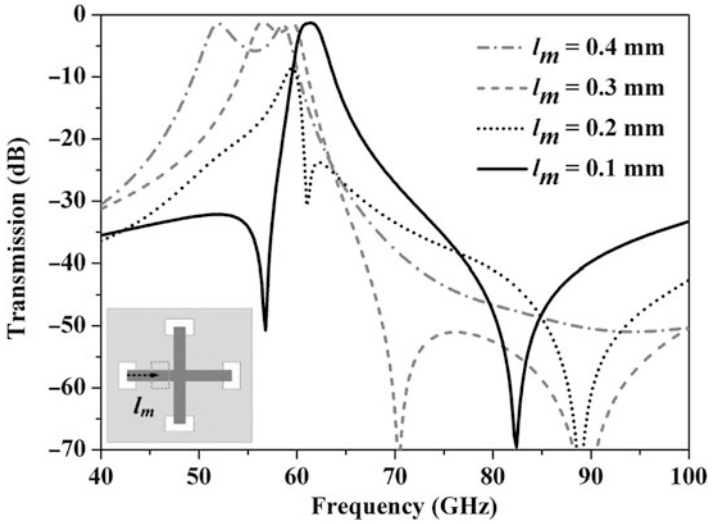


Fig. 6 Effects of locations of the coupling apertures on the transmission coefficient of the FSS

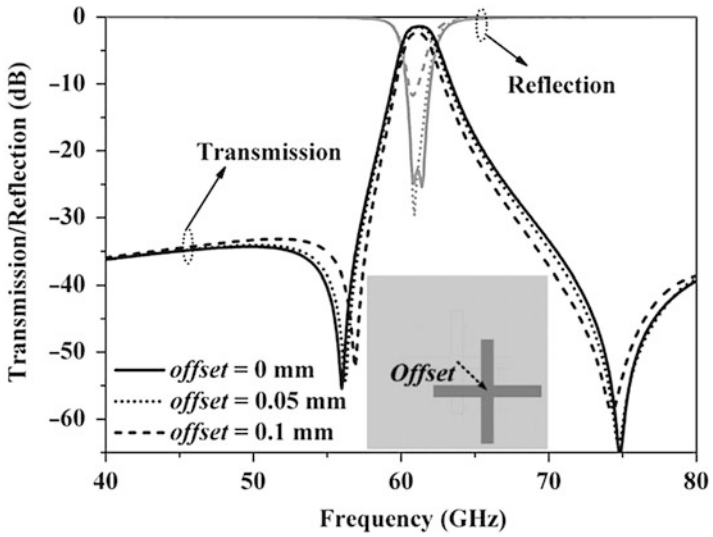


Fig. 7 Effects of the offset between resonators at different layers on frequency response of the FSS structure

Stability and Higher-Order Characteristics

Stability Analysis

As shown in Fig. 2, since the proposed FSS needs to be realized by stacking three metal layers supported by two substrates, there could be a misalignment between the top and bottom layers. Figure 7 shows effects of the offset between resonators at different layers on the frequency response of the FSS, in which the parameter offset is clearly defined and shown in the inset. It can be found that the structure is still stable until the offset goes up to 0.1 mm.

The simulated transmission coefficients of the FSS illuminated by plane waves with incident angles of 0° and 30° for both TE and TM polarizations are plotted in Fig. 8. In the figure, "theta" represents the angle between the propagation vector of the incident wave and the normal direction to the surface. As observed, the operating frequency and bandwidth are very stable in case of oblique incidence for the two

DOI 10.1007/978-981-4560-75-7_23-1

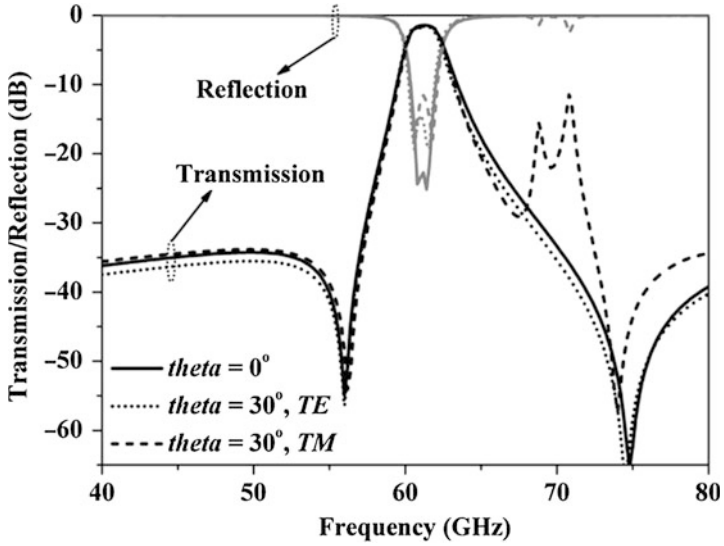


Fig. 8 Simulated transmission coefficients of the FSS with different incident angles and polarization states

polarization states. For an incident angle of 30°, some unwanted resonances emerge at around 70 GHz. Fortunately, the amplitudes are below −12 dB. The frequency response of the structure is relatively stable to variations of incident angle and polarization state.

Higher-Order Characteristics

In order to further improve the frequency selectivity, a higher-order FSS is realized by stacking two second-order FSS structures shown in Fig. 2 with an air spacing of 1.3 mm, as shown in Fig. 9a. The stacked FSSs, respectively, have dominant electric and magnetic couplings, causing that the higher-order FSS would produce multiple TZs nearby the passband simultaneously. Figure 9b shows the simulated frequency responses of the higher-order FSS. From the figure, one TZ at the lower sideband and two TZs at the upper sideband are found, which improve the frequency selectivity to a large extent. Furthermore, comparison of the simulated responses of the original second-order FSS and the higher-order FSS illustrates that stacking FSSs can greatly improve the filter response.

SIW FSS with Multiple TZs

Substrate integrated waveguide (SIW) technology has already been applied to design high-performance FSS structures (Luo et al. 2005, 2006, 2007). Resulting from multiple coupling paths, a transmission zero (TZ) nearby the passband can be obtained, therefore improving the upper or lower stopband of the frequency responses. In order to further enhance the frequency selectivity, two FSS structures individually with one TZ have been combined to obtain multiple TZs (Luo et al. 2008). However, this combined FSS suffers from a complicated structure that it is difficult to design and practically use.

In this section, substrate integrated waveguide (SIW) technology is similarly applied to design FSS structures. Due to the existing multiple coupling paths through multiple resonators, including slot resonator, cavity resonator, and Fabry-Perot resonator, multiple TZs can be obtained to suppress the lower and upper stopbands.

FSS Structure

Figure 10 shows the unit cell of the proposed SIW FSS with a square lattice. As seen in the unit cell, two identical square loop patches are located horizontally on the top and bottom layers, respectively. Metallic

DOI 10.1007/978-981-4560-75-7_23-1

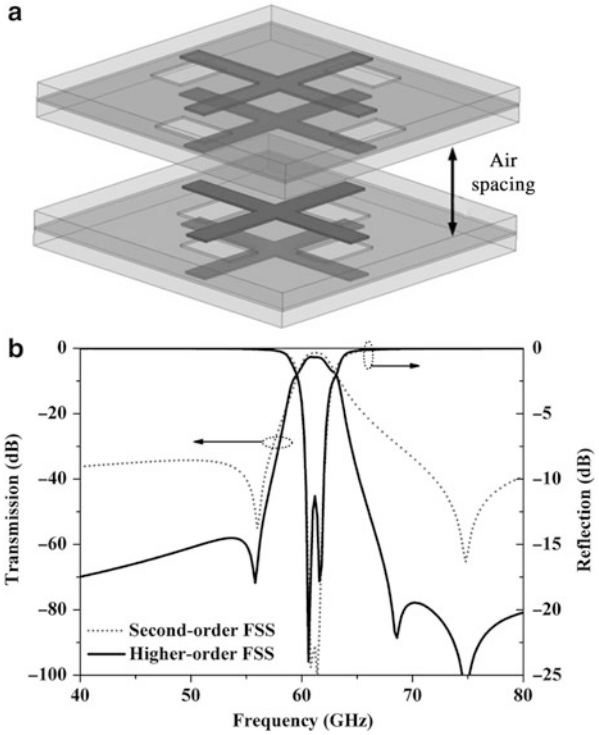


Fig. 9 Higher-order FSS. (**a**) Unit cell and (**b**) frequency responses

Table 2 Geometrical parameters of the FSS independently with second-order and third-order responses

Parameters		l_p	l_s	d	d_p	h	g	t
Second-order response	Values (mm)	2.6	1.35	0.2	0.3	1.575	0.1	0.018
Third-order response		4.3	1.5	0.2	0.3	1.575	0.1	0.018

vias forming vertical sidewalls are located next to outer edges of the square loop patches, in which a certain center-to-center spacing between adjacent vias (d_p as given in Table 2) needs to be reached, as described in Zhang et al. (2005). Such unit cells are arranged periodically to construct an SIW FSS. In this structure, a Fabry-Perot resonator exists in the substrate zone between adjacent sidewalls. The rectangular slots surrounded by the loop patches on the top and bottom layers can construct a slot resonator, and, in contrast, the metallic sidewalls provide a cavity resonator. Actually, it is not difficult to conclude that the Fabry-Perot resonator is mainly determined by the thickness of the used substrate and the gap between adjacent sidewalls (Lima and Parker 1996), and the slot resonator and cavity resonator are mainly determined by the edge lengths of the rectangular slots and inner edge lengths of the SIW cavity, respectively. Therefore, by adjusting the related geometrical parameters, one can achieve two different frequency responses, which are the second-order and third-order frequency characteristics.

Second-Order Response

When the size of the cavity constructed by the metallic via array is small so that the resonant frequency of this cavity resonator is far from the interested frequency range, the cavity can actually be viewed as a Faraday cage (Xu et al. 2008). The Faraday cage at this time can reduce the interferences between unit cells and provide a Fabry-Perot resonator located between adjacent Faraday cages in a doubly periodic

DOI 10.1007/978-981-4560-75-7_23-1

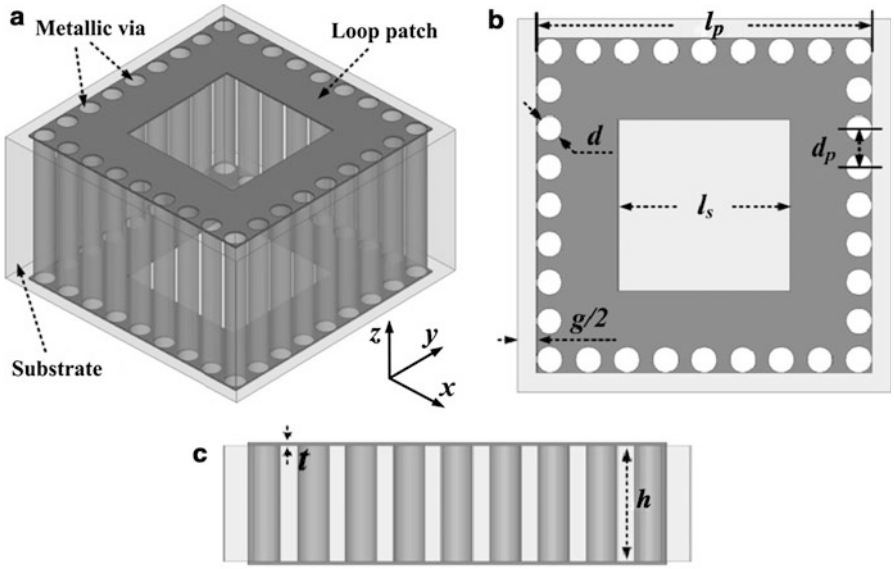


Fig. 10 SIW FSS unit cell. (**a**) Perspective view, (**b**) top view, and **c** side view

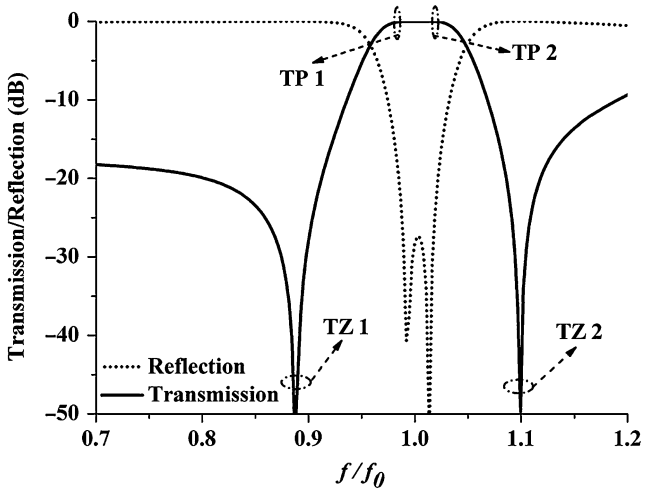


Fig. 11 Frequency responses of the second-order FSS

array, as described above. Therefore, in the frequency range of interest, there should be two transmission poles generated by the slot resonator and the Fabry-Perot resonator, respectively. The simulated frequency response curves are plotted in Fig. 11. In the following simulation models, the via array is replaced by a square solid wall with inner edge length $l_c = l_p - 2d$ and outer edge length l_p (where l_p and d are shown in Fig. 10), and the obtained results are all for an x- or y-polarized normal incidence and normalized. From Fig. 11, one can see that there are two TPs in the passband, which are labeled as TP 1 and TP 2. Two TZs at the lower and upper sidebands are also obtained, which are labeled as TZ 1 and TZ 2, respectively. TZ 1 is caused by the couplings between slot resonant mode and evanescent modes in free space, while TZ 2 is by the couplings between slot resonant mode and Fabry-Perot resonant mode. These produced TZs improve the frequency selectivity of the FSS structure to a great extent, and their generation mechanism will be analyzed in the next subsection in detail. However, the upper stopband is not successfully suppressed, especially at frequencies over the 1.2 f_0, in which f_0 is the central frequency. The reason causing the

DOI 10.1007/978-981-4560-75-7_23-1

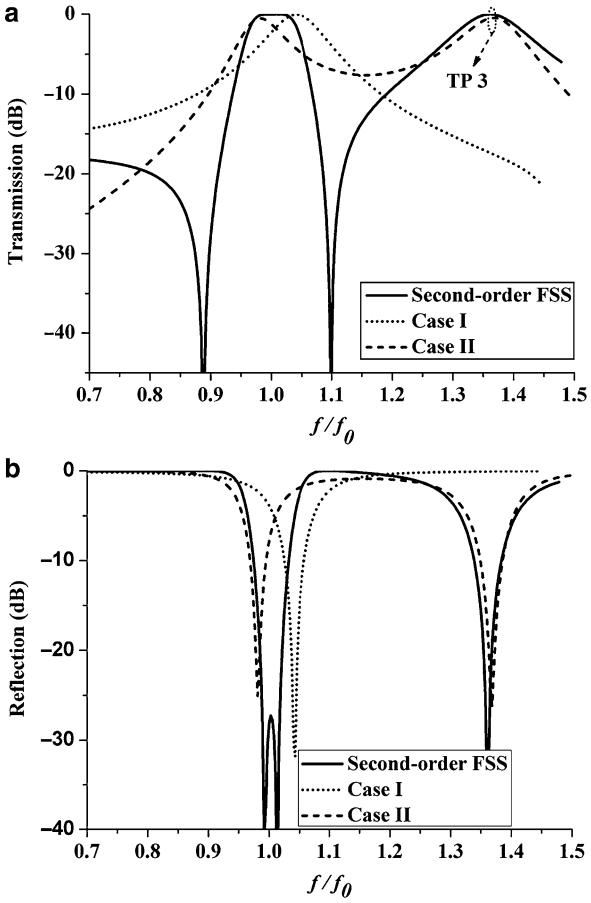


Fig. 12 Frequency responses of the second-order FSS and the structure in two cases: (**a**) transmission and (**b**) reflection

imperfect upper stopband can be found from the frequency responses over a wide frequency range. As shown in Fig. 12a, an additional TP labeled as TP 3 emerges at around 1.4 f_0, which is caused by the Faraday cage as a cavity resonator. This undesired TP worsens the frequency performance of the upper stopband, and one solution is to reduce the size of the Faraday cage and keep its resonant frequency far away from the frequency range of interest.

In order to clearly show the working mechanism of the produced TPs, including TP 1, TP 2, and TP 3, the different components of the FSS structure shown in Fig. 10 are investigated individually. Frequency performance analysis of the Fabry-Perot resonator alone can be realized by filling the rectangular slots of the original FSS structure with metal, called Case I here. Filling the gaps between adjacent sidewalls can help to independently analyze the slot resonator and the cavity resonator constructed by the via array, designated as Case II. From Fig. 12a, b, TP 1 and TP 3 of the original second-order FSS coincide with the ones of Case II, which are caused by the slot resonator and the cavity resonator. TP 2 coincides with that of Case I, which is caused by the Fabry-Perot resonator. The frequencies of all the three TPs of the original FSS structure accurately agree with the ones of the TPs of Cases I and II, validating the different components of the FSS structure which can be adjusted independently to obtain the desired frequency response.

Figure 13 shows the E-field distributions on the FSS structure at frequencies of the three transmission poles. As observed, the E-field distributions once again verify the conclusion that the three transmission poles TP 1, TP 2, and TP 3 are caused by the three resonators, i.e., the slot resonator, Fabry-Perot resonator, and cavity resonator, respectively. Based on this conclusion, one can change the frequencies of

DOI 10.1007/978-981-4560-75-7_23-1

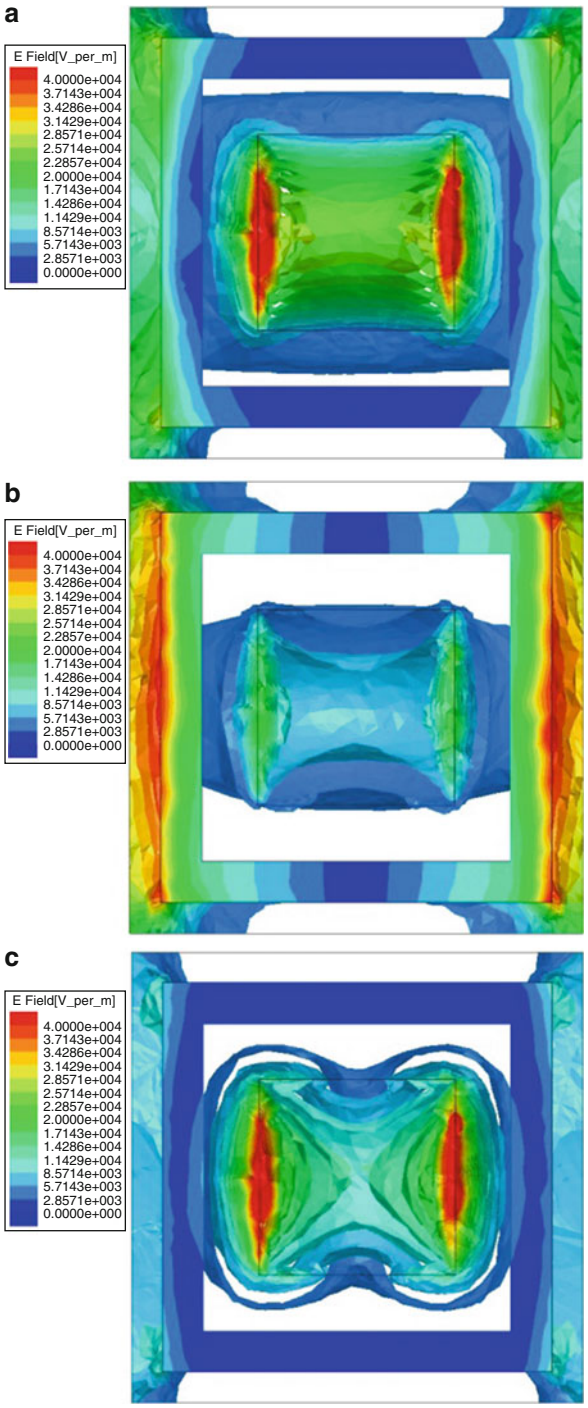


Fig. 13 E-field distributions of the second-order FSS structure at frequencies of three transmission poles: (**a**) TP 1, (**b**) TP 2, and (**c**) TP 3

the transmission poles in the passband by separately adjusting the geometrical parameters l_s and g. Meanwhile, in order to shift the undesired resonance of the TP 3 to a higher frequency, one can reduce the inner edge length of solid walls l_c labeled in the inset of Fig. 14c as much as possible. Figure 14 clearly shows the frequency responses of the FSS structure with various values of the related parameters l_s, g, and l_c, which can guide us to design an FSS with the desired responses. It should be reiterated that the metallic via array is replaced by solid metallic walls in the simulation models.

DOI 10.1007/978-981-4560-75-7_23-1

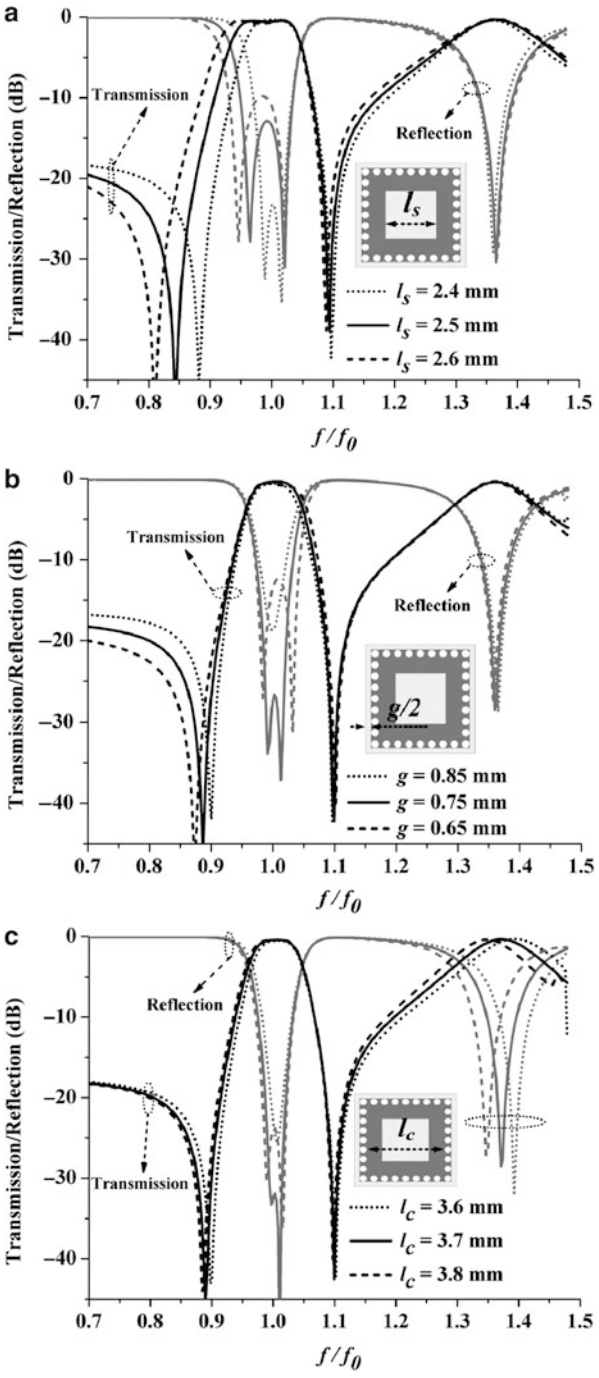


Fig. 14 Frequency responses of the FSS with different parameters: (**a**) l_s, (**b**) g, and (**c**) l_c

In order to validate the proposed design, the second-order FSS is designed in V band and its simulated frequency responses for an x- or y-polarized normal incidence are shown in Fig. 15. In the simulation model, metallic vias with a certain center-to-center spacing (d_p as given in Table 2) are used to construct the sidewalls, and Rogers Duroid 5880 is applied as the substrate to support the structure because of its stable dielectric constant of 2.2 and low loss tangent of 0.004 at 60 GHz (Li and Luk 2014). Table 2 lists values of the related parameters. From the figure, the central frequency of the FSS passband is 56.5 GHz. There are two TPs in the passband with a 3-dB bandwidth of 5.5 % (3.1 GHz) and two TZs located at

DOI 10.1007/978-981-4560-75-7_23-1

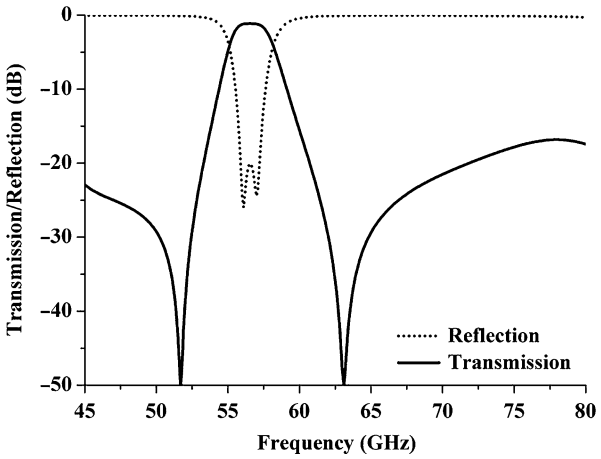


Fig. 15 Frequency responses of the second-order FSS in V band for normal incidence

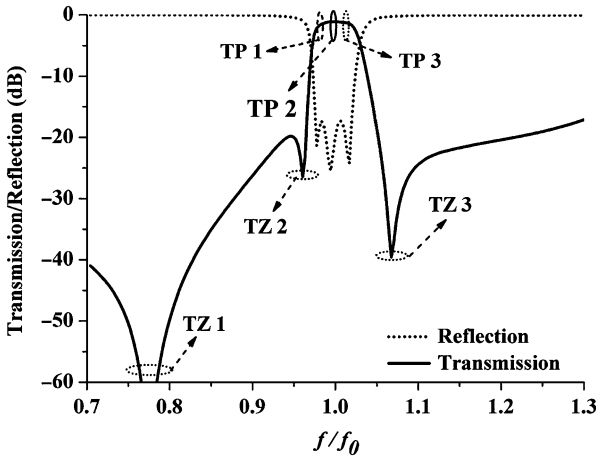


Fig. 16 Frequency responses of the third-order FSS

51.7 and 63 GHz, respectively. Over the frequency ranges from 45 to 53 GHz and from 61 to 80 GHz, the stopband suppression is better than 20 dB. In the passband, the insertion and return losses are about 1.4 and 20 dB, respectively. Actually, the negative impacts of the cavity on the upper stopband have already been lowered to a certain extent.

Third-Order Response

As described in Luo et al. (2007), the cavity resonator constructed by the via array would couple with the slot resonator to produce an additional TZ. In this part, the cavity is enlarged to realize a third-order response. Figure 16 depicts the normalized frequency curves of the third-order FSS. Three TPs in the passband, labeled as TP 1, TP 2, and TP 3, and three TZs around the passband, labeled as TZ 1, TZ 2, and TZ 3, can be found. TZ 1 and TZ 2 are located at the lower sideband, and TZ 3 is located at the upper sideband, thereby improving the frequency selectivity of the FSS structure. Similar to the second-order response, the three transmission poles, TP 1, TP 2, and TP 3, are generated by the Fabry-Perot resonator, slot resonator, and cavity resonator, respectively, which can be verified by the E-field distributions shown in Fig. 17.

DOI 10.1007/978-981-4560-75-7_23-1

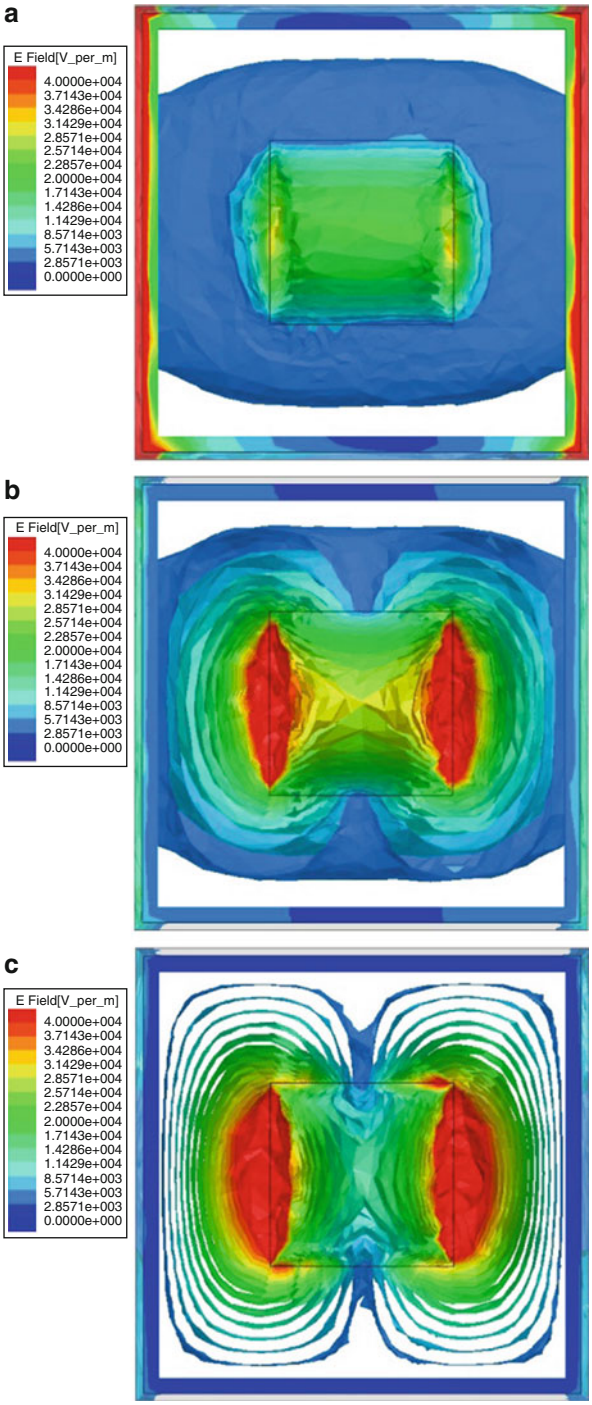


Fig. 17 E-field distributions of the third-order FSS at frequencies of three transmission poles: (**a**) TP 1, (**b**) TP 2, and (**c**) TP 3

The produced TPs have already been analyzed above; in this subsection, the generation mechanism of the three produced TZs is reiterated. The diagrams of couplings causing the three TZs are shown in Fig. 18. TZ 1 at the lower frequency is caused by the couplings between slot resonant mode and evanescent modes in free space (Rashid et al. 2012), while the TZ 2 and TZ 3 are caused by the couplings between slot resonant mode and Fabry-Perot resonant mode and the couplings between slot resonant mode and cavity resonance mode, respectively. Similarly, for a further demonstration, the different

DOI 10.1007/978-981-4560-75-7_23-1

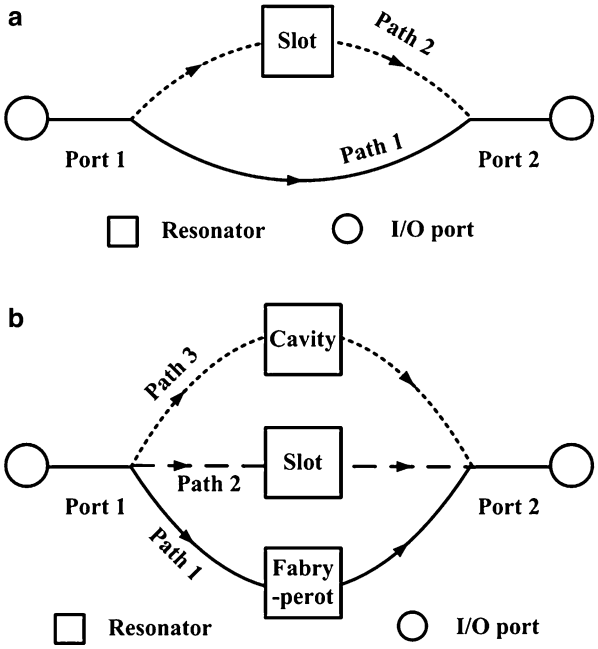


Fig. 18 Diagrams of couplings causing three TZs: (**a**) TZ 1, (**b**) TZ 2, and TZ 3

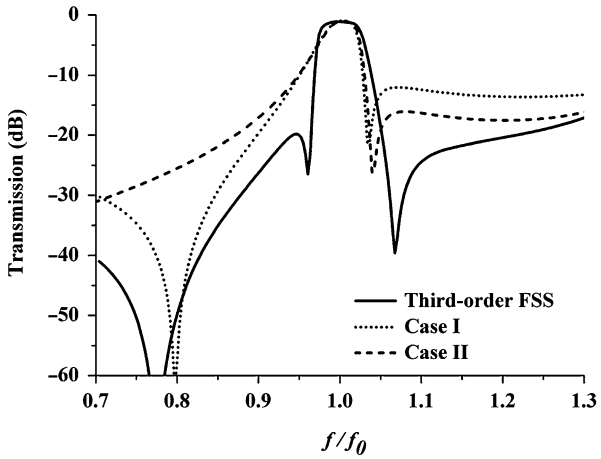


Fig. 19 Normalized transmission responses of the third-order FSS and the structure in two cases

components of the FSS structure shown in Fig. 10 are independently investigated. Replacing the substrate between adjacent sidewalls by air can remove the Fabry-Perot resonator, and this case is called Case I here. Filling the air gaps between adjacent sidewalls using metals can independently study the slot resonator and the cavity resonator, designated Case II. The simulated responses of the original third-order FSS and the structure in the two cases are plotted in Fig. 19. In Case I, there are two TZs at finite frequencies, i.e., TZ 1 and TZ 3. In Case II, there is only one TZ at the upper sideband, namely, TZ 3. Based on the comparison results shown in Fig. 19, it is quite easy to draw the conclusion described at the beginning of this paragraph. In comparison with the second-order response, these three TZs of the third-order response further improve the frequency selectivity, especially over the lower stopband.

In order to validate the third-order frequency characteristics, the third-order FSS is designed in V band and its simulated frequency responses for normal incidence are shown in Fig. 20. Similarly, in the

DOI 10.1007/978-981-4560-75-7_23-1

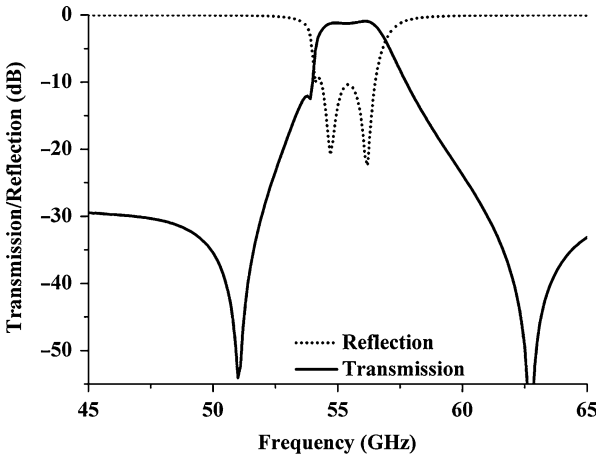


Fig. 20 Frequency responses of the third-order FSS in V band for normal incidence

simulation model, metallic vias and Rogers Duroid 5880 are used to construct the cavities and support the whole structure, respectively. Dimensions of the FSS are shown in Table 2. As shown in Fig. 20, the central frequency of the FSS is 55.5 GHz. There are three TPs in the passband with a 3-dB bandwidth of 4.9 % (2.7 GHz) and three TZs around the passband. In the passband, the insertion and return losses are about 0.9 and 10 dB, respectively.

Summary for SIW FSS

In section "SIW FSS with Multiple TZs," a kind of SIW FSSs has been presented and its second-order and third-order responses are investigated, respectively. Because of multiple coupling paths, the FSS provides multiple TZs, thus greatly improving the high frequency selectivity. The produced TPs have also been analyzed by showing the E-field distributions and comparing the frequency characteristics of the FSS in different cases. Two practical designs in V band have been provided, which verify the described second-order and third-order responses. The prototypes of the V-band FSSs could be fabricated by the standard printed-circuit-board (PCB) and plated through-hole technologies. However, due to so many small metallic vias, fabrication would become a big problem when the operating frequency continues going up, particularly into the THz regime.

FSS with 3D Unit-Cell Elements

3D FSS Introduction

In the previous section, FSSs are composed of periodic screens supported by planar dielectric layers. When the unit-cell element is 3D, they could be referred to as frequency selective volumes (FSVs) (Yang et al. 1997) or more fashionably as metamaterials (Engheta and Ziolkowski 2006). When the 3D unit cell does not have many fine details, typical FEM solver like HFSS would be sufficient. In contrast, when the number of unknowns is kept increasing, one needs to resort to other efficient algorithms, for example, the integral-equation-based multilevel Green's function interpolation method (MLGFIM) developed for the analysis of arbitrary 3D composite dielectric and conducting objects with double periodicity in which the periodic Green's function is calculated by applying Ewald's transformation (Shi and Chan 2010). The use of multilevel scheme substantially reduces the periodic Green's function points, and therefore, the matrix-fill time can also be reduced. The volume-surface integral equation for dielectric and conducting objects is formulated with conformal basis functions defined on curvilinear hexahedral and quadrilateral

DOI 10.1007/978-981-4560-75-7_23-1

elements which are used to discretize the 3D unit cell. Periodic boundary conditions are introduced to the unknowns on the opposite boundaries of the unit cell. The resulting matrix equation is solved iteratively using the MLGFIM. This approach scales linearly with the number of unknowns for both CPU time per iteration and the computer memory storage.

For the examples presented here, the 3D unit cells can be effectively modeled by HFSS, and readers are referred to its manual for detail construction of the unit-cell model. However, it should be mentioned that a rhombic box enclosing the unit cell can be chosen to account for the two axes of periodicity that are not necessarily orthogonal. One boundary surface of the box is designated as a master boundary, and the other boundary opposite to it is the slave. Fields at the master and slave boundaries satisfy the Floquet conditions. A Floquet port allows the flexible choices of polarization and angle of incidence.

The fabrication of FSVs or metamaterials with 3D unit cell is not as straightforward as the planar FSSs using PCB technology. The emergence of inexpensive 3D printing allows us to realize exotic unit-cell geometrical designs that are previously unimaginable. In this section, the effectiveness of commercial software like HFSS has been demonstrated for the analysis of periodic structures with 3D unit cells which can be conveniently fabricated by inexpensive 3D printing. The simulated spectral responses are validated by experiments.

Measurement Setup

Figure 21 shows the measurement setup, in which the two ports of the Agilent E8361A PNA network analyzer (10 MHz–67 GHz) are connected to a pair of transmitting and receiving horns, individually providing a 23.8-dBi gain. Alignment of the setup is made using a laser source. The FSS sample under test is mounted on a microwave absorbing panel with an opening window slightly smaller than the size of the sample. In the 60 GHz band, the measured transmission response through the microwave absorber has an insertion loss larger than 50 dB. The transmission response of the FSS is obtained by taking the difference of the measured $|S_{21}|$ with and without the sample.

3D FSSs Fabricated by 3D Printing

3D FSS samples can be fabricated using a Stratasys Objet30 Scholar 3D printer, which is convenient and inexpensive. Figures 22 and 23 show the 3D printer and some fabricated FSS samples, respectively. As shown, the 3D printer can easily fabricate various FSSs with previously unimaginable 3D structures

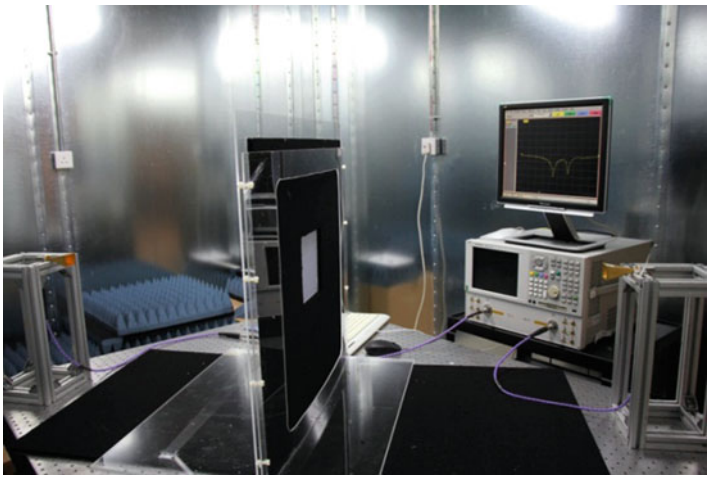


Fig. 21 Measurement setup of FSS structure

DOI 10.1007/978-981-4560-75-7_23-1

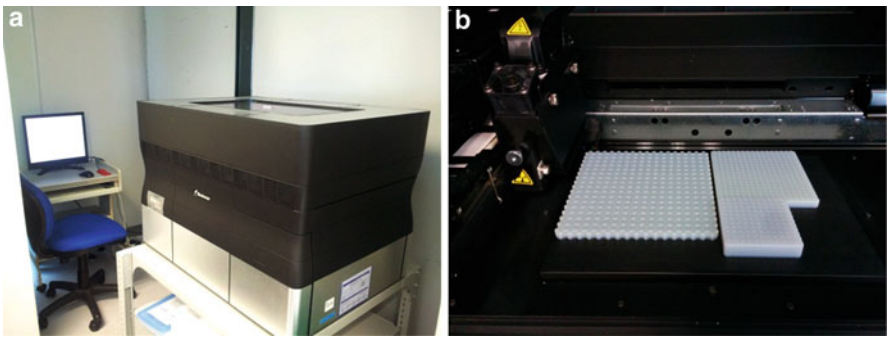


Fig. 22 A typical low-cost 3D printer. (**a**) The whole system and (**b**) the printing platform

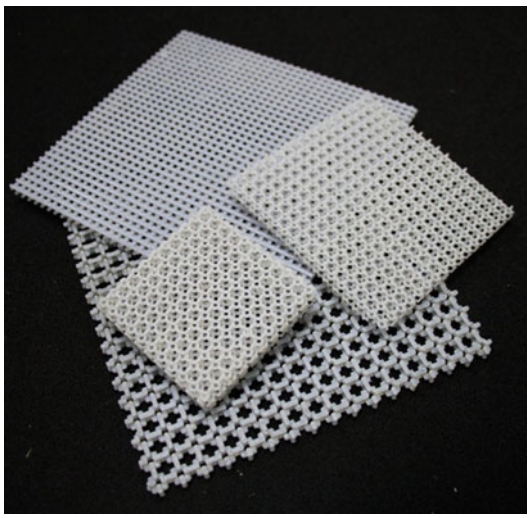


Fig. 23 Samples of FSSs with 3D unit cells fabricated by 3D printing

which cannot be realized by the standard PCB processing technology. As examples, two kinds of 3D FSSs, including woodpile FSS and torus FSS, will be introduced in detail in the following subsections.

Woodpile FSS

The first example of FSSs with 3D unit cells is a simple four-layer woodpile structure. Each layer has a thickness of 1.3 mm, and the size of the whole unit cell is 4.5 mm by 4.5 mm. The rods are 1.25 mm in width and height and the center-to-center separation is 4.5 mm, and the third layer is stacked with the rods shifted by 2.25 mm with respect to the first layer. Figure 24a–c shows the top view and side view of the FSS structure and its unit cell, respectively, and Fig. 24d shows the fabricated woodpile FSS. The woodpile was printed with polymer material which has an estimated dielectric constant of 2.95 and a loss tangent of 0.01. Slabs of different thickness with gloss and matt formats were first made, and time-domain spectroscopy measurement was conducted. Material electrical property was then extracted out from the measured data. After each layer of printing with 25 μm resolution, the surface is cured by ultraviolet light. During the printing process, air voids of the FSS were filled with water-soluble material which was washed off after the fabrication was completed.

Figure 25 shows the measured transmission response versus simulated results with different relative permittivities ranging from 2.5 to 3.5 with the same loss tangent of 0.01. In the measurement, the first and second resonances occur at 53.5 and 58.2 GHz, respectively. For the relative permittivity of 2.95, the

DOI 10.1007/978-981-4560-75-7_23-1

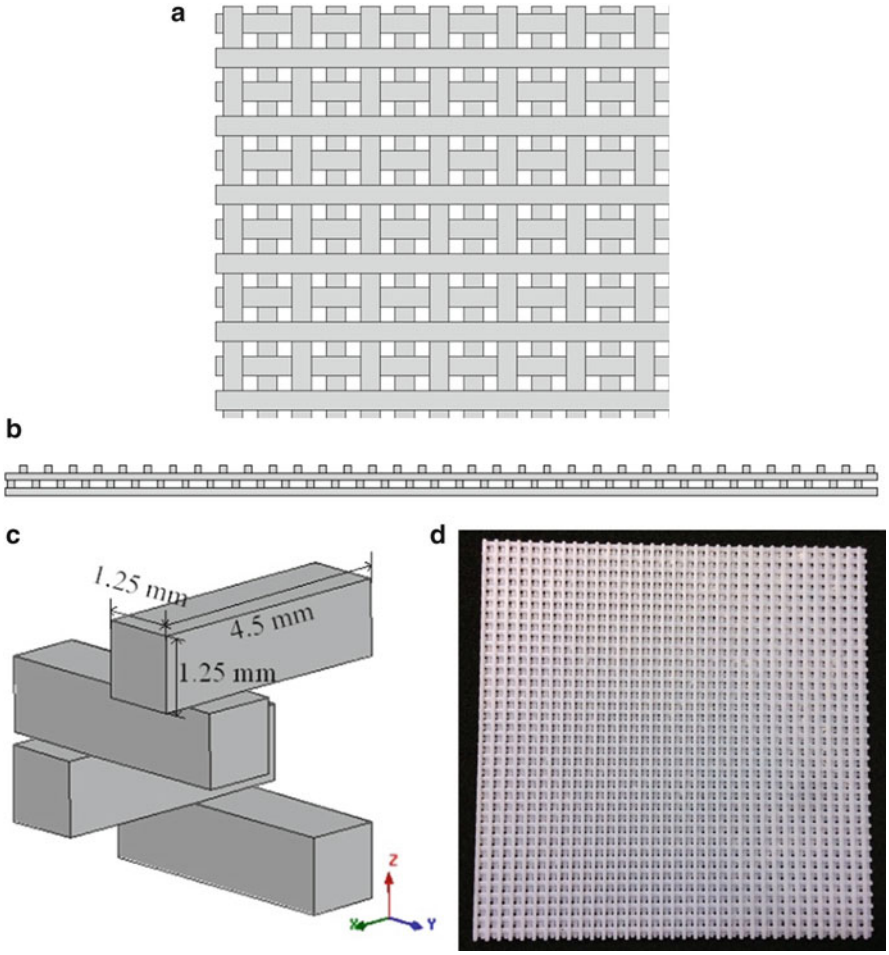


Fig. 24 Four-layer woodpile FSS. (**a**) Top view, (**b**) side view, (**c**) unit cell, and (**d**) fabricated sample

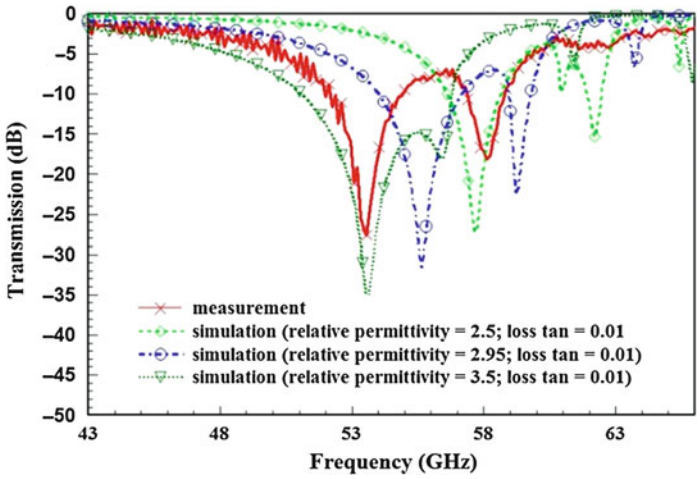


Fig. 25 Frequency responses of the four-layer woodpile FSS with different material properties

simulated first resonance occurs at 55.5 GHz and the second resonance occurs at 59.2 GHz, having a respective error of 3.8 % and 1.7 %. If one tunes the relative permittivity to 3.5, the first resonance matches that of the measured result, while the second resonance shifts to 56.4 GHz with an error of 3.1 %.

DOI 10.1007/978-981-4560-75-7_23-1

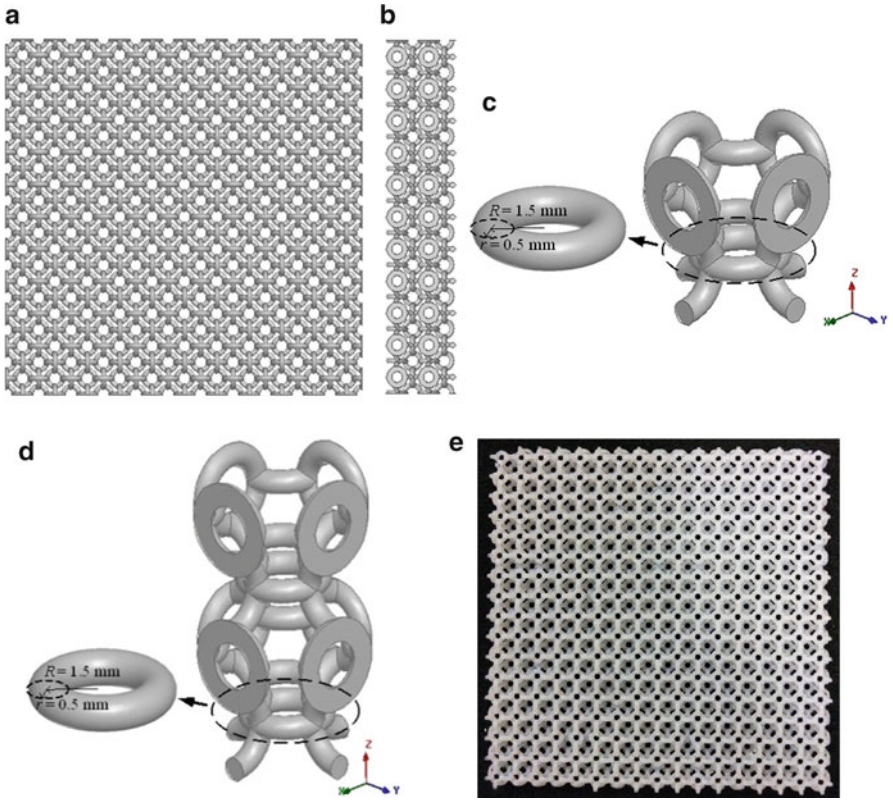


Fig. 26 Torus FSSs with the following parameters: major radius $R = 1.5$ mm, minor radius $r = 0.5$ mm, and center-to-center separation is 3 mm (**a**) Top view (two layers) (**b**) Side view (four layers) (**c**) Unit cell (two layers) (**d**) Unit cell (four layers) (**e**) Fabricated prototype

Torus FSS

In this subsection, a kind of FSS structure consisting of a torus array, called torus FSS, is investigated. It has more complex 3D unit cells compared to the woodpile FSS described above. Figure 26 shows the two-layer and four-layer torus FSS structures, in which Fig. 26a shows the top view of the two-layer torus FSS, and Fig. 26b shows the side view of the four-layer torus FSS. Figure 26c, d shows the unit cells of the two FSSs. At last, a four-layer torus FSS example is fabricated and its prototype is shown in Fig. 26e.

For the relative permittivity of 2.95 and loss tangent of 0.01 in the HFSS models, the measured results are compared with the simulated ones, as shown in Fig. 27. For the four-layer torus FSS, the resonance occurs at 45 GHz for both the measurement and simulation. For the two-layer torus FSS, however, the measured resonant frequency is shifted to 45.48 GHz while the simulated resonant frequency is 46.52 GHz. The predicted error is around 2.3 %. It should be noted that the simulated resonances are not as deep as the measured ones, and therefore, the loss tangent is increased from 0.01 to 0.035 and repeated the simulations while keeping the relative permittivity of 2.95 fixed. Figure 28 shows the comparison of the measured and simulated results. It is shown that the resonance location remains the same for both two-layer and four-layer torus FSSs, but the levels of transmission loss between the simulations and measurements have much better agreement when a loss tangent of 0.035 is used. As there is some degree of rotational symmetry of the torus FSS, it is expected that the resonant frequency would not be seriously affected by rotating the FSS screen with respect to the plane of the incident field. For verification, the fabricated two-layer FSS shown in Fig. 29a is measured for different rotation angles, and the measured results are depicted in Fig. 29b. As seen, the resonance location is slightly shifted by about 1 % for rotating the two-layer torus FSS by 45°.

DOI 10.1007/978-981-4560-75-7_23-1

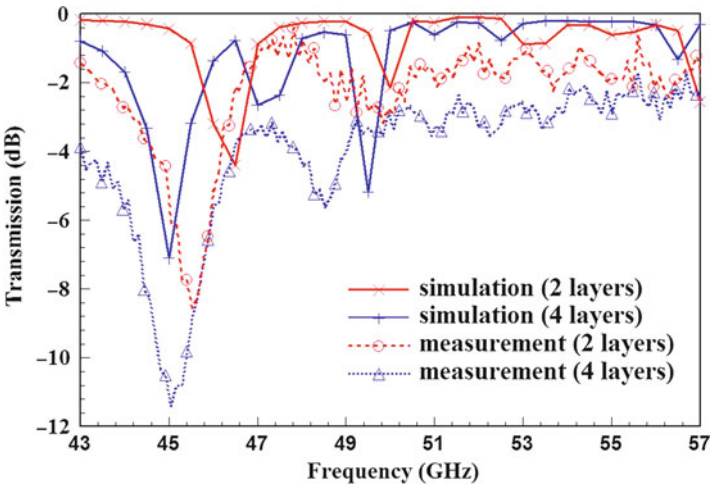


Fig. 27 Transmission responses of the torus FSSs when the loss tangent is 0.01

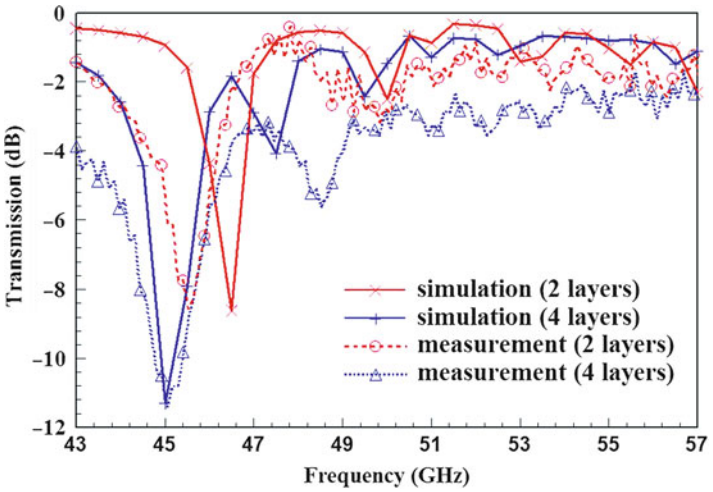


Fig. 28 Transmission responses of the torus FSSs when the loss tangent is changed to 0.035

Discussion on the FSSs Fabricated by 3D Printing

In this section, it has been demonstrated that commercial software like HFSS indeed yields accurate results in modeling FSS with complex 3D unit cells, which can be conveniently fabricated by 3D printing. It has also been demonstrated that the relative permittivity of the printing material needs to be reasonably characterized. For the woodpile and torus FSSs, the relative permittivity of the material is not high and the unit cells are partially filled with air, and therefore, the resonant frequency is not very sensitive to the slight changes of the relative permittivity. However, the level of transmission would mainly depend on the value of the loss tangent of the used material.

Design of Antennas Using Results from Periodic Structures

In sections "Multilayered FSS Filter" and "FSS with 3D Unit-Cell Elements," FSSs with planar and 3D unit cells in which the periodic elements are identical have been analyzed. In this section, the information calculated for the periodic case is used to design dielectric lens antennas and reflectarrays in which the

DOI 10.1007/978-981-4560-75-7_23-1

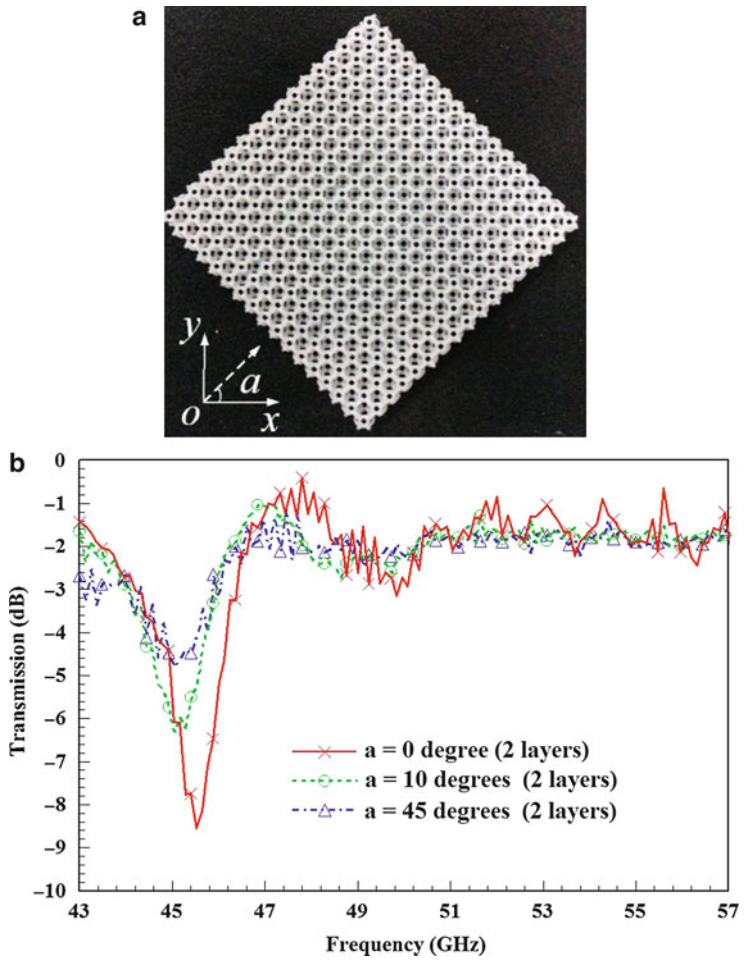


Fig. 29 Transmission responses of the two-layer torus FSS with different rotation angles (**a**) Fabricated FSS prototype (**b**) Measured results

array elements are different. Lens antennas and reflectarrays are suitable candidates for millimeter-wave applications as they have space-fed structures, leading to lower feeding loss when compared to patch antenna arrays. The antenna elements are carefully designed such that their total transmission or reflection phase can completely compensate for the phase difference due to different path lengths from the antenna feed to the position of the corresponding unit cell. For the dielectric lens antenna operating at 60 GHz, the low-cost 3D printing technology is used. In contrast, for the microstrip reflectarrays, conventional printed-circuit-board technology is employed for the construction of the wideband and dual-band antennas.

Lens Antennas

In making use of 3D printing technology for the lens antenna, it is assumed that the lens antenna is composed of a periodic array of square dielectric columns of variable column heights (Yi et al. 2014). The height of the column is determined such that the total electric length from the source to the bottom of the column, through the column, and from the top of the column to a common horizontal reference plane is constant.

Matching Layers

Knowing that reflection will occur at the interface of two different materials and, in this case, air and the 3D printing polymer material, two matching layers are added near the top and bottom surfaces of the

DOI 10.1007/978-981-4560-75-7_23-1

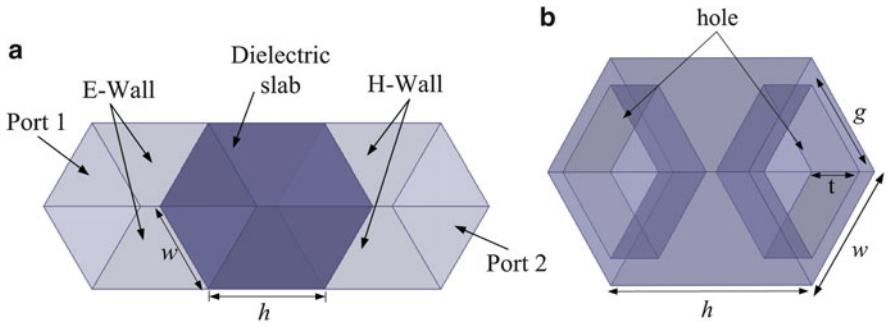


Fig. 30 (**a**) Simulation setup of the dielectric column without matching layers. (**b**) Dielectric column with two matching layers. $w = 2.5$ mm, $g = 1.9$ mm, and $t = 0.95$ mm

column to reduce the reflection. Figure 30a shows the simulation setup of the dielectric column without the matching layer. When incorporated with the matching layers, the dielectric column in Fig. 30a is replaced by the column with two matching layers in Fig. 30b. The matching layers are formed by introducing two square voids near the top and bottom surfaces of the dielectric column. All the dimensions of the column and voids are integral multiples of the printing resolution of the 3D printer at 25 μm. The relative dielectric constant and loss tangent at 60 GHz are $\varepsilon_r = 2.95$ and $\tan\delta = 0.01$, respectively. The dimensions of the matching layer are determined as a quarter-wave transformer to match the air region with the dielectric layer. The effective dielectric constant needed for the matching layer is given by $\varepsilon_e = \sqrt{\varepsilon_r \times 1}$ and the thickness $t = \frac{\lambda_e}{4} = \frac{\lambda}{4\sqrt{\varepsilon_e}}$, where λ is the wavelength in air and λ_e is the effective wavelength in the matching layer. The width of the square void is optimized such that the effective dielectric constant is equal to $\varepsilon_e = \sqrt{2.95} \approx 1.72$, and this results in $g = 1.9$ mm.

The columns with and without the matching layers are simulated using HFSS with the periodic boundary conditions at normal incidence. The simulated magnitudes of the reflection coefficient ($|S_{11}|$) and transmission coefficient ($|S_{21}|$) against the column height at 60 GHz are shown in Fig. 31a, b, respectively. With the matching layers, $|S_{11}|$ is below -30 dB for the whole range of column height from 2 to 10 mm. As a consequence, $|S_{21}|$, the total insertion loss due to reflection and material loss only slightly increases to about 0.7 dB at $h = 10$ mm. Nonetheless, the loss is not very significant.

Lens Antennas With and Without Matching Layers

Both the lens antennas with and without the matching layers are designed and simulated at 60 GHz. Both models have 19 × 19 square dielectric columns with various heights to provide the required phase compensation. The antenna is fed by a point source from the center of planar surface of the lens with a focus-to-diameter ratio (F/D) of 0.41. Figures 32 and 33 show the simulation models without and with the matching layers, respectively. When viewing from the top, the model without the matching layers resembles a grid of 19 × 19 squares (Fig. 32a) which are of different heights. When viewing from the bottom, it is flat plate as shown in Fig. 32b. The 3D view of the model is shown in Fig. 32c. On the other hand, the model with the matching layers when viewing from the top (Fig. 33a) and from the bottom (Fig. 33b) resembles an array of 19 × 19 square patches. Each patch is in fact a square void of size g by g with a depth of t. The 3D model is shown in Fig. 33c. Figure 34 shows the comparison of the simulated directivities of the two models. It shows that the directivity of the lens antenna with the matching layers is 0.7 dB higher than that of lens antenna without the matching layers throughout the whole frequency band from 50 to 70 GHz which demonstrates the effectiveness of the matching layers.

DOI 10.1007/978-981-4560-75-7_23-1

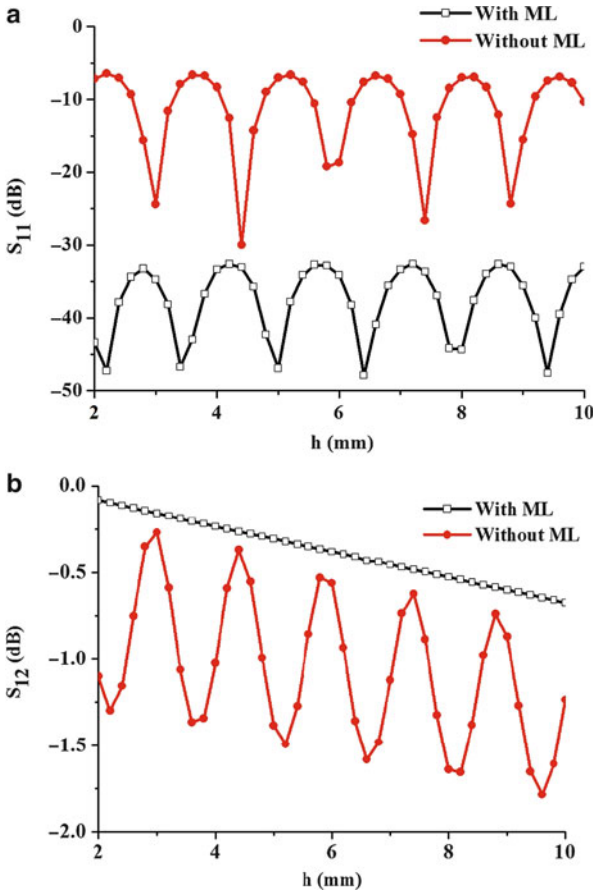


Fig. 31 Simulated $|S_{11}|$ and $|S_{21}|$ against column height with and without matching layers. (**a**) $|S_{11}|$ and (**b**) $|S_{21}|$

Lens Antenna Measurement

The lens antenna with matching layers was fabricated using the 3D printer, and the prototype is shown in Fig. 35a. The prototype was excited by a waveguide and mounted on an NSI near-field measurement system as shown in Fig. 35b for gain and pattern measurements. Figure 36 shows the comparison of the measured and simulated peak gains of the lens antenna. In general, there is about maximum 1 dB difference between the simulated and measured gains from 50 to 67 GHz covering the whole 60 GHz band. The measured peak gain is 23.5 dBi at 61 GHz. This is possibly due to the imprecise dielectric constant and loss tangent, but it could also be due to the measurement setup and alignment.

Figure 37 shows the comparisons of the measured and simulated radiation patterns of the lens antenna prototype at 60 GHz. The side-lobe levels are below −18 dB and −14 dB in H and E planes, respectively. Excellent agreement between the simulated and measured results is obtained.

In this subsection, it has been successfully demonstrated that design of a dielectric lens antenna is based on the transmission phase information computed for a doubly periodic structure with 3D unit-cell geometry. The antenna gain is enhanced by reducing the reflection loss due to material discontinuity at the lens and air interfaces using two matching layers. A low-cost fabricated prototype of the high-gain millimeter-wave antenna at 60 GHz using 3D printing technology has also been demonstrated.

Microstrip Reflectarray

In the previous subsection, the use of utilizing the transmission phase of 3D element of different column heights for a lens antenna design has been demonstrated. In this subsection, the reflection phase of a

DOI 10.1007/978-981-4560-75-7_23-1

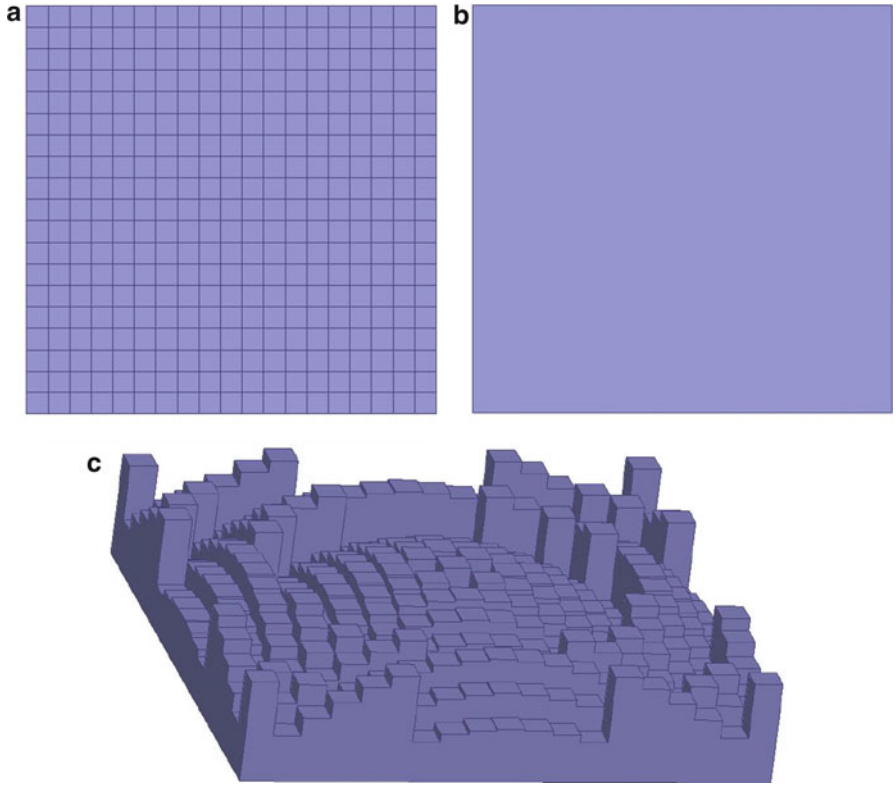


Fig. 32 Simulation model for the lens antenna without the matching layers. (**a**) Top view, (**b**) bottom view, and (**c**) 3D view

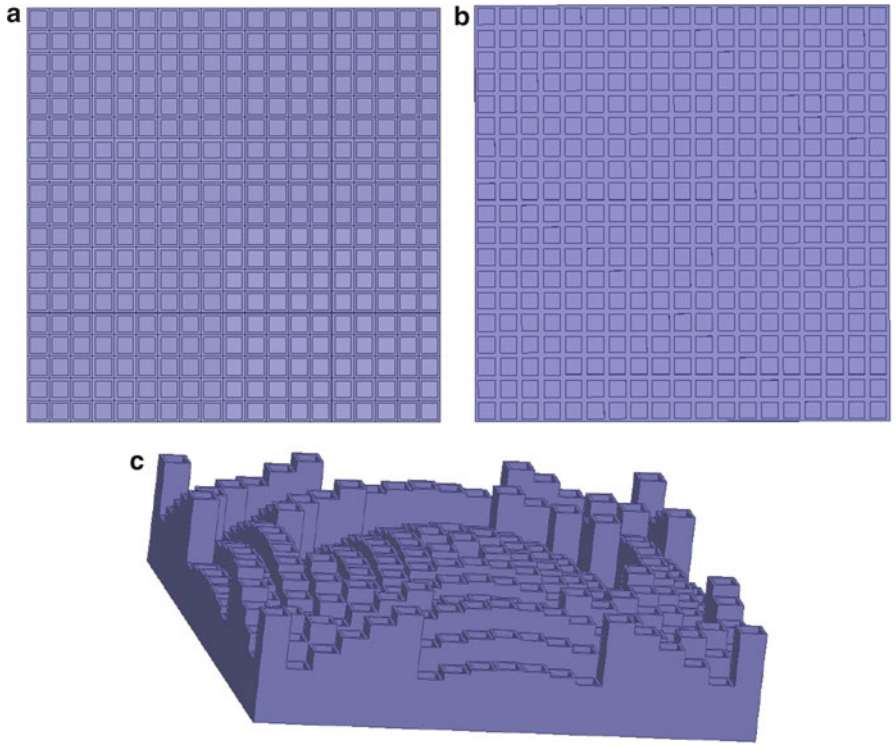


Fig. 33 Simulation model for the lens antenna with the matching layers. (**a**) Top view, (**b**) bottom view, and (**c**) 3D view

DOI 10.1007/978-981-4560-75-7_23-1

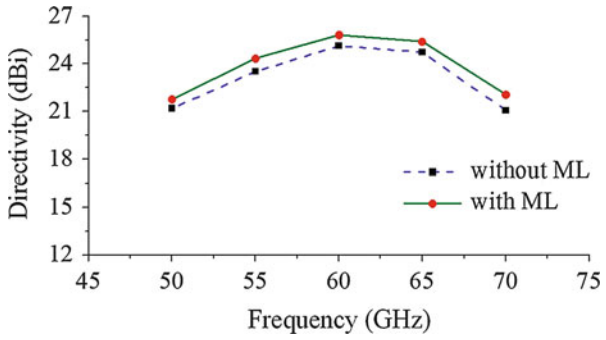


Fig. 34 Simulated directivities of the lens antennas with and without matching layers (ML)

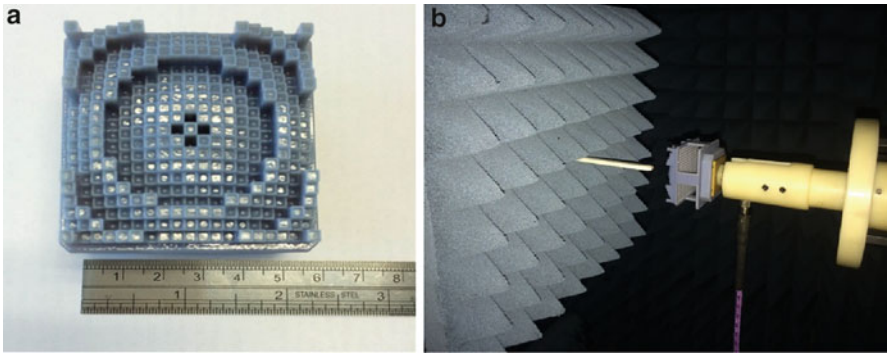


Fig. 35 (**a**) Fabricated prototype and (**b**) NSI system for gain and pattern measurements

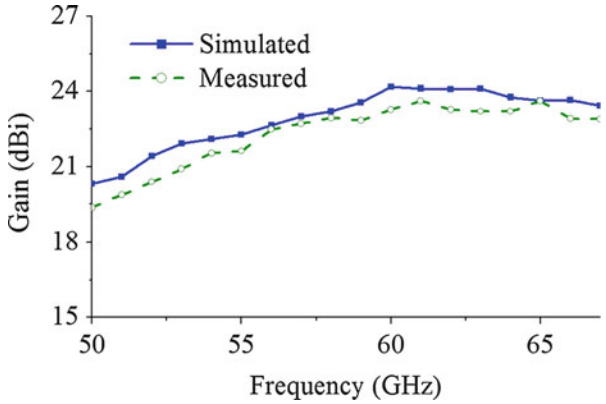


Fig. 36 Comparison of measured and simulated peak gains versus frequency

periodic array of unit-cell element will be exploited for the design of reflectarrays by varying the geometrical parameters of the unit cell (Wu et al. 2014).

Working Principle of Reflectarray

Figure 38 shows the working principle of a typical reflectarray. The reflectarray can be single layered or multiple layered. A single-layered reflectarray is preferred for its lower fabrication cost and lower weight. The phase difference $\Delta\varnothing$ due to the path length difference of is $S_2 - S_1 = \Delta S$ compensated by the reflection phase of the reflecting element such that all the reflected fields are radiating in phase in a

DOI 10.1007/978-981-4560-75-7_23-1

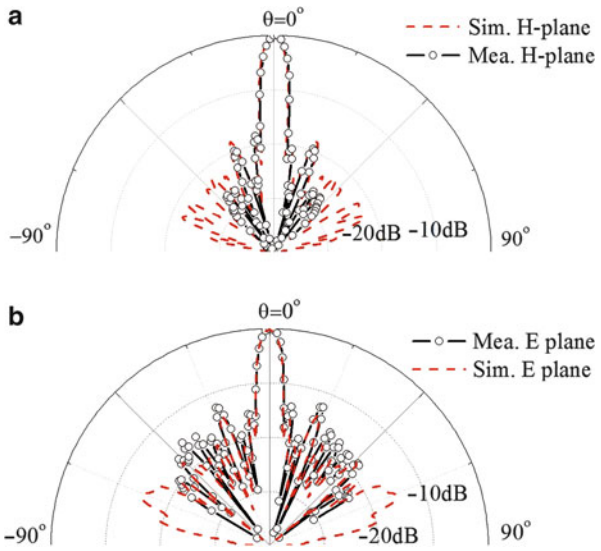


Fig. 37 Comparison of measured and simulated radiation pattern at 60 GHz. (**a**) H plane and (**b**) E plane

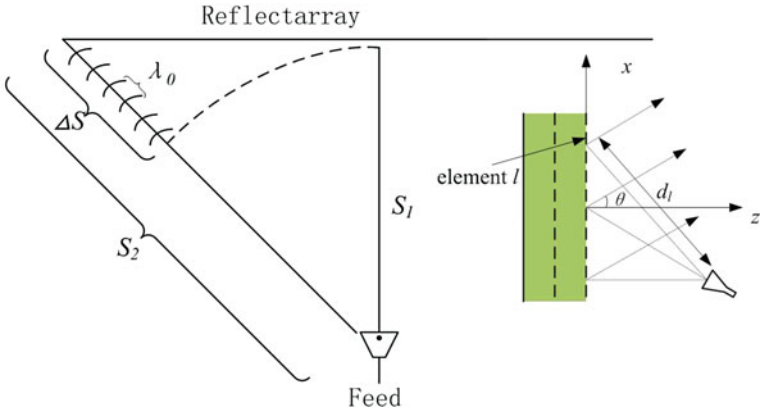


Fig. 38 Working principle of a reflectarray

designated direction. The maximum phase difference would be $\frac{\Delta S}{\lambda_o} 2\pi$, where λ_o is the free-space wavelength. Successful design of the reflectarray requires the reflected phase covering the whole range of the 360°, and the reflected phase variation with respect to the changing geometrical parameters should be as linear as possible (Huang and Encinar 2007). Conventional unit cells of square patches with different dimensions or square patches with fixed dimensions attached with a stub of different lengths cannot meet both of these requirements as shown in Fig. 39, resulting in lower antenna gain, distorted radiation pattern, and narrow antenna bandwidth.

Dual-Resonance Unit Cell

Unit-cell geometry with multiple resonances would allow us to generate a phase curve that varies smoothly with changing parameters. Figure 40 shows a square unit cell with a ring and an I-shaped dipole (Chen et al. 2013). Each side of the unit cell is L=10 mm and the outer radius of the ring is fixed at $R_o = 4.2$ mm. Four degrees of freedom are provided, namely, the gap width W_S between the circular ring and the I-shaped dipole, the inner radius of the ring R_i, and the ratios of $N = 0.5\frac{W_G}{(R_i - W_S)}$ and

DOI 10.1007/978-981-4560-75-7_23-1

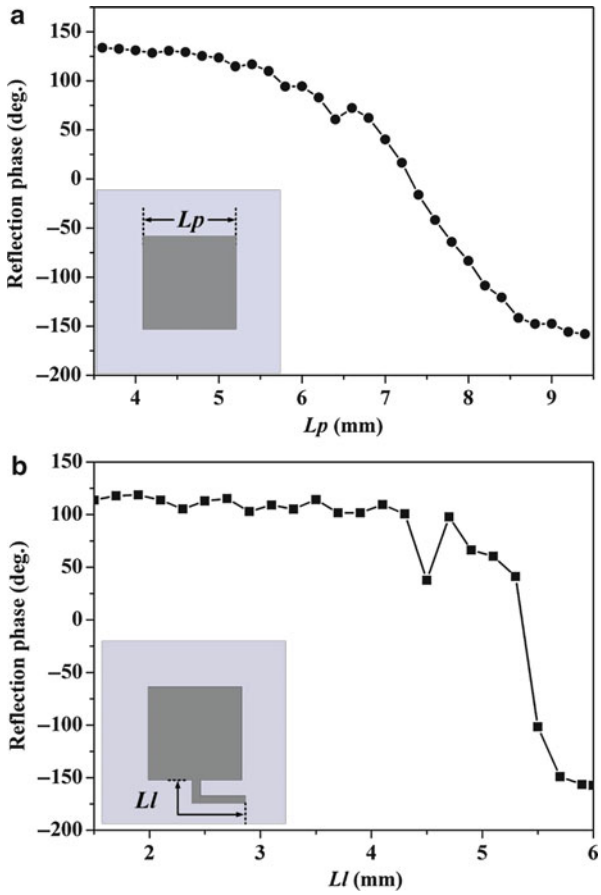


Fig. 39 Reflection phase of (**a**) square patch of various dimensions *Lp* and (**b**) square patch of fixed dimensions ($Lp = 5$ mm) with various stub lengths *Ll* (10.5 GHz, vertical polarization; substrate thickness $h = 1.5$ mm; dielectric constant $\varepsilon_r = 2.2$; side length of unit cell $L = 10$ mm)

$M = 0.5\frac{W_B}{(R_i - W_S)}$. Here W_G and W_B are the length and width of the center portion of the I-shaped dipole, respectively. The dielectric constant of the substrate is $\varepsilon_r = 2.2$, and its thickness is $h = 1.5$ mm. Also shown in Fig. 40 is the reflection phase comparison of the I-shaped dipole, the ring, and the combination of the two versus frequency. A uniform plane wave polarized to the x direction impinges on the periodic surface at normal incidence. While both the reflection phases of periodic surfaces composed of the ring and the I-shaped dipole alone change smoothly, they do not cover the entire 360°. In contrast, the double resonance provided by the combined structure covers the whole range of phase delay needed with slowly changing phase gradient. Similar phase behavior will be obtained if one varies the parameters at a fixed frequency.

Figure 41a shows the smooth changing phase of FSS with the ring and I-shaped dipole combined unit cell at 10.5 GHz. It covers over 400° phase change, obviously better than those obtained by conventional elements shown in Fig. 39. This is achieved by changing the inner radius of the ring from 1.5 to 3.5 mm while other parameters are optimal. Figure 41b on the other hand shows the reflection phase variations at different frequencies from 9.5 to 11.5 GHz. At each frequency, the whole 360° reflection phase is covered by changing the inner radius of the ring while keeping W_S, M, and N unchanged. The length and width of the central portion of the I-shaped dipole will proportionally change according to the varying R_i. The resulting phase-frequency curves are approximately linear from 9.5 to 11.5 GHz. Figure 42 shows the current distributions on the combined ring/I-shaped dipole FSS for $R_i = 2.7$ mm at 9.5, 10.5, and

DOI 10.1007/978-981-4560-75-7_23-1

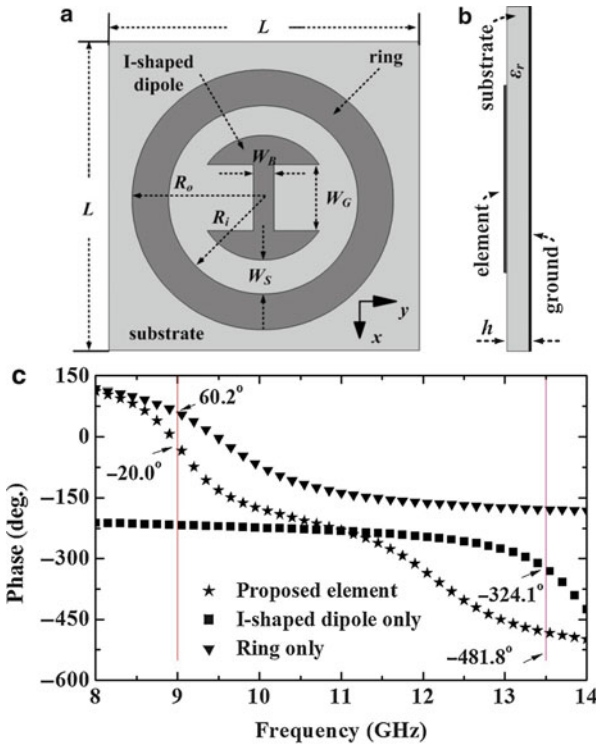


Fig. 40 Unit cell with double resonances. (**a**) Top view, (**b**) side view, and (**c**) phase delay comparison of the different elements ($M = 0.1$, $N = 0.8$, $W_s = 0.15$ mm, $R_i = 2.7$ mm)

11.5 GHz. By controlling the four parameters, one can effectively change the current distributions and resonant frequencies which in turn change the amplitude and gradient of the phase delay versus frequency. The resonant frequency of the ring by itself is around 9.2 GHz while that of the I-shaped dipole is around 22.3 GHz. The large separation of these two resonance frequencies allows us to have more linear phase variation by tuning different parameters.

In the conventional reflectarray design, the desired phase delay at a particular array element is compensated by tuning the element geometrical parameters at the center frequency. However, the reflection phases at the edge frequencies may be deviated from the desired phases significantly, therefore limiting the operating bandwidth of the reflectarray. Instead of compensating the phase delay at a single frequency, the almost linear phase curves as appear in Fig. 41b would allow us to compensate the phase delay for a band of frequencies if one can create a reflection phase database by tuning the four parameters of R_i, W_s, M, and N. The parametric sweeping was done in HFSS according to the tuning range of the four parameters: N from 0.4 to 0.6, M from 0.1 to 0.2, W_s from 0.15 to 0.35 mm, and R_i from 1.5 to 3.5 mm with an incremental step of 0.05. It should be noted that the value of the outer radius is kept at $R_o = 4.2$ mm so that the mutual coupling among the array elements is more or less the same.

Design of Wideband Reflectarray

By exploiting the linear phase-frequency response or this dual-resonance element depicted in Fig. 40, one can minimize the difference between the desired and realized phases at the lower and upper edges of the frequency band denoted by f_l and f_u, respectively. The desired phase Φ delays should fulfill

$$\Phi(f_l)(n) - \Phi(f_l)(n+1) = \Phi(f_u)(n) - \Phi(f_u)(n+1) \tag{1}$$

DOI 10.1007/978-981-4560-75-7_23-1

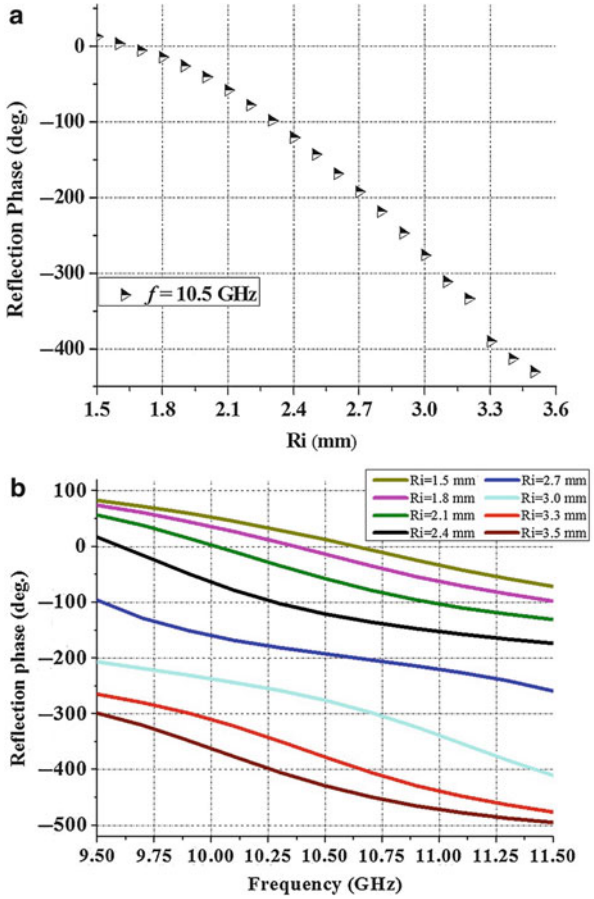


Fig. 41 Reflection phase (**a**) at a fixed frequency versus R_i and (**b**) versus frequency for different values of R_i ($M = 0.15$, $N = 0.5$, $W_s = 0.25$ mm, $R_o = 4.2$ mm)

where n is the nth element and n+1 is the (n+1)th element as shown in Fig. 43. The value of (1) is the desired phase delay difference between the nth and (n+1)th elements at center frequency. The optimal parameters of R_i, W_s, M, and N can be obtained such that the achieved phases are as close to the desired values as possible. Base on the generated database for the phase compensation, a 6 × 10 element reflectarray with a scattering angle of 30° for normal incidence of plane waves was designed and fabricated (Chen et al. 2013) and the model for simulation is shown in Fig. 44 in which symmetric boundary condition is assigned to the plane with a normal direction of –x axis. A simulated 3-dB directivity-drop bandwidth of 22.3 % is achieved when the reflectarray is centered at 10.5 GHz and the predetermined f_l and f_u are 10 and 11 GHz, respectively. The computed aperture efficiency at 10.5 GHz is 38.5 %. Comparison of the normalized simulated and measured radiation patterns is shown in Fig. 45, and a good agreement is obtained.

Unit Cell with Two Split Rings and an I-Shaped Dipole

Another unit cell that can provide a wideband linear phase delay was introduced by Chen et al. (2012) which is given in Fig. 46. The dimensions of the unit cell and the substrate are the same as in those Fig. 40. The outer ring of Fig. 40 is now split into two coupled open rings which generate three resonances. The two variable gaps of the two rings provide more flexibility in tuning the reflection phase. Again, the reflection phase of the FSS with each set of geometrical parameters is simulated by using HFSS under periodic boundary conditions with normal incidence and x-polarized electric fields. By varying two

DOI 10.1007/978-981-4560-75-7_23-1

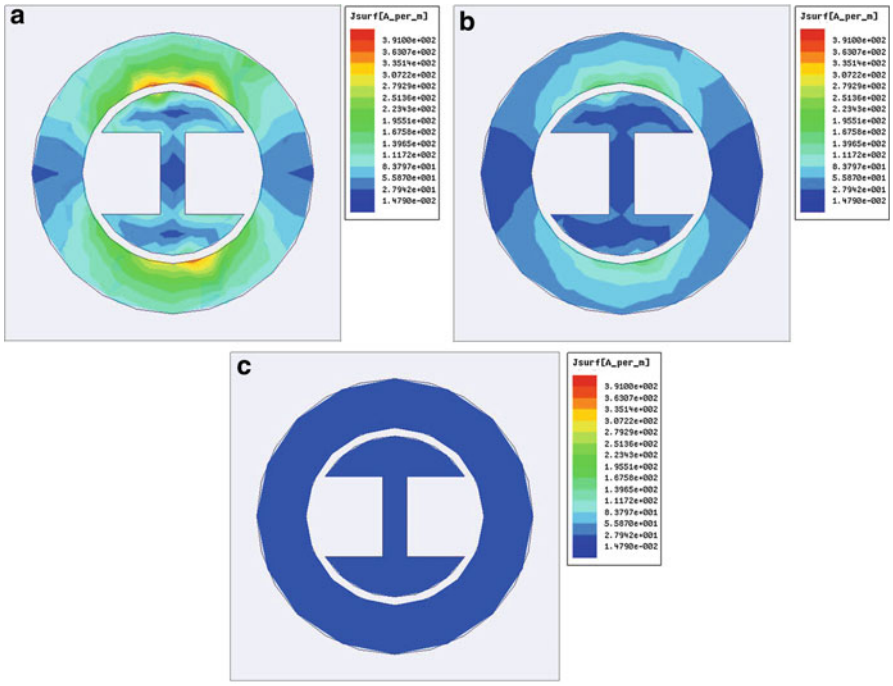


Fig. 42 Current distributions on the ring and I-shaped dipole (**a**) 9.5 GHz (**b**) 10.5 GHz (**c**) 11.5 GHz

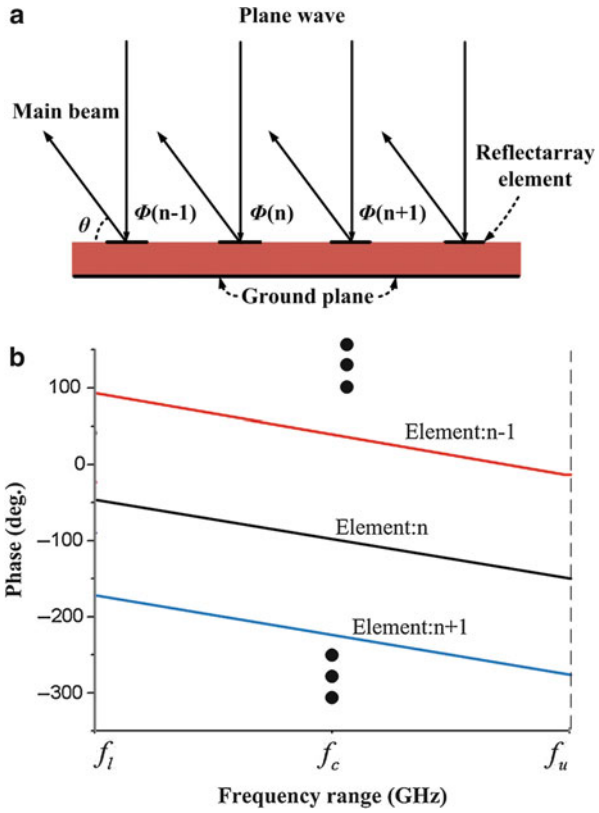


Fig. 43 Phase difference between two adjacent elements. (**a**) Profile view of the reflectarray and (**b**) phase versus frequency for three typical elements

DOI 10.1007/978-981-4560-75-7_23-1

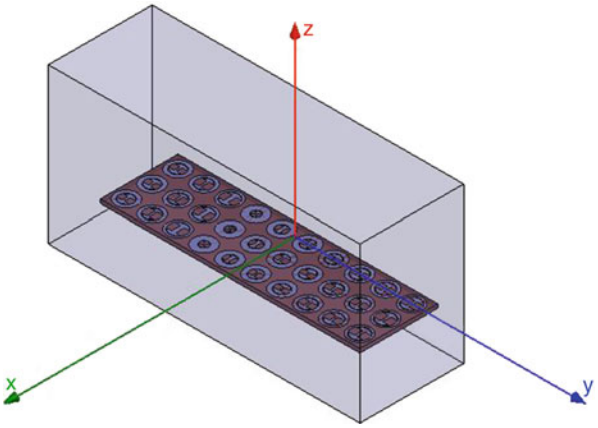


Fig. 44 HFSS model of the 6 × 10 element reflectarray. Symmetric boundary condition is assigned to the plane with a normal direction of the –x axis

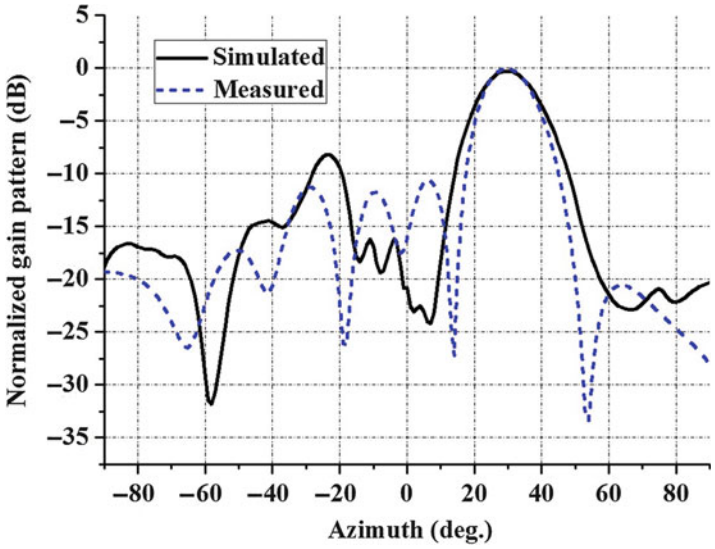


Fig. 45 Comparison of the simulated and measured radiation patterns at 10.5 GHz.

parameters, W_2 and M, while fixing the remaining parameters as given in Fig. 46, linear phase-frequency curves covering a large range of angles are obtained.

Two classes of elements with different geometrical parameters are employed to generate the phase curves in Fig. 47. In Class 1, W_2 varies from 2.0 to 2.0 mm when M is fixed at 0.5 mm. In contrast, for Class 2, W_2 is fixed at 4.0 mm when M varies from 0.4 to 4 mm. By combining these two classes, both range and linearity requirements for the phase delay can be satisfied from 12 to 14 GHz. Based on these reflection phase curves and design condition in (1), a 20 × 20 element reflectarray was designed (Chen et al. 2012) and the measurement setup is shown in Fig. 48. The heights of the feed horn and the reflectarray are $h_f = 100$ mm and $h_r = 100$ mm, respectively, and the elevation angle of the horn is $\alpha = 17°$. The distance from the feeder to the reflectarray is $d = 470$ mm.

The simulated and measured radiation patterns in the H plane are shown in Fig. 49, and they compare very well at 12, 13, and 14 GHz. Only the simulated E-plane patterns at 12, 13, and 14 GHz are shown in Fig. 50 as measurement could not be conducted due to the obstruction of the supporting structure. It is observed that the main beams at 12 and 14 GHz are slightly shifted from that of the center frequency.

DOI 10.1007/978-981-4560-75-7_23-1

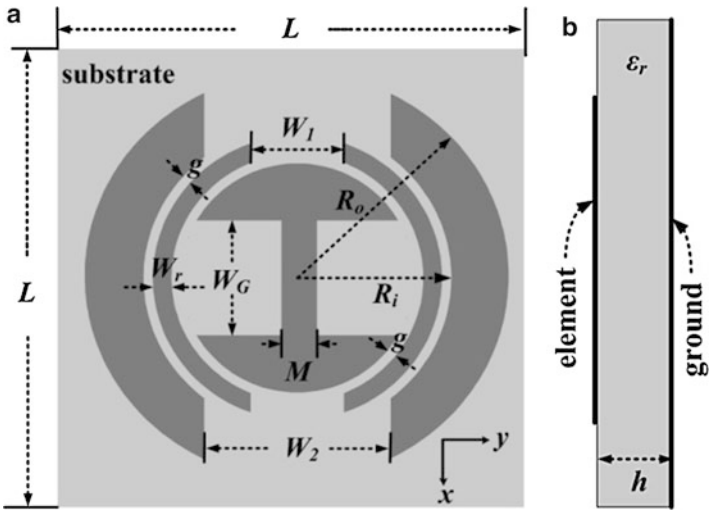


Fig. 46 Geometry of the unit cell with two split rings and an I-shaped dipole. (**a**) Top view and (**b**) side view ($R_i = 3.3$ mm, $g = 0.2$ mm, $W_r = 0.4$ mm, $W_1 = 1$ mm, $W_G = 2.5$ mm, $\varepsilon_r = 2.2$, $h = 1.5$ mm)

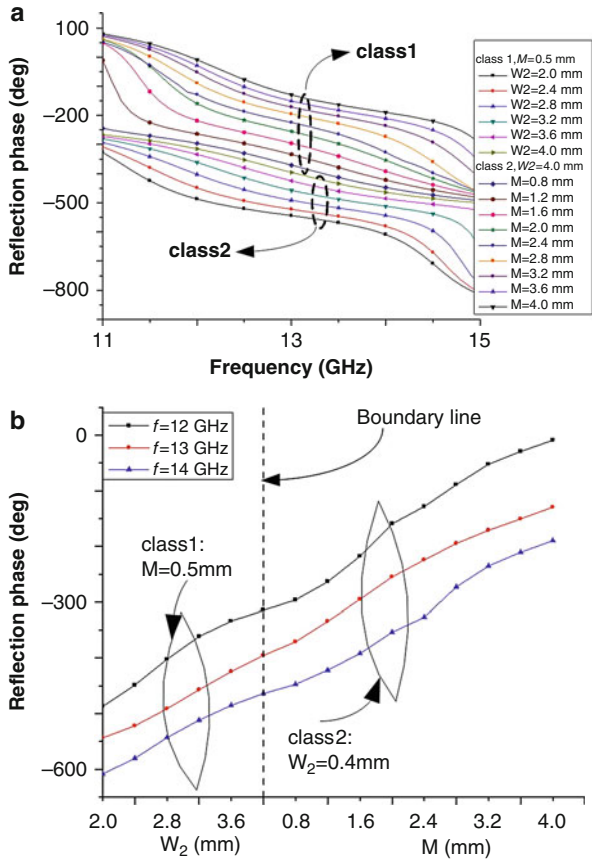


Fig. 47 Reflection phase versus (**a**) frequency and (**b**) dimensional parameters of two classes of elements

This is attributed to the parallel phase characteristic of the elements and the use of design condition (1). The simulated 3-dB directivity-drop bandwidth is 22.7 % from 11.75 to 14.75 GHz. The directivity at 13 GHz is 26.68 dBi which is equivalent to 50 % aperture efficiency.

DOI 10.1007/978-981-4560-75-7_23-1

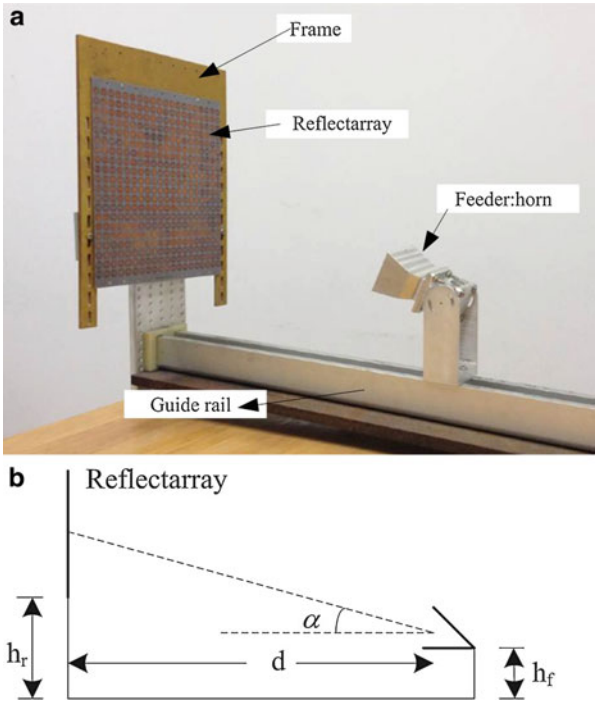


Fig. 48 Fabricated 20 × 20 reflection with elements given in Fig. 46. (**a**) Photograph and (**b**) reflectarray configuration

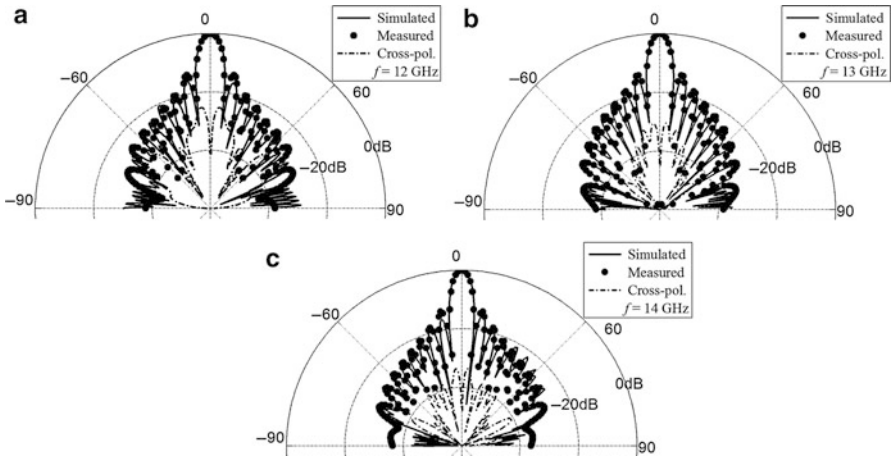


Fig. 49 Comparison of simulated and measured H-plane patterns at (**a**) 12 GHz, (**b**) 13 GHz, and (**c**) 14 GHz

Phase-Phase Distribution of Dual-Resonance Unit Cell for Dual-Band Reflectarray

The unit-cell elements in Figs. 40 and 46 allow us to design reflectarrays operating at 9–11 GHz and 12–14 GHz, respectively. It would be interesting to see how they can be combined to design a dual-band reflectarray that can operate at both 9 and 13.5 GHz. As discussed in Tsai and Bialkowski (2003), the achievable reflection phase range for a traditional reflectarray element is less than 360° for a single-layer substrate. This can be approximately calculated by $2\pi \times (1 - kh/\pi)$ for a small substrate thickness h, where k is the wave number in the substrate (Qu et al. 2014a). The combined ring/I-shaped dipole unit cell in Fig. 40 has two resonances, and some of the required phase compensation at both 9 and 13.5 GHz cannot be simultaneously satisfied, leading to design failure. Figure 51 shows the phase-phase distribution

DOI 10.1007/978-981-4560-75-7_23-1

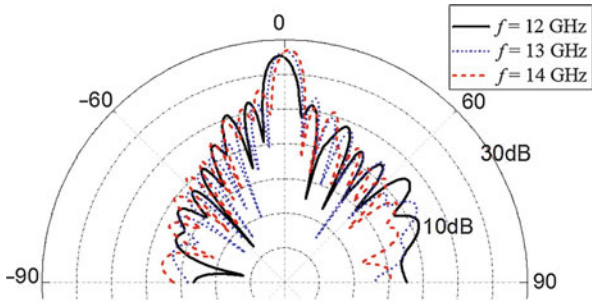


Fig. 50 Simulated E-plane patterns at 12, 13, and 14 GHz

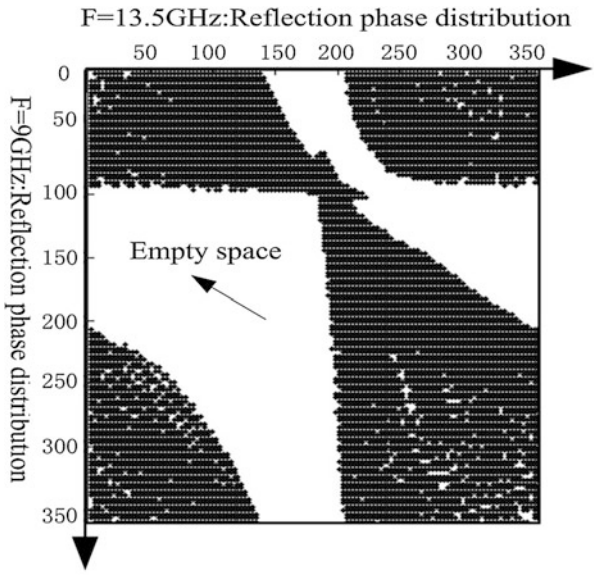


Fig. 51 Phase-phase distribution of the dual-resonance element at 9 and 13.5 GHz

of the dual-resonance element at 9 and 13.5 GHz. Each dot in the plot indicates that a combination for a phase compensation of p_1 degree at 9 GHz and p_2 degree at 13.5 GHz within an error of less than 3° is achievable by a set of parameters (R_i, W_s, M, N). However, both p_1 and p_2 can be any value within the range of 0–360° in general cases. The empty space in Fig. 51 indicates that within the range of parametric tuning, no specific unit cell can satisfy this ($p1$, $p2$) combination simultaneously. Therefore, another unit-cell element as shown in Fig. 52 was introduced in Qu et al. (2014a), in which four resonances are provided.

Four-Resonance Unit Cell

The element depicted in Fig. 52 is designed by combining the unit cells in Figs. 40 and 46 with a circular ring, two concentric open rings, and an I-shaped dipole. As shown in Fig. 46, the gap widths W_1 and W_2 are for tuning the phase delay at the upper frequencies. The circular ring is mainly for the tuning of the lower frequencies. The parameters W_1, W_2, M and are tuned to build the database for designing the reflectarray where $M = 0.5W_B/(R_i - 0.9mm)$. The outer radius of the ring is fixed at $R_o = 4.2$ mm. The widths of the two centric rings are 0.3 mm and the gap size g is 0.1 mm, and $N = \frac{0.5W_G}{(R_i - 0.9\ mm)}$ is fixed at 0.6. The tuning ranges of W_1, W_2, M, and R_1 are 3.5–4.0, 0.5–3.0, 0.4–3.0, and 0.05–0.5 mm and in step sizes of 0.5, 0.5, 0.2, and 0.05, respectively. Figure 53 shows the phase-phase distribution of the

DOI 10.1007/978-981-4560-75-7_23-1

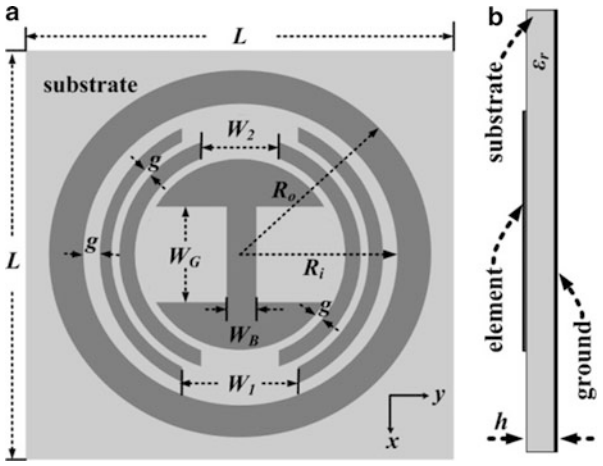


Fig. 52 Four-resonance unit cell. (**a**) Top view and (**b**) side view

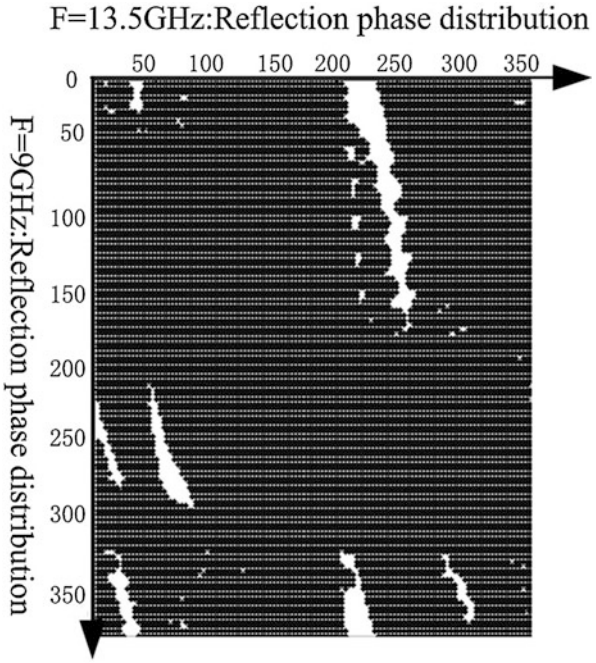


Fig. 53 Phase-phase distribution of the dual-resonance element at 9 and 13.5 GHz

four-resonance element at 9 and 13.5 GHz. The empty space in Fig. 53 is substantially smaller than that in Fig. 51. The unfilled areas in Fig. 53 are mainly attributed to the high sensitivity of the phase variation to the dimensions, meaning that the fabrication tolerance of the unit-cell element may affect the performance of the designed reflectarray. Therefore, the phase information of this element should be used to complement that of Fig. 51. That is, only when one cannot find a corresponding set of parameters in Fig. 51 using the element in Fig. 40, the phase information in Fig. 53 using the element in Fig. 52 should be used.

A 20 × 20 Dual-Band Reflectarray Using Hybrid Elements

Following the design procedure as discussed above, a 20 × 20 reflectarray with two types of elements in Figs. 50 and 52 has been designed, and the measurement setup is similar to that in Fig. 48 with the same values for the parameters $\propto, h_f, h_r$, and d. The focus-to-diameter (F/D) ratio of the reflectarray is about

DOI 10.1007/978-981-4560-75-7_23-1

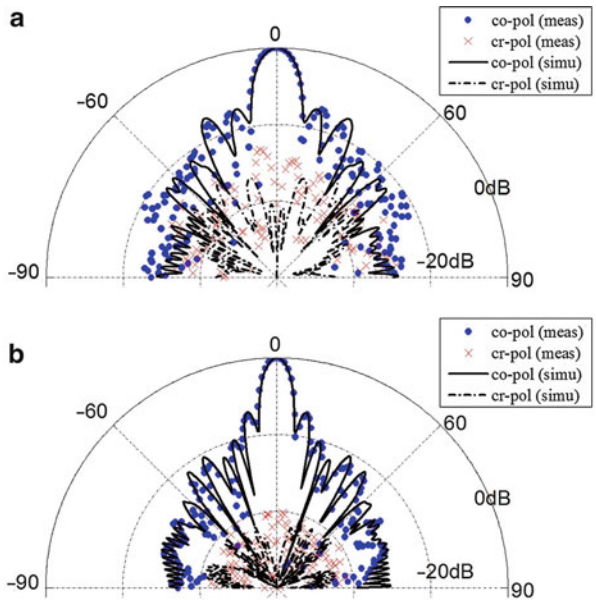


Fig. 54 Comparison of the simulated and measured radiation patterns in the H plane at (**a**) 9 GHz and (**b**) 13.5 GHz of the reflectarray with hybrid elements

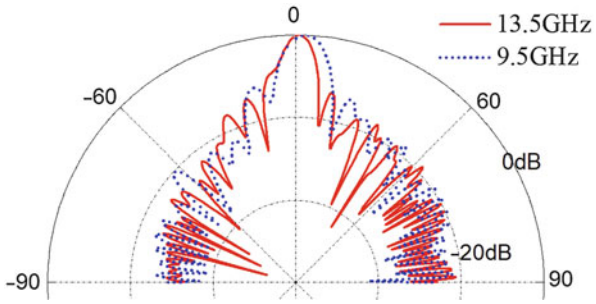


Fig. 55 Simulated radiation patterns in the E plane of the reflectarray with hybrid elements at 9 and 13.5 GHz

2.35. There are 278 dual-resonance elements and 122 four-resonance elements. Figure 54 shows the comparison of the simulated and measured radiation patterns in the H plane. Very good agreements are obtained for both 9 and 13.5 GHz. The measured side lobes are around −15 dB for both frequencies, and the measured cross-polarizations (cr-pols) are below −16 and −40 dB at 9 and 13.5 GHz, respectively. As mentioned earlier, due to the presence of the supporting structure, measurement in the E plane cannot be performed, and therefore, only simulated radiation patterns in the E plane are shown in Fig. 55. It is noted that there is a slight shift of the main beam from +z direction by about 2–1° at 9 and 13.5 GHz, respectively. The simulated cr-pols are below −40 dB for both frequencies. The comparison of the simulated and measured gains versus frequency is shown in Fig. 56. As the direction of the peak gain slightly shifted with frequency, the simulated and measured gains are compared at $\theta = 1°$. There is about 0.5-dBi difference between the simulated and measured results in both frequency bands. The measured gains are 18 and 24 dBi at 9 and 13.5 GHz, respectively.

The bandwidth of the reflectarray can be widened with the use of unit-cell elements that have multiple resonances as illustrated by the three examples discussed above. This design methodology can be extended to higher frequencies into the terahertz regime (Qu et al. 2014b). An operating bandwidth of over 36 % for the wideband design and over 40 % for the frequency-scanning design with beam scanning

DOI 10.1007/978-981-4560-75-7_23-1

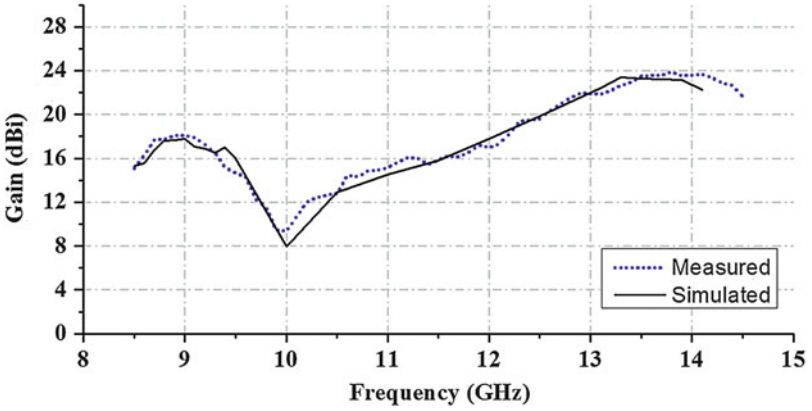


Fig. 56 Comparison of the simulated and measured gains versus frequency of the 20 × 20 reflectarray with hybrid elements

angle of −35° to −5° from 0.2 to 0.3 THz are achieved. It should be noted that for these multiple resonant structures with fine details, the simulation time would be substantially increased, whether integral equation method or finite-element solution is used. For HFSS simulation of the dual-resonance unit cell shown in Fig. 40, each simulation for a set of parameters at one frequency point costs 0.24 GB memory and 80 s on average on a computer with 4-GB memory and Intel Core i3 CPU@2.93. The memory and CPU time increase to 0.64 GB and 158 s for the four-resonance unit cell in Fig. 52..

Conclusion

This chapter began with brief descriptions of some of the earlier works on FSSs, mainly on the physical interpretations of the observed scattering phenomena associated with periodic structures. Numerical modeling of the FSS started 50 years ago first with the variation method and later with the method of moments. Unit-cell geometries of the FSS are limited as restricted by the choices of suitable basis functions.

With the modeling of flexibility of the subdomain basis functions, a variety of numerical techniques were developed in the 1980s and 1990s to speed up the convergence of the doubly infinite summation associated with the integral equation formulation of the FSS. The FSS structures were still mostly in the form of planar types in which FSS screens can be fabricated conveniently with PCB technologies.

With the advent of computer technology both in computational speed and larger memory storage, partial-differential equation-based solvers, e.g., HFSS, become popular tools in analyzing FSSs. The time-consuming computation of the periodic Green's function in the integral equation formulation can be effortlessly replaced by enforcing the periodic boundary conditions in the FEM formulation. The use of 3D unit-cell elements in FSS greatly enhances the design flexibility. The focus of FSS research gradually shifted from developing fast numerical methods to the applications of commercial solvers for new FSS structures.

The solutions to periodic structure can be exploited to design array antennas as well as switchable and tunable arrays when active devices are incorporated. With the convenience of commercial solvers, new unit-cell geometries can be designed to meet new challenges in array antennas. Some more recent research developments on FSS in the past decade have also been included.

A key function of the FSSs is used as a spatial filter. The design of FSS filter is not different from that of a filter in microwave circuits. A good filter needs to have small insertion loss and sharp roll-off at the band

skirts. In this book chapter, introduction of transmission zeros in the spectral response of the FSS to enhance the selectivity of the FSS filter has been presented. Higher-order filters are also discussed.

The increasing popularity of 3D printing allows us to fabricate FSS with 3D unit cells conveniently. The effective use of 3D printing has been demonstrated to construct 3D FSSs that cannot be easily realized otherwise. The spectral responses of the fabricated 3D FSSs compared favorably with simulated results even in the millimeter-wave frequency band. The uncertainty of the dielectric constant created some sight shift in the resonant frequency.

The transmission and reflection phases of the FSS can be exploited for array antenna designs. In this chapter, the transmission phase is used for a lens antenna design. There is only 1-dBi difference in the simulated and measured peak gains. The difference can be attributed to the fabrication tolerance or measurement alignment but also the accuracy in determining the dielectric constant and loss tangent of the 3D printing materials.

FSSs with double-resonance and four-resonance unit cells are discussed for which their reflection phase covers a wide range of angles and linear phase-frequency curves can be obtained. These phase curves have been successfully used to design wideband and dual-band reflectarrays. The beam tilt in the E plane due to frequency shift from the center frequency can be exploited to design frequency-scanning reflectarrays (Qu et al. 2014b).

The emerging fields of THz science and technology call for novel research activities and technological developments such as THz tunable sources, filters, antennas, and detectors, and it is envisaged that researches on FSS will continue to expand into THz regime.

Acknowledgement

The authors would like to express their gratitude to H. Yi of the University of Electronic Science and Technology of China and D. Q. Liu, P. Zhao, and K. B. Ng of the City University of Hong Kong for their contributions in the design, simulation, and measurement of some of the frequency selective surfaces, lens antennas, and reflectarrays illustrated in this chapter. This work was supported partly by a GRF grant from the Hong Kong Research Grants Council Project No. CityU 110713 and partly by the Natural Science Foundation of China (NSFC) Project No. 61371051.

Cross-References

► Lens Antennas
► Reflectarray Antennas

References

Al-Joumayly M, Behdad N (2009) A new technique for design of low-profile, second-order, bandpass frequency selective surfaces. IEEE Trans Antennas Propag 57:452–459

Barton JH, Rumpf RC, Smith RW, Kozikowski CL, Zellner PA (2012) All-dielectric frequency selective surfaces with few number of periods. Prog Electromagn Res B 41:269–283

Barton JH, Garcia CR, Berry EA, May RG, Gray DT, Rumpf RC (2014) All-dielectric frequency selective surface for high power microwaves. IEEE Trans Antennas Propag 62:3652–3656

DOI 10.1007/978-981-4560-75-7_23-1

Bayatpur F, Sarabandi K (2008) Multipole spatial filters using metamaterial-based miniaturized-element frequency selective surfaces. IEEE Trans Microw Theory Tech 56:2742–2747

Behdad N, Al-Joumayly M, Salehi M (2009) A low profile third-order bandpass frequency selective surface. IEEE Trans Antennas Propag 57:460–466

Bossard JA, Werner DH, Mayer TS, Drupp RP (2005) A novel design methodology for reconfigurable frequency selective surfaces using genetic algorithms. IEEE Trans Antennas Propag 53:1390–1400

Bossard JA, Werner DH, Mayer TS, Smith JA, Tang YU, Drupp RP, Li L (2006) The design and fabrication of planar multiband metallodielectric frequency selective surfaces for infrared applications. IEEE Trans Antennas Propag 54:1265–1276

Bushbeck MD, Chan CH (1993) A tuneable, switchable dielectric grating. IEEE Microw Guid Wave Lett 3:296–298

Capolino F (2009) Theory and phenomena of metamaterials, CRC Press. Boca Raton, Florida

Chan CH, Mittra R (1990) On the analysis of frequency selective surfaces using subdomain basis functions. IEEE Trans Antennas Propag 38:40–50

Chan CH, Tardy I, Yee JS (1992) Analysis of three closely coupled frequency selective surface. Arch Elek Ubertragung 46:321–327

Chang CL, Wang WC, Lin HR, Hsieh FJ, Pun YB, Chan CH (2013) Tunable terahertz fishnet metamaterial. Appl Phys Lett 102:151903–151903-4

Chen CC (1970) Transmission through a conducting screen perforated periodically with apertures. IEEE Trans Microw Theory Tech 18:627–632

Chen QY, Qu SW, Zhang XQ, Xia MY (2012) Low-profile wideband reflectarray by novel elements with linear phase response. IEEE Antennas Wirel Propag Lett 11:1545–1547

Chen QY, Qu SW, Li JF, Chen Q, Xia MY (2013) An X-band reflectarray with novel elements and enhanced bandwidth. IEEE Antennas Wirel Propag Lett 12:317–320

Chin SK, Nicorovici NA, Mcphedran RC (1994) Green's function and lattice sums for electromagnetic scattering by a square array of cylinders. Phys Rev E Stat Phys Plasmas Fluids Relat 49:4590–4602

Chiu CN, Chang KP (2009) A novel miniaturized-element frequency selective surface having a stable resonance. IEEE Antennas Wirel Propag Lett 8:1175–1177

Chow YL, Yang JJ, Fang DG, Howard GE (1991) A closed-form spatial Green's function for the thick microstrip substrate. IEEE Trans Microw Theory Tech 39:588–592

Chu QX, Wang H (2008) A compact open-loop filter with mixed electric and magnetic coupling. IEEE Trans Microw Theory Tech 56:431–439

Cwik TA, Mittra R (1987) Scattering from a periodic array of free-standing arbitrarily shaped perfectly conducting or resistive patches. IEEE Trans Antennas Propag 35:1226–1234

Lima AC deC, Parker EA (1996) Fabry-Perot approach to the design of double layer FSS. Proc Inst Elect Eng Microw Antennas Propag 143:157–162

Dickie R, Cahill R, Gamble HS, Fusco VF, Schuchinsky A, Grant N (2005) Spatial demultiplexing in the sub-mmwave band using multilayer free-standing frequency selective surfaces. IEEE Trans Antennas Propag 53:1903–1911

Dickie R, Cahill R, Gamble HS, Fusco VF, Henry M, Oldfield ML, Huggard PG, Howard P, Grant N, Munro Y, de Maagt P (2009) Submillimeter wave frequency selective surface with polarization independent spectral responses. IEEE Trans Antennas Propag 57:1985–1994

Dickie R, Cahill R, Fusco VF, Gamble HS, Mitchell N (2014) THz frequency selective surface filters for earth observation remote sensing instruments. IEEE Trans THz Sci Technol 1:450–461

Eibert TF, Volakis JL, Wilton DR, Jackson DR (1999) Hybrid FE/BI modeling of 3-D doubly periodic structures utilizing triangular prismatic elements and an MPIE formulation accelerated by the Ewald transformation. IEEE Trans Antennas Propag 47:843–850

DOI 10.1007/978-981-4560-75-7_23-1

Engheta N, Ziolkowski R (2006) Metamaterials: physics and engineering explorations. Wiley-Interscience. Hoboken, New Jersey

Epp L, Chan CH, Mittra R (1992) Periodic structures with time-varying nonlinear loads. IEEE Trans Antennas Propag 40:251–256

Foroozesh A, Shafai L (2010) Investigation into the effects of the patch-type FSS superstrate on the high-gain cavity resonance antenna design. IEEE Trans Antennas Propag 58:258–270

Genovesi S, Costa F, Monorchio A (2012) Low-profile array with reduced radar cross section by using hybrid frequency selective surfaces. IEEE Trans Antennas Propag 60:2327–2335

Hamdy SMA, Parker EA (1982) Current distribution on the elements of a square loop frequency selective surface. Electron Lett 18:624–626

Harms P, Mittra R, Ko W (1994) Implementation of the periodic boundary condition in the finite-difference time-domain algorithm for FSS structures. IEEE Trans Antennas Propag 42:1317–1324

Huang J, Encinar JA (2007) Reflectarray antennas. Wiley-IEEE Press. Hoboken, New Jersey

Jazi MN, Denidni TA (2010) Frequency selective surfaces and their applications for nimble-radiation pattern antennas. IEEE Trans Antennas Propag 58:2227–2237

Jazi MN, Denidni TA (2013) Electronically sweeping-beam antenna using a new cylindrical frequency-selective surface. IEEE Trans Antennas Propag 61:666–676

Jordan KE, Richter GR, Sheng P (1986) An efficient numerical evaluation of the Green's function for the Helmholtz operator on periodic structures. J Comput Phys 63:222–235

Jorgenson RE, Mittra R (1990) Efficient calculation of the free-space periodic Green's function. IEEE Trans Antennas Propag 38:633–642

Jorgenson RE, Mittra R (1991) Scattering from structured slabs having two-dimensional periodicity. IEEE Trans Antennas Propag 39:151–156

Kieburtz RB, Ishimaru A (1961) Scattering by a periodically aperture conducting screen. IRE Trans Antennas Propag 9:506–514

Kipp RA, Chan CH (1994) A numerically efficient technique for the method of moments solution to planar periodic structures in layered media. IEEE Trans Microw Theory Tech 42:635–643

Lee SW (1971) Scattering by dielectric-loaded screen. IEEE Trans Antennas Propag 19:656–665

Li MJ, Luk KM (2014) A wideband circularly polarized antenna for microwave and millimeter-wave applications. IEEE Trans Antennas Propag 62:1872–1879

Liu HL, Ford KL, Langley RJ (2009) Design methodology for a miniaturized frequency selective surface using lumped reactive components. IEEE Trans Antennas Propag 57:2732–2738

Luo GQ, Hong W, Hao ZC, Liu B, Li WD, Chen JX, Zhou HX, Wu K (2005) Theory and experiment of novel frequency selective surface based on substrate integrated waveguide technology. IEEE Trans Antennas Propag 53:4035–4043

Luo GQ, Hong W, Tang HJ, Wu K (2006) High performance frequency selective surface using cascading substrate integrated waveguide cavities. IEEE Microw Wirel Components Lett 16:648–650

Luo GQ, Hong W, Lai QH, Wu K, Sun LL (2007) Design and experimental verification of compact frequency-selective surface with quasi-elliptic bandpass response. IEEE Trans Microw Theory Tech 55:2481–2487

Luo GQ, Hong W, Lai QH, Sun LL (2008) Frequency-selective surfaces with two sharp sidebands realized by cascading and shunting substrate integrated waveguide cavities. IET Microw. Antennas Propag 2:23–27.

Ma K, Ma JG, Yeo KS, Do MA (2006) A compact size coupling controllable filter with separate electric and magnetic coupling paths. IEEE Trans Microw Theory Tech 54:1113–1119

Mittra R, Hall RC, Tsao CH (1984) Spectral-domain analysis of circular patch frequency selective surfaces. IEEE Trans Antennas Propag 32:533–536

Mittra R, Chan CH, Cwik T (1988) Techniques for analyzing frequency selective surfaces-a review. IEEE Proc 76:1593–1615

Munk BA (2000) Frequency selective surfaces: theory and design, Wiley, New York. ISBN: 978-0-471-37047-5, Apr

Ohira M, Deguchi H, Tsuji M, Shigesawa H (2004) Multiband single-layer frequency selective surface designed by combination of genetic algorithm and geometry-refinement technique. IEEE Trans Antennas Propag 52:2925–2931

Ohira M, Deguchi H, Tsuji M, Shigesawa H (2005) Novel waveguide filters with multiple attenuation poles using dual-behavior resonance of frequency-selective surfaces. IEEE Trans Antennas Propag 53:3320–3326

Ott RH, Kouyoumjian RG, Peters L Jr (1967) Scattering by a two-dimensional periodic array of narrow plates. Radio Sci 2:1347–1359

Parker EA, Hamdy SMA (1981) Rings as elements for frequency selective surfaces. Electron Lett 17:612–614

Pelton EL, Munk BA (1979) Scattering from periodic arrays of crossed dipoles. IEEE Trans Antennas Propag 27:323–330

Pous R, Pozar DM (1991) A frequency-selective surface using aperture couples microstrip patches. IEEE Trans Antennas Propag 39:1763–1769

Qu SW, Chen QY, Xia MY, Zhang XY (2014a) Single-layer dual-band reflectarray with single linear polarization. IEEE Trans Antennas Propag 62:199–205

Qu SW, Wu WW, Ng KB, Chen BJ,Chan C H, Pun EYB (2014b) Wideband terahertz reflectarrays with fixed/frequency-scanning beams. In: 2014 XXXIth URSI General Assembly and Scientific Symposium, Beijing, China

Rahmat-Samii Y, Michelssen E (eds) (1999) Electromagnetic optimization by genetic algorithms. Wiley-Interscience, New York

Rashid AK, Shen ZX (2010) A novel band-reject frequency selective surface with pseudo-elliptic response. IEEE Trans Antennas Propag 58:1220–1226

Rashid AK, Shen ZX (2011) Scattering by a two-dimensional periodic array of vertically placed microstrip lines. IEEE Trans Antennas Propag 59:2599–2606

Rashid AK, Shen ZX, Li B (2012) An elliptical bandpass frequency selective structure based on microstrip lines. IEEE Trans Antennas Propag 60:4661–4669

Rashid AK, Li B, Shen Z (2014) An overview of three-dimensional frequency-selective structures. IEEE Trans Antennas Propag 56:43–67

Roden JA, Gedney SD, Kesler MP, Maloney JG, Harms PH (1998) Time-domain analysis of periodic structures at oblique incidence: orthogonal and nonorthogonal FDTD implementations. IEEE Trans Microw Theory Tech 46:420–427

Rubin BJ, Bertoni HL (1983) Reflection from a periodically perforated plane using a subsectional current approximation. IEEE Trans Antennas Propag 31:829–836

Sanz-Izquierdo B, Parker EA (2014) Dual polarized reconfigurable frequency selective surfaces. IEEE Trans Antennas Propag 62:764–771

Sanz-Izquierdo B, Parker EA, Batchelor JC (2010) Dual-band tunable screen using complementary split ring resonators. IEEE Trans Antennas Propag 58:3761–3765

Sanz-Izquierdo B, Parker EA, Batchelor JC (2011) Switchable frequency selective slot arrays. IEEE Trans Antennas Propag 59:2728–2731

Sarabandi K, Behdad N (2007) A frequency selective surface with miniaturized elements. IEEE Trans Antennas Propag 55:1239–1245

Schimert TR, Koch ME, Chan CH (1990) Analysis of scattering from frequency-selective surfaces in the infrared. J Opt Soc Am A 7:1545–1553

Schimert TR, Brouns AJ, Chan CH, Mittra R (1991) Investigation of millimeter-wave scattering from frequency selective surface. IEEE Trans Microw Theory Tech 39:315–322

Shi Y, Chan CH (2010) MLGFIM analysis of 3-D frequency selective structures using volume/surface integral equation. J Opt Soc Am A 27:308–318

Singh S, Singh R (1990) On the use of Shanks' transform to accelerate the summation of slowly converging series. IEEE Trans Microw Theory Tech 39:608–610

Stroke GW (1967) Diffraction gratings. Encycl Phys 5/29:426–754

Tsai FCE, Bialkowski ME (2003) Designing of a 161-element Ku-band microstrip reflectarray of variable size patches using an equivalent unit cell waveguide approach. IEEE Trans Antennas Propag 51:2953–2962

Tsao CH, Mittra R (1984) Spectral-domain analysis of frequency selective surfaces comprised of periodic arrays of cross dipoles and Jerusalem crosses. IEEE Trans Antennas Propag 32:478–486

Vardaxoglou JC (1977) Frequency selective surface: analysis and design, Research Studies Press, Taunton, England, June

Veysoglu ME, Shin RT, Kong JA (1993) A finite-difference time-domain analysis of wave scattering from periodic surfaces: oblique incidence case. J Electron Waves Appl 7:1595–1607

Wood RW (1902) On a remarkable case of uneven distribution of light in a diffraction grating spectrum. Philos Mag 4:396–402

Wu TK (1995) Frequency selective surface and grid array. Wiley-Interscience, New York. ISBN ISBN-13: 978-0471311898

Wu WW, Qu SW, Zhang XQ (2014) Single-layer reflectarray with novel elements for wideband applications. Microw Opt Technol Lett 56:950–954

Xu RR, Zhao HC, Zong ZY, Wu W (2008) Dual-band capacitive loaded frequency selective surfaces with close band spacing. IEEE Microw Wirel Components Lett 18:782–784

Yan M, Qu S, Wang J, Zhang J, Zhang A, Xia S, Wang W (2014) A novel miniaturized frequency selective surface with stable resonance. IEEE Antennas Wirel Propag Lett 13:639–641

Yang HYD, Diaz R, Alexopoulos NG (1997) Reflection and transmission of waves from multilayer structures with planar implanted periodic material blocks. J Opt Soc Am B 14:2513–2521

Yi H, Qu SW, Ng KB, Chan CH (2014) 3-D printed discrete dielectric lens antennas with matching layer. ISAP, Kaohsiung

Yu YX, Chan CH (2000) Efficient hybrid spatial and spectral techniques in analyzing planar periodic structures with non-uniform discretizations. IEEE Trans Microw Theory Tech 48:1623–1627

Yu YM, Chiu CN, Chiou YP, Wu TL (2014) A novel 2.5-dimensional ultraminiaturized-element frequency selective surface. IEEE Trans Antennas Propag 62:3657–3663

Zhang YL, Hong W, Wu K, Chen JX, Tang HJ (2005) Novel substrate integrated waveguide cavity filter with defected ground structure. IEEE Trans Microw Theory Tech 53:1280–1287

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_25-1

Linear Wire Antennas

Kazuhiro Hirasawa*
Institute of Information Sciences and Electronics, University of Tsukuba, Tsukuba, Ibaraki, Japan

Abstract

This chapter shows the basic characteristics of a linear wire antenna in time-harmonic electromagnetic fields. The application of the method of moments is explained briefly to obtain the current distribution on the antenna. For a transmitting antenna, input impedances, current distributions, and radiation patterns are shown for four typical antenna lengths. For a receiving antenna, current distributions and reradiation patterns are shown when the complex conjugate of the input impedance is loaded at the receiving point. The received and reradiated power of the loaded receiving antenna is shown for three different plane-wave incident angles. Also the application of the Thevenin equivalent circuit is discussed for the calculation of the received power. Characteristics of a Yagi-Uda antenna are shown where the optimum gain of a reactance-loaded Yagi-Uda antenna is compared with that of an end-fire array.

Keywords

The method of moments; Linear wire antenna; Transmitting antenna; Receiving antenna; Thevenin equivalent circuit; Received power; Reradiated power; Scattered power; Reactance loading; Yagi-Uda antenna; End-fire array; Optimization; Biquadratic programming method

Introduction

In this chapter time-harmonic electromagnetic fields are only considered. In section "History," the brief history of a linear dipole antenna development and analysis is introduced. The application of the method of moments (MoM) to a linear wire antenna is explained in section "Analysis" where an electric-field boundary value problem is solved to obtain the current distribution on the transmitting antenna. In section "Transmitting Dipole Antenna," transmitting antenna characteristics such as input impedances, current distributions, and radiation patterns are shown for some antenna lengths. In section "Monopole Antenna," the equivalence between a monopole antenna on a large ground plane and a dipole antenna is explained. In section "Receiving Dipole Antenna," an MoM treatment of a receiving antenna with a plane-wave incidence is shown where the input power, received power, and reradiated power are derived from the incident voltage and the induced current on the antenna. Also the application of the MoM and the Thevenin equivalent circuit is discussed for a receiving dipole antenna. In section "Received Power and Reradiated Power," the received power and the reradiated power are shown for a receiving antenna when the complex conjugate of the input impedance is loaded. In section "Yagi-Uda Antenna," characteristics of a Yagi-Uda antenna are presented where the optimum gain of a reactance-loaded Yagi-Uda antenna is compared with that of an end-fire array.

*Email: hirasawa@ieee.org

DOI 10.1007/978-981-4560-75-7_25-1

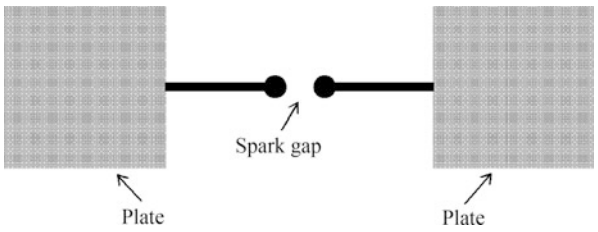


Fig. 1 A plate-loaded transmitting dipole antenna for Hertz's experiment

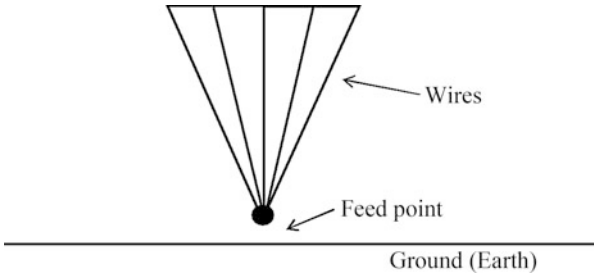


Fig. 2 A transmitting monopole antenna on the earth for Marconi's transatlantic wireless communication

History

In 1886 H. R. Hertz (Hertz 1962; Weeks 1968) used a plate-loaded dipole antenna (about 1.5 m long) shown in Fig. 1 to transmit sparked electromagnetic waves in his experiment and showed the existence of electromagnetic waves predicted in 1864 by J. C. Maxwell (1873). In 1901 G. Marconi built a fan-type monopole antenna (about 45 m high) on the earth shown in Fig. 2. It consists of 50 vertical wires to transmit the first transatlantic wireless signals from England to Canada (Weeks 1968). The wires are supported by a horizontal wire with two vertical masts.

In 1906 de Forest invented the first triode vacuum tube that can continuously amplify higher-frequency signals than those used by G. Marconi. Since then a half-wavelength dipole antenna has been used for wireless communications. A Yagi-Uda antenna consisting of wires of about a half-wavelength is invented by H. Yagi (1928) and S. Uda (1926).

E. Hallén, L. V. King, H. C. Pocklington, and R. W. P. King are early contributors to calculate the current distribution on a linear wire antenna (Pocklington 1897; King 1937, 1956; Hallén 1938). The most difficult part of the antenna analysis is to find the current distribution.

With the development of computers, frequency-domain numerical analysis methods such as the method of moments (MoM) (Harrington 1993) have been widely used for the analysis of various antennas (Hirasawa and Haneishi 1991; Balanis 2005; Stutzman and Thiele 2012; Kraus and Marhefka 2002).

Analysis

In this chapter time-harmonic electric and magnetic fields $\boldsymbol{E}$ and $\boldsymbol{H}$ are only considered and $e^{j\omega t}$ is assumed where $\omega = 2\pi f$ (f, operating frequency). A transmitting straight wire antenna of length h and radius a in free space (ε_0, μ_0) is considered as shown in Fig. 3. It is assumed that the antenna is lossless, $a = 0.001\lambda$ (λ is a free-space wavelength), and the feed gap is infinitesimal. Voltage V applied at the feed point produces the current on the antenna. Once the current is known, it is easy to calculate the antenna

DOI 10.1007/978-981-4560-75-7_25-1

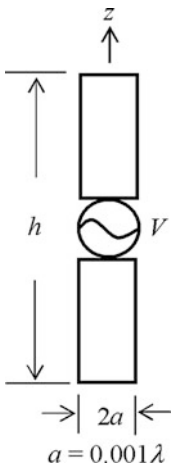


Fig. 3 A dipole antenna on the z-axis with length h, radius a, and feed voltage V

characteristics such as input impedance and gain. A FORTRAN program (Hirasawa and Haneishi 1991) may be used to obtain transmitting antenna characteristics shown in this chapter.

An equivalent theorem (Harrington 2001) is used to obtain the current distribution on a linear wire antenna. In Fig. 4, a closed surface C is assumed just outside the antenna and the equivalent problem is considered where $\boldsymbol{E}$ and $\boldsymbol{H}$ outside C are the same as the original problem and $\boldsymbol{E}$ and $\boldsymbol{H}$ inside C are forced to be zero. This situation is realized by using impressed electric field $\boldsymbol{E}^{imp}$ and equivalent electric current density $\boldsymbol{J}_s$

$$\boldsymbol{J}_s = \hat{n} \times \boldsymbol{H} \tag{1}$$

on C. The unit vector $\hat{\mathbf{n}}$ is normal and outward to C. In Eq. 1, $\boldsymbol{H}$ is a magnetic field on C and $\boldsymbol{E}^{imp}$ on C is to produce voltage V at the feed gap of the antenna. In the equivalent problem, the electric field inside C is forced to be zero and conductors such as antennas inside C can be deleted. Therefore, $\boldsymbol{J}_s$ can be in free space and the free-space Green's function can be used to solve the boundary value problem to obtain $\boldsymbol{J}_s$ with known $\boldsymbol{E}^{imp}$.

In the equivalent problem shown in Fig. 4, the electric field just inside C becomes

$$E_z(J_z) + E_z^{imp} = 0 \tag{2}$$

where it is assumed that J_z is an equivalent line current and E_z^{imp} is constant just inside and outside C. On the other hand, just outside C the boundary condition becomes

$$E_z^{imp} = \begin{cases} \dfrac{V}{d_0} \cdots \text{feed} \\ 0 \cdots \text{elsewhere} \end{cases} \tag{3}$$

In Eqs. 2 and 3, the boundary condition is only considered on the z-component of the electric fields due to the thin wire assumption ($a = 0.001\lambda$). Also the impressed electric field E_z^{imp} is nonzero just outside the feed point, V is a known feed voltage, and d_0 is the distance of an infinitesimal feed gap.

Next an antenna with one impedance load Z_L is considered as shown in Fig. 5. The total electric field $E_z + E_z^{imp}$ becomes zero inside C as Eq. 2, but at the infinitesimal loading point $E_z + E_z^{imp}$ is equal to E_z^L due to Z_L:

DOI 10.1007/978-981-4560-75-7_25-1

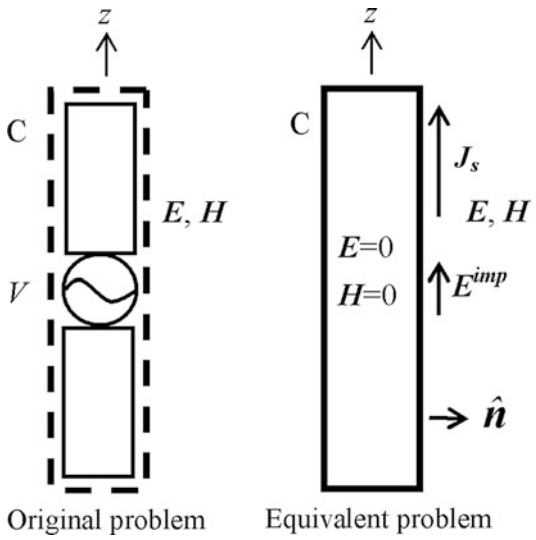


Fig. 4 An original and the equivalent problem: both have the same electromagnetic fields outside region *C*

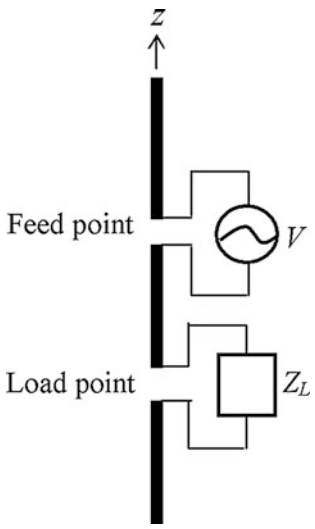


Fig. 5 A transmitting antenna with feed voltage *V* and load Z_L

$$E_z(J_z) + E_z^{imp} = \begin{cases} E_z^L \cdots \text{load} \\ 0 \cdots \text{elsewhere} \end{cases} \tag{4}$$

Before the MoM is applied to the boundary conditions (Eqs. 2, 3, and 4), line current J_z with unknown I_n is assumed as

$$J_z(z') = \sum_{n=1}^{N} I_n J_{zn}(z') \tag{5}$$

where N is the number of the expansion functions and the nth expansion function is

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_25-1

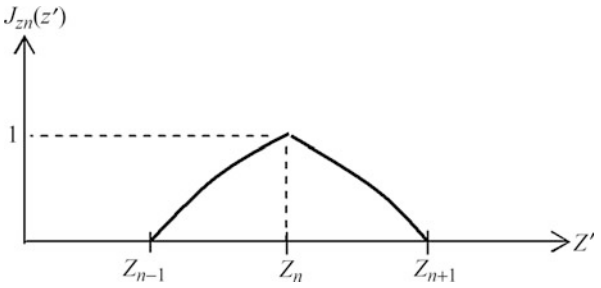


Fig. 6 An expansion function for the current on the wire

$$J_{zn}(z') = \begin{cases} \dfrac{\sin k(z' - z_{n-1})}{\sin k(z_n - z_{n-1})} \cdots z_{n-1} \le z' \le z_n \\ \dfrac{\sin k(z_{n+1} - z')}{\sin k(z_{n+1} - z_n)} \cdots z_n \le z' \le z_{n+1} \\ 0 \ldots\ldots\ldots\ldots\ldots \text{elsewhere} \end{cases} \tag{6}$$

and $\Delta z = h/(N+1), z_{n+1} - z_n = z_n - z_{n-1} = \Delta z, k = 2\pi/\lambda$ (Fig. 6) (Stutzman and Thiele 2012). Current coefficient I_n is a complex number and is determined to satisfy the boundary conditions in Eqs. 2, 3, and 4.

For current J_{zn}, vector potential $\boldsymbol{A_n}(x,y,z)$ has only a z-component and electric field E_{zn} is expressed as

$$E_{zn} = -j\omega\mu_0 A_{zn} + \frac{1}{j\omega\varepsilon_0} \cdot \frac{\partial^2 A_{zn}}{\partial z^2} \tag{7}$$

$$A_{zn}(x,y,z) = \int_{-\Delta z}^{\Delta z} \frac{J_{zn}(\boldsymbol{r}')e^{-jk|\boldsymbol{r}-\boldsymbol{r}'|}}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|} dz' \tag{8}$$

Position vectors $\boldsymbol{r}(= x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}} + z\hat{\boldsymbol{z}})$ and $\boldsymbol{r}' = (x'\hat{\boldsymbol{x}} + y'\hat{\boldsymbol{y}} + z'\hat{\boldsymbol{z}})$ with unit vectors $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{z}}$ are from the origin to an observation point (x, y, z) on C and to a point (x', y', z') on the line current, respectively. Electric field $E_{zn}(z)$ on the surface of cylinder C with radius a is due to current J_{zn} on the z-axis and becomes

$$E_{zn}(z) = -\frac{j30}{\sin(k\Delta z)}\left[\frac{e^{-jkr_{n-1}}}{r_{n-1}} - \frac{2\cos(k\Delta z)e^{-jkr_n}}{r_n} + \frac{e^{-jkr_{n+1}}}{r_{n+1}}\right] \tag{9}$$

where r_{n-1}, r_n, and r_{n+1} are the distances from points z_{n-1}, z_n, and z_{n+1} on the z'-axis to a point z on the cylinder C with radius a (Fig. 7) (Jordan and Balmain 1968).

An inner product

$$\langle \boldsymbol{E}_n(z), \boldsymbol{J}_w(z) \rangle = \int_{-\Delta z}^{\Delta z} \boldsymbol{E}_n(z) \cdot \boldsymbol{J}_w(z) dz \tag{10}$$

is chosen where $\boldsymbol{E}_n(z)$ is an electric field on C produced by current $J_{zn}(z')$ in Eq. 6. $\boldsymbol{J}_w(z)$ is a weighting function on C.

DOI 10.1007/978-981-4560-75-7_25-1

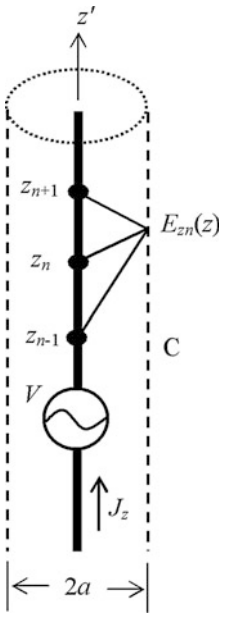


Fig. 7 Electric field $E_{zn}(z)$ on C (radius a) due to line current $J_{zn}(z')$

At first a transmitting antenna without impedance load Z_L in Fig. 3 is considered to obtain the current distribution on the antenna. A weighting function $\boldsymbol{J}_w(z)$ is chosen to be the same as the expansion function in Eq. 6, and the MoM is applied to the boundary condition in Eqs. 2 and 3 with the inner product of Eq. 9. Then, the simultaneous linear equations are derived as

$$\sum_{n=1}^{N} Z_{mn} I_n = V_m \ (m = 1, 2, \cdots, N) \tag{11}$$

where element Z_{mn} of matrix $[Z]$ is a mutual impedance between two small dipoles (section $z_{n-1} - z_{n+1}$ on the z'-axis and section $z_{m-1} - z_{m+1}$ on the surface C). The infinitesimal feed points are z_n and z_m. Mutual impedance Z_{mn} is obtained from Eq. 10 as

$$Z_{mn} = -\int_{z_{m-1}}^{z_m} \frac{\sin k(z - z_{m-1})}{\sin k(z_m - z_{m-1})} E_{zn}(z)dz - \int_{z_m}^{z_{m+1}} \frac{\sin k(z_{m+1} - z)}{\sin k(z_{m+1} - z_m)} E_{zn}(z)dz \tag{12}$$

Equation 12 can be calculated numerically by using a Gauss quadrature (Abramowitz and Stegun 1965), and Eq. 11 can be written in a matrix form as

$$[Z][I] = [V] \tag{13}$$

where impedance matrix $[Z]$ is $N \times N$ and current matrix $[I]$ and feed voltage matrix $[V]$ are $N \times 1$. Unknown current matrix $[I]$ is obtained by solving the simultaneous linear equations.

Matrix $[V]$ has only one nonzero element corresponding to a feed point. The feed point has to be chosen to coincide with the peak of an expansion function (Fig. 6). Then, input impedance Z_{in} becomes

$$Z_{in} = \frac{V}{I} \tag{14}$$

DOI 10.1007/978-981-4560-75-7_25-1

where V and I are the elements of $[V]$ and $[I]$ in Eq. 13 corresponding to the feed point.

For a transmitting antenna with impedance load Z_L shown in Fig. 5, the matrix equation is obtained from the boundary conditions (3) and (4) as

$$[Z][I] = [V] + \left[V^L\right] \tag{15}$$

where Z_L is loaded at the load point corresponding to the peak of an expansion function (Fig. 6). The mth element of load voltage matrix $[V^L](N \times 1)$ has

$$V_m^L = -\int_{z_{m-1}}^{z_{m+1}} E_z^L J_{zm}(z)dz \tag{16}$$

$$= -Z_L I_n (n = m) \tag{17}$$

where I_n is the nth element of the current matrix. From Eqs. 15 and 17, the matrix equation

$$\left[Z + Z^L\right][I] = [V] \tag{18}$$

is derived where load matrix $[Z^L]$ is $N \times N$ and the nth diagonal element Z_{nn}^L is only nonzero. In Eqs. 13 and 18, it is assumed that $[V]$ is known and unknown $[I]$ is obtained numerically.

Electric field E_θ is the only component of the far electric field for the linear dipole (length h) located in the z-axis direction and is obtained from Eq. 7 as

$$E_\theta = j\omega\mu_0 A_z \sin\theta \tag{19}$$

$$A_z = \frac{e^{-jkr}}{4\pi r} \int_{-\frac{h}{2}}^{\frac{h}{2}} J_z(z) e^{jkz\cos\theta} dz \tag{20}$$

Gain G is defined as

$$G = 4\pi r^2 \frac{|E_\theta|^2}{\eta_0 P} \tag{21}$$

In Eq. 21 $\eta_0 = \sqrt{\mu_0/\varepsilon_0}$, r is the distance from the antenna to the far electric field E_θ and input power P of an antenna is expressed by voltage V and current I at the feed point as

$$P = \mathrm{Re}(VI^*) \tag{22}$$

where Re and * denote "real part" and "complex conjugate," respectively. Since E_θ is proportional to $1/r$ as shown in Eq. 20, G is constant with respect to r in the far field. Then, higher G gives higher $|E_\theta|$ with constant P as shown in Eq. 21.

DOI 10.1007/978-981-4560-75-7_25-1

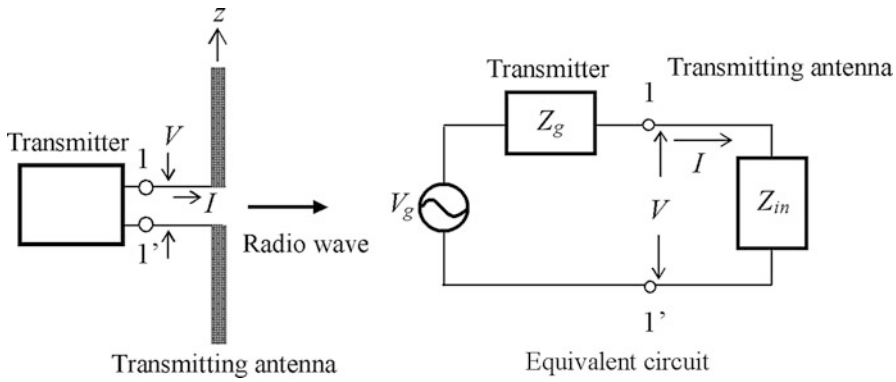


Fig. 8 An antenna with a transmitter and the equivalent circuit where current I is at the feed point

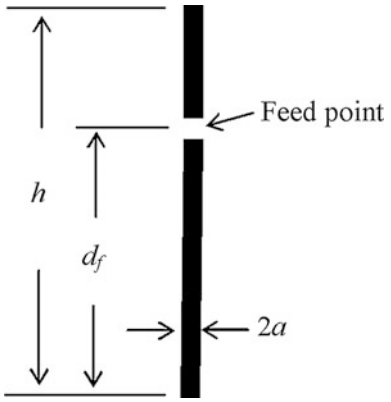


Fig. 9 A dipole antenna with an offset feed

Transmitting Dipole Antenna

Figure 8 shows a transmitting dipole antenna with a transmitter and the equivalent circuit. Equivalent circuits are often used to explain some antenna functions from the circuit point of view without using $\boldsymbol{E}$ and $\boldsymbol{H}$. The left-hand side of 1-1′ is the Thevenin equivalent circuit of the transmitter that can be used to obtain current I flowing terminal 1. Voltage V_g is the open-circuit voltage of the transmitter when output terminals 1-1′ are open circuited. Impedance Z_g is the input impedance of the transmitter when the source switch is off. Then, the complicated transmitter circuit is simply expressed by the voltage source and the impedance if only current I flowing on the antenna is desired without knowing the details of the transmitter. Impedance Z_{in} $(= R_{in} + jX_{in})$ is the input impedance of the antenna, and R_{in} is called the radiation resistance closely related to the radiated power of the antenna. Thus, a transmitting dipole antenna with feed voltage V is considered in this chapter as shown in Fig. 3 and the transmitter is omitted for simplicity. Also it is recognized that voltage V is equal to V_g when $Z_g = 0$.

An offset-feed dipole antenna ($a = 0.001\lambda$) is shown in Fig. 9 to investigate the effects of the position of the feed point. Length d_f is the distance to the feed from the bottom end of the dipole. The input impedance of a half-wavelength dipole antenna is shown in Fig. 10 where d_f is changed from 0 to 0.5λ. R_{in} is very large near the end of the dipole due to the small current and becomes a minimum in the center due to the current maximum. X_{in} becomes capacitive when the feed is near the end due to the large charge accumulation at the end and inductive or the least capacitive at the center due to the current maximum.

DOI 10.1007/978-981-4560-75-7_25-1

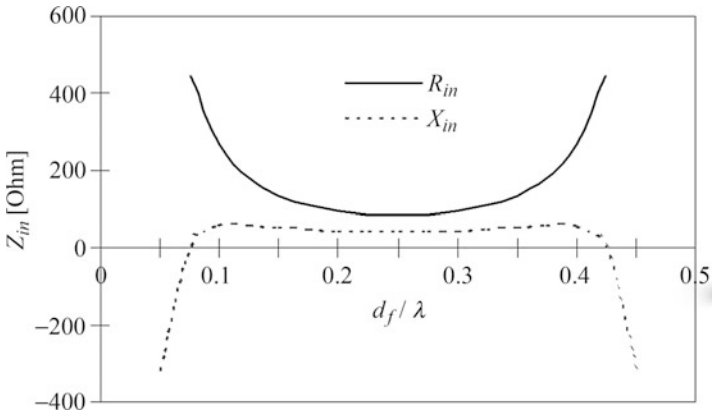


Fig. 10 Input impedance Z_{in} ($= R_{in} + jX_{in}$) of a half-wavelength dipole antenna with an offset feed

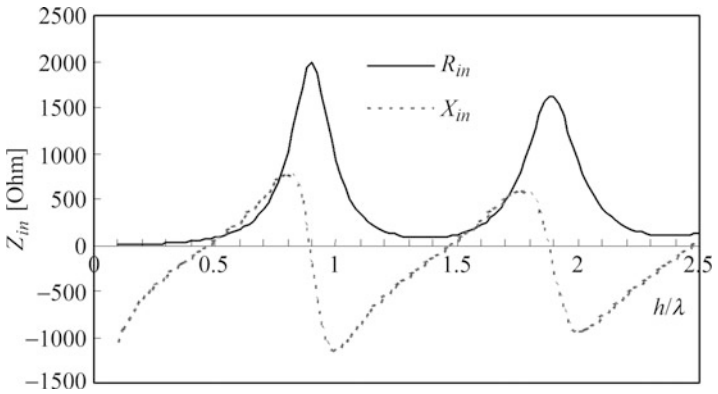


Fig. 11 Input impedances Z_{in} ($= R_{in} + jX_{in}$) of a center-fed dipole antenna of length h

From the impedance matching point of view, it is recognized that the best position of the feed point is near the center for the half-wavelength dipole, and this is one of the reasons why the center-fed half-wavelength dipole is widely used.

Figure 11 shows input impedance Z_{in} of a center-fed dipole antenna ($d_f = h/2$, $a = 0.001\lambda$) with respect to h/λ. The real part R_{in} is maximum around $h = 0.9\lambda$ and 1.9λ. The imaginary part X_{in} is capacitive when h is less than 0.5λ. X_{in} becomes zero around 0.5λ and 0.9λ, 1.5λ, 1.9λ, and 2.5λ. When X_{in} is zero, it is called that the antenna is resonant. There are two kinds of resonance: series resonance (0.5λ, 1.5λ, and 2.5λ) and parallel resonance (0.9λ and 1.9λ). For series resonance, R_{in} is not large, but R_{in} is very large for the parallel resonance, and the dipole antenna is usually used around the series resonance, especially around the first series resonance ($h = 0.5\lambda$). Input resistance R_{in} is 70–100 [Ohm] around $h = 0.5\lambda$ and good impedance matching can be obtained easily with the connected transmission line.

Figure 12 shows current distributions on dipoles ($d_f = h/2$) of $h = 0.5\lambda$, λ, 1.5λ, and 2λ where $V = 1$ [V]. The current amplitude has a peak every half-wavelength. When $h = 0.5\lambda$ and λ, the phase is almost constant except the feed point. When $h = 1.5\lambda$ and 2λ, the phase change is about 180° [deg.] except around the feed and the phase transition.

Figure 13 shows the normalized vertical radiation patterns $|E_\theta|$ corresponding to the currents in Fig. 12. The horizontal radiation patterns $|E_\theta|$ are omnidirectional and are not shown here. When the phase of the current is constant on a dipole antenna such as a half-wavelength or a one-wavelength dipole, there is only one radiation peak, but there are more than one radiation peak when the current phase changes on dipoles

DOI 10.1007/978-981-4560-75-7_25-1

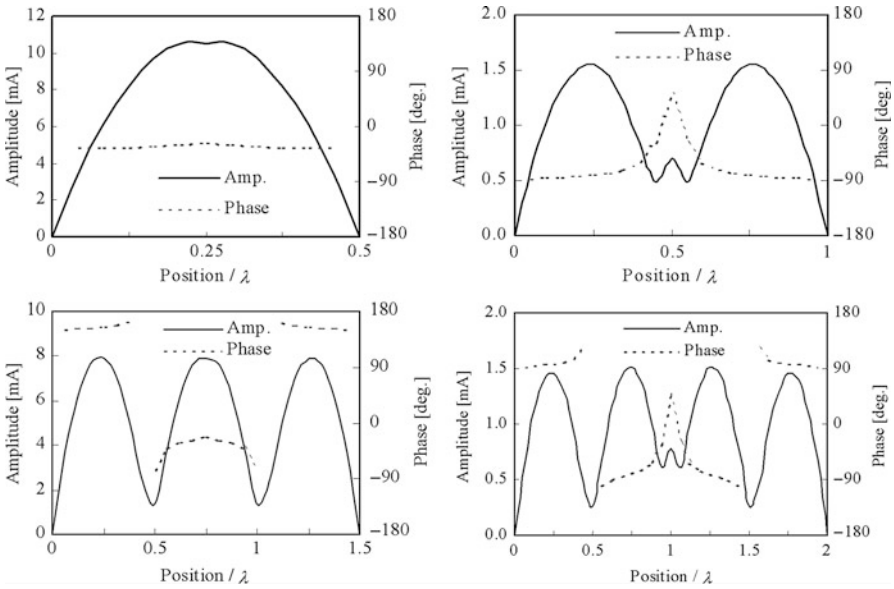


Fig. 12 Currents (amplitude and phase) on center-fed dipole antennas for $h = 0.5\lambda$, 1λ, 1.5λ, and 2λ ($V = 1$[V])

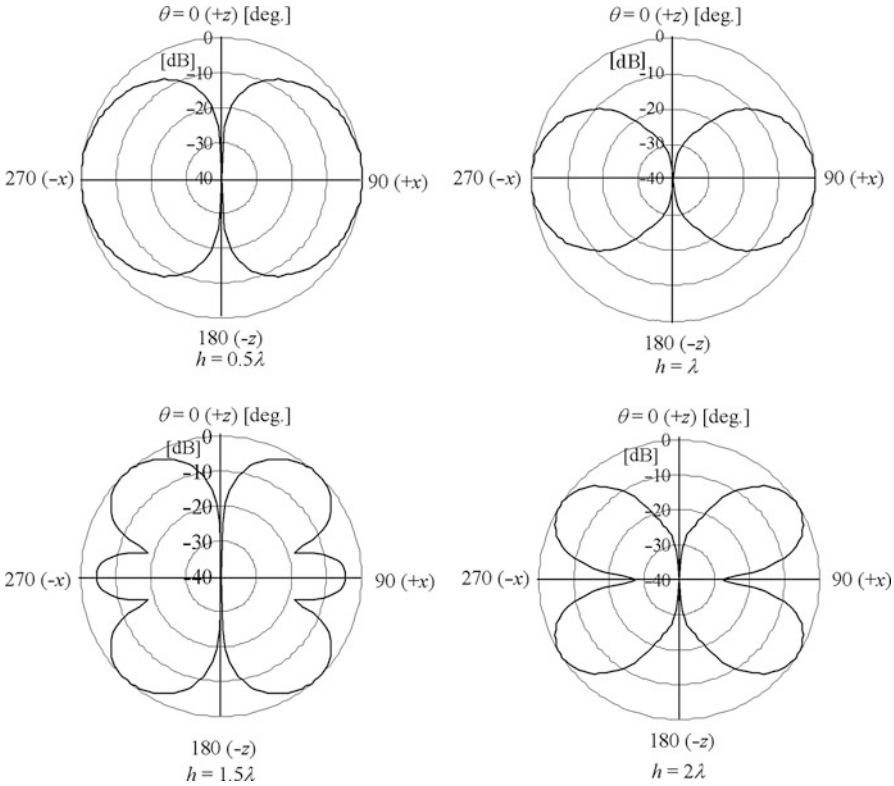


Fig. 13 Normalized vertical radiation patterns of the center-fed dipole antennas corresponding to the currents in Fig. 12

of $h = 1.5\lambda$ and 2λ. Then, the radiated energy is distributed into more than one vertical direction. This causes inefficient radiation in the desired direction and undesired radiation in the other directions. Therefore, a half-wavelength dipole antenna is used widely due to its unidirectional vertical radiation and easy impedance matching.

DOI 10.1007/978-981-4560-75-7_25-1

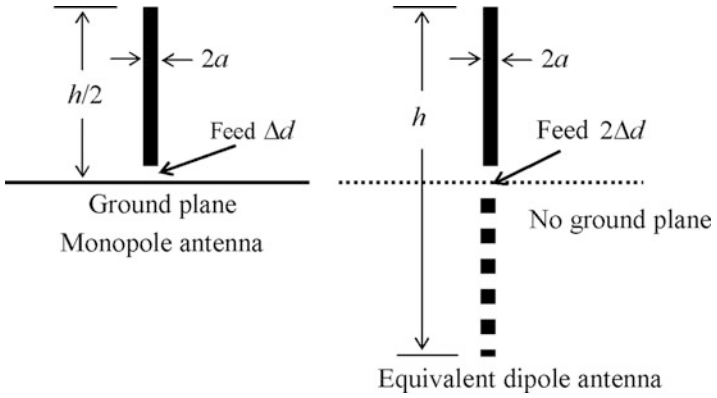


Fig. 14 A monopole antenna on an infinite ground plane and the equivalent dipole antenna

Monopole Antenna

A monopole antenna on a large ground plane such as the earth and building roofs can be used to make antennas smaller without sacrificing communication quality. For the first transatlantic wireless communication, G. Marconi built a monopole antenna (45 m) where the earth is used as a large ground plane (Fig. 2). The operating frequency of the transmitter is under 100 kHz and the antenna is still small in wavelength (Weeks 1968).

A monopole antenna on an infinite ground plane is equivalent to a dipole antenna shown in Fig. 14. The currents on the monopole and the electric field are the same as those of the dipole in the upper half-space. Since the infinitesimal feed distance of the monopole is Δd and that of the dipole is $2\Delta d$, voltage $V^m = E_z^{imp}\Delta d$ at the monopole feed becomes half of $V^d = E_z^{imp}(2\Delta d)$ at the dipole feed where E_z^{imp} is the impressed electric field shown in Fig. 4. Thus, their input impedances Z_{in}^m and Z_{in}^d have the following relation:

$$Z_{in}^m = \frac{2V^m}{2I^m} = \frac{V^d}{2I^d} = \frac{Z_{in}^d}{2} \tag{23}$$

Since $V^m = V^d/2$ and $I^m = I^d$, $P^m = P^d/2$ from Eq. 22. Then, from Eq. 21, the relation between monopole and dipole gain G^m and G^d is obtained as

$$G^m = 2G^d \tag{24}$$

Receiving Dipole Antenna

For a z-directed receiving antenna and the equivalent circuit shown in Fig. 15, it is assumed that a plane wave $\boldsymbol{E}^{imp}$ is incident and impedance Z_L is loaded at the receiving point on the antenna. Also Z_{in} and V_o are the input impedance and the open-circuit voltage of the receiving antenna, respectively. In Figs. 4 and 15, the position of $\boldsymbol{E}^{imp}$ is on the feed point for the transmitting antenna and is outside for the receiving antenna. Therefore, for the receiving antenna, the boundary condition is the same as Eq. 4 and the matrix equation to obtain the current distribution becomes Eq. 18. It is assumed that the receiving point is infinitesimal and corresponds to the peak of the expansion function (Fig. 6).

The incident plane wave is assumed as

DOI 10.1007/978-981-4560-75-7_25-1

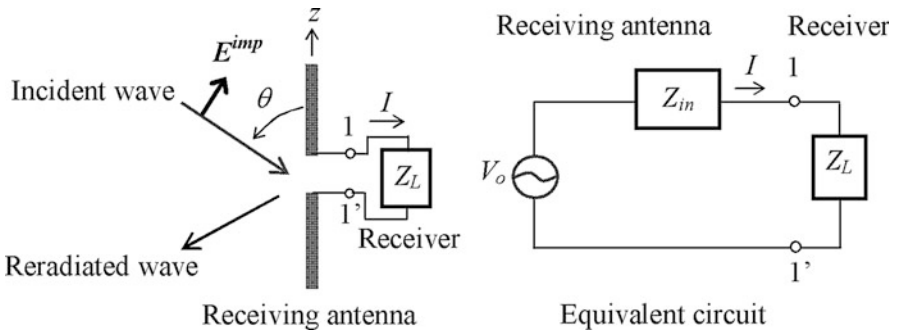


Fig. 15 A receiving antenna with a receiver and the equivalent circuit where current I is at the receiving point

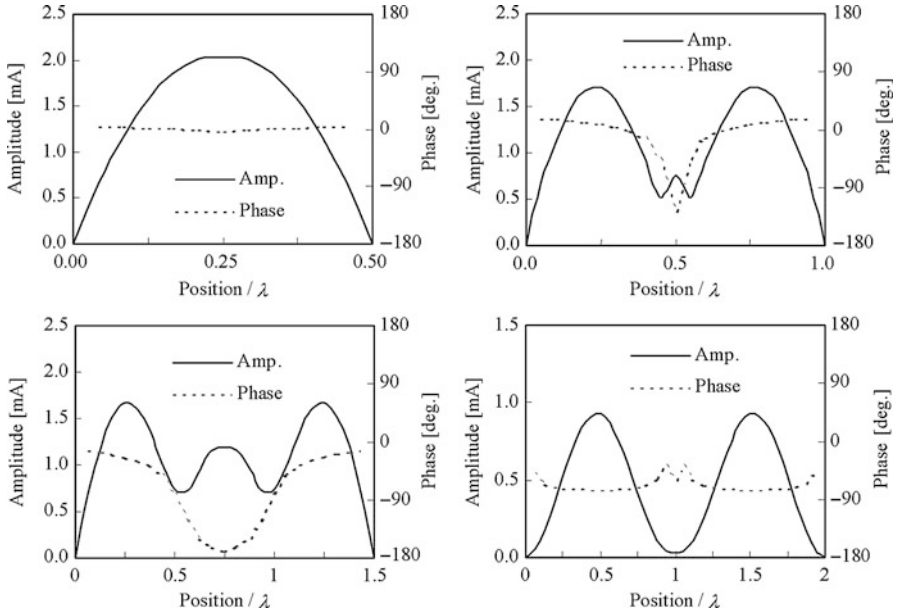


Fig. 16 Currents (amplitude and phase) on center-loaded receiving dipole antennas for $h = 0.5\lambda$, 1λ, 1.5λ, and $2\lambda \left(Z_L = Z_{in}^*, E_\theta{}^{imp} = -1[\mathrm{V/m}], \theta = 90^\circ\right)$

$$\boldsymbol{E}^{\boldsymbol{imp}}(z) = \hat{u}_{\boldsymbol{i}} \boldsymbol{e}^{-j\boldsymbol{k}_i \cdot \hat{z} z} \tag{25}$$

where unit vector $\hat{u}_{\boldsymbol{i}}$ specifies the polarization. Wave-number vector $\boldsymbol{k}_{\boldsymbol{i}}$ is in the direction of the incident wave and $|\boldsymbol{k}_{\boldsymbol{i}}| = 2\pi/\lambda$. Then, the element of incident voltage matrix $[V]$ becomes

$$V_m = \int_{z_{m-1}}^{z_{m+1}} J_{zm} \hat{z} \cdot \hat{u}_{\boldsymbol{i}} \boldsymbol{e}^{-j\boldsymbol{k}_i \cdot \hat{z} z} dz \tag{26}$$

Current distributions on center-loaded receiving dipoles of $h = 0.5\lambda$, λ, 1.5λ, and 2λ are shown in Fig. 16 where it is assumed that electric field $E_\theta{}^{imp}$ $(= -1$ [V/m]) is normally incident ($\theta = 90^\circ$) and load $Z_L = Z_{in}$* at the receiving point. Figure 16 can be compared to the current distributions of the transmitting antennas in Fig. 12. The amplitude peak appears every half-wavelength that is similar to that of the transmitting antenna for $h = 0.5\lambda$, λ, 1.5λ. When $h = 2\lambda$, there are only two amplitude peaks for the receiving case instead of the four peaks for the transmitting one. This is because the current distribution as well as the receiving characteristics is dependent on the incident wave and the load value at the receiving

DOI 10.1007/978-981-4560-75-7_25-1

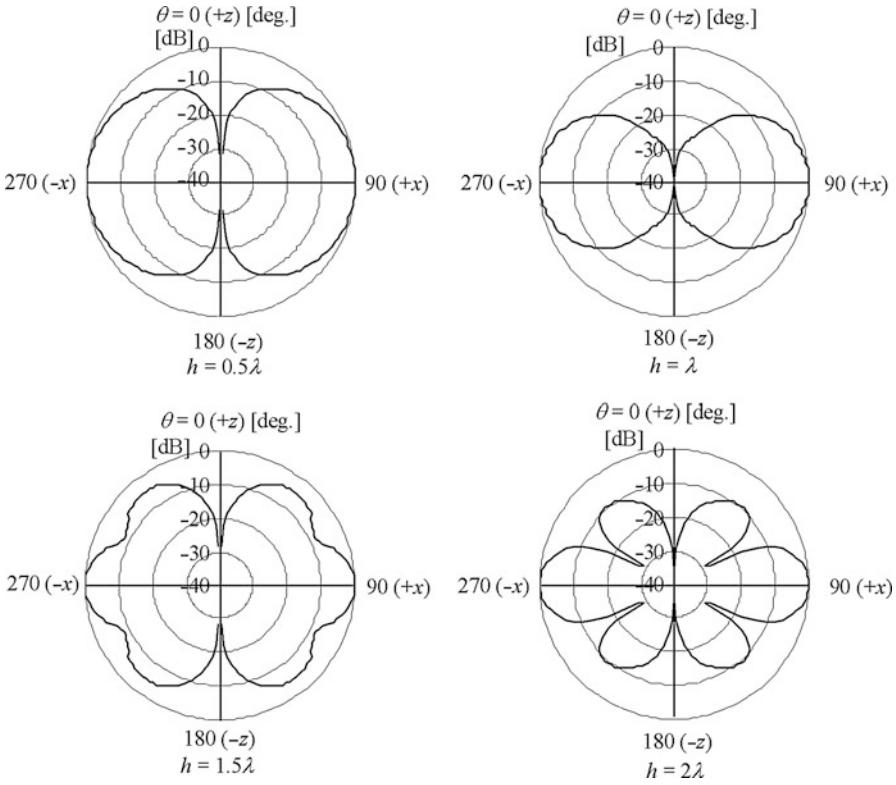


Fig. 17 Normalized vertical reradiation patterns of the center-loaded dipole antennas corresponding to the currents in Fig. 16

point (Hirasawa 1987). Figure 17 shows the normalized vertical reradiation pattern $|E_\theta|$ corresponding to the current in Fig. 16. If the reradiation and radiation patterns in Figs. 17 and 13 are compared, they are similar for $h = 0.5\lambda$ and λ where there is only one peak ($\theta = 90°$). When $h = 1.5\lambda$, the maximum is in the $\theta = 90°$ direction in Fig. 17 instead of $\theta = 45°$ in Fig. 13. Also there are no deep nulls in the reradiation pattern that are quite different from the radiation pattern. When $h = 2\lambda$, there is a peak in the $\theta = 90°$ direction instead of the null in the radiation pattern. The horizontal reradiation patterns $|E_\theta|$ are omnidirectional and are not shown here. Also the receiving and the radiation patterns are the same for an antenna due to the reciprocity theorem (Stutzman and Thiele 2012), and the receiving pattern is not discussed here.

Open-circuit voltage V_o in Fig. 15 can be obtained from short-circuit current I_s by the MoM. The element of $[I]$ corresponding to the receiving point in Eq. 18 becomes I_s when the element of $[Z_L]$ is set to zero:

$$V_o = Z_{in} I_s \tag{27}$$

The Thevenin equivalent circuit on the left-hand side of terminals 1-1′ in Fig. 15 is valid only for the calculation of the current and received power at Z_L (Silver 1949). Therefore, in general, the power consumed at Z_{in} does not show the correct reradiated power. Similarly it is recognized that the electromagnetic fields in Fig. 14 are the same only above the upper half-space for the monopole and the dipole problem. It is important to keep in mind the valid region of the equivalent problem.

The incident, received, and reradiated powers are respectively expressed as

DOI 10.1007/978-981-4560-75-7_25-1

$$P_{in} = \sum_{n=1}^{N} \mathrm{Re}\left(V_n I_n^*\right) \tag{28}$$

$$P_{re} = \mathrm{Re}(Z_L)|I|^2 \tag{29}$$

$$P_{ra} = P_{in} - P_{re} \tag{30}$$

Reradiated power P_{ra} can be also obtained by integrating power density ($|E_\theta|^2/\eta_0$) over a large sphere. In Eq. 29, I is the current at the receiving point where impedance Z_L is loaded. Powers in Eqs. 28–30 are all obtained by using the MoM. Also P_{re} can be obtained correctly by using the Thevenin equivalent circuit shown in Fig. 15 as discussed above.

Received Power and Reradiated Power

Received and reradiated (or scattered) powers on the receiving dipole antenna have been investigated with the relation to the MoM and the Thevenin equivalent circuit since 1994. The reradiated power and the received power of a center-loaded dipole antenna and a Yagi-Uda antenna of load $Z_L = Z_{in}{}^*$ are calculated by the MoM and the limitation of the Thevenin equivalent circuit is shown (Hirasawa et al. 1994, 1997). Since then various discussions have been reported for a receiving antenna on the Thevenin equivalent circuit and the reradiated power (Collin 2003; Onuki et al. 2007; Best and Kaanta 2009).

In this section, incident, received, and reradiated powers with respect to three incident angles $\theta = 90°$, 60°, and 30° are shown for a center-loaded receiving dipole antenna of load $Z_L = Z_{in}{}^*$. With this load impedance, the maximum power is received at the load. It is also assumed that incident electric field $\left|E_\theta{}^{imp}\right| = 1[\mathrm{V/m}]$. Figure 18 shows received power P_{re} with respect to dipole length h. Also received power P_{re} is obtained from the Thevenin equivalent circuit (Fig. 15) once the open-circuit voltage V_o is obtained from Eq. 27. Figure 19 shows reradiated power P_{ra} from Eq. 30 with respect to h. P_{ra} can be also obtained by integrating power density ($|E_\theta|^2/\eta_0$) over a large sphere. In general P_{ra} cannot be calculated from the Thevenin equivalent since the circuit on the left-hand side of terminals 1-1′ in Fig. 15 is equivalent to the original problem only for the calculation of current I and P_{re} (Silver 1949; Collin 2003). Figure 20 shows P_{re}/P_{ra}. The reradiated power becomes larger than the received power as

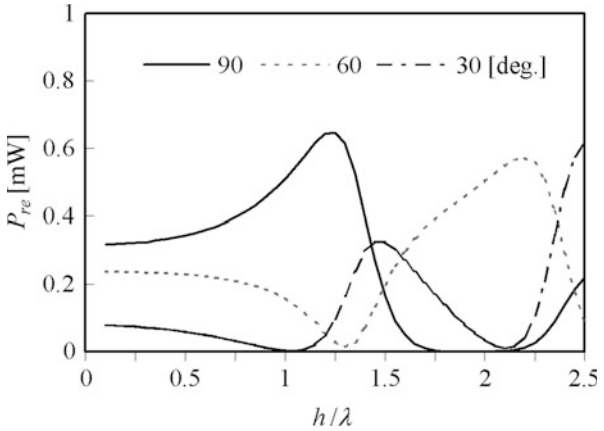


Fig. 18 Received power P_{re} with $\left|E_\theta{}^{imp}\right| = 1[\mathrm{V/m}]$ and $Z_L = Z_{in}{}^*$ for three incident angles $\theta = 90°$, 60°, and 30°

DOI 10.1007/978-981-4560-75-7_25-1

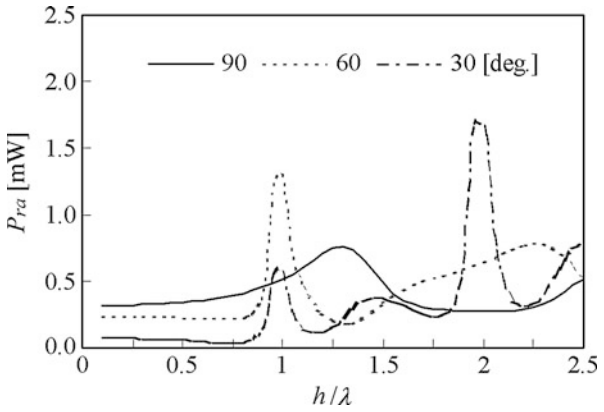


Fig. 19 Reradiated power P_{ra} with $\left|E_\theta^{imp}\right| = 1[\mathrm{V/m}]$ and $Z_L = Z_{in}$* for three incident angles $\theta = 90°$, 60°, and 30°

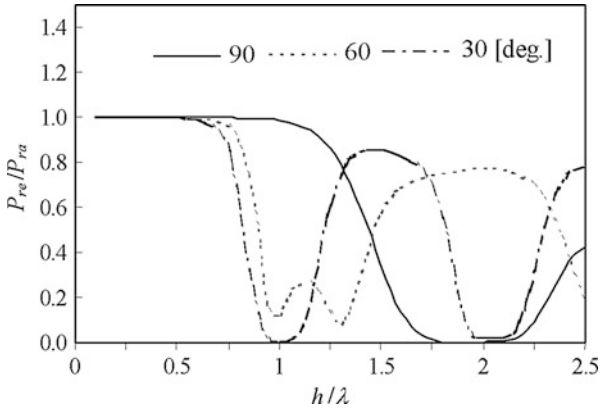


Fig. 20 P_{re}/P_{ra} with $\left|E_\theta^{imp}\right| = 1[\mathrm{V/m}]$ and $Z_L = Z_{in}$* for three incident angles $\theta = 90°$, 60°, and 30°

h becomes larger than 0.9λ, 0.6λ, and 0.55λ for incidence angles $\theta = 90°$, 60°, and 30°, respectively. Figure 21 shows the current distribution (amplitude) for $h = \lambda$ and the current amplitude becomes almost zero at the receiving point for $\theta = 30°$. Thus, the corresponding P_{re} and P_{re}/P_{ra} becomes very small.

Figures 18, 19, and 20 show that the receiving antenna can receive and reradiate 50 % of the incident power at best and receive less than 50 % of the incident power when h is larger than a certain length. Also this can be explained by separating the current distribution on the receiving dipole antenna into the coupling component and the non-coupling component. Then, the former contributes to the received power at the load but the latter only contributes to the reradiated power (Onuki et al. 2007).

Yagi-Uda Antenna

A typical Yagi-Uda antenna of six elements (Yagi 1928; Uda 1926) is shown in Fig. 22. It consists of one feed dipole, a reflector, and a director. The reflector is a wire located behind the feed dipole and is usually a little longer than the feed dipole. The director consists of a few wires in front of the feed dipole, and they are a little shorter than the feed dipole. The wire element spacing is 0.25λ–0.375λ where the mutual coupling is strong between the elements. The reflector and director guide electromagnetic fields efficiently in the x-axis direction. The antenna has a high gain in the x-axis direction with only one feed. The gain

DOI 10.1007/978-981-4560-75-7_25-1

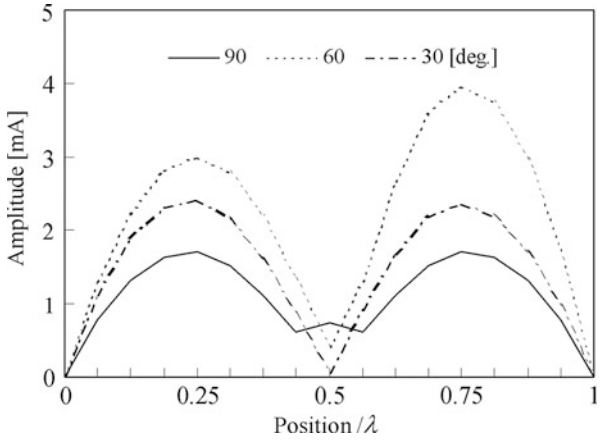


Fig. 21 Current amplitudes on the center-loaded receiving dipole antenna ($h = 1\lambda$) with $\left|E_\theta{}^{imp}\right| = 1[\mathrm{V/m}]$ and $Z_L = Z_{in}{}^*$ for three incident angles $\theta = 90°$, 60°, and 30°

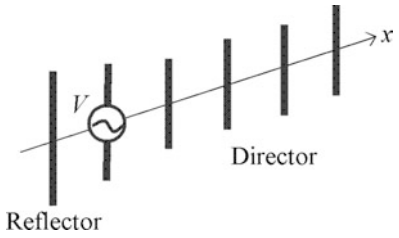


Fig. 22 A six-element Yagi-Uda antenna

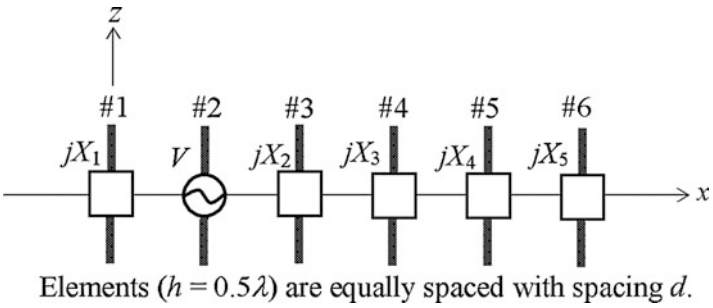


Fig. 23 An equally spaced Yagi-Uda antenna with six half-wavelength elements and five reactance loads

increases gradually with the increase of the number of director wires. Due to the simple structure, it is lightweight and easy to make. Thus, it is widely used as TV and FM receiving antennas in many countries.

Figure 23 shows a reactance-loaded Yagi-Uda antenna where the elements are equally spaced. Each element is a half-wavelength wire with $a = 0.001\lambda$. The reactance loads and the feed are located in the center of the element. The current distribution is changed by loading a reactance component on a wire except the feed wire. In Fig. 24 the optimum gain in the x-axis direction ($\theta = 90°$, $\phi = 0°$) is compared with that of the end-fire array shown in Fig. 25 where all elements are fed. The loaded reactance values for the optimum gain are obtained by using a biquadratic programming method (Hirasawa 1980, 1987, 1988). The difference between the optimum gains by reactance loads and feeds is 0.8 dB for element spacing $d = 0.35\lambda$ as shown in Fig. 24. Figure 26 shows the normalized horizontal radiation patterns ($\theta = 90°$)

DOI 10.1007/978-981-4560-75-7_25-1

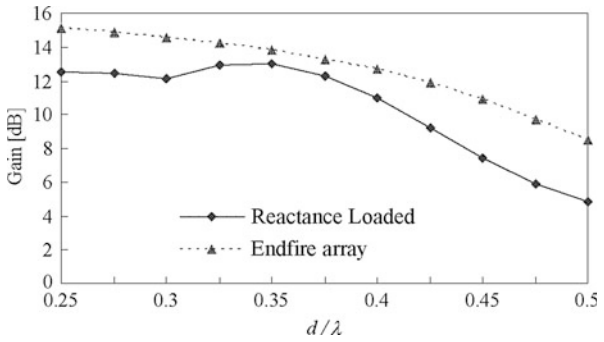


Fig. 24 Optimum gains of a six-element Yagi-Uda antenna with optimum loads compared with those of an end-fire array

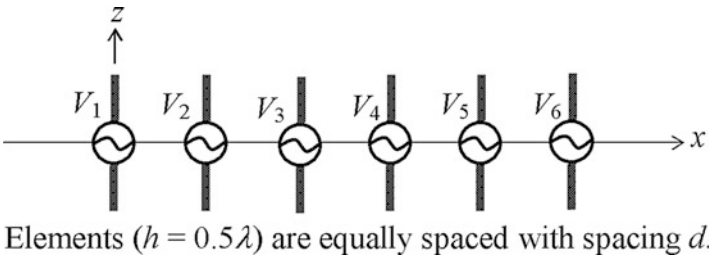


Fig. 25 An equally spaced end-fire array with half-wavelength elements

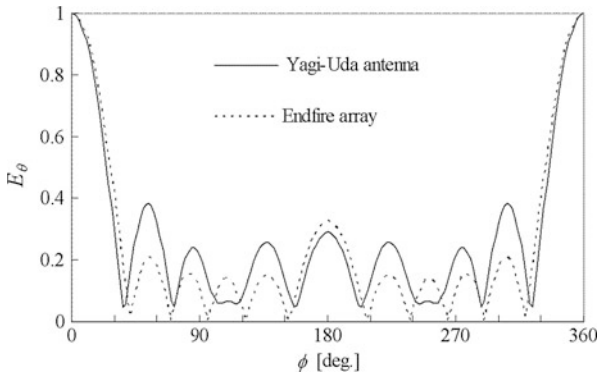


Fig. 26 Normalized horizontal radiation patterns of the optimum Yagi-Uda antenna and end-fire array when $d = 0.35\lambda$

for a reactance-loaded Yagi-Uda antenna and the end-fire array when $d = 0.35\lambda$. The sidelobes of the end-fire array are a little bit lower than those of the optimized Yagi-Uda antenna and the slightly higher gain is obtained by the end-fire array. Figures 27 and 28 show the input impedance and the loaded reactance values for the optimum gain. As d gets smaller, R_{in} becomes smaller. Therefore, $d = 0.3\lambda - 0.375\lambda$ may be good in the point of gain and impedance matching. The loaded reactance values for the optimum gain are all capacitive as shown in Fig. 28. It is quite difficult to realize the optimum gain by adjusting six feed voltages accurately, but it is much easier to realize the optimum gain of the Yagi-Uda antenna by using five capacitors.

DOI 10.1007/978-981-4560-75-7_25-1

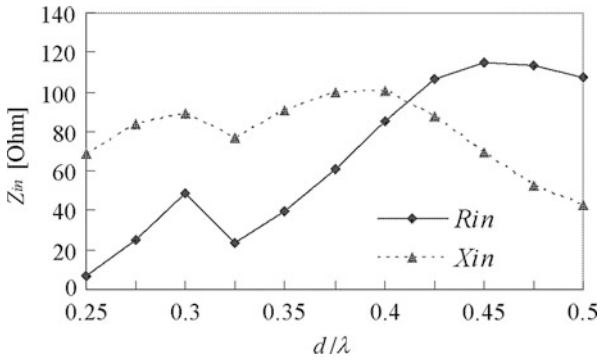


Fig. 27 Input impedance Z_{in} ($= R_{in} + jX_{in}$) versus inter-element spacing d for the optimum gain of the Yagi-Uda antenna

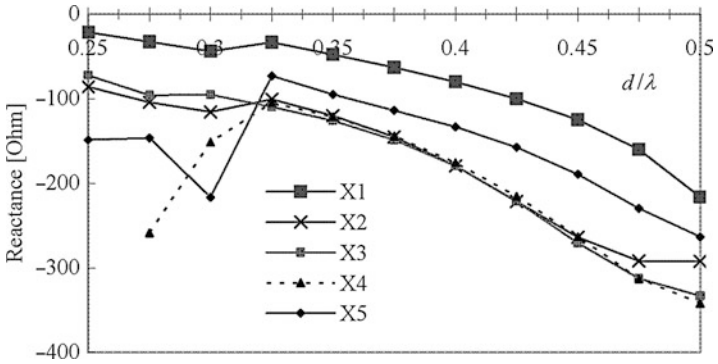


Fig. 28 Optimum reactance loads for the optimum gain of the Yagi-Uda antenna

Conclusion

A brief history is shown for a linear wire antenna development and its frequency-domain analysis methods. Then, the MoM is explained with the electric-field boundary conditions for a transmitting and receiving linear wire antenna. By using the MoM, transmitting linear wire antenna characteristics such as input impedances, current distributions, and radiation patterns are calculated. For receiving antennas, current distributions, reradiation patterns, received power, and reradiated power are shown when a plane wave is incident and the complex conjugate of the input impedance is loaded at the receiving point. Also the application of the Thevenin equivalent circuit is discussed for the receiving antenna. At the end the gain of a reactance-loaded Yagi-Uda antenna is shown and compared with that of an end-fire array.

Cross-References

- ▶ HF, VHF and UHF Antennas
- ▶ Impedance Matching and Baluns
- ▶ Numerical Modeling in Antenna Engineering
- ▶ Optimization Methods in Antenna Engineering

DOI 10.1007/978-981-4560-75-7_25-1

References

Abramowitz M, Stegun IA (1965) Handbook of mathematical functions. Dover, New York, pp 916–919
Balanis CA (2005) Antenna theory: analysis and design, 2nd edn. Wiley, New York
Best SR, Kaanta BC (2009) A tutorial on the receiving and scattering properties of antennas. IEEE AP Mag 51:26–37
Collin RE (2003) Limitations of the Thevenin and Norton equivalent circuits for a receiving antenna. IEEE AP Mag 45:119–124
Hallén E (1938) Theoretical investigation into the transmitting and receiving qualities of antennae. Nova Acta Regiae Soc Sci Upsaliensis Ser VI 11:1–44
Harrington RF (1993) Field computation by moment methods. IEEE, New York
Harrington RF (2001) Time-harmonic electromagnetic fields. IEEE, New Jersey, Chapter 3
Hertz HR (1962) Electric waves. Dover, New York
Hirasawa K (1980) Optimum gain of reactively loaded Yagi-Uda antenna. IEICE Trans J63-B:121–127
Hirasawa K (1987) Reduction of radar cross section by multiple passive impedance loadings. IEEE J Ocean Eng (Spec Issue Scattering) OE-12:453–457
Hirasawa K (1988) The application of a biquadratic programming method to phase only optimization of antenna arrays. IEEE Trans Antennas Propag AP-36:1545–1550
Hirasawa K, Haneishi M (eds) (1991) Analysis, design and measurement of small and low-profile antennas. Artech House, Dedham, Chapter 2
Hirasawa K, Shintaku M, Morishita H (1994) Received and scattered power of receiving antenna. IEEE Antennas Propag Int Symp 1:205–208
Hirasawa K, Sato A, Ojiro Y, Morioka T, Shibasaki S (1997) Thevenin equivalent circuit and scattered power. Prog Electromag Research Symp 1:63–63
Jordan EC, Balmain KG (1968) Electromagnetic waves and radiating systems, 2nd edn. Prentice-Hall, Englewood Cliffs, pp 333–338
King LV (1937) On the radiation field of a perfectly conducting base-insulated cylindrical antenna over a perfectly conducting plane earth, and the calculation of the radiation resistance and reactance. Phil Trans R Soc 236:381–422
King RWP (1956) Theory of linear antennas. Harvard University Press, Cambridge, MA
Kraus JD, Marhefka RJ (2002) Antennas for all applications, 3rd edn. McGraw-Hill, New York
Maxwell JC (1873) A treatise of electricity and magnetism. Oxford University Press, Oxford
Onuki H, Umebayashi K, Kamiya Y, Hirasawa K., Suzuki Y (2007) A study on received and re-radiated power of a receiving antenna. EuCap2007
Poklington HC (1897) Electrical oscillations in wires. Proc Camb Philos Soc 9:324–332
Silver S (ed) (1949) Microwave antenna theory and design. Mc-Graw-Hill, New York, Chapter 2
Stutzman WL, Thiele GA (2012) Antenna theory and design. Wiley, New Jersey
Uda S (1926) On the wireless beam of short electric waves. J IEE (Japan) 46:273–282
Weeks WK (1968) Antenna engineering. McGraw-Hill, New York
Yagi H (1928) Beam transmission of ultra short waves. Proc IRE 16:715–740

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_26-1

Loop Antennas

P. J. Massey[a]*, P. Fellows[a], D. Mirshekar-Syahkal[b], A. Pal[c] and A. Mehta[c]
[a]Her Majesty's Government Communications Centre, Milton Keynes, UK
[b]University of Essex, Colchester, UK
[c]College of Engineering, Swansea University, Swansea, UK

Abstract

This chapter on loop antennas covers implementation aspects of both electrically small and electrically large loops which have a wide range of applications. On one side, they are still a part of the old broadcasting systems such as analog radios, and, on the other side, they are increasingly gaining attention for the contemporary high-definition video transmission and reception. The chapter begins with electrically small loops and coil antennas and covers its different aspects of radiation and losses, tuning, quality factor, and matching. Subsequently, resonant full-wave loops and their polarization are presented. A technique is shown on how loops can be placed in closed vicinity of a communicating device conducting surface without losing efficiency. In the final part, single-element beam steering loop antennas and their arrays for enabling devices to achieve high throughput and high-gain wide scanning range are discussed.

Keywords

Loop antennas; Circularly polarized antennas; Coil antennas; Electrically small antennas; Directive antennas; Ferrite antennas; Receiving antennas; Man-made noise; Multi-feed antennas; Resonant antennas; Rhombic antennas; Switched feeds; Reconfigurable loop antennas and their arrays; Wide scan high-gain adaptive arrays

Introduction

A loop antenna is an antenna where the radiating conductor forms a loop. Loop antennas have existed and have continuously evolved since the earliest days of radio. Consequently, "loop antennas" is now a subject area that is too large for all its features to be covered in just a chapter or even in any single book. Instead, three particular aspects are covered in this chapter:

1. Electrically small loop antennas and their close relations: electrically small multi-turn coils
2. Full-wave self-resonant loops and how they can be modified to cover different frequency bands and to radiate different polarizations and to radiate with high gain
3. Multi-feed loops where the radiation pattern is steered by switching between feeds

A common feature of all these aspects is that they are ways of saving physical space. Electrically small loop and coil antennas save space by being small. Full-wave loops can be used to save space by sharing that space with conducting structures already occupying the space. They can also be combined with other

*Email: peterma@hmgcc.gsi.gov.uk

DOI 10.1007/978-981-4560-75-7_26-1

loops to create compact directive radiators. And multi-feed loops save space by reusing the same radiating element (the loop) to generate several beams.

Electrically Small Loop and Coil Antennas

Introduction

What an electrically small loop or coil antenna consists of. An electrically small antenna is one that is much less than a wavelength across. In addition for the loops and coil antennas discussed below, it is assumed that the distance around the loop and along the length of the coil is short enough that the current can be approximated as near constant.

Because of the low conversion factor between currents on a small structure and radiation and because of the way the current contributions to radiation cancel out in loop and coil antennas (gone into in detail in subsection "Radiation and Losses"), electrically small antennas generally require tuning components to resonate, so that the amount of current carried, and hence the coupling to radiation, is maximized. For electrically small coil and loop antennas, this means tuning with one or more capacitors across the ends of the inductive coil or loop. In addition, they require some connection to their transceiver. There are many ways of achieving this, and some of them are discussed in subsection "Matching." However, other than ensuring that the antenna is impedance and/or noise matched to the transceiver circuitry, the method of matching usually has little effect upon the antenna performance, which is more dependent upon the antenna geometry and upon the environment surrounding the antenna.

Applications. The main application for electrically small loop and coil antennas has been as the antennas in medium- and long-wave broadcast radio receivers. The earlier antennas had the interior of the antenna filled with air. High-permeability soft ferrite was introduced in the 1950s, and coil antennas wrapped around a ferrite core have now become predominant as antennas for medium and long-wave reception. (Subsection "Receiving Antennas and Man-Made Electrical Noise" gives the reasons why they are widely used for this application.)

In the second half of the twentieth century, electrically small loop and coil antennas found another major application within radio pagers, operating at up to UHF (ultra high frequency band; 300MHz to 3GHz). The significant advantages over alternative antenna types are their relatively small size and their relative immunity to detuning due to the pager being carried on a user's body, e.g., mounted on a belt, clipped to a shirt, or placed in a pocket (subsection "Proximity Effects and Detuning" discusses proximity detuning).

Section overview. This section is arranged as follows. First, in subsection "Radiation and Losses," the radiation and losses are calculated, both for electrically small loops and coils and also for comparison for electrically small dipoles. These calculations reveal that electrically small loops and coils are less efficient radiators than dipoles of comparable size. Then, it is followed by subsections "Receiving Antennas and Man-Made Electrical Noise," "Near Field Noise," "Far Field Noise," "Tuning Components and Quality Factors," "Quality Factors," "Finding the Inductance of an Electrically Small Loop or Coil Antenna," and "Proximity Effects and Detuning," which are on the reasons why, despite their inefficiencies, loops and coils became popular for radio reception and for near-human body antennas. These subsections cover (a) the effects of man-made electrical noise on receiving antenna performance, (b) losses due to tuning components and how to find the component values, and (c) detuning from coupling to human bodies and other nearby objects. The last subsection, "Matching", discusses those matching issues that have not already been covered in earlier sections.

There is little discussion of an antenna's impedance bandwidth. This is because electrically small loops and coils tend to be used for narrowband communications. For electrically small loops and coils, the

DOI 10.1007/978-981-4560-75-7_26-1

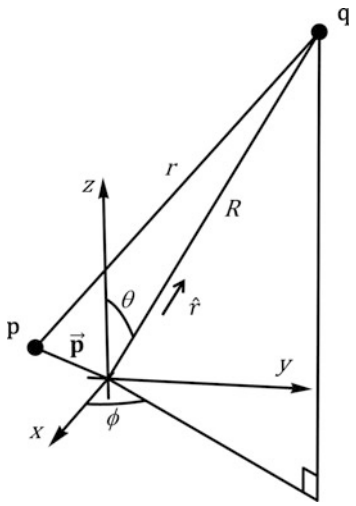


Fig. 1 Coordinates used in calculating radiation

antenna bandwidth tends to be primarily determined by the Ohmic losses of the antenna. For most designs the antenna bandwidth has a lower limit of around 1/2 % and is often around 1–2%, which is adequate for their applications.

Radiation and Losses

This subsection discusses radiation and losses by calculating them for a transmitting antenna. That the efficiency of the antenna when receiving is the same follows from the reciprocity theorem [e.g., see Collin (2001) for a discussion and derivation of the reciprocity theorem].

Radiation in terms of current. In order to calculate the radiation from a loop or coil antenna, one can use the Schelkunoff formula for the far field due to currents (Schelkunoff 1939) [see also Lee (1984)]. This formula comes from the formula for electric field in terms of magnetic vector potential **A** and electric scalar potential Φ. In general, with

$$\mathbf{A} = \frac{\mu}{4\pi}\iiint_v \mathbf{J}\frac{e^{-jkr}}{r}dv\,\Phi = \frac{1}{4\pi\varepsilon}\iiint_v \rho\frac{e^{-jkr}}{r}dv \tag{1}$$

where **J** is the current density, ρ is the charge density, r is the distance between current or charge source at point **p** and the point **q** at which **A** or Φ is being evaluated, and v denotes integration over volume (Fig. 1), then the electric field **E** is given by

$$\mathbf{E} = -j\omega\mathbf{A} - \nabla\Phi \tag{2}$$

See, for instance, Collin (2001), for a derivation of the above formulae.

Schelkunoff showed that when the electric field is due to the far field, the following simplifications apply:

$$\frac{e^{-jkr}}{r} \cong \frac{e^{-jkR}}{R}e^{jk\vec{\mathbf{p}}\cdot\hat{\boldsymbol{r}}} \tag{3}$$

where R is the distance between the antenna's center and **q**, $\hat{\boldsymbol{r}}$ is the direction vector pointing from the center of the antenna to **q** in the far field, and $\vec{\mathbf{p}}$ is the position vector of point **p**. Then, in an r, θ, ϕ

DOI 10.1007/978-981-4560-75-7_26-1

spherical coordinate system, the contribution of Φ to **E** is to cancel out with the radial component of **A**, and **E** becomes

$$\mathbf{E} = -j\omega\{0, A_\theta, A_\phi\} \tag{4}$$

where **A** simplifies to

$$\mathbf{A} \cong \mathbf{A}_{FF} = \frac{\mu}{4\pi}\frac{e^{-jkR}}{R}\iiint_v \mathbf{J} e^{jk\vec{\mathbf{p}}\cdot\hat{\mathbf{r}}}\, dv \tag{5}$$

Electrically small dipole. Before applying the above formula to calculating the radiation from a small loop or coil, it will be useful to apply it to the simpler case of a small dipole. Suppose the electrically small dipole is oriented along the *z*-axis with its terminals at the origin of the coordinate system. The current at the terminals is denoted I_0. The current's magnitude decays toward the open ends of the arms. The way it decays depends on the shape of the arms, but in practice for small antennas one can approximate the current distribution as linear, with it decaying to zero at the tip of the arms. Therefore, the current *I* can be modeled as

$$I(z) = I_0\left(1 - \frac{|z|}{\frac{l_0}{2}}\right) \quad \text{for } -\frac{l_0}{2} \le z \le \frac{l_0}{2} \tag{6}$$

where the dipole's overall length is l_0 and each of the dipole arms is $l_0/2$ long. In the above, the magnitude of *I* is taken as the root-mean-square (rms) value of the current, so as to avoid the need for factors of ½ in the formulae for radiated power and Ohmic losses. The peak magnitude is $\sqrt{2}$ times the rms value (Harrington 2001).

The exponential part of the formula for **A** can be expanded as

$$e^{jk\vec{\mathbf{p}}\cdot\hat{\mathbf{r}}} = 1 + s + \frac{s^2}{2!} + \frac{s^3}{3!} + \ldots \quad \text{where } s = jkz\cos\theta \tag{7}$$

For small distances where $kz \ll 1$, only the leading term is significant. Substituting this approximation and the formula for $I(z)$ into the equations for the far field gives

$$\mathbf{A} = \frac{\mu}{4\pi}\frac{e^{-jkR}}{R}\frac{I_0 l_0}{2}\hat{\boldsymbol{z}} \tag{8}$$

$$E_\theta = \frac{j\omega\mu}{4\pi}\frac{e^{-jkR}}{R}\frac{I_0 l_0 \sin\theta}{2} \qquad E_\phi = 0 \tag{9}$$

where $\hat{\boldsymbol{z}}$ is the unit direction vector in the *z*-direction. The radiated power density is given by

$$P_{\text{Rad dipole}} = \frac{|\mathbf{E}|^2}{Z_0} = \frac{\left(\frac{\omega\mu|I_0|l_0\sin\theta}{8\pi R}\right)^2}{Z_0} = Z_0\left(\frac{|I_0|l_0\sin\theta}{4\lambda R}\right)^2 \tag{10}$$

DOI 10.1007/978-981-4560-75-7_26-1

The total radiated power is the integration of the radiated power density over the sphere of radius R:

$$P_{\text{Rad}T\text{ dipole}} = 2\pi R^2 \int_0^{\pi} Z_0 \left(\frac{|I_0| l_0 \sin\theta}{4\lambda R} \right)^2 \sin\theta \, d\theta = \frac{Z_0 \pi}{6} \left(\frac{|I_0| l_0}{\lambda} \right)^2 \tag{11}$$

The directivity is given by

$$D_{\text{dipole}} = \frac{4\pi \, P_{\text{Rad dipole}}}{P_{\text{Rad}T\text{ dipole}}} = \frac{3 \sin^2\theta}{2} \tag{12}$$

It is useful to compare the radiated power with the power lost through resistive losses. Assuming that the dipole is made of a wire with resistivity ρ_w per meter, then the losses are given by

$$P_{\Omega\text{ dipole}} = 2\rho_w \int_0^{\frac{l_0}{2}} |I_0|^2 \left(1 - \frac{|z|}{\frac{l_0}{2}} \right)^2 dz = \frac{1}{3} \rho_w |I_0|^2 l_0 \tag{13}$$

The radiation efficiency *R.E.* is given by

$$R.E. = \frac{P_{\text{Rad}T}}{P_{\text{Rad}T} + P_{\Omega}} = \frac{P.F.}{P.F. + 1} \tag{14}$$

where the power factor *P.F.* is

$$P.F. = \frac{P_{\text{Rad}T}}{P_{\Omega}} \tag{15}$$

When $P.F. \ll 1$, then $R.E. \approx P.F.$

For the dipole,

$$P.F._{\cdot\text{dipole}} = \frac{Z_0 \pi l_0}{2\rho_w \lambda^2} \tag{16}$$

So the power factor, and therefore for small values the radiation efficiency, is proportional to the dipole length. For example, if the dipole is 1/20th wavelength long, $P.F._{\cdot\text{dipole}} \cong 0.079 Z_0/(\rho_w \lambda)$, and if the dipole is 1/100th wavelength long, $P.F._{\cdot\text{dipole}} \cong 0.016 Z_0/(\rho_w \lambda)$.

Electrically small wire loop. In order to compute radiation from the currents in a small loop, first the case of a small rectangular loop will be considered. Then the results will be extended to loops that are planar but otherwise of arbitrary shape.

Consider an electrically small rectangular loop whose sides are parallel to the x and y axes. For the moment, consider only the far electric field in the plane where $\phi = 0$. Then the direction vector to the far field has the form $\hat{\boldsymbol{r}} = \hat{\boldsymbol{x}} \sin\theta + \hat{\boldsymbol{z}} \cos\theta$ where $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{z}}$ are the unit length direction vectors in the x-, y-, and z-directions. Therefore, $\vec{\boldsymbol{p}} \cdot \hat{\boldsymbol{r}} = p_x \sin\theta + p_z \cos\theta$, and

$$e^{jk\,\vec{\mathbf{p}}\cdot\hat{\mathbf{r}}} = e^{jk(p_x \sin\theta + p_z \cos\theta)} = e^{jkp_x \sin\theta} e^{jkp_z \cos\theta} \tag{17}$$

DOI 10.1007/978-981-4560-75-7_26-1

Substituting this into the equation for $\mathbf{A}_{FF}$ gives

$$\begin{aligned}\mathbf{A}_{FF} &= \frac{\mu}{4\pi}\frac{e^{-jkR}e^{jkp_z\cos\theta}}{R}\iiint_v \mathbf{J}e^{jkp_x\sin\theta}\,dv \\ &= \frac{\mu}{4\pi}\frac{e^{-jkR}e^{jkp_z\cos\theta}}{R}\left(\int_{x_1}^{x_2} dx\left(\hat{\boldsymbol{x}}I_0e^{jkx\sin\theta} - \hat{\boldsymbol{x}}I_0e^{jkx\sin\theta}\right) + \int_{y_1}^{y_2} dy\left(\hat{\boldsymbol{y}}I_0e^{jkx_2\sin\theta} - \hat{\boldsymbol{y}}I_0e^{jkx_1\sin\theta}\right)\right)\end{aligned} \tag{18}$$

where the sides parallel to the x-axis run between x_1 and x_2 and the sides parallel to the y-axis run between y_1 and y_2, and as with the dipole derivation case, I_0's magnitude corresponds to the rms value.

The integration with respect to x is for the two arms parallel to the x-axis, and the contributions from these arms cancel with each other. This leaves the contribution from the two arms parallel to the y-direction. The exponent remaining within the integral can be evaluated using a Taylor expansion:

$$e^{jkp_x\sin\theta} = 1 + s_x + \frac{{s_x}^2}{2!} + \frac{{s_x}^3}{3!} + \ldots \qquad \text{where } s_x = jkp_x\sin\theta \tag{19}$$

The first term of the expansion cancels in the integrand. This leaves the second term as the most significant, and the integration reduces to

$$\int_{y_1}^{y_2} dy\left(I_0e^{jkx_2\sin\theta} - I_0e^{jkx_1\sin\theta}\right) \cong \int_{y_1}^{y_2} dyI_0jk(x_2 - x_1)\sin\theta = I_0jk\sin\theta(y_2 - y_1)(x_2 - x_1) \tag{20}$$

$(y_2 - y_1)(x_2 - x_1)$ is the product of the length and width of the rectangular loop and is equivalent to the loop's area $A_{\text{rectangular loop}}$. So $\mathbf{A}_{FF}$ simplifies to

$$\mathbf{A}_{FF} = \frac{jk\mu}{4\pi}\frac{e^{-jkR}e^{jkp_z\cos\theta}}{R}I_0\sin\theta A_{\text{rectangular loop}}\hat{\boldsymbol{y}} \tag{21}$$

Substituting this back into the equation for the electric far field gives

$$E_\theta = 0\; E_\phi = \frac{\omega\mu k}{4\pi}\frac{e^{-jkR}e^{jkp_z\cos\theta}}{R}I_0\sin\theta A_{\text{rectangular loop}} = c_{\text{rect}}I_0\sin\theta A_{\text{rectangular loop}} \tag{22}$$

where c_{rect} represents the factors that are independent of I_0, θ, and loop dimensions. It can be rearranged by expressing ω and k in terms of the speed of light c and wavelength λ and expressing $c\,\mu$ in terms of Z_0:

$$\frac{\omega\mu k}{4\pi} = \frac{4\pi^2 c\mu}{4\pi\lambda^2} = \frac{\pi Z_0}{\lambda^2} \tag{23}$$

$$c_{rect} = \frac{e^{-jkR}e^{jkp_z\cos\theta}}{R}\frac{\pi Z_0}{\lambda^2} \tag{24}$$

DOI 10.1007/978-981-4560-75-7_26-1

Now consider the far field at a point **q** in the *x-z* plane from a loop L_A of arbitrary outline in a plane parallel to the *x-y* plane with a current I_0 circulating around it in the anticlockwise direction. To calculate the far field, one can conduct a thought experiment where the area enclosed by the loop is filled with small rectangular loops all of which have I_0 circulating around them in the anticlockwise direction. On the boundaries between adjacent rectangular loops, the currents cancel, and so the net current within the interior of the tessellation of rectangular loops is 0, the same as for L_A. The only places where the current does not cancel out are at the exterior boundary of the tessellation, where here it is I_0, the same as on the arbitrary outlined loop. Therefore, the radiated field from the arbitrary outlined loop is the same as from the tessellation of rectangular loops. The radiation from the tessellation of rectangular loops is given by summing the fields of the individual loops, which for E_ϕ is the sum of $c_{rect}\sin\theta$ times their areas, or

$$E_\phi = c_{rect} I_0 A_{\mathrm{LA}} \sin\theta \tag{25}$$

where A_{LA} is the total area enclosed by the loop.

The above was calculated for far field points where $\phi = 0$. But one can argue that the result applies for any angle of ϕ by noting that as the shape of the loop in the x-y plane is arbitrary, one could rotate the coordinate system about the z-axis by $-\phi$ and arrive at the same result. So the above equation applies for all ϕ.

The radiated power density is given by

$$P_{\mathrm{Rad\ loop}} = \frac{|\mathbf{E}|^2}{Z_0} = \frac{|c_{rect}|^2 (|I_0| A_{\mathrm{LA}} \sin\theta)^2}{Z_0} = Z_0 \left(\frac{\pi |I_0| A_{\mathrm{LA}} \sin\theta}{\lambda^2 R} \right)^2 \tag{26}$$

The total power radiated is given by

$$P_{\mathrm{Rad}T\ \mathrm{loop}} = 2\pi R^2 \int_0^\pi Z_0 \left(\frac{\pi |I_0| A_{\mathrm{LA}} \sin\theta}{\lambda^2 R} \right)^2 \sin\theta \, d\theta = \frac{8 Z_0 \pi^3}{3} \left(\frac{|I_0| A_{\mathrm{LA}}}{\lambda^2} \right)^2 \tag{27}$$

The Ohmic losses are given by

$$P_{\Omega\mathrm{loop}} = \rho_w \oint dl \, |I_0|^2 = \rho_w p_{\mathrm{LA}} \, |I_0|^2 \tag{28}$$

where p_{LA} is the perimeter of the loop.

The power factor is given by

$$P.F._{\cdot \mathrm{wire\ loop}} = \frac{P_{\mathrm{Rad}T\ \mathrm{loop}}}{P_{\Omega\mathrm{loop}}} = \frac{8 Z_0 \pi^3}{3 \rho_w p_{\mathrm{LA}}} \left(\frac{A_{\mathrm{LA}}}{\lambda^2} \right)^2 \tag{29}$$

To compare with the result for a dipole, consider a square loop, with sides of length l_0. If $l_0 = \lambda/20$, then $P.F._{\mathrm{loop}} \cong 0.0026\, Z_0/(\rho_w \lambda)$, and if $l_0 = \lambda/100$, $P.F._{\cdot loop} \cong 2.1 \times 10^{-5}\, Z_0/(\rho_w \lambda)$. So for antennas made out of the same wire and a 20th of a wavelength across, the dipole's power factor is 30 times better than the loop's, and for antennas a hundredth of a wavelength across the dipole's power factor is 760 times larger than that of the loop.

DOI 10.1007/978-981-4560-75-7_26-1

For later comparison with the strip metal loop case, it is useful to consider when the wire diameter is much greater than the skin depth, and therefore the relation $\rho_w = \rho_s/(\pi d)$ applies, where ρ_s is the surface resistivity. Then

$$P.F._{\text{wire loop}} = \frac{8Z_0\pi^4 d}{3\rho_s p_{\text{LA}}}\left(\frac{A_{\text{LA}}}{\lambda^2}\right)^2 \tag{30}$$

Electrically small loop of strip metal. Some electrically small loop antennas are constructed from strips of metal rather than wires, in order to reduce the Ohmic losses of the loop. The computation of the radiated field can be thought of as a summation of the field from a collection of electrically small wire loops stacked along the z-axis. So the field calculation proceeds as for the calculation of a small wire loop above, except that there is an integration over the current density with z. That is, Eq. 25 is modified to

$$E_\phi = \int_{z_1}^{z_2} dz\, c_{rect}(z)\, I(z) A_{\text{LA}}(z) \sin\theta \tag{31}$$

The z dependant factor in c_{rect} is $e^{jkz\cos\theta}$, which can be expanded by Taylor expansion:

$$e^{jkz\cos\theta} = 1 + s_z + \frac{s_z^2}{2!} + \frac{s_z^3}{3!} + \ldots \qquad \text{where } s_z = jkz\cos\theta \tag{32}$$

and for small kz the leading constant term dominates and the other terms can be neglected. Also usually the loop has constant cross-sectional area A_{LA}. Then the right-hand side of Eq. 31 becomes the same as Eq. 25, and the formulae for the radiated power and total power radiated are as given in Eqs. 26 and 27.

The formula for the Ohmic losses is

$$P_{\Omega\ \text{strip loop}} = \oint dl \iint_S dS\rho\, |J|^2 = \rho\, p_{\text{LA}} \iint_S dS\rho\, |J|^2 \tag{33}$$

where ρ is the bulk resistivity, J is the current density, p_{LA} is the perimeter of the loop, and the double integral over S denotes integration across a cross section of the strip. The variation in current density depends on the loop geometry; because of induction, current tends to be pulled towards the inner side of the strip. It also depends upon z as the current concentrates near the corners of the strip. Therefore, the double integral is difficult to evaluate. However, for many applications, the thickness of the strip is significantly greater than the skin depth, and then the formula can be written as

$$P_{\Omega\ \text{strip loop}} = \rho_S\, p_{\text{LA}} \left(\int_{z_1}^{z_2} dz \left(|J_{S\ \text{inner}}|^2 + |J_{S\ \text{outer}}|^2 \right) + \int_{x_{\text{inner}}}^{x_{\text{outer}}} dx \left(|J_{S\ \text{top}}|^2 + |J_{S\ \text{bottom}}|^2 \right) \right) \tag{34}$$

where ρ_s is the surface resistivity in Ohms per square, J_s is the surface current density, "inner" and "outer" signify the surfaces to the inside $x = x_{\text{inner}}$ and outside $x = x_{\text{outer}}$ of the strip, and "top" and "bottom" signify the surfaces at the top $z = z_2$ and bottom $z = z_1$ of the strip. The above integrals are difficult to estimate analytically but, for a wide range of cases applicable to VHF (very high frequency band; 30 to

DOI 10.1007/978-981-4560-75-7_26-1

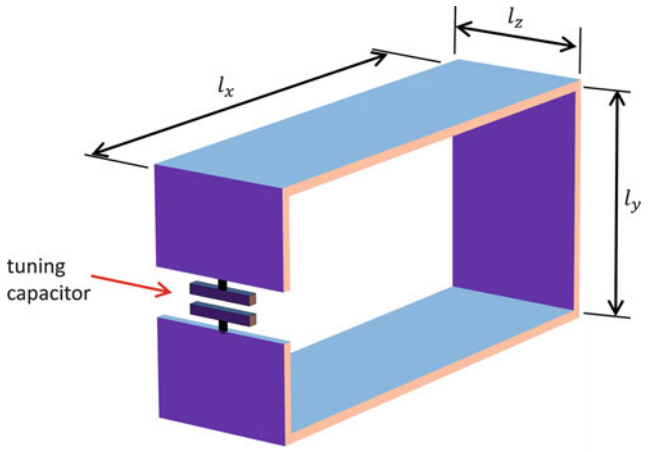


Fig. 2 Dimensions of a rectangular loop. While the tuning capacitor is shown in its often preferred position in one of the shorter sides (see subsection "Proximity Effects and Detuning"), its location is immaterial to the validity of Eq. 39

300MHz) and UHF body-borne loops (e.g., for radio pagers), have been examined numerically. These numerical studies have shown that the following approximation can be made:

$$P_{\Omega \text{ strip loop}} = \rho_S \, p_{\text{LA}} \, \beta \left(\int_{z_1}^{z_2} dz \, 2|J_S|^2 \right) = 2\rho_S \, p_{\text{LA}} \, \beta \, l_z |J_S|^2 = \rho_S \, p_{\text{LA}} \, \beta \frac{|I_0|^2}{2 \, l_z} \tag{35}$$

where $l_z = (z_2 - z_1)$ is the strip's width. In other words, one can approximate the power lost by calculating the Ohmic losses assuming that the current density is constant and along the inner and outer sides. The additional losses due to the current preferring the inner side and collecting at the edges of the strip and due to the current at the top and bottom surfaces are taken care of by introducing a factor β where $1 < \beta < \sim 1.6$. Combining this with Eq. 27 gives

$$P.F._{\text{strip loop}} = \frac{16 \, l_z Z_0 \pi^3 \left(\frac{A_{\text{LA}}}{\lambda^2} \right)^2}{3 \, \rho_S \, p_{\text{LA}} \, \beta} \tag{36}$$

The difference between this and Eq. 30 is

$$\frac{P.F._{\text{strip loop}}}{P.F._{\text{wire loop}}} = \frac{2 l_z}{d \pi \beta} \tag{37}$$

In radio pagers, there is often enough space available to use a 5 mm width strip metal, or one could use a relatively thick wire (e.g., 1 mm diameter) to save material costs. The power factor improvement from using a 5 mm wide strip instead of a 1 mm diameter wire would be $10/(\pi\beta)$ which (depending on the value of β) is between ~2 and ~3 or between 3 and 4.8 dB. And because of Eq. 14 for small values of *P.F.*, this gives a similar difference in the radiation efficiency. This may seem a very useful improvement, however as discussed in subsection "Far Field Noise," often such differences in antenna performance make little difference to the overall performance of a receiver.

DOI 10.1007/978-981-4560-75-7_26-1

For body-borne VHF and UHF loop applications, a rectangular loop is particularly of interest (see subsection "Proximity Effects and Detuning"). Then as $p_{\mathrm{LA}} = 2\left(l_x + l_y\right)$ where l_x is the length across the loop and l_y is the width across the loop (see Fig. 2), the formula for Ohmic losses simplifies to

$$P_{\Omega \text{ rectangular strip loop}} = \left(l_x + l_y\right) \rho_S \, \beta \frac{|I_0|^2}{l_z} \tag{38}$$

Combining this with Eq. 26 gives

$$P.F._{\text{ rectangular strip loop}} = \frac{\dfrac{8Z_0\pi^2}{3}\left(\dfrac{|I_0|l_x l_y}{\lambda^2}\right)^2}{\left(l_x + l_y\right)\rho_S \, \beta \dfrac{|I_0|^2}{l_z}} = \frac{1.48 \times 10^{-23} f^{\frac{7}{2}} l_z \left(l_x l_y\right)^2}{\beta\sqrt{\rho_r}\left(l_x + l_y\right)} \tag{39}$$

where f is the frequency (in hertz); l_x, l_y, and l_z are the loop dimensions in meters; ρ_r is the bulk resistivity relative to that of annealed copper (1.72×10^{-8} Ω m); and the formula for surface resistivity in terms of bulk resistivity ρ has been applied:

$$\rho_S = \sqrt{\pi f \rho \mu} \tag{40}$$

Equation 39 has been used in industry for early evaluation of design choices in pager receiver design. In this context, it has been found that the losses associated with the tuning capacitors can also be accounted for by increasing the estimate for β to around 3 (Massey 2001).

Electrically small coils. The electric field from coils can be approximated by considering them as a collection of small loops spaced along the coil axis parallel to the z-axis. For an N-turn coil, the radiated field from the coil is given by

$$E_{\phi\text{coil}} = N c_{\text{rect}} I_0 A_{\text{LA}} \sin\theta \tag{41}$$

and

$$P_{\text{Rad coil}} = \left(\frac{N\pi|I_0|A_{\text{LA}}\sin\theta}{\lambda^2 R}\right)^2 \tag{42}$$

The total power radiated is given by

$$P_{\text{Rad}T\text{ coil}} = \frac{8Z_0\pi^3}{3}\left(\frac{NI_0A_{\text{LA}}}{\lambda^2}\right)^2 \tag{43}$$

The directivity is given by

$$D_{\text{coil}} = \frac{4\pi \, P_{\text{Rad coil}}}{P_{\text{Rad}T\text{ coil}}} = \frac{3\sin^2\theta}{2} \tag{44}$$

This is the same as the formula for the directivity of an electrically small dipole. As the above formula applies for all N, including $N = 1$, a loop has the same directivity as a coil. And as for loops and coils, P_{Rad}

DOI 10.1007/978-981-4560-75-7_26-1

and $P_{\mathrm{Rad}T}$ depend only on the total current I_0 and are independent of conducting wire or strip's cross section; the directivity is also independent of wire or strip size and cross section.

The Ohmic losses are

$$P_{\Omega\ \mathrm{coil}} = N\rho_w p_{\mathrm{LA}}\ |I_0|^2 \tag{45}$$

So,

$$P.F._{\mathrm{coil}} = \frac{8NZ_0\pi^3}{3\rho_w p_{\mathrm{LA}}}\left(\frac{A_{\mathrm{LA}}}{\lambda^2}\right)^2 \tag{46}$$

At first sight, Eq. 46 appears to state that coils have the advantage over loops that their power factor is improved by a factor of N, where N is the number of turns. While this is true if one was restricted to only using a particular wire, to see that this is not generally true, consider if instead of using a loop with one strand of wire in the turn, one had N strands. The total current I_0 stays the same, and Eq. 27 holds. However, assuming that the wire diameter is significantly more than the skin depth, then the Ohmic losses are proportional to the inverse of the perimeter of the conductor cross section and are reduced by a factor of N. So an N-turn coil made with one strand of wire has no performance advantage over a loop made from N strands of wire.

In practice loop antennas are rarely made from multiple strands of wire. Using metal strips often makes better use of the space that would be occupied by the multiple wire strands.

The reason that coils are popular is that their impedance is higher than loops. This reason was particularly important for wireless receiver applications when valves were used. The alternative to using a coil antenna for those applications was to include an impedance transformer, which is itself made using coils and is relatively expensive.

Because the radiation from an electrically small loop or coil is proportional to the loop or coil cross section's area squared, while the radiation from a dipole is proportional to the dipole length squared, electrically small loops rapidly become more inefficient than dipole antennas as the antenna dimensions are reduced. However, despite the better efficiency of the dipole-type structure, loop antennas find uses because they have a number of advantages. These advantages are related to:

- **Man-made electrical noise**: The near fields of human-generated interference are primarily electrical. Therefore, antennas which are insensitive to electrical near fields pick up less noise.
- **Quality factors of tuning components**: Electrically small loop antennas require capacitors for tuning components. Generally these have higher quality factors than the inductors used by electrically small dipole antennas and therefore introduce less loss.
- **Proximity detuning**: Nearby objects such as metal plates and human bodies can detune an electrically small antenna. The amount of detuning depends upon the type and orientation of the antenna, and the shape and preferred orientations of some loop antennas are better for use in many VHF and UHF body-mounted applications.

The following sections go into these topics in more detail.

Receiving Antennas and Man-Made Electrical Noise

When Marconi made his first successful transatlantic communications in 1901, for the receiving antenna he used a long wire antenna (effectively an electric monopole) suspended from a kite. Yet by the 1930s, loop antennas were used in practically all consumer radio receiver sets. The reason for this change in

DOI 10.1007/978-981-4560-75-7_26-1

antenna type was that in the interim three decades, man-made electrical noise had increased enormously. When this noise is from nearby sources and the receiver is in the near field of the sources, the noise usually has a much higher electric field, and loop antennas have a clear advantage because of their insensitivity to electric fields – see subsection "Near Field Noise." Noise sources further away reduce the performance advantages of larger and less lossy antennas – see subsection "Far Field Noise."

Near Field Noise

For noise sources closer than the order of a wavelength away, the noise is predominantly electric field. Therefore, a loop antenna, being more sensitive to the magnetic field, picks up less of the near field noise (Blok and Rietveld 1955). The excess of electric field in near field noise is particularly strong at long and medium wavelengths and is sometimes called the "aerial effect."

In suppressing near field noise, it is very important to ensure that the antenna is truly behaving as a loop antenna and that it is not inadvertently coupling to nearby conductors in such a way to form a dipole antenna. This issue is discussed in subsection "Proximity Effects and Detuning."

Far Field Noise

Electrically small loop and coil antennas are generally less efficient (more lossy) than long wire antennas. However, by the 1930s, the high levels of man-made environmental noise were making the performance reductions due to antenna losses much less significant. A full analysis of this would require a discussion of noise figures and their application to system noise analysis [see, e.g., ITU (2001)], which is beyond the scope of this chapter. However, the mechanism by which high levels of environmental noise reduce the effects of antenna losses can be explained with some examples.

The ability of a receiver to accurately detect a signal transmission is limited by how much the signal's power level P_{SO} is at the receiver input compared to the random noise power level P_{NO} at the input. If the antenna were lossless, then these powers would correspond to the signal power P_{SI} and the noise power P_{NI} picked up by the antenna. However, antenna losses degrade the ratio P_{SO}/P_{NO}. Therefore, it is of interest to examine how the signal-to-noise changes with antenna losses. One is interested in

$$f_a = \frac{\dfrac{P_{SI}}{P_{NI}}}{\dfrac{P_{SO}}{P_{NO}}} = \frac{P_{SI}P_{NO}}{P_{SO}P_{NI}} \tag{47}$$

(If the input and output powers of a two-port circuit were being discussed, the ratio f_a would be called the noise factor, and its value in dB would usually be called the noise figure. However, the term "antenna noise figure" is normally applied to different quantity that describes the excess of environmental noise over thermal noise, and so to avoid confusion, the terms "antenna noise factor" and "antenna noise figure" will not be used in the remainder of the subsection.)

With a perfect antenna, the signal incident upon the antenna would be transferred from space to the antenna terminals, and no additional noise would be introduced: $P_{SO} = P_{SI}$, $P_{NO} = P_{NI}$, and $f_a = 1$. Unfortunately, real antennas have attenuation due to resistive losses. This degrades both the signal and the noise from the atmosphere. It also introduces noise that is proportional to the temperature of the antenna. Consequently, for real antennas, $f_a > 1$. If the attenuation is represented by $g < 1$, then $P_{SO} = gP_{SI}$, and the noise power at the output transferred from space is $= gP_{NI}$. The total noise power at the output is that transferred from space plus the noise created by the antenna resistance: $P_{NO} = gP_{NI} + (1-g)kT_0B$ where k is Boltzmann's constant, T_0 is the temperature of the antenna

DOI 10.1007/978-981-4560-75-7_26-1

structure, and B is the bandwidth across which the power is being measured. Substituting into the formula for f_a gives

$$f_a = \frac{1}{g}\frac{gP_{NI} + (1-g)kT_0B}{P_{NI}} = 1 + \frac{(1-g)kT_0B}{gP_{NI}} \tag{48}$$

Suppose a well-matched receiving antenna is lossy – e.g., it is only 10 % efficient. Then $g = 0.1$ and $f_a = 1 + 9kT_0B/P_{NI}$. If the incident environmental noise power is comparable to the antenna structure's noise temperature: $P_{NI} = kT_0B$, then $f_a = 1 + 9 = 10$, or expressed in dB, the antenna losses reduce the signal-to-noise ratio by 10 dB. So for this low environmental noise example, the signal is attenuated by 10 dB, and the 10 dB attenuation of the environmental noise attenuation is made up by the noise introduced by the antenna resistance.

Now consider the same receiving antenna in a high environmental noise environment. If the environmental noise level $P_{NI} = 9kT_0B$ then $f_a = 1 + 9 \times (1/9) = 2$. Expressed in dB, the signal-to-noise level has degraded by 3 dB. The presence of high environmental noise reduces the increase in signal-to-noise level due to the antenna. In order to achieve the same 3 dB degradation level with $P_{NI} = kT_0B$, the antenna efficiency would have to be 50 %.

In both urban and rural areas throughout the populated parts of North America, Europe, and in many other regions, man-made electromagnetic noise has reached significant levels. At 100 MHz in these regions, the noise power that an antenna captures is often a hundred times or 20 dB above kT_0B. For a 10 % efficient antenna in this environment, the signal-to-noise ratio degrades by just 0.37 dB. In order to achieve this limited degradation with $P_{NI} = kT_0B$, the antenna efficiency would have to be 92 %.

For the reason explained in the next paragraph, man-made noise power contributing to P_{NI} increases as the frequency reduces. So the contribution of man-made noise to reducing the performance degradation caused by low antenna efficiency is more marked in the lower frequency bands. Conversely, for the upper UHF band and higher frequencies, man-made noise power is so low that the performance of a receiving antenna is strongly affected by its antenna efficiency – i.e., an ndB antenna loss increases the signal-to-noise level by ndB.

Atmospheric man-made radio frequency noise has a power density spectrum (measured in watt/m^2/Hz) with a frequency dependency that varies with location. However, the amount of man-made noise power (measured in watt/Hz) picked up by a receiving antenna is the product of both the power density spectrum and the antenna's capture area, and P_{NI} is the noise power received before the antenna losses are accounted for. The reciprocity relations between a transmitting and a receiving antenna and the Friis equation show that a lossless receiving antenna's capture area is proportional to the wavelength squared or equivalently to the inverse of the frequency squared. This inverse square dependency attenuates P_{NI} with increasing frequency faster than any growth in power density spectrum with frequency and means that regardless of location, the amount of man-made noise power P_{NI} increases quickly as the frequency is reduced. See Wagstaff and Merricks (2003) for a range of examples of how P_{NI} varies with frequency.

Tuning Components and Quality Factors

This subsection is about quality factors of components and circuits and why they are significant in discussing the performances of electrically small dipoles and loops. It also contains a discussion of how the inductance of the electrically small loop antenna and its tuning capacitance is found.

Quality Factors

Inductors are required to tune electrically small dipoles, and capacitors are required to tune electrically small loop antennas. The resistive losses associated with these tuning components have a detrimental

DOI 10.1007/978-981-4560-75-7_26-1

effect upon the antenna performance. Component manufacturers specify the losses in terms of quality factors as follows. The quality factors Q_c of a capacitor and Q_L of an inductor are defined as:

$$Q_C = \frac{2\omega E_C}{P_C} = \frac{1}{\omega C R_{SC}}\ ,\ Q_L = \frac{2\omega E_L}{P_L} = \frac{\omega L}{R_{SL}} \tag{49}$$

where E_C is the time-averaged energy stored in the capacitor, P_C is the time-average power dissipated by the capacitor, C is the capacitance, R_{SC} is the resistance that dissipates power losses P_C when put in series with a lossless capacitance C, E_L is the time-averaged energy stored in the inductor, P_L is the time-average power dissipated by the inductor, L is the inductance, and R_{SL} is the resistance that dissipates power P_L when put in series with an ideal lossless inductance L.

The quality factor of a circuit is the ratio of its stored energy to the energy dissipated per cycle. Since for a series or parallel LC circuit the total power dissipated P_T is

$$P_T = P_C + P_L \tag{50}$$

Then

$$\frac{P_T}{\omega\, E_T} = \frac{1}{Q_T} = \frac{1}{Q_C} + \frac{1}{Q_L} \tag{51}$$

where E_T is the energy stored within the circuit, and as this energy moves between the inductor and capacitor, $E_T = 2E_C = 2E_L$.

The LC circuit is of significance because:

- An electrically small dipole behaves like a capacitor, which has Ohmic and radiation loss resistances associated with it. It is tuned using an inductor.
- An electrically small loop or coil behaves like an inductor, which has Ohmic and radiation loss resistances associated with it. It is tuned using a capacitor.

With this in mind, Eq. 50 can be written as

$$P_T = P_{\text{antenna}} + P_{\text{tune}} = P_\Omega + P_{\text{Rad}T} + P_{\text{tune}} \tag{52}$$

where P_{tune} is the Ohmic losses of the tuning component, and as in subsection "Radiation and Losses," P_Ω is the Ohmic loss of the antenna and $P_{\text{Rad}T}$ is the radiated power from the antenna. Then, Eq. 52 can be rewritten as

$$\frac{1}{Q_T} = \frac{1}{Q_{\text{antenna}}} + \frac{1}{Q_{\text{tune}}} = \frac{1}{Q_\Omega} + \frac{1}{Q_{\text{Rad}T}} + \frac{1}{Q_{\text{tune}}} \tag{53}$$

Then, the ratio of the increase in Ohmic losses due to the tuning component to the Ohmic losses with a perfect tuning component is

$$\frac{P_\Omega + P_{\text{tune}}}{P_\Omega} = Q_\Omega\left(\frac{1}{Q_\Omega} + \frac{1}{Q_{\text{tune}}}\right) = 1 + \frac{Q_\Omega}{Q_{\text{tune}}} \tag{54}$$

DOI 10.1007/978-981-4560-75-7_26-1

The quality factors of most commercially available capacitors and inductors are quoted in the manufacturers' data sheets, and commercially available capacitors typically have quality factors of a few thousand. Commercially available inductors typically have quality factors ranging from 20 to around 200. Because electrically small antennas are usually significantly bigger than tuning components, the currents are spread across larger conducting surfaces, and so the Ohmic loss quality factor Q_Ω of electrically small dipole is usually a little better than Q_C of a capacitor, and factor Q_Ω of electrically small loop or coil is usually a little better than Q_L of an inductor. So one finds that for an electrically small dipole,

$$\frac{Q_\Omega}{Q_{\text{tune}}} \approx \frac{\sim 2000}{20 \text{ to } 200} \text{ and } \frac{P_\Omega + P_{\text{tune}}}{P_\Omega} > \sim 11 \tag{55}$$

And for an electrically small loop,

$$\frac{Q_\Omega}{Q_{\text{tune}}} \approx \frac{20 \text{ to } 200}{2000} \text{ and } \frac{P_\Omega + P_{\text{tune}}}{P_\Omega} \sim 1 \tag{56}$$

So the capacitor used to tune an electrically small loop or coil usually increases the losses much less than the inductor used to tune an electrically small dipole.

Finding the Inductance of an Electrically Small Loop or Coil Antenna

There are several methods for finding inductance. There are formulae for a variety of loop shapes and coil configurations, e.g., see Balanis (2005) for a simple circular cross section wire loop and Grover (2009) for a large selection of loop and coil configurations. In recent years, electromagnetic simulators have been useful and often more accurate than formulae in calculating inductance and can model realistic loops in their environment, e.g., noncircular shapes and varying cross sections that fit within the space available within their receiver, accounting for the effect of nearby metalwork within the receiver. However, often in practice, none of these theoretical methods account for all the contributions to the loop/coil's inductance, which may include manufacturing tolerances and construction variations. Consequently, measurement is the most reliable method of determining the inductance. At radio and microwave frequencies, a practical method of determining the inductance of the antenna loop/coil is to place tuning capacitance across the ends of the coil and search using a network analyzer for the lowest frequency resonance of the antenna. The inductance L is then calculated from the resonant frequency f_0 and the known capacitance value of the tuning capacitor(s) C:

$$L = \frac{1}{(2\pi f_0)^2 C} \tag{57}$$

The way in which the network analyzer is used to determine the resonant frequency depends on what connections are available on loop/coil antenna. If the network analyzer is connected across the ends of the loop/coil, then a search is done for the lowest frequency at which the magnitude of the impedance peaks. N.B. in this situation, it is very important that the connection between loop/coil and network analyzer uses a balanced line or balun; otherwise, the results will be contaminated by the currents traveling down the outside conducting surfaces of an unbalanced feed.

Alternatively, if a matching circuit is already in place (see subsection "Matching" for discussion of matching circuit types), then the network analyzer can be used to find the lowest frequency at which there is a peak in the insertion loss.

DOI 10.1007/978-981-4560-75-7_26-1

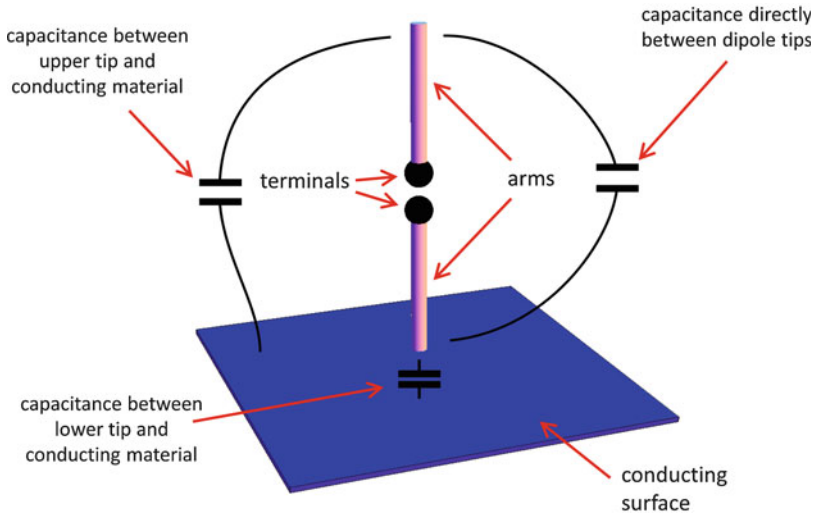


Fig. 3 Dipole perpendicular to the surface of a conducting body. The *black lines* show the approximate route of field lines between the tips and from the tips to the conducting surface. The distance along the field lines across space from one tip to the other tip via the conducting surface (the path on the *left* + the path in the *lower center*) is greater than the distance across space running directly between the tips. This indicates that the additional capacitance between the tips due to the presence of the conducting surface is less than the capacitance between the tips in free space

The accuracy of the resonance search method described above to determine the antenna inductance is often limited by the tolerances on the value of the tuning capacitor(s). These vary with capacitor type and series but typically lie between 2 % and 10 %. These tolerances also limit the accuracy to which the operating frequency of an electrically small loop/coil antenna can be preset using off-the-shelf fixed capacitors. For fixed frequency antennas, it is a good practice to use an adjustable trimmer capacitor in parallel with a fixed capacitor, so that the antenna's frequency can be adjusted prior to deployment.

Proximity Effects and Detuning

In this section, proximity detuning for a dipole is discussed first in order to set the scene for explaining the benefits of a loop antenna.

For both electrically small dipoles and for electrically small loops and coil antennas, detuning can occur through:

1. Capacitive coupling to nearby objects – This additional capacitance increases the tuning capacitance.
2. Inductive coupling to nearby objects – Nearby metal surfaces support currents that move in opposition to currents on the loop and act to reduce the inductance.

So capacitive proximity coupling lowers the resonant frequency, and inductive proximity coupling raises the resonant frequency. Some examples are described below.

All the following examples show a conducting sheet in their diagrams. These are included to represent the surface of a conductor, which could be a thin sheet (as shown) or a thick conducting body such as a block of metal. As the human body is composed of high dielectric lossy conducting fluids and tissues, it also has a similar effect upon the antennas.

Figure 3 shows a dipole perpendicular to the surface of a conducting body. The body couples capacitively to the dipole via the high electric fields at the dipole's lower tip. However, this capacitance has little effect upon the tuning as the conducting body is relatively far away from the upper arm.

DOI 10.1007/978-981-4560-75-7_26-1

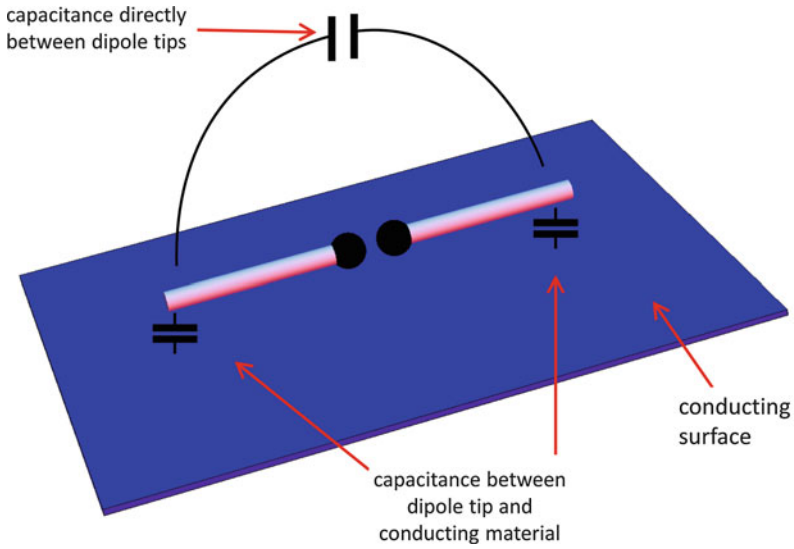


Fig. 4 Dipole parallel to the surface of a conducting body. The *black lines* show the approximate route of field lines between the tips and from the tips to the conducting surface. The distance along the field lines across space from one tip to the other tip via the conducting surface (the path on the *left* + the path on the *right*) is much shorter than the distance across space running directly between the tips. This indicates that the additional capacitance between the tips due to the presence of the conducting surface is much greater than the capacitance between the tips in free space

In contrast, Fig. 4 shows the situation when a dipole is placed parallel to the surface of a conducting body. The capacitance between each of the tips and the conducting body is much higher than the capacitance between the tips. Therefore, the presence of the conducting body provides a high capacitance path between the tips that detunes the antenna. Moreover, the current excited in the conducting body flows in the opposite direction to the current in the dipole and so cancels out its radiation.

The consequence of the above is that dipole-type antennas are suitable for use near conducting bodies, providing that they are oriented perpendicular to the surface of the body. For human body-mounted devices, this would mean that the antenna sticks out, which is generally considered undesirable. As is discussed below, loop antennas offer an alternative that gives a low profile above the body.

Figures 5 and 6 show some examples of proximity coupling in electrically small loops. The figures use a loop that is long and relatively narrow, which is a common shape as it is suited for fitting in devices such as pagers that fit in pockets next to a user's body. See Fig. 9 for an example on how this type of loop is installed in a radio pager.

When mounting loop antennas in body-worn devices, to avoid the detrimental effects of the inductively induced currents discussed in Fig. 5, it is important that the plane of the loop is perpendicular to the body's surface. And to avoid the capacitive detuning seen in the arrangement shown in Fig. 6, it is important that the tuning capacitor is mounted at the end of the long side of the loop. Figure 7 shows the preferred location for the tuning capacitor and for the orientation of the loop when next to a conducting body.

Besides the issues of detuning and of reduction in radiation, proximity coupling can also increase the coupling to noise. This can occur in two ways. The first is if the nearby structures are carrying noise. Mains cables are a common example; besides carrying 50–60 Hz mains signals, their currents can be contaminated with radio frequency noise picked up from connected equipment. The second way is if capacitive coupling adds a dipole-type antenna structure to the loop/coil. Then the dipole structure will pick up electric near field, and the near field noise-reducing advantage of using a loop (section "Near Field Noise") is diminished.

The capacitance available for capacitive proximity coupling is proportional to the size of the loop/coil. The effect of the proximity capacitance is dependent upon the inductance of the loop/coil and its tuning

DOI 10.1007/978-981-4560-75-7_26-1

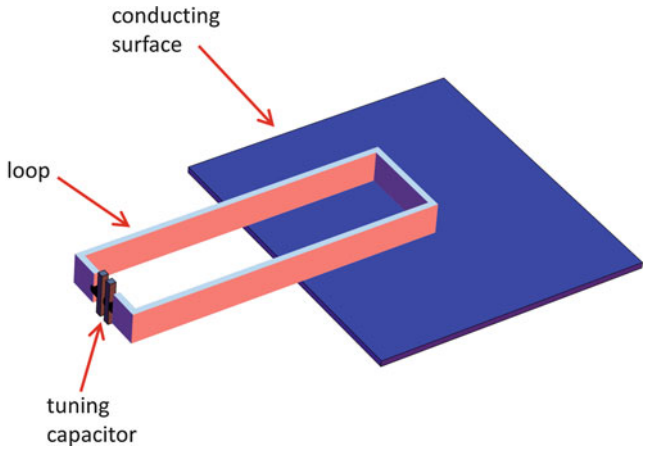


Fig. 5 Inductive proximity coupling to loops. The presence of a conducting sheet parallel to the plane of the loop (or perpendicular to the axis of a coil) reduces the loop's (or coil's) inductance because currents are excited in the sheet that travel in the opposite direction to those in the loop. Therefore, the antenna's resonant frequency increases

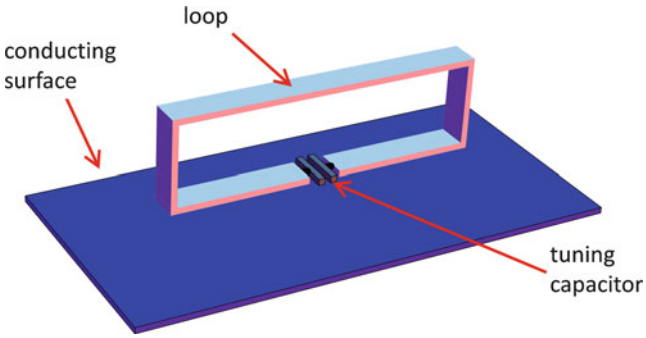


Fig. 6 Capacitive proximity coupling to loops. Capacitance between the loop and the conducting sheet adds to that of the tuning capacitor, reducing the resonant frequency. Placing the tuning capacitor in the middle of a long side of a loop antenna is the worst location for it, because of the propensity of the antenna to suffer from detuning due to capacitive coupling, by the same mechanism as shown in Fig. 4. This can even happen if the sheet is placed next to the long side without the tuning capacitor. For this reason, the tuning capacitor is normally placed at or near a narrow side

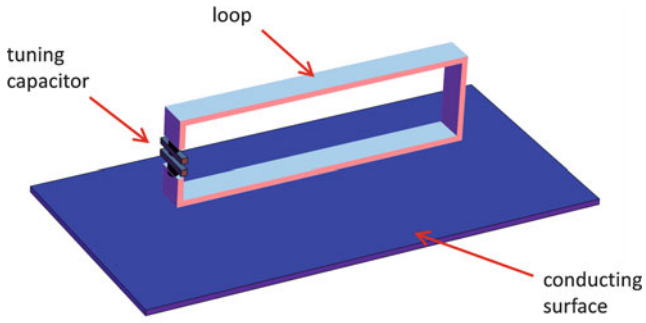


Fig. 7 Preferred configuration and orientation of electrically small loop next to conducting body

capacitance. If it is a large loop or a coil with a large number of turns, then the tuning capacitance will be less, and the proximity capacitance will be more significant. The large amount of metal wire associated with the high number of turns also increases the proximity capacitance compared with a few turns or a single turn of the same gauge of wire.

DOI 10.1007/978-981-4560-75-7_26-1

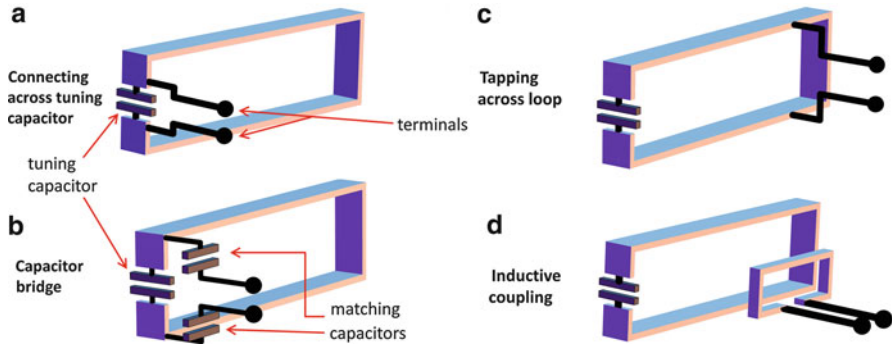


Fig. 8 Some matching arrangements for a loop or coil antenna

In the early twentieth century, multi-turn coils were used to provide the maximum voltage and high impedance across the tuning capacitor in order to directly drive the grid of the valve of a broadcast radio receiver. It was known that using a single-turn loop would reduce the effects of the proximity capacitance and reduce the amount of near field noise. However, in order to meet the impedance requirements for the input stage of the valve, a transformer would have been required, which would have added cost to the consumer product [see Blok and Rietveld (1955)]. The incorporation of a soft ferrite high-permeability core permitted a reduction in the diameter of the coil, which reduced the amount of proximity capacitance.

Advice on the use of soft ferrite cores can be found in Blok and Rietveld (1955) and Snelling (1988).

Soft ferrite has been successfully used to boost the performance of coil and loop antennas at frequencies up to VHF. However, the permeability of soft ferrite reduces, and the losses increase with increasing frequency (van der Zaag 1999), and beyond VHF air-cored loops and coils give the best performance.

Matching

Figure 8 shows some possible matching arrangements. It can be shown that most of these have very similar properties [see Massey (2009)], and therefore the choice of matching arrangement is determined by considerations other than which gives the best radio frequency (RF) performance. As examples:

- The direct connection across the tuning capacitor shown in Fig. 8a gives the highest impedance.
- The capacitor bridge shown in Fig. 8b is easily implemented as a circuit on a printed circuit board (PCB) to which the end of the loops is attached. See Fig. 9.
- Directly tapping across the loop or coil in Fig. 8c is very practical for experimental work, as the impedance match can be adjusted by moving the tap points. See Fig. 10.
- Inductively coupling into the loop using a small coupling loop in Fig. 8d builds the balun into the coupling structure. This is usually achieved at the cost of a small loss (~1 dB or less typically) due to the imperfect trapping of the magnetic flux between the antenna and the coupling loop.

The first three of these examples require either that they are connected to a balanced port or that a balun is used.

Full-Wave Loops

Introduction

The electrically small antennas of the last section always have to be tuned with a capacitor to make them work at the required frequency. This section discusses loops that are large enough to be self-resonant, i.e.,

DOI 10.1007/978-981-4560-75-7_26-1

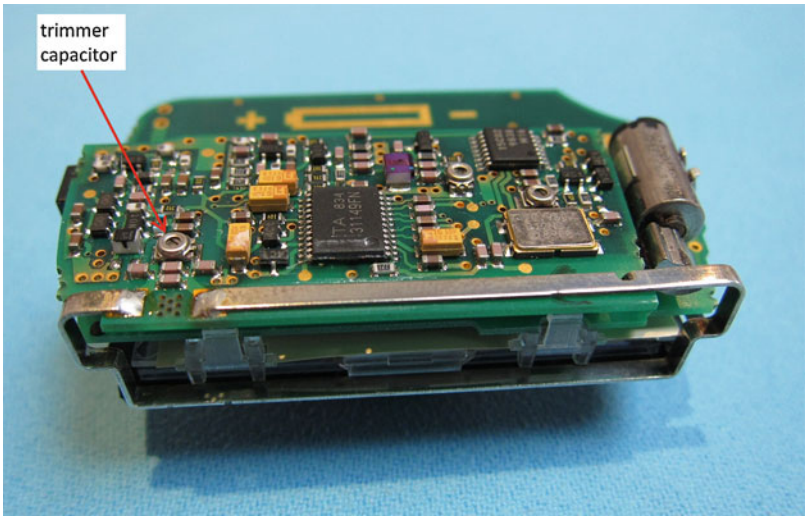


Fig. 9 Implementation of a capacitor bridge circuit in a pager receiver. The picture shows a radio pager with its casing and battery removed. The loop antenna is silver plated for maximum conductivity. The matching circuit is basically that shown in Fig. 8b, and the receiver has a balanced input port. However, in order to allow the tuning capacitance and matching capacitors to be mounted on the upper PCB, the ends of the loop are soldered to tracks on the PCB. The tuning capacitance consists of fixed capacitors in parallel with the trimmer capacitor. This is done because **a** fixed capacitors usually have greater Q than adjustable capacitors, and, **b** as most of the tuning capacitance is supplied by the fixed capacitors, the tuning range just covers the range necessary for adjustment (rather than a range of perhaps twice the center frequency, which would the case if the trimming capacitor was relied on to provide all the tuning capacitance). Therefore, the change in tuning frequency with rotation angle of the trimmer is less, and it is easier to bring the antenna into tune. The vibrator motor and its weight to the right of the PCB are mounted with the weight nearest the loop. This is because the permanent magnet in the motor introduces significant losses if placed close to the loop antenna

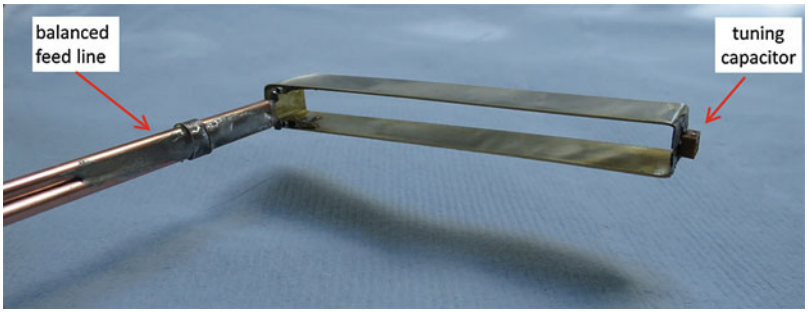


Fig. 10 Directly tapping across a loop using a balanced feed. The balanced line consists of two 50 Ω coaxial semirigid cables soldered together. The inners of the cables are fed from a 3 dB 180° splitter that converts an unbalanced 50 Ω feed into two 50 Ω signals that are 180° out of phase with each other, giving a balanced impedance of 50 + 50 = 100 Ω. To enable easy connection and disconnection between line and antenna, some sockets have been soldered onto the antenna, and wires which fit into the sockets have been soldered to the ends of the inner conductors of the coaxial cables. This arrangement is used for measuring the performance of different shapes of loops and for measuring the effects of bringing components close to the loop. The single adjustable tuning capacitor gives a very large tuning range but is very sensitive. For mass production, tuning using fixed capacitors in parallel with an adjustable capacitor is preferred. See Fig. 9

have a finite resistance and very low reactance between terminals that are across a gap in the loop. As the loop size is increased, self-resonance first occurs when the perimeter of the loop is around one wavelength long, i.e., the full-wave loop. In contrast to electrically small loops, full-wave loops generally have good efficiency and can be used for efficient transmission as well as for reception.

DOI 10.1007/978-981-4560-75-7_26-1

This section is arranged as follows. Subsection "Resonant Full-Wave Loop" discusses the basic full-wave loop structure, emphasizing that its radiation resistance is typically ~200 or more Ohms. The subsequent subsections detail how other features can be added to optimize full-wave loop behavior and tune the performance.

Subsection "Resonant Full-Wave Loop in Close Proximity to Metallic Structures" describes how the full-wave loop's natural resonant impedance can be reduced to values suitable for directly connecting to low-impedance (e.g., 50 Ω) ports by placing the loop around a conducting structure. To demonstrate the principle, a flat conducting structure, such as the electronics of a tablet computer, is used as an example. Then methods for frequency tuning (subsection "Tuning Resonant Loop Antennas") and creating circular polarization (subsection "Circularly Polarizing the Resonant Loop Antenna") are demonstrated.

Next, parasitic loops are introduced to modify the frequency coverage (subsection "Dual Banding the Resonant Loop Antenna"). These can be combined with the method for creating circular polarization as described in subsection "Dual Banding and Circular Polarizing the Resonant Loop Antenna."

The remaining subsections are devoted to describing methods for creating directive antennas with loops. Firstly, a loop over a ground plane is considered in subsection "Resonant Loop Over Planar Reflector." Subsection "Directive Resonant Loop Array" discusses the spacing of parasitic elements to create a Yagi-Uda like array with high directivity.

Another method of creating an efficient and relatively compact directive antenna is to mount shared feed loops about a quarter wavelength above a ground plane. Both linear (also known as biquad, described in subsection "Dual Rhombic/Circular Loop Antenna Over Planar Reflector (Biquad Antenna)") and circular polarized versions (described in subsection "Circularly Polarized Dual Rhombic/Circular Loop Antenna Over Planar Reflector") are detailed. Finally, methods for bandwidth extension are described in subsection "Wideband Circularly Polarized Dual Rhombic Loop Antenna with Parasitic Loops."

Resonant Full-Wave Loop

A full-wave loop structure can be almost any shape, for example, circular, square, rhombic, and triangular, as illustrated in Fig. 11. The full-wave loop can be constructed in three dimensions (e.g., from wire or a tube); it can have a planar structure, for example, being manufactured from a sheet of metal or etched onto the conductor of a printed circuit board. The full-wave resonant loop antenna is often approximated by a pair of (often bent) half-wave dipoles connected in parallel, as their current and voltage distributions are comparable to the full-wave loop. The dipole pairs resonate in phase and are connected at their tips where the current is almost zero. In Fig. 11, these points are denoted by red dots and labeled as $I_{\min}$.

Resonant Full-Wave Loop in Close Proximity to Metallic Structures

The moderately large characteristic impedance of the resonant loop at its feed point (denoted as the loop characteristic impedance from here onward) can be exploited to create an efficient radiator of significant bandwidth and low volumetric profile, by intentionally coupling the loop to surrounding structures such as circuit boards or metallic cases. To explain how this works, first consider a full-wave large loop antenna in isolation designed to operate at 500 MHz, fed from a 50 Ω feed. The top left diagram in Fig. 12 shows a rectangular loop example. In the figure, the loop is a track shown in yellow against an FR4 printed circuit board shown in green, and the location of the antenna feed is shown in red. The printed circuit board is 200 mm long by 100 mm high.

At 500 MHz, it can be seen that the feed impedance is over 150 Ω, the return loss is > -5 dB, and the realized gain is suboptimal at about +1 dBi. The large mismatch between the antenna impedance and the 50 Ω impedance of the feed accounts for a significant proportion of the low realized gain. (Realized gain is defined as radiated power in a given direction compared to the power radiated by an ideal perfectly

DOI 10.1007/978-981-4560-75-7_26-1

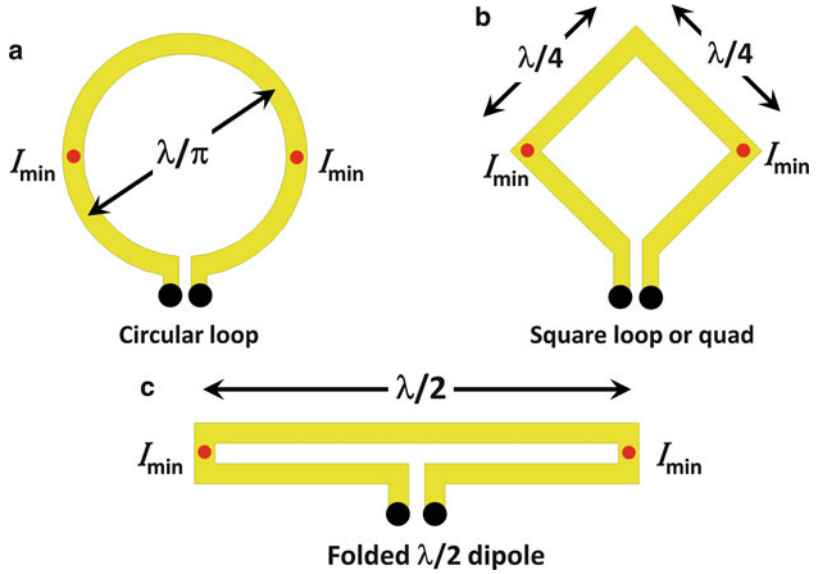


Fig. 11 Some examples of full-wave loop antennas. Around resonance, the radiation resistances of the circular and square loop antennas are approximately 200 Ω, and the radiation resistance of the folded half-wave dipole is approximately 300 Ω. The directivity is usually at a maximum perpendicular to the antenna plane (perpendicular to the plane of the figure). For all the antennas depicted, the radiation pattern is doughnut shaped, with its axis of minimum radiation running horizontally across the figure. The radiation varies slightly with shape, and consequently the maximum directivities are around 1.5 (1.76 dBi) for the circular loop, 1.7 (2.3 dBi) for the square loop, and 1.64 (2.15 dBi) for the folded half-wave dipole

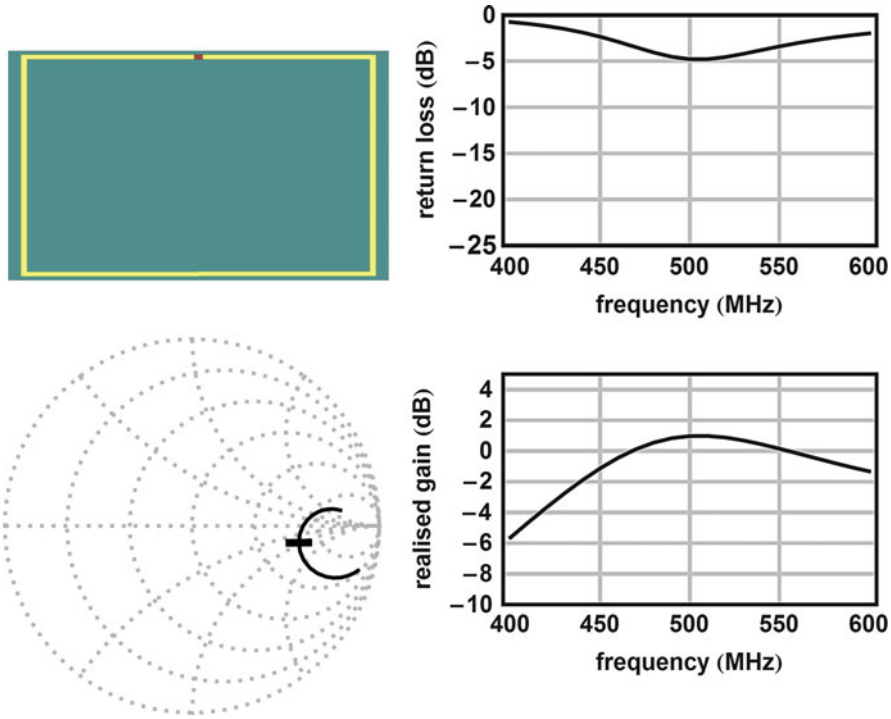


Fig. 12 Basic rectangular full-wave loop on FR4 sheet. The figure shows, in clockwise order from top left, the antenna (see text for description), the return loss response, the realized gain perpendicular to the antenna's board, and the reflection coefficient response between 400 and 600 MHz on a Smith chart whose center is normalized to 50 Ω. In the Smith chart, a marker has been put at 500 MHz

DOI 10.1007/978-981-4560-75-7_26-1

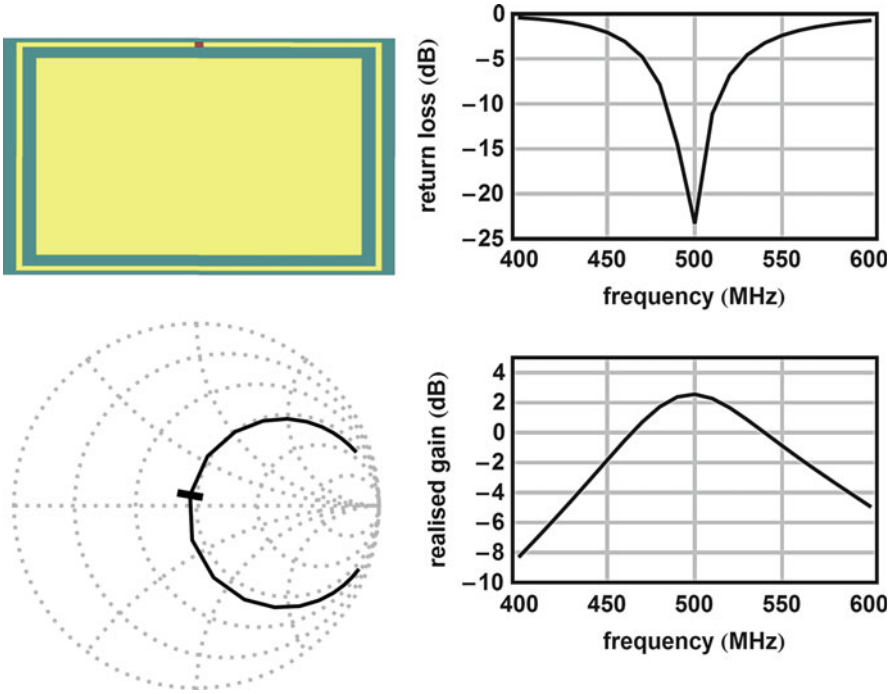


Fig. 13 Basic full-wave loop on FR4 sheet with impedance modification. For key to diagrams, see text below Fig. 12

matched ideal isotropic antenna fed with the same signal.) It should also be noted that the size of the antenna at 500 MHz is such that it takes up a significant area.

Now consider the same antenna operating in the presence of other components. As an example, Fig. 13 shows a copper ground plane (represented by the large yellow rectangle) filling much of the antenna's interior, which in this case represents a circuit board. However, this structure could also be a metallic case or other component part which is conductive in nature. The coupling which occurs between the loop and in this case the copper ground plane results in modification of the loop characteristic impedance. It is therefore possible to use this additional structure to improve the loop antenna's performance; hence, it is considered to be a cooperative structure.

Considering the example of Fig. 13 and comparing the responses to Fig. 12, it can be seen that the addition of the ground plane has led to a significant modification in loop characteristic impedance. The resulting impedance is now much closer to the desired 50 Ω of the feed, and thus the cooperative structure results in an improvement in both match and overall performance of the antenna. Examining the remaining results of Fig. 13 shows that at the design frequency of 500 MHz, the resonance is significantly stronger resulting in a return loss of < -20 dB. The realized gain has also increased to over 2.5 dBi.

Application Example

Applying the principle of what is illustrated in Fig. 13 to a real-life example, consider a small tablet computer. For this example, a single-turn resonant loop antenna has been inserted around the perimeter of the tablet. For this instance, the loop antenna is designed to operate at around 400 MHz. The loop has been positioned to closely couple to the remaining structures in the tablet, in order to modify the impedance appropriately. A simplified computer aided design (CAD) view of the antenna incorporated into the tablet computer is illustrated in Fig. 14, where the loop is highlighted in yellow and closely follows the perimeter of the tablet computer's case.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_26-1

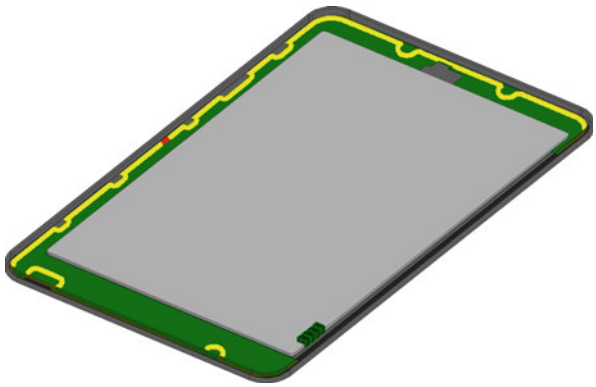


Fig. 14 Resonant loop antenna around the screen of a tablet computer. This diagram is taken from CAD file. The tablet's rear cover has been removed to show the loop, which is highlighted in *yellow*, and its width has been exaggerated (doubled) to aid visibility. The loop travels around the interior of the perimeter of the tablet's PCB but is hidden from view along most of the two nearest sides by the lip of the tablet's front cover. The feed location is shown in *red*. A *silver-colored rectangular sheet* represents the aluminum can covering the screen driver circuitry. The tablet's screen cannot be seen as it is on the far side of the view. See Fig. 15 for a more detailed CAD model with the main board populated with circuitry and battery and Fig. 16 for the real-life demonstration piece and its measured results

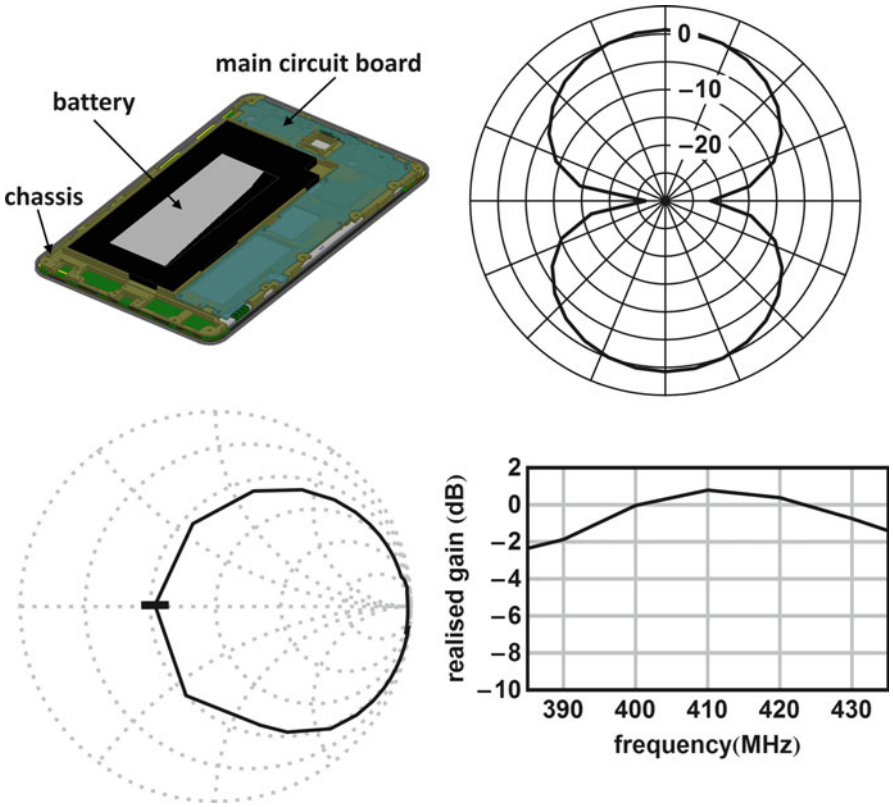


Fig. 15 Simulation results of a full-wave antenna around the screen of a tablet computer. The figure shows, in clockwise order from *top left*, (*i*) the CAD model of the tablet computer with its back removed, (*ii*) the realized gain pattern in the plane perpendicular to the plane of the tablet which cuts the tablet along its long axis, (*iii*) the realized gain in the direction perpendicular to the tablet plane versus frequency, and (*iv*) the S11 response between 380 and 440 MHz plotted on a Smith chart whose center is normalized to 50 Ω. In the Smith chart, a marker has been at 410 MHz, which corresponds to the resonant frequency realized by this antenna configuration

DOI 10.1007/978-981-4560-75-7_26-1

For this demonstration example, the loop has been designed to operate in the presence of the touch screen, battery, circuit board, and the rear case containing Wi-Fi antennas and a large NFC (Near Field Communication) antenna.

The top left-hand corner in Fig. 15 shows a more detailed CAD image of the tablet computer with the main board populated with circuitry and battery. The "chassis" component shown in Fig. 15 was in its original unmodified incarnation manufactured in a lightweight metallic alloy; however, to implement the perimeter loop antenna, a structural plastic has been used for this component, which prevents shorting out of the loop antenna's fields. The remainder of Fig. 15 shows the simulated results obtained from the installed antenna.

In this demonstration example, and without additional matching, the impedance of the antenna is approximately 25 Ω at the center frequency (410 MHz), and the realized gain is 0.7 dBi, with a −3 dB gain bandwidth in excess of 50 MHz. Four hundred and ten megahertz corresponds to the natural resonant frequency for this particular antenna. Tuning to other frequencies is possible and is discussed in subsection "Tuning Resonant Loop Antennas."

Tested Example and Range Plots

A real-life demonstration example was manufactured consisting of a tablet computer with a retro-fitted loop antenna, as illustrated in Fig. 16a. The tablet computer is shown with its back cover and NFC antenna removed for clarity. The realized gain of the installed antenna configuration is measured and illustrated in Fig. 16b. Figure 16a also indicates the orientation of the axes used in the measurement. For the purposes of the measurement, the antenna is mounted "upside down" compared to the image in Fig. 16a so that its test cable hangs down and runs to the chamber's feed cable without crossing the loop antenna.

The actual resonant frequency of this structure occurs at 425 MHz which is slightly higher in frequency than predicted. This is because several simplifications have been made in the simulation, in order to be able to realize an initial design. Much of the detail, which is present in practice, has not been replicated in simulation. However, the measured magnitudes of the return loss and gain responses (not shown) are similar to those of the simulated model. For the purposes of demonstrating the principle of cooperative structures, the realized gain patterns at 425 MHz are illustrated in Fig. 16b.

It can be seen from both the simulation results in Fig. 15 and the measurement results in Fig. 16 that the full-wave loop structure with its naturally high impedance is ideally suited to applications where an antenna is asked to coexist with metallic structures. These metallic structures can be used in a cooperative manner with loop structures to enable a good installed antenna performance. For other more conventional antennas such as patches or printed inverted F antennas (PIFAs), the presence of metallic structures has a more detrimental impact on the installed antenna performance.

Tuning Resonant Loop Antennas

It has already been shown that a resonant loop antenna can be considered as behaving like a pair of connected dipole antennas (see subsection "Resonant Full-Wave Loop"). This concept can be thought of as the motivation for the following strategy to reduce the resonant frequency of a large loop antenna. The two high-current nodes of the loop antenna are in the centers of its constituent dipole antennas. As illustrated in Fig. 17, i.e., they are at the feed and at a point about halfway around the antenna from the feed.

Placing an inductor at the high-current node halfway around the loop from the feed increases the electrical length of the antenna and lowers its resonant frequency. An example of this arrangement is shown in Fig. 18. If the inductor had been placed closer to the left or right sides of the example in Fig. 18, its effects upon the current distribution and electrical length would have been less.

DOI 10.1007/978-981-4560-75-7_26-1

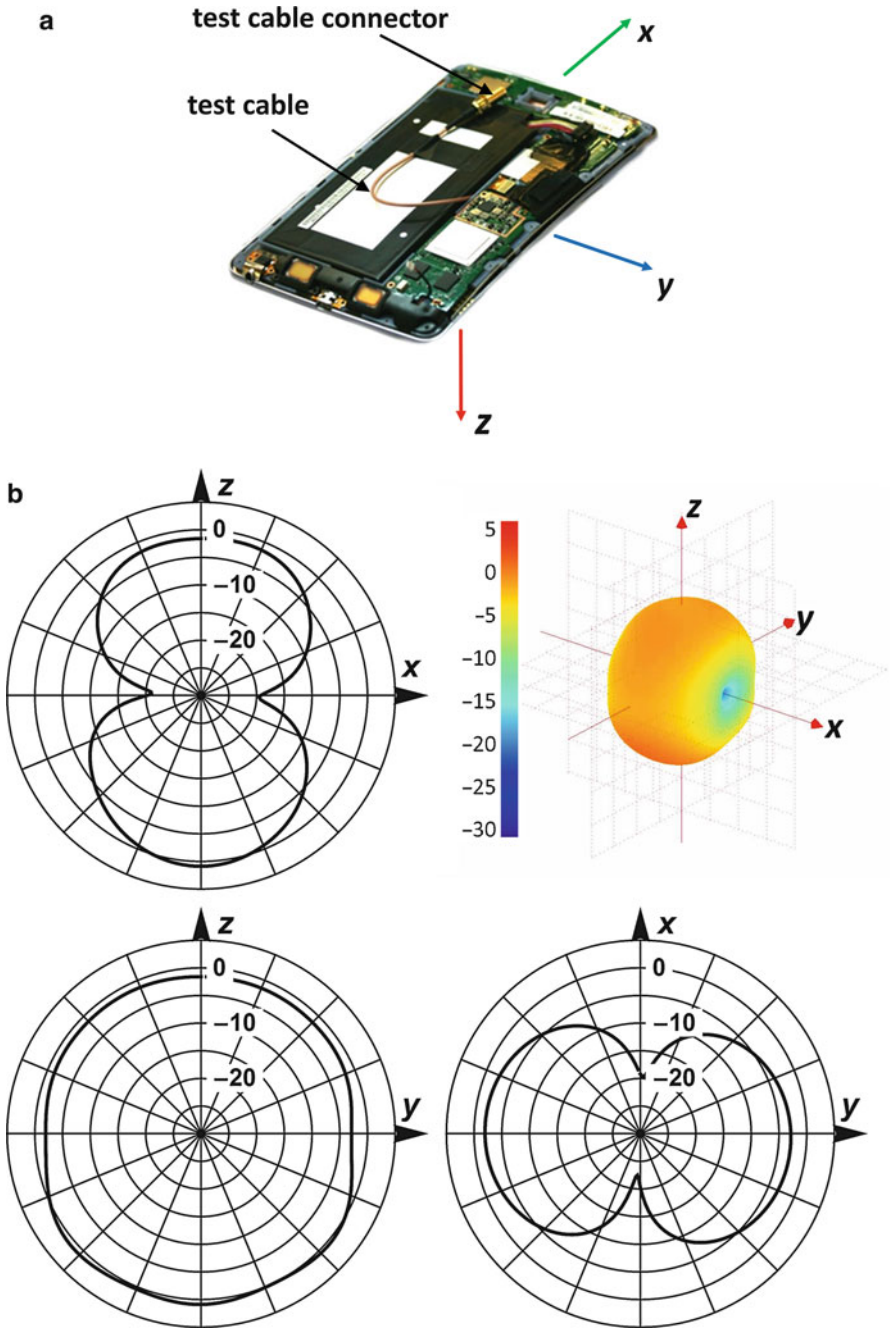


Fig. 16 Tablet computer and its radiation measurement results. See text for description of the figure

The effect of inductor values upon the return loss of the loop is illustrated in Fig. 19. A change in the resonant frequency of the antenna is clearly seen.

It can be seen in Fig. 19 that the impedance modified full-wave loop structure responds well to inductive loading, retaining useful matched bandwidth at all illustrated tuned frequencies. Of course, in a real-life example, the loop to structure coupling would be set with a best match at the center of the tunable frequency range in order to ensure the best possible match at all operating frequencies.

DOI 10.1007/978-981-4560-75-7_26-1

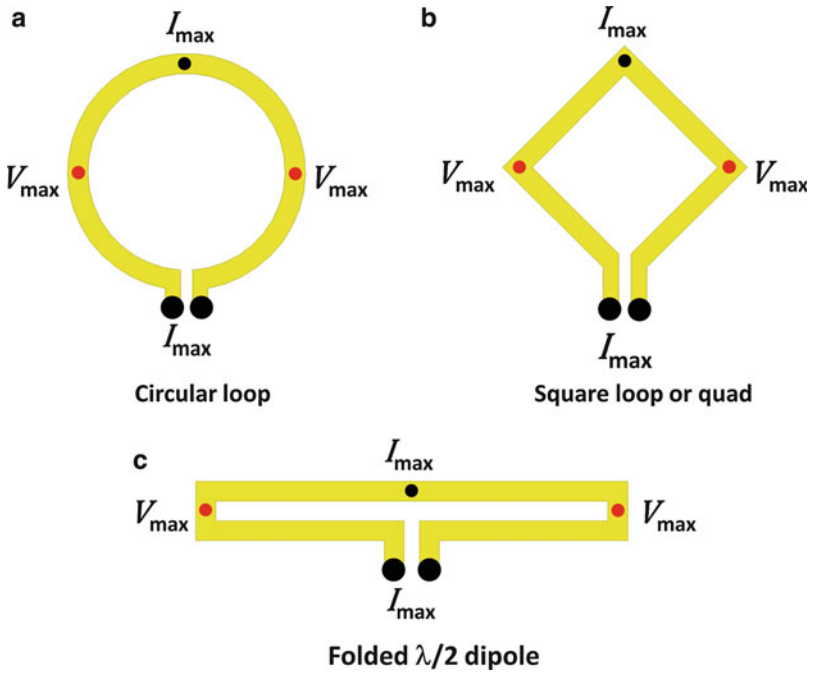


Fig. 17 High voltage and current points on full-wave dipole antennas

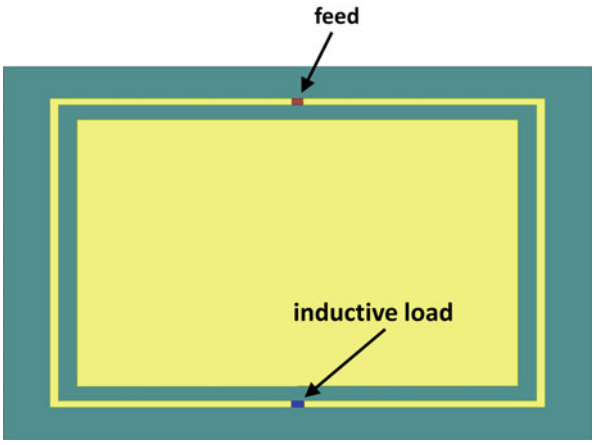


Fig. 18 Location of inductive load at high-current node

Circularly Polarizing the Resonant Loop Antenna

The resonant loop antenna can be further adapted to include a circular polarization response by the inclusion of capacitive gaps in the tracking as illustrated in the example of Fig. 20. These capacitive breaks in the loop structure give rise to a traveling wave current distribution of constant amplitude and linearly changing phase. Circular polarization is normally achieved by breaking the tracking in one place in the loop. For the case shown in Fig. 20, the single break version would retain the break that is closest to the feed point. However, Fig. 20 shows a variation with a pair of breaks as this allows the designer greater scope for adjustment. In this uncluttered example, the antenna's loop is almost exactly one wavelength in circumference, and the capacitors' respective positions are typically at −1/6 λ and +1/3 λ from the feed point, i.e., λ/2 apart. It is important to note that these locations will be significantly influenced by the

DOI 10.1007/978-981-4560-75-7_26-1

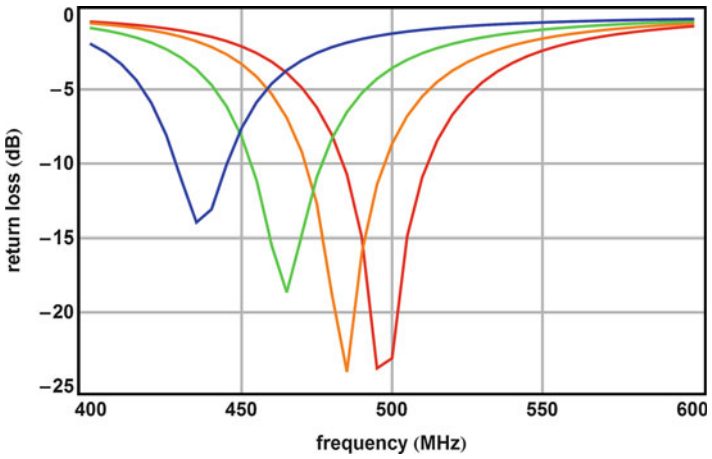


Fig. 19 Tuning response of a resonant loop antenna with switched inductors. Key, 0.1 nH response is in *red*, 10 nH response is in *orange*, 32 nH response is in *green*, and 47 nH response is in *blue*

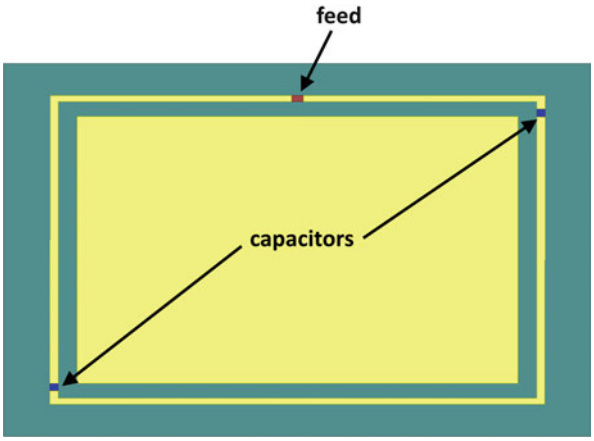


Fig. 20 Capacitor placement for a circularly polarized loop antenna

coupling of the loop to the structure and also by the impact of the substrate and surrounding materials, so should be taken as guidance only.

It is possible to flip the polarization between right and left handed by mirroring of the capacitive breaks in the loop structure from the lower to the upper half of the loop and vice versa. The circularly polarized response can be tuned simply by varying the value of the capacitance and is achieved electronically using varactor diodes or similar devices. The gain and reflection coefficient plots for the circularly polarized loop illustrated in Fig. 20 are shown in Fig. 21.

Dual Banding the Resonant Loop Antenna

One method for dual banding the resonant loop antenna is to proximity couple one or sometimes more secondary loops to the primary driven loop. Secondary loops can be placed either inside, outside, or both inside and outside the driven loop and effectively act as parasitic resonant structures. The example shown below builds on the cooperative structure-loaded loop described at the start of subsection "Resonant Full-Wave Loop in Close Proximity to Metallic Structures" and uses a secondary loop outside the primary loop that has been inductively loaded to further reduce the frequency of the secondary band.

Note that the secondary loop is coupled predominantly along the high-current sections of the antenna, with the high-voltage (low current) sections to the driven loop more widely separated from the secondary

DOI 10.1007/978-981-4560-75-7_26-1

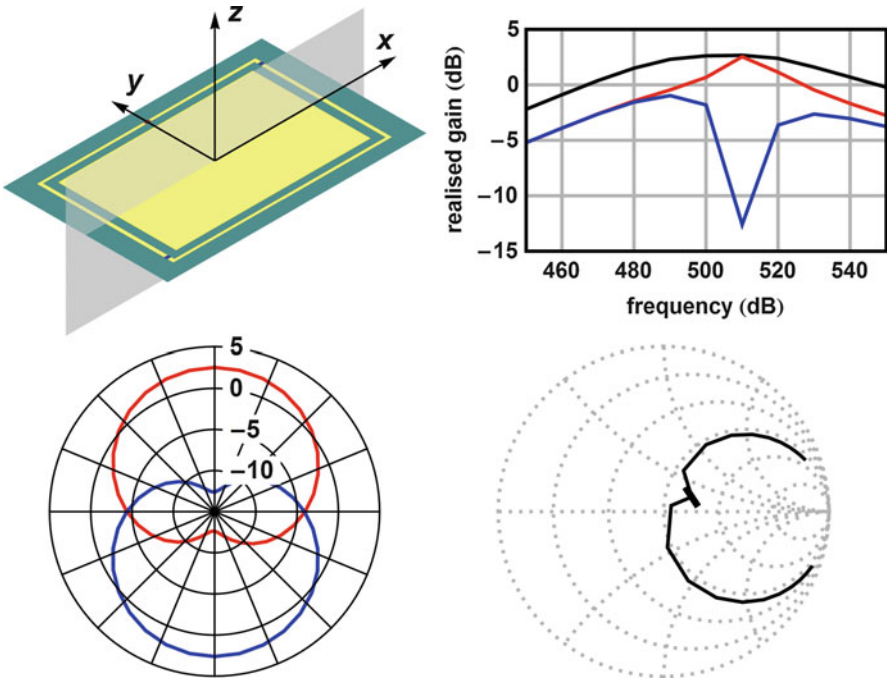


Fig. 21 Circularly polarized response of the capacitive loaded resonant loop antenna. The figure shows, in clockwise order from *top left*, (***i***) a three-dimensional image of the antenna that is shown in plan view in Fig. 20, with the cut plane for the radiation patterns in (***iv***) shown in *gray*; (***ii***) a plot of realized gain in the z-direction versus frequency; (***iii***) the reflection coefficient response plotted from 450 to 550 MHz on a Smith chart whose center is normalized to 50 Ω; and (***iv***) the right- and left-hand circularly polarized components of realized gain at 510 MHz in the cut plane shown in *gray* in (*i*). *Blue* represents left-hand circularly polarized and *red* represents right-hand circularly polarized. In (*ii*) the total realized gain is shown in *black*. In the Smith chart, the *curve* runs from 450 to 550 MHz, and a marker is placed at 510 MHz

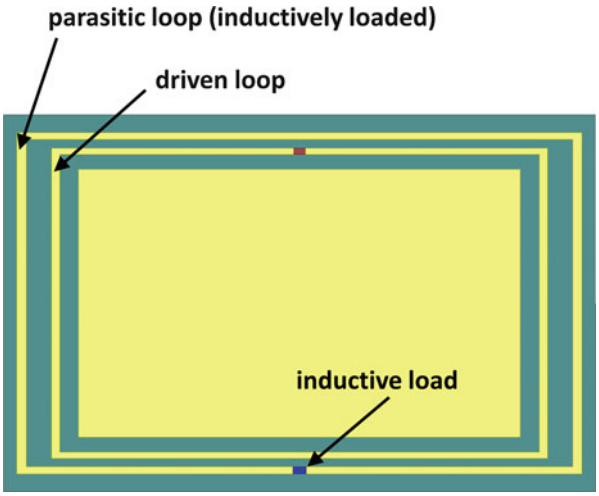


Fig. 22 Dual banding the resonant loop antenna

loop; this prevents the collapse of the electric field at the high-voltage nodes of the primary loop. Figure 23 illustrates the realized gain of the dual loop structure, compared with the single-loop response taken from Fig. 13.

It can be seen in the gain versus frequency plot in Fig. 23 that the lower-frequency response is typically of a higher quality factor (has a narrower peak) than that of the higher frequency band. This is due not only

DOI 10.1007/978-981-4560-75-7_26-1

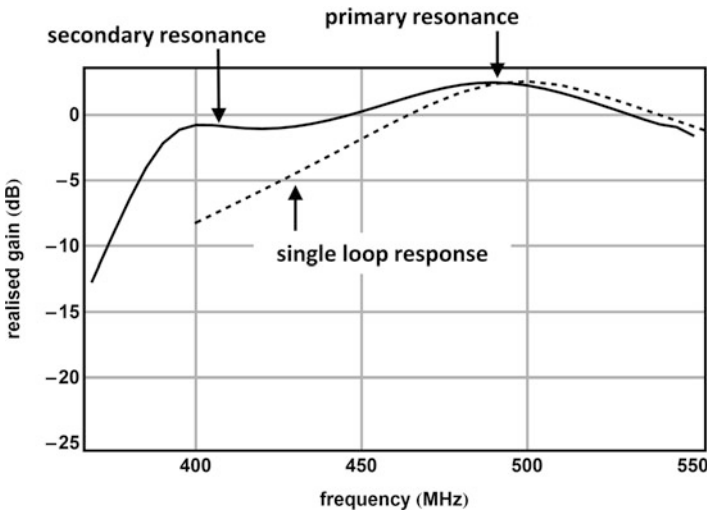


Fig. 23 Realized gain of single and dual loops showing the two resonances realized by the dual loop configuration shown in Fig. 22. The single-loop response is taken from the results shown in Fig. 13

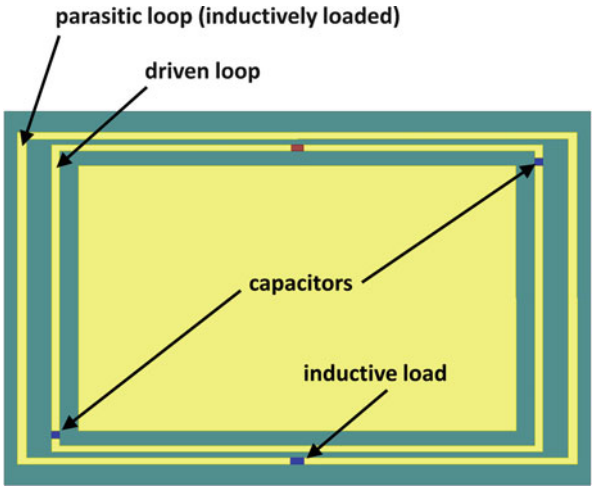


Fig. 24 Dual banding and circularly polarizing the resonant loop antenna

to a reduction in the antenna's electrical size (size in free-space wavelengths) at lower frequencies but is also a function of the coupling action between the loops.

Dual Banding and Circular Polarizing the Resonant Loop Antenna

It is possible to combine two or more of the methods described in the preceding subsections, namely, inductive loading, dual banding, and circularly polarizing. An example is illustrated in Fig. 24. It should be noted here that the parasitic loop has not been capacitively loaded, as the circularly polarized wave from the primary loop couples strongly into the secondary (parasitic) loop. In other instances to obtain circular polarization where the coupling is less strong, it will be necessary to capacitively load the parasitic loop.

Figure 25 illustrates the gain versus frequency response of the antenna illustrated in Fig. 24. The gain is reported for the direction out of the plane of the antenna in Fig. 24 towards the viewer. If the plot were done for the opposite direction into the plane of the antenna, the red and blue traces would be interchanged.

DOI 10.1007/978-981-4560-75-7_26-1

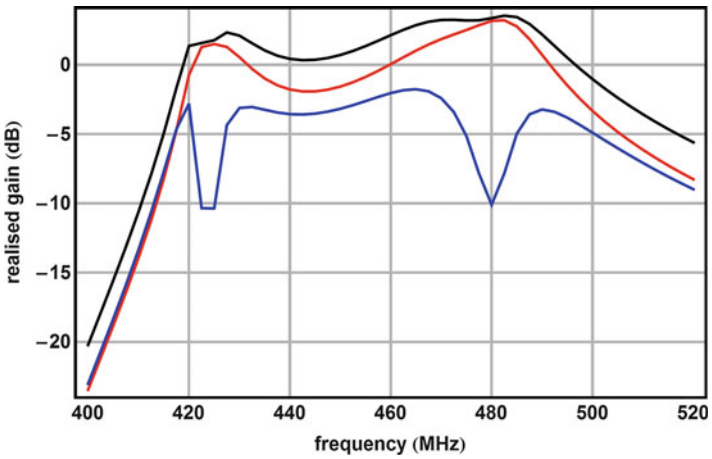


Fig. 25 Dual band circularly polarized response of the primary and secondary loops. Key, *black* = total gain, *blue* = left-hand circular polarization, and *red* = right-hand circular polarization

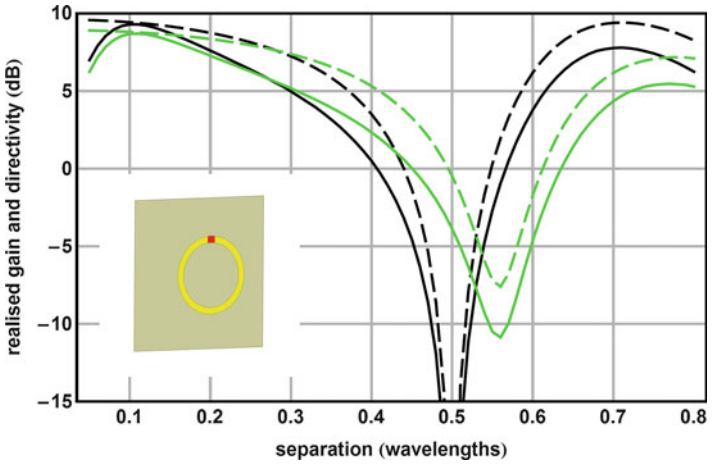


Fig. 26 Directivity and realized gain of resonant loop over infinite ground plane and over planar reflector. See text for the key to the plots and for details of the antenna and reflecting plane

Resonant Loop over Planar Reflector

The full-wave resonant loop element is often used over a ground plane in order to increase its performance in both directivity and gain. The moderately large real input impedance of the loop element in free space is useful, as the proximity of the reflector can be used to adjust the feed impedance of the antenna without adjusting the loop circumference. This allows easy matching to 50 Ω, and thus to being fed from a standard coaxial cable. Figure 26 shows the gain and directivity perpendicular to the reflecting plane, where the radiation is plotted for the loop side of the reflecting plane, in the direction perpendicular to the plane. The directivity is depicted using dashed lines and the realized gain with solid lines. As can be seen in Fig. 26, the single-turn resonant loop has a significantly improved directivity when placed over a reflecting surface.

In Fig. 26, the response over an infinite ground plane is shown in black, and the response over a 2λ/3 long square planar reflector is shown in green. The loop is a flat ring with interior diameter of 0.28λ and exterior diameter of λ/3. The loop and reflecting surface have parallel planes, which are separated by the separation distance, shown on the figure's horizontal axis. Some interesting phenomena can be seen in Fig. 26:

DOI 10.1007/978-981-4560-75-7_26-1

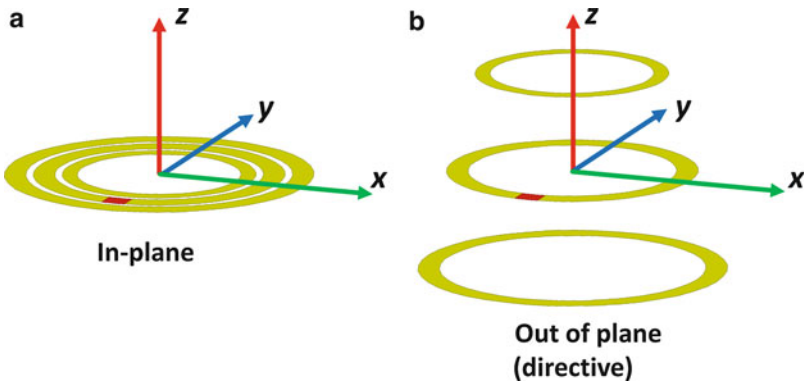


Fig. 27 In-plane coupled loops and the directive out-of-plane derivative

- The loop is best matched to 50 Ω when it is around 0.11 wavelengths from the reflecting surface.
- At small separations, the realized gains and the directivities of the infinite plane and finite reflector cases are very similar, but for a given separation the infinite plane values are slightly higher.
- At half-wavelength separation, the radiation vanishes for the infinite ground plane.
- The radiation of the finite ground case has a minimum at a slightly higher separation but never completely vanishes.
- With increasing separation, the next maximum in gains and directivities is at just over 0.7λ separation. For the infinite ground plane, its magnitude is similar to that of the 0.11 wavelength peak, but in the finite reflector case, it is somewhat reduced.

Directive Resonant Loop Array

So far, only coupled loops lying in the same plane have been discussed, such as that shown on the left-hand side in Fig. 27, which shows a central loop as the driven element, with the outer loop being the low-frequency parasitic element and the inner loop being the higher-frequency parasitic element. Moving the parasitic loops away from the driven loop, as shown on the right-hand side in Fig. 27, gives opportunities for enhanced gain. In a manner similar to parasitic elements in conventional Yagi-Uda arrays of dipoles, the lower-frequency loops become reflectors and the higher-frequency loops act as directors. The reflector can also take the form of a plate (as illustrated in Fig. 26).

It is possible to give some typical rule of thumb sizes for the circumferences of the different elements in order to achieve a directive resonant loop array. They are:

- Reflector: 1.05 wavelengths
- Driven element: 1 wavelength
- Directors: 0.95 wavelengths

The reflector to driven element spacing and driven element to director spacing are highly dependent on the number of elements, and the spacing follows closely the pattern set for the design of a conventional Yagi-Uda antenna of linear elements. In the Yagi-Uda configuration, the reflector is separated from the driven element by approximately 0.15 wavelengths, and there is approximately 0.2 wavelength separation between both adjacent directors and also between the driven loop and its nearest director.

Dual Rhombic/Circular Loop Antenna over Planar Reflector (Biquad Antenna)

The biquad takes several forms, the most notable of which are the parallel- and series-fed variety, shown in Fig. 28. These antennas have a higher gain than the equivalent area patch antenna, and, because of their

DOI 10.1007/978-981-4560-75-7_26-1

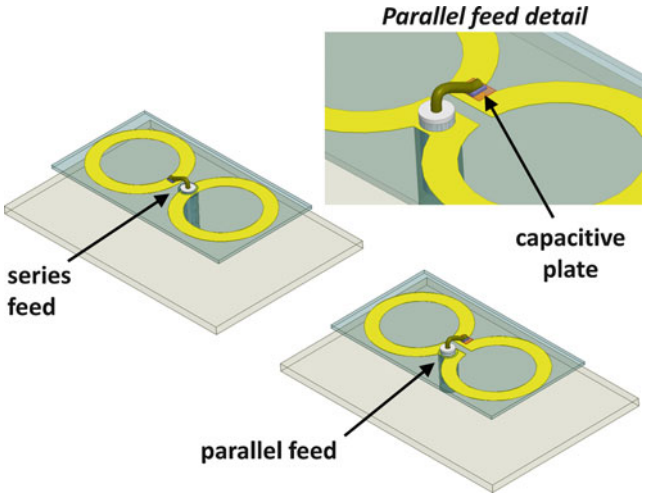


Fig. 28 Series- and parallel-fed versions of the biquad antenna

depth (i.e., distance between the loops and their ground reflector, which is set to optimize the feed impedance), they have significant bandwidths. This coupled with their ease of construction has contributed to their widespread adoption by the hobbyist radio community for the Wi-Fi band. Each loop is typically one wavelength in circumference (measured around the outside of the loop) at the lowest design frequency, with the inner circumference setting the higher part of the resonance band.

Figure 28 also shows two possible feed configurations: a directly connected feed as shown in the series configuration and a capacitive feed as illustrated in the expansion of the parallel configuration. In the capacitive feed arrangement, the center pin of the feed coax is connected to a small capacitive plate, which is located on the opposite side of the board to the metallization which forms the loop (underside of the PCB in Fig. 28). Thus the loop is fed by capacitive coupling to this plate. This configuration provides additional capacitance at the feed point which can be useful for canceling out the additional inductance which results from the exposed center conductor of the feed cable. This additional inductance can be particularly substantial at higher frequencies; thus the capacitive feed variant is highly applicable here. It should be noted that either feed configuration can be used with both the series and parallel versions of the biquad.

Circularly Polarized Dual Rhombic/Circular Loop Antenna over Planar Reflector

The rhombic version of the full-wave loop antenna has four linear sides each measuring about 0.3 wavelengths long. To induce circular polarization, capacitive breaks are introduced into the loop structure, similar to those discussed in subsection "Circularly Polarizing the Resonant Loop Antenna," and a traveling wave current distribution of constant amplitude and linearly changing phase is realized. Capacitor position is typically at $\pm 1/6$th loop length from the feed point (depending on the polarization required). The capacitance can be realized by breaks in tracking or wiring or by using lumped components depending upon construction techniques being employed.

Two circularly polarized rhombic loops are combined to form a dual rhombic loop. The antenna is illustrated in Fig. 29, which shows two possible feed configurations. For clarity, the examples are depicted without showing the ground plane that lies below the loops.

In Morishita et al. (1998), axial ratio bandwidths (<2 dB) of more than 20 % are achieved for both parallel and series feed antennas, with gains of 10 dBi for the parallel-fed version and 11 dBi for the series-fed antenna.

DOI 10.1007/978-981-4560-75-7_26-1

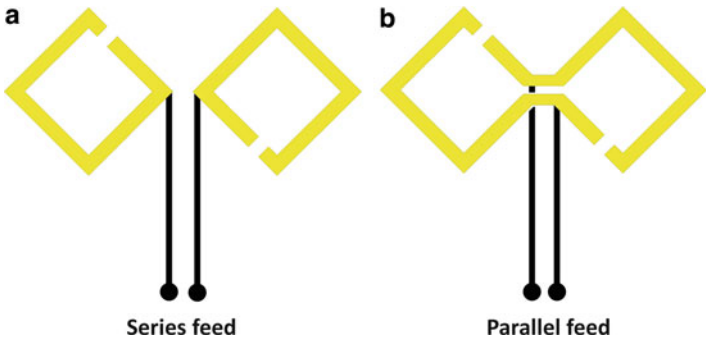


Fig. 29 Circularly polarized biquad concept. See text for description

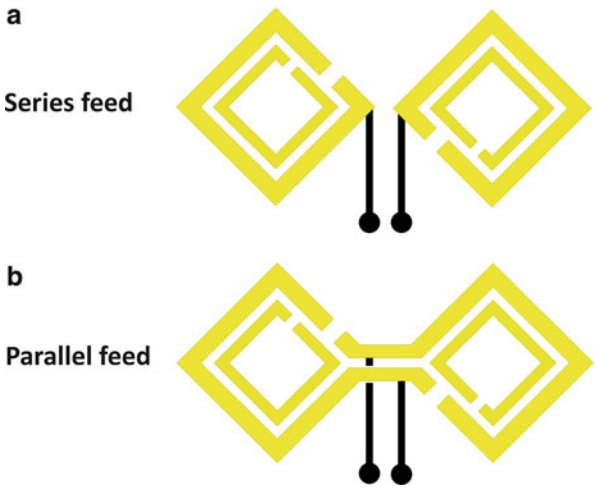


Fig. 30 Wideband circularly polarized resonant loop antennas

Wideband Circularly Polarized Dual Rhombic Loop Antenna with Parasitic Loops

This antenna design expands on the concept of the circularly polarized dual rhombic loop discussed above and introduces a set of smaller parasitic elements within the main driven loops, shown in Fig. 30 (see Li et al. 2005). The parallel-fed structure has a significantly wider circular polarization bandwidth (measured as the −2 dB axial ratio) of 50 %, compared to 25 % for the series-fed arrangement.

Reconfigurable Multi-feed Loops and Their Arrays

Introduction

The previous section on full-wave loop antennas and their derivatives described how loops can be fitted around or near the conducting structures of a device to increase efficiency while adding very little extra volume to the device. This section describes another way in which space can be saved, by reusing the radiating structure of a loop antenna for several different radiation patterns by switching between feeds. In order to have a directive high-gain steering performance, the loop is placed above a ground plane. Further, this section discusses how high-impedance surfaces can be used to reduce the distance between the loop and the ground plane, while keeping good efficiency and controlling the beam shape. Finally, the latest research on array of beam steering loop antennas for achieving high-gain scanning range will be presented.

DOI 10.1007/978-981-4560-75-7_26-1

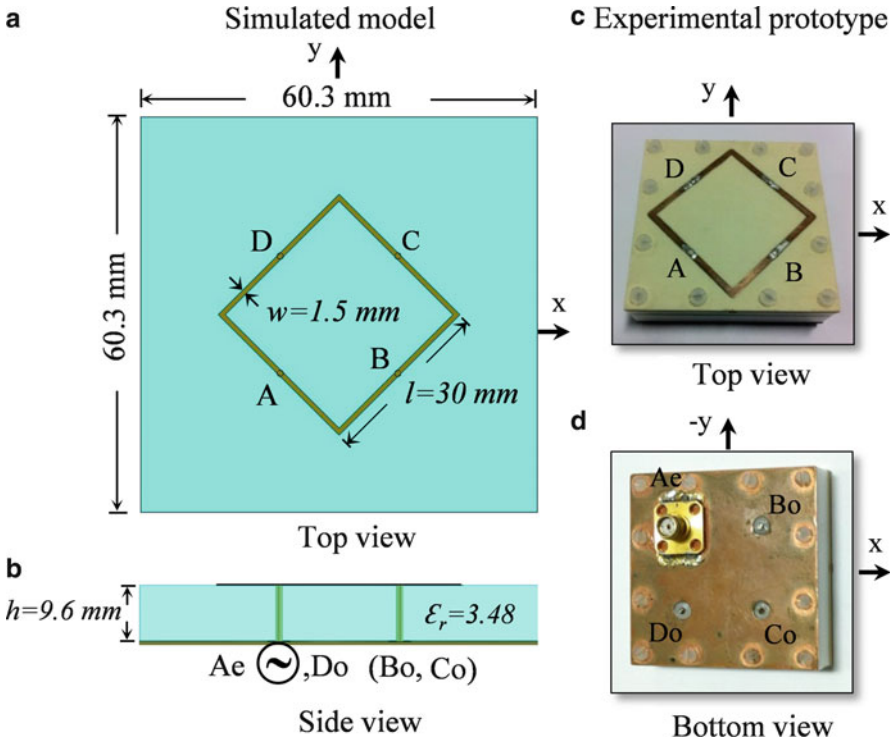


Fig. 31 Square loop antenna with four feeding ports. The simulated antenna is shown to the left, and the experimental prototype is shown on the right. (a) is a top view of the simulated antenna, (b) is a side view of the simulated antenna, (c) is a top view of the experimental antenna, and (d) is a bottom view of the experimental antenna. (d) shows one active SMA feed and three other open circuited feeds

Background

As mobile communications migrate up from 4G systems to 5G systems and beyond with a promise of enabling consumer end devices (phones, tablets, etc.) to work in excess of 1 giga bit per second, there is a tremendous focus on the development of beam steering antenna technology in order to solve some of the channel problems. The pattern reconfigurable antennas have the ability to generate multiple radiation patterns and are capable of directing the beam toward the intended direction of signal arrival and evade the noise sources. From a transceiver perspective, this means a higher signal-to-noise system capable of offering a higher data throughput capacity. The follow-on benefits include mitigation of multipath effects, reduction in transmitted power levels, longer battery life, range enhancement, reduction in co-channel interference (jamming avoidance), and faster handover. Single-element reconfigurable antennas as presented here are of special significance as they do not require multiple antenna elements with complex phase shifters as in the case of a conventional phased array antenna (Liberti and Rappaport 1999; Pozar 1986). For a near equivalent performance scenario, a single-element adaptive antenna compared with a conventional phased array antenna normally is 75 % smaller, ten times faster, and 95 % cheaper.

Early research and development on single-element beam steering/beam switched/pattern reconfigurable antennas started around 2003–2004 (Huff et al. 2003; Mehta and Mirshekar-Syahkal 2004). Various architectures (Huff and Bernhard 2006; Jung et al. 2006; Mehta et al. 2006), mainly traveling wave types, to realize the single-element steerable antennas were explored. These antennas employed multiple switches to vary the current distribution along the peripheral length for achieving reconfigurable radiation patterns. These early designs had three major restrictions. Firstly, due to the different current distributions, the antenna polarization differs from one switch configuration to another,

DOI 10.1007/978-981-4560-75-7_26-1

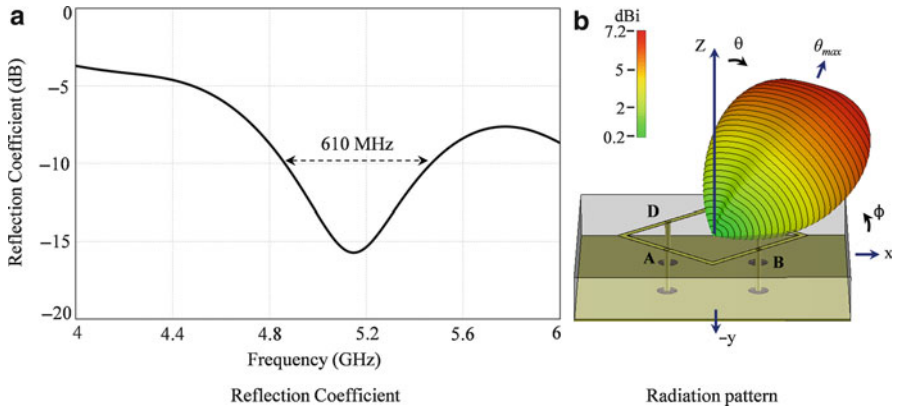


Fig. 32 Frequency response of reflection coefficient of the SLA and radiation pattern at 5.2 GHz

which in effect causes polarization randomness. In situations where a link is polarization sensitive, the polarization randomness can simply make beam steering futile. Secondly, some configurations had many switches which introduced significant RF loss. And finally, none of these designs offered radiation uniformity from one pattern to another.

To overcome these limitations, another form of traveling antenna called symmetric planar square loop antenna (SLA) for switched beam steering applications (Mehta and Mirshekar-Syahkal 2007; Pal et al. 2008) was proposed (Fig. 31). Its structural symmetry offered reconfigurable radiation pattern without any polarization and pattern variation from one beam to another. Since then, that particular SLA has evolved into a much more compact and efficient array configuration. A step-by-step evolution of the antenna for achieving beam steering is presented next.

Radiation Mechanism for Beam Steering

On a ground backed square loop antenna (SLA) of a periphery length greater than 2 × guide wavelength (λ_g), the traveling wave phenomenon dominates, and the antenna radiates a tilted beam ($\theta_{\max} \approx 32^\circ$) in the direction of flow of current. The larger is the periphery in terms of guide wavelength, λ_g, the higher is the beam tilt. For current flowing on a strip backed by substrate of relative permittivity ε_r on one side and air on the other, $\lambda_g = \lambda_0/\sqrt{(1+\varepsilon_r)/2}$ where λ_0 is the free-space wavelength. Figure 31 shows the SLA with four feeding points designed for beam steering applications for the 5 GHz Wi-Fi band. The center and test frequency of the loop is 5.2 GHz. The metal square loop has a periphery length of 120 mm and is etched on Rogers 4350 B substrate whose relative permittivity ε_r is 3.48. The antenna is backed by a ground plane which is at a distance of 9.6 mm. For the center (test) frequency of 5.2 GHz, $\lambda_0 = 57.6$ mm, and for current flowing on 120 mm metal loop, $\lambda_g = 38.7$ mm. Thus, the loop periphery is $3.1\lambda_g$. The antenna has a height of 9.6 mm with planar dimensions of 60.3 × 60.3 mm.

For achieving steering at one time, the antenna is excited at one of the four points, labeled as Ae, and the other three ports (Bo, Co, and Do) are left open circuited. This feeding arrangement creates an asymmetric traveling wave current distribution along the square loop with respect to the whole antenna structure and generates a tilted beam directing to the opposite space quadrant ($0^\circ \leq \phi \leq 90^\circ$) of the excited feed. This is shown in Fig. 32 which also shows that the impedance bandwidth of the antenna is over 12 % (610 MHz). This particular property of tilted beam generation is exploited for beam steering. Since the four feeds are symmetric to each other, by exciting and switching one feed at a time, the tilted beam can be moved in four quadrants of the space in front of the antenna. This concept is depicted in Fig. 33.

DOI 10.1007/978-981-4560-75-7_26-1

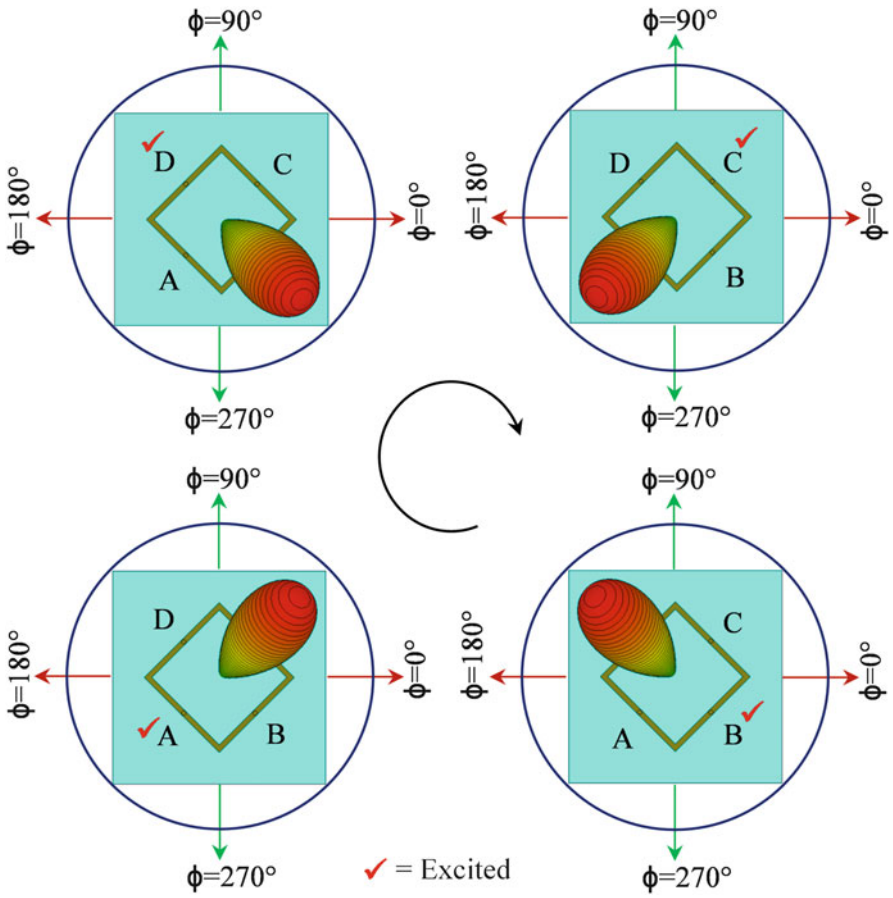


Fig. 33 Demonstration of beam steering by four feeds of SLA. Beam points away from the excited feed

Using High-Impedance Surfaces for Height Reduction

For many modern devices, antennas must be low profile. However, with a height of 9.6 mm, the example presented above is quite thick for implementation in most of the modern portable wireless devices. To overcome this problem, a hybrid high-impedance surface (HHIS) is used for reducing the height by a factor of approximately 50 % (Deo et al. 2010). HHIS is essentially an EBG (electromagnetic bandgap) structure that allows the ground plane to be kept close to the square metal loop without affecting the antenna radiation performance. For this, a two-layer stratified model of substrate is used. The top layer has a height 1.5 mm and is of Rogers's substrate which has $\varepsilon_r = 3.48$ and has the square loop printed on it. This layer is placed over a HHIS substrate layer. That HHIS substrate layer has a height of 3.2 mm and is composed of a 6×6 array of square metal patches on its top surface. Each square has a side length of 8.8 mm with a gap of 1.5 mm between two neighboring patches. The squares on the outer periphery of the HHIS structure are shorted to the ground by metal vias, each having a diameter of 3 mm. The vias are inserted only to the outermost patches to reduce the side lobes emanating from the sides which can cause reduction in directivity. The HHIS layer is then backed by the ground plane and from there four feeding pins go vertically upward toward the loop strip at the top for excitation. This architecture not only reduces the height of the original SLA from 9.6 mm down to 4.7 mm, but it also provides extra bandwidth and higher directivity of 8.8 dBi at $\theta_{\max} = 32°$ and $\phi_{\max} = 45°$. This is shown in Fig. 34.

DOI 10.1007/978-981-4560-75-7_26-1

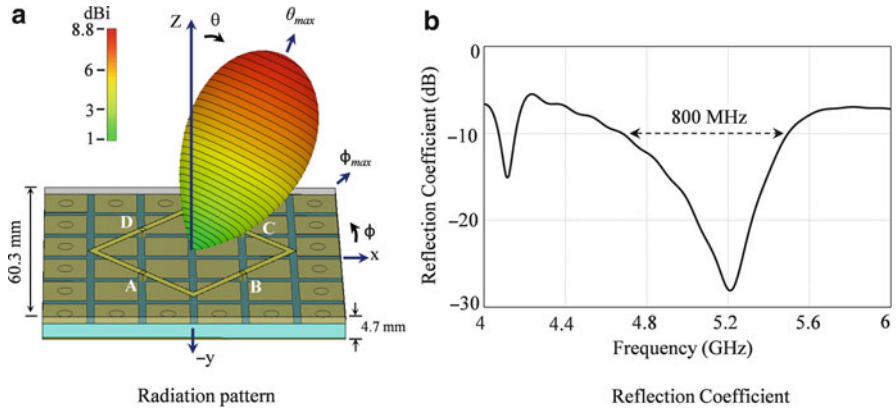


Fig. 34 Reflection coefficient and radiation pattern (at 5.2 GHz) of the HHIS-based SLA with a height of 4.7 mm

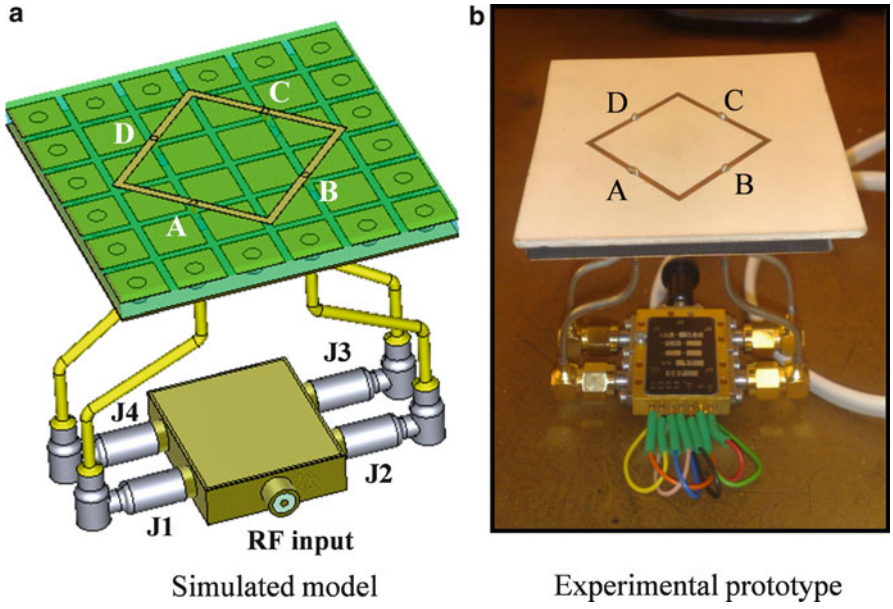


Fig. 35 Simulated and experimental model of the HHIS-based SLA integrated with RF switch (Deo et al. 2011)

Incorporating Electronic RF Switching Circuitry for Electronic Beam Switching

Figure 35 shows the experimental and simulated model for the HHIS-based SLA with the RF switching network. Experimentally, a PIN diode-based RF switch (Deo et al. 2011) is used to connect RF main input to any of the four ports (J1–J4). The selection of one of the four SMA (subminiature version A) connector ports generally can be done via binary control logic from microcontroller providing switching DC voltages linked with a desktop computer running the c algorithm/logic code. For multipath mitigation, the algorithm for the port selection could be fast scan all four ports, measure the received signal/interference (S/I) on all the four ports, compare all four S/I and lock on the port with the highest S/I, and keep monitoring. As mentioned above, by selecting one feed at a time, the SLA can steer its tilted radiation beam in four different space quadrants ($0° \leq \phi \leq 90°$, $90° \leq \phi \leq 180°$, $180° \leq \phi \leq 270°$, and $270° \leq \phi \leq 360°$). With the current developments in PIN diodes, there are PIN diodes available that have: switching speeds in the tens of nanoseconds, low insertion loss (1–2 dB), and high power handling capabilities.

DOI 10.1007/978-981-4560-75-7_26-1

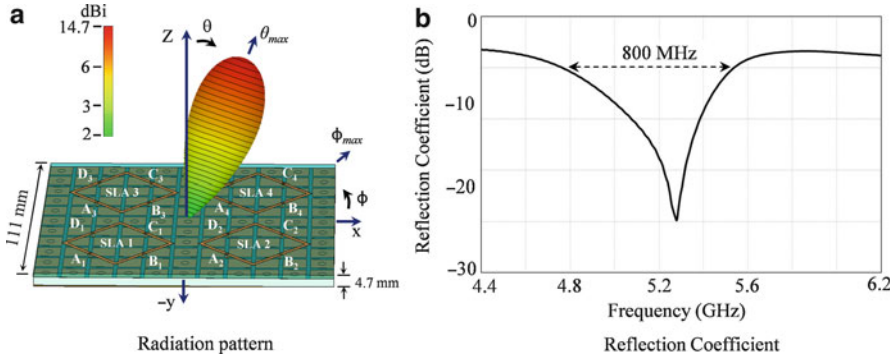


Fig. 36 Radiation pattern of 2 × 2 array of SLA and reflection coefficient for a unit antenna with port A_1 excited and others open circuited

In addition to the four tilted beams in four quadrants, the SLA can also generate a circularly polarized axial (Pal et al. 2013) and a semi-doughnut beam (Pal et al. 2011). For axial, the four ports of the SLA need to be simultaneously excited with quadrature phase shifts. The axial beam would be useful for communicating with vertically located communication systems such as satellites, unmanned aerial vehicles (UAVs), planes, etc. The semi-doughnut can be generated when the four ports of the SLA are simultaneously excited with the same phases. The semi-doughnut pattern is suitable for applications where signal is coming from all the directions but from the top and bottom. Examples for this include receiving radio and digital television from terrestrial station in a moving car/vehicle.

Array of SLAs for High-Gain Wide-Angle Steering, Wide Null Scanning, and Low Grating Lobes

In a phased array antenna system, multiple antenna elements are excited coherently with variable phase to steer the beam in a particular direction and to suppress radiation in undesired directions. They have been traditionally used in several military applications for many years; however, recent advances in wireless communication are drawing attention for utilizing array antenna systems for commercial applications such as 5G, 60 GHz, satellite TV, and cellular base stations as well. The conventional phased array antenna system deploys a grid of ground backed patch or dipole antennas with a fixed radiation pattern, generally axial with typically broadside directivities of 6–7 dBi for patch antennas. According to the phased array theory, the net pattern of an array system is given by multiplication of the unit pattern with the array factor. Thus, the scanning angle and radiation gain of phased array systems are determined by the unit radiation pattern of the single antenna element. The radiation gain of the array decreases by at least 3 dB when the scanning angle of the array increases beyond the half power beam width (HPBW) of the unit pattern. Due to this phenomenon, for such axial systems, the typical 3 dB scanning range is confined within the range from $-40°$ to $+40°$ from the zenith (including a grating lobe < -10 dB criteria). This limited scanning range inhibits the extensive commercial and defense applications of phased array systems which need high-gain signals at off-boresight (tilted) angles as well. This problem can be solved by deploying a pattern reconfigurable SLA as a unit element, which is capable of producing a tilted steerable beam as its fundamental pattern. In essence, the SLA array consisting of pattern reconfigurable unit elements can provide an extra degree of freedom for net pattern formation. Recently, some good work has been published to exploit pattern adaptive behavior of unit element for various applications (Roach et al. 2007; Wu et al. 2010).

As seen from Figs. 32b and 34a, the SLA main radiation is tilted in one direction (high gain on one side and low gain in opposite direction); therefore, its array system will offer a much higher gain in off boresight. Further, in arrays grating lobes start to come into play as the system steers the beam away from

DOI 10.1007/978-981-4560-75-7_26-1

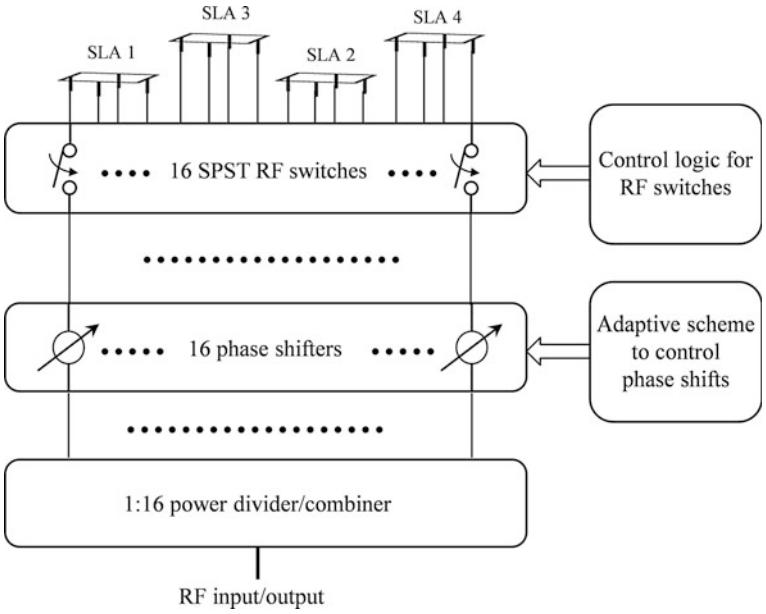


Fig. 37 Block diagram of the feeding network of 2 × 2 array

the unit pattern. For a patch antenna with axial unit pattern, the grating lobes start to get dominant for steering beyond $\theta_{max} > 40°$. However, in the SLA-based array, its antenna elements have tilted beams to begin with: $\theta_{max} \approx 32°$, and hence the array's grating lobes would only appear when the beam is steered beyond $\theta_{max} > 60°$. Besides providing high gain in tilted direction, the SLA can also adapt for boresight steering. As mentioned earlier, Pal et al. (2013) show that the SLA can change its fundamental pattern from tilted to a circularly polarized axial beam by sequential phasing of the four ports. Hence, by combining the SLA tilted beam scanning and axial beam, a complete high-gain scanning range can be achieved, which is not possible with conventional patch or dipole antenna elements. This concept is elaborated next using array of HHIS-based SLA.

HHIS-Based SLA for Demonstrating Gain Beam Steering Using Tilted and Axial Modes

A similar configuration to that described below is presented in Pal et al. (2014).

Figure 36 shows a 2 × 2 array for four HHIS-based square loops. Each SLA loop has the same specification as that of shown in Fig. 34. Again, for side lobe suppression, vias of 3 mm run around the four sides and through the middle of the array for shorting the HHIS patches to the ground. In addition, there are also two via lines that go through the middle of the structure which separates the four individual SLAs. The net planar size of the array is 111 × 111 mm with height staying the same at 4.7 mm. Furthermore, with four feeding ports for each SLA, the net array system has a total of 16 ports which are fed from the bottom side. These ports are labeled as (A_1, B_1, C_1, D_1), (A_2, B_2, C_2, D_2), (A_3, B_3, C_3, D_3), and (A_4, B_4, C_4, D_4). The antenna array operates over 5 GHz Wi-Fi band, and, for port A_1 excited, it has a reflection coefficient bandwidth of 800 MHz (16 %), as shown in Fig. 36b.

This array has two modes of operation: one is for producing steerable high-gain off-boresight tilted beams and the other is for producing a steerable high-gain axial beam. For tilted beam steering, depending upon the quadrant, only four ports are excited at a time and others are left open circuited. For example, for generating a steerable beam in the first quadrant of $0° < \phi < 90°$, the four A ports (A_1–A_4) need to be excited. For axial beam mode, all of the 16 ports need to be excited at the same time, with condition that the four ports of every individual SLA are in phase quadrature with each other. The selection of both these modes can be achieved using a feeding network shown below (Fig. 37). For axial mode, all of the 16 SPST

DOI 10.1007/978-981-4560-75-7_26-1

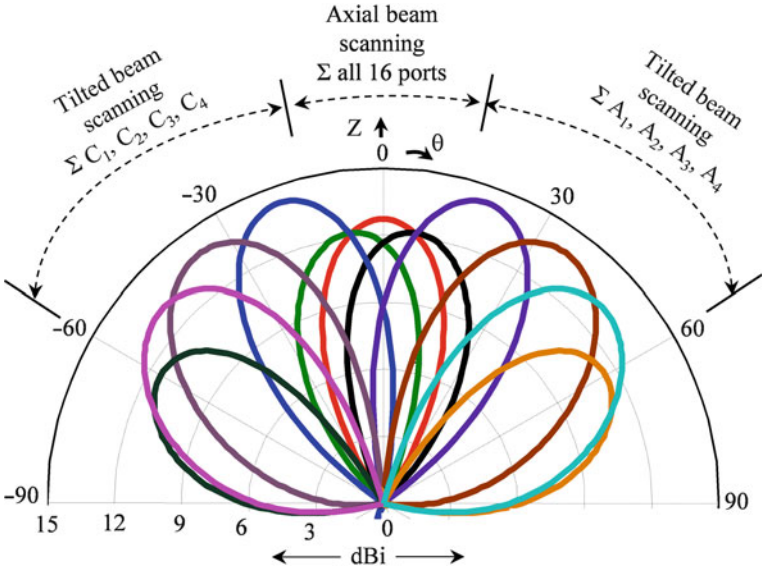


Fig. 38 Block diagram of the scanning of the radiation beam at $\phi_{max} = 45°$

(single pole single throw) RF switches should be on, and for tilted mode, only four SPST should be on and others open circuited. The automated selection of the switch settings and the phase shifts can be done using computer control.

For this 2 × 2 configuration, the antenna array factor is given by Balanis (2005):

$$AF(\theta,\phi) = \frac{1}{4}\left\{\frac{\sin(\psi_x)}{\sin\left(\frac{\psi_x}{2}\right)}\right\}\left\{\frac{\sin\left(\psi_y\right)}{\sin\left(\frac{\psi_y}{2}\right)}\right\} \tag{58}$$

Where:

$$\psi_x = kd_x \sin\theta\cos\phi + \beta_x$$
$$\psi_y = kd_y \sin\theta\sin\phi + \beta_y$$
$$k = 2\pi/\lambda$$
$$d_x = \text{Distance between two elements in } x \text{ direction}$$
$$dy = \text{Distance between two elements in } y \text{ direction}$$
$$\beta_x = \text{Phase shift between two elements in } x \text{ direction}$$
$$\beta_y = \text{Phase shift between two elements in } y \text{ direction}$$
$$\lambda = \text{Wavelength}$$

The maximum radiation beam and grating lobes are located at

$$\begin{aligned} (kd_x \sin\theta\cos\phi + \beta_x) &= \pm 2m\pi \qquad m = 0,\ 1,\ 2,\ \ldots \\ \left(kd_y \sin\theta\sin\phi + \beta_y\right) &= \pm 2n\pi \qquad n = 0,\ 1,\ 2,\ \ldots \end{aligned} \tag{59}$$

DOI 10.1007/978-981-4560-75-7_26-1

To direct the main beam along $\theta = \theta_{\max}$ and $\phi = \phi_{\max}$, the phase shift between the two elements in x- and y-directions must be equal to

$$\begin{aligned} \beta_x &= -kd_x \sin\theta_{\max} \cos\phi_{\max} \\ \beta_y &= -kd_y \sin\theta_{\max} \sin\phi_{\max} \end{aligned} \tag{60}$$

For the 2×2 SLA array, shown in Fig. 36, $d_x = d_y = 51\ \text{mm} = 0.9\lambda_0$. Using the equations above, both the tilted and axial beam maximums were steered in the space for various maximums at different $\theta_{\max}$, and the results for elevation cut at $\phi_{\max} = 45^\circ$ are shown in Fig. 38. It was found that, for tilted beam configuration, a maximum directivity of 14.7 dBi was achieved at $\theta_{\max} = 32^\circ$, and, for an axial beam configuration, a directivity of 13.6 dBi was achieved in the boresight. Hence, together both tilted and axial mode can provide a near constant high-gain (14.7–12.7 dBi) scanning range from -60° to $+60^\circ$ (for pattern having $\leq$10 dB grating lobes). This high-gain wide scanning range is extremely difficult to achieve from a non-reconfigurable phased array system. For example, an array using patch antenna elements (6dBi gain) would need approximately 9 elements to provide a similar boresight gain but still will not be able to provide 14dBi gains at tilted angles. Hence, the proposed system here is of a huge significance when the systems are deployed for covering wide sky, especially for airborne and marine applications.

Conceptual Examples

Problem 1 The concept of radiation resistance R_r is very useful in various applications especially in matching wire antennas including dipole and wire loop antennas to transmitter or receiver circuit. It is defined by the following expression:

$$R_r = \frac{P_{\text{Rad}T}}{|I_0|^2} \tag{61}$$

where I_0 is the rms current and $P_{\text{Rad}T}$ is the power radiated by the antenna, as denoted in subsection "Radiation and Losses."

(a) Considering that the total power radiated by a single-turn small loop antenna is given by

$$P_{\text{Rad}T\text{loop}} \frac{8Z_0\pi^3}{3} \left(\frac{|I_0|A_{\text{LA}}}{\lambda^2}\right)^2 \tag{62}$$

show that the radiation resistance of the loop antenna can be calculated using

$$R_{r\text{loop}} = 31171 \left(\frac{A_{\text{LA}}}{\lambda^2}\right)^2 \tag{63}$$

or using

DOI 10.1007/978-981-4560-75-7_26-1

$$R_{r\text{loop}} = 20\pi^2\left(\frac{C}{\lambda}\right)^4 \tag{64}$$

where C = $2\pi a$ is the perimeter of the loop of radius a.

What would be the expression for the radiation resistance of an n-turn small loop antenna?

(b) Considering that the radiation efficiency of an antenna (Eq. 14) is defined as

$$R.E. = \frac{P_{\text{Rad}T}}{P_{\text{Rad}T} + P_\Omega} \tag{65}$$

show that, for the loop antenna, it reduces to

$$R.E. = \frac{R_{r\text{loop}}}{R_{r\text{loop}} + R_{\Omega\text{loop}}} \tag{66}$$

where $R_{\Omega\text{loop}}$ is known as the Ohmic loss resistance of the loop antenna.

(c) Show that the Ohmic loss resistance of a single-turn loop antenna of radius a and wire diameter d larger than the current skin depth δ (i.e., $d > \delta$ where $\sqrt{2/(\omega\mu_0\sigma)}$) is given by

$$R_{\Omega\text{loop}} = \left(\frac{a}{d}\right)\rho_S \tag{67}$$

where $\rho_S = \sqrt{\omega\mu_0/(2\sigma)}$ is the surface resistance. Note σ is the conductivity of the loop metal/alloy and related to the resistivity ρ through $\sigma = 1/\rho$.

(d) Calculate the radiation resistance, the loss resistance, the efficiency, the Q factor, and the input impedance of a single-turn loop antenna made of a silver wire of conductivity $\sigma = 6.3 \times 10^7$S/m. This loop antenna has a radius of $a = \lambda/30$ and a wire radius of $d = \lambda/3000$, and is designed to operate at 200 MHz.

(e) In order to operate (resonate) the loop antenna at 200 MHz, a capacitor is connected in parallel across the two terminals of the loop. Calculate the value of the capacitor and the input impedance of the antenna at resonance. What would happen to the resonant frequency and the input impedance of the antenna if the capacitor is connected in series with the loop?

Solutions

(a) Inserting $P_{\text{Rad}T\text{loop}} = 8Z_0\pi^3\left(|I_0|A_{\text{LA}}/\lambda^2\right)^2/3$ in the radiation resistance definition leads to

$$R_r = \frac{P_{\text{Rad}T}}{|I_0|^2} = \frac{8Z_0\pi^3}{3}\left(\frac{A_{\text{LA}}}{\lambda^2}\right)^2 \tag{68}$$

i. Since the intrinsic impedance of air $Z_0 = 120\pi = 377\Omega$, the expression for R_r transforms to the following equation: $R_r = 8 \times 120\pi \times \pi^3\left(A_{LA}/\lambda^2\right)^2/3 = 31171\left(A_{\text{LA}}/\lambda^2\right)^2$.

ii. Since the loop area $A_{\text{LA}} = \pi a^2$, the expression for R_r can be written as follows:

DOI 10.1007/978-981-4560-75-7_26-1

$$R_r = \frac{8Z_0\pi^3}{3}\left(\frac{\pi a^2}{\lambda^2}\right)^2 = \frac{120\pi \times \pi}{2 \times 3}\left(\frac{2\pi a}{\lambda}\right)^4 \tag{69}$$

Substituting for the perimeter, $C = 2\pi a$, in the above equation gives $R_r = 20\pi^2(C/\lambda)^4$

When the loop antenna is n-turn, the current can be assumed to have weighted by a factor of n, i.e., $I_0 \rightarrow nI_0$. Therefore, the radiated power increases by a factor of n^2 (see the expression for radiated power). This means that R_r of an n-turn loop antenna is n^2 of that of the single-turn loop antenna.

(b) Using the definition of the radiation resistance $R_r = P_{\mathrm{Rad}T}/|I_0|^2$, the total power radiated by the loop antenna is

$$P_{\mathrm{Rad}T} = R_{r\mathrm{loop}}|I_0|^2 \tag{70}$$

Similarly, the total Ohmic power loss born by the single loop antenna is

$$P_\Omega = R_{\Omega\mathrm{loop}}|I_0|^2 \tag{71}$$

Inserting these equations in the expression for radiation efficiency, $R.E. = P_{\mathrm{Rad}T}/(P_{\mathrm{Rad}T} + P_\Omega)$, leads to

$$R.E. = \frac{P_{\mathrm{Rad}T}}{P_{\mathrm{Rad}T} + P_\Omega} = \frac{R_{r\mathrm{loop}}|I_0|^2}{R_{r\mathrm{loop}}|I_0|^2 + R_{\Omega\mathrm{loop}}|I_0|^2} = \frac{R_{r\mathrm{loop}}}{R_{r\mathrm{loop}} + R_{\Omega\mathrm{loop}}} \tag{72}$$

(c) The surface resistance, ρ_S, is in Ohms per square (Omega/sq). As the current flows along the loop, the loss resistance of the loop is obtained as follows:

$$R_{\Omega\mathrm{loop}} = \frac{\rho_S}{2\pi d}2\pi a = \frac{a}{d}\rho_S \tag{73}$$

It should be noted that the loss resistance of a loop antenna with n turns is not simply n times of that of its single loop. This is because there is an interaction between the fields of the loops, changing the current distribution and hence the value of the loss resistance. However, if the loops are sufficiently separated, the n-time approximation can be used.

(d) To find the radiation resistance of the antenna, the radius of the loop, $a = \lambda/30$, is used to find the area of the antenna:

$$A_{\mathrm{LA}} = \pi a^2 = \pi\left(\frac{\lambda}{30}\right)^2 \tag{74}$$

Then, this area is inserted in the expression for the radiation resistance:

DOI 10.1007/978-981-4560-75-7_26-1

$$R_{r\text{loop}} = 31171\left(\frac{A_{\text{LA}}}{\lambda^2}\right)^2 = 31171\left[\frac{\pi\left(\frac{\lambda}{30}\right)^2}{\lambda^2}\right]^2 = 31171\left(\frac{\pi}{30^2}\right)^2 = 0.38\,\Omega \tag{75}$$

To calculate the loss resistance, first the surface resistivity is determined:

$$\rho_S = \sqrt{\frac{\omega\mu_0}{2\sigma}} = \sqrt{\frac{\pi \times 200 \times 10^6 \times 4\pi \times 10^{-7}}{6.3 \times 10^7}} = 3.54 \times 10^{-3}\,\Omega \text{ per square.} \tag{76}$$

Then, the surface resistance; the loop radius, $a = \lambda/30$; and the loop wire radius, $d = \lambda/3000$, are inserted in the expression for the loss resistance:

$$R_{\Omega\text{loop}} = \left(\frac{a}{d}\right)\rho_S = [(\lambda/30)/(\lambda/3000)] \times 3.54 \times 10^{-3} = 0.35\Omega \tag{77}$$

Therefore, the radiation efficiency of the antenna is

$$R.E. = \frac{R_{r\text{loop}}}{R_{r\text{loop}} + R_{\Omega\text{loop}}} = \frac{0.38}{0.38 + 0.35} = 52\% \tag{78}$$

To calculate the Q factor of the antenna, the value of the inductance of the loop is required. This inductance, L, is the sum of the external inductance, $L_{ext\text{loop}}$, and the internal inductance, $L_{int\text{loop}}$, of the loop:

$$L = L_{ext\text{loop}} + L_{int\text{loop}} \tag{79}$$

The external inductance, associated with the storage of the magnetic field external to the loop wire, is obtained using an approximate expression (Johnson and Jasik 1993):

$$L_{ext\text{loop}} = \mu_0 a\left[\ln\left(\frac{8a}{d}\right) - 2\right] = 4\pi \times 10^{-7}\left(\frac{1.5}{30}\right)\left[\ln\left(\left(\frac{\lambda}{30}\right)/\left(\frac{\lambda}{3000}\right)\right) - 2\right] = 2.94 \times 10^{-7}\text{H} \tag{80}$$

The internal inductance, associated with the storage of the magnetic field within the loop wire, is obtained using

$$L_{int\text{loop}} = \frac{1}{\omega}\left(\frac{a}{d}\right)\rho_S = \frac{1}{\omega}R_{\Omega\text{loop}} = \frac{0.35}{2 \times \pi \times 200 \times 10^6} = 2.8 \times 10^{-10}\text{H} \tag{81}$$

As can be seen, $L_{ext\text{loop}} \gg L_{int\text{loop}}$ and this is normally the case and hence $L_{int\text{loop}}$ can be ignored. Therefore, the Q factor of the loop antenna is

$$Q = \frac{\omega L}{R_{r\text{loop}} + R_{\Omega\text{loop}}} \cong \frac{2\pi \times 200 \times 10^6 \times \left(2.94 \times 10^{-7}\right)}{0.38 + 0.35} = 506 \tag{82}$$

To calculate the input impedance of the antenna, the equivalent circuit of the loop antenna is used. In

DOI 10.1007/978-981-4560-75-7_26-1

this circuit, the radiation and loss resistance are in series with the inductance of the loop antenna. Therefore,

$$Z_{in\text{loop}} = R_{r\text{loop}} + R_{\Omega\text{loop}} + j\omega L \cong 0.38 + 0.35 + j2\pi \times 200 \times 10^6 \times 2.94 \times 10^{-7} = 0.73 + j369.5\Omega \tag{83}$$

As expected, the input impedance is very inductive.

(e) To find the capacitance C of the capacitor, the input admittance of the antenna is formed as the capacitor is connected in parallel with the loop:

$$Y_{in} = jC\omega + \frac{1}{R + jL\omega} \tag{84}$$

where $R = R_{r\text{loop}} + R_{\Omega\text{loop}}$ and $L = L_{ext\text{loop}} + L_{int\text{loop}}$. This can be written as follows:

$$Y_{in} = \frac{R + j\left\{C\omega\left[R^2 + (L\omega)^2\right] - L\omega\right\}}{R^2 + (L\omega)^2} \tag{85}$$

At resonance, the imaginary part of the input admittance vanishes. Therefore,

$$C\left[R^2 + (L\omega)^2\right] - L = 0 \tag{86}$$

from which the value of capacitance is obtained:

$$C = \frac{L}{\left[R^2 + (L\omega)^2\right]} \cong \frac{1}{L\omega^2} = \frac{1}{2.94 \times 10^{-7} \times \left(2\pi \times 200 \times 10^6\right)^2} = \frac{1}{0.464 \times 10^{12}} = 2.15\text{pF} \tag{87}$$

In the above and following calculations, the resistance is ignored as it is much smaller than the reactance.

The input impedance of the antenna at resonance is found using the input admittance at resonance:

$$Y_{in} = \frac{R}{R^2 + (L\omega)^2} \tag{88}$$

$$Z_{in} = \frac{R^2 + (L\omega)^2}{R} \cong \frac{(L\omega)^2}{R} = \frac{\left(2.94 \times 10^{-7} \times 2\pi \times 200 \times 10^6\right)^2}{0.38 + 0.35} = 187\,\text{k}\Omega \tag{89}$$

When the capacitor is connected in series with the loop, the resonant frequency of the antenna stays nearly the same as before (200 MHz). However, the input impedance at resonance changes drastically to a small value and is equal to

$$R = R_{r\text{loop}} + R_{\Omega\text{loop}} = 0.38 + 0.35 = 0.73\Omega \tag{90}$$

DOI 10.1007/978-981-4560-75-7_26-1

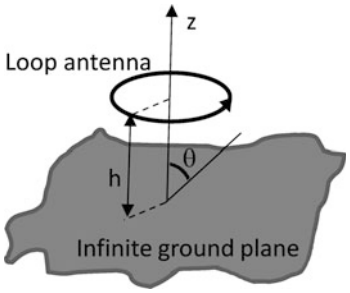


Fig. 39 Loop antenna parallel to and above an infinite perfectly conducting plane, discussed in problem 2

Comments

1. Capacitors have quality factors Q_C that are typically in the range of ~200 (for low-cost variable capacitors) to ~2,000 for high-quality capacitors. For the 2.15 pF capacitor of this example, using the relation $Q_C = 1/(\omega C\, R_{Sc})$ where R_{SC} is the effective series resistance of the capacitor, gives R_{SC}~2 Ω (for Q_C~200) to R_{SC}~0.2 Ω (for Q_C~2,000). These values are comparable to those of $R_{r\text{loop}}$ and $R_{\Omega\text{loop}}$ of the loop antenna and shows that in practice the losses associated with nonideal capacitors should be considered in the design calculations. The details of the mathematics for this are beyond the scope of these examples. However, by an extension of the analysis described above, it can be shown that, in general, the effect of the finite quality factor of a real-life capacitor is to reduce the resonant impedance of the loop tuned with a capacitor in parallel and to increase the resonant impedance of the loop tuned with a capacitor in series.
2. Transmitters and receivers generally require impedances that lie between the two extremes calculated above for series and parallel capacitor tuning. This observation can be thought of as an inspiration for the combined series and parallel capacitor tuning arrangement shown in Fig. 8b, which is capable of giving any impedance between the very low-impedance case with series capacitors only (parallel capacitance vanishes) to the very high-impedance case with the parallel capacitor providing practically all of the tuning and the series capacitors having such high capacitances that they appear to have negligible reactance.

Problem 2 Consider a small loop antenna horizontally placed at a distance h over an infinite size perfect conductor (Fig. 39). Assuming that the radiation pattern of this antenna in the absence of the ground plane varies according to the following expression:

$$E_\phi = k \sin(\theta) \tag{91}$$

where k is constant, find an expression for the radiation pattern of this antenna when the ground plane is present.

Solution As the ground is a perfect conductor, it can be replaced by the image of the loop antenna. This image is another loop antenna at distance h from the surface of the ground plane, and it carries a current opposite to that of the original loop antenna. In other words, there is a phase shift of π between the currents of the two elements. The system can be assumed as a linear array of two loop antennas. In un-normalized form, the array factor for a linear array of N elements is

DOI 10.1007/978-981-4560-75-7_26-1

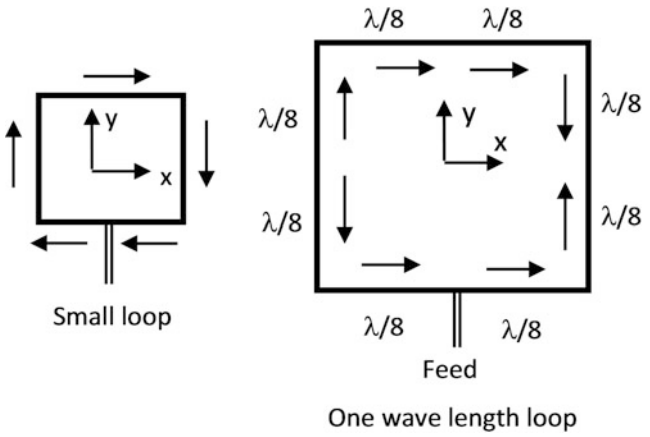


Fig. 40 Current distribution on a square loop antenna

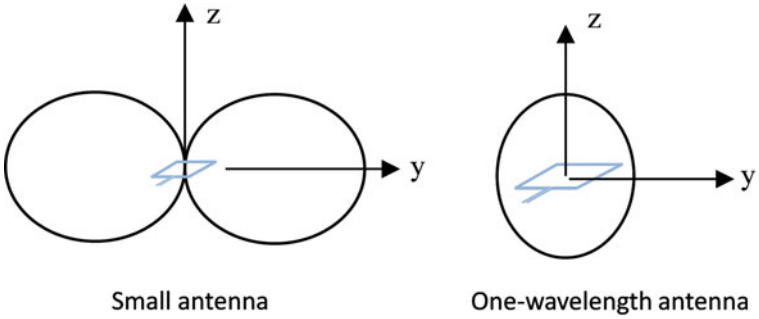


Fig. 41 The yz-cut of the radiation pattern of square loop antenna

$$AF = \frac{\sin\left(\frac{N}{2}\Psi\right)}{\sin\left(\frac{1}{2}\Psi\right)} \tag{92}$$

where $\Psi = \beta s \cos\theta + \alpha$. In this case, the separation between the elements $s = 2h$, the number of elements $N = 2$, and the progressive phase $a = \pi$. Therefore, the array factor simplifies to

$$AF = \frac{\sin\left[\frac{2}{2}\left(\frac{2\pi}{\lambda}2h\cos(\theta)+\pi\right)\right]}{\sin\left[\frac{1}{2}\left(\frac{2\pi}{\lambda}2h\cos(\theta)+\pi\right)\right]} = \frac{\sin\left[\frac{2\pi}{\lambda}2h\cos(\theta)+\pi\right]}{\sin\left[\frac{1}{2}\left(\frac{2\pi}{\lambda}2h\cos(\theta)+\pi\right)\right]}$$
$$= 2\cos\left[\frac{1}{2}\left(\frac{2\pi}{\lambda}2h\cos(\theta)+\pi\right)\right] \tag{93}$$

Then, the overall radiation pattern is found using the pattern multiplication rule:

$$E_{\phi T} = \text{antenna (element) pattern} \times \text{array factor} = 2k\sin(\theta)\cos\left[\left(2\frac{h}{\lambda}\cos(\theta)+\frac{1}{2}\right)\pi\right] \tag{94}$$

DOI 10.1007/978-981-4560-75-7_26-1

From this expression, several conclusions can be made:

i. $E_{\phi T} = 0$ along the loop axis ($\theta = 0$).
ii. $E_{\phi T} = 0$ over the ground plane ($\theta = \pi/2$).
iii. $E_{\phi T}$ has beam direction and directivity which varies with h/λ.

Problem 3 Assume that a square loop antenna fed as shown in Fig. 40 lays in the x-y plane with its center at the origin of the coordinate system. Sketch the radiation pattern of the antenna in the y-z plane for the two cases of small square loop antenna and one-wavelength square loop antenna.

Solution Since the current is approximately uniform on the wire of the small square loop antenna (Fig. 40; left), the radiation pattern is similar to that of the small circular loop antenna. Therefore, the pattern is omnidirectional with its yz-cut shown in Fig. 41 (left).

The pattern of the one-wavelength square loop antenna is more complex if one considers the current distribution of this antenna at the first resonance as shown in Fig. 40 (right). Since the currents in the left and right arms of the antenna are opposite, they contribute little to the radiation. The currents in the top and bottom arms are the main contributors. These arms can be considered as the array of two dipole antennas separated by a distance of quarter wavelength. In this case, the yz-cut of the pattern has the maximum radiation occurring along the z-axis (due to the constructive interference of the fields of the two dipoles) (Fig. 41; right). In the y-direction, the radiation strength is about half of that in the z-direction (due to the vector sum of the two fields of the dipoles with 90° phase shift) (Fig. 41; right). The cuts of the radiation pattern in the two other principle planes can also be predicted. In the x-z plane, the radiation has nulls in the ±x-directions and is maximum in the ±z-directions. In the x-y plane, the radiation has nulls along the x-axis and has a value about half of that along the z-axis (as explained earlier). Also, see Fig. 16 of subsection "Tested Example and Range Plots" for the radiation pattern of an antenna example similar to that in this problem.

Conclusion

This chapter describes three classes of loop antennas in order of their size in wavelengths: electrically small loop (and coil) antennas, full-wave loop antennas, and reconfigurable multi-feed loops. This ordering also reflects the historical development of loop antennas, with the electrically small loop antennas coming early in the history of wireless development and reconfigurable multi-feed loops still being a current research topic.

Throughout loop antenna history, their ubiquity has inevitably led them being combined with other innovations. For example, electronically small loop antennas were among the first to be used in a commercial product with antenna diversity – by the early 1950s, radios were sold with a pair of loops angled away from each other, and the user could switch between loops for the best reception (Blok and Rietveld 1955). After World War II, electrically small loop antennas were combined with the then very new soft ferrite material to give smaller antennas with similar sensitivity and less noise reception through reducing the aerial effect. Today, reconfigurable multi-feed loops are placed above high-impedance surfaces, to reduce the distance between the loop and its backing ground plane, and also because the high-impedance surface is found to improve isolation between a loop's feeds.

Loop antenna research continues to be a very active field. A search of the IEEE Xplore database for articles published between the beginning of 2014 and June 2015 (the date when the search was conducted) revealed 123 papers with "loop antenna" in their titles, and over three times this number in which loop

DOI 10.1007/978-981-4560-75-7_26-1

antennas played a significant role. Over the last few years, popular research and application areas have included:

1. Creating loops from metamaterial structures (e.g., Nakano et al. 2013a; Zhang et al. 2015)
2. Loop antennas in medical on-body and implant applications (e.g., Alrawashdeh et al. 2014; Lee and Jung 2015)
3. Loop antennas in handheld multiband mobile phone devices (e.g., Ban et al. 2015)
4. Using loops as circularly polarized elements in arrays (e.g., Nakano et al. 2013b; Hirose et al. 2012, 2015a, b)

In view of the recent nature of the newer publications, it is difficult to predict how these topics would develop and which developments would prove to be the most significant. However, what is certain is that electrically small loop antennas will continue to have a significant role in antennas for products designed to be worn against the user's body and in other portable devices. In addition, the unique ability of traveling wave loops to provide broad impedance bandwidth and adaptive antenna element pattern will enable the realization of wideband reconfigurable arrays for communications and sensor networks.

Cross-References

- ▶ Impedance Matching and Baluns
- ▶ Low-Profile Antennas
- ▶ Metamaterials and Antennas
- ▶ Omnidirectional Antennas
- ▶ Phased Arrays
- ▶ Radiation Efficiency Measurement of Small Antennas
- ▶ Wheeler-Fano Impedance Matching Principles

References

Alrawashdeh R, Huang Y, Sajak AAB (2014) A flexible loop antenna for biomedical bone implants. In: 8th European conference on antennas and propagation, pp 861–864

Balanis CA (2005) Antenna theory, analysis and design, 3rd edn. Wiley, Hoboken, Chapter 5: Loop antennas. ISBN 0-471-66782-X

Ban Y-L, Qiang Y-F, Chen Z, Kang K, Guo J-H (2015) A dual-loop antenna design for hepta-band WWAN/LTE metal-rimmed smartphone applications. IEEE Trans Antennas Propag 63(1):48–58

Blok H, Reitveld JJ (1955) Inductive aerials in modern broadcast receivers. Philips Tech Rev 16(7):181–212

Collin RE (2001) Foundations for microwave engineering. Wiley, New Jersey, Chapter 2

Deo P, Mehta A, Mirshekar-Syahkal D, Massey PJ, Nakano H (2010) Thickness reduction and performance enhancement of steerable square loop antenna using hybrid high impedance surface. IEEE Trans Antennas Propag 58(5):1477–1485

DOI 10.1007/978-981-4560-75-7_26-1

Deo P, Pant M, Mehta A, Mirshekar-Syahkal D, Nakano H (2011) Implementation and simulation of commercial rf switch integration with steerable square loop antenna. Electron Lett 47(12):686–687

Grover FW (2009) Inductance calculations, working formulas and tables. Dover, New York

Harrington RF (2001) Time harmonic electromagnetic fields. Wiley, New York

Hirose K, Shibasaki T, Nakano H (2012) Fundamental study on novel loop-line antennas radiating a circularly polarized wave. IEEE Antennas Wirel Propag Lett 11:476–479

Hirose K, Shinozaki K, Nakano H (2015a) A loop antenna with parallel wires for circular polarization—its application to two types of microstrip-line antennas. IEEE Antennas Wirel Propag Lett 14:538–586

Hirose K, Shinozaki K, Nakano H (2015b) A comb-line antenna modified for wideband circular polarization. IEEE Antennas Wirel Propag Lett 14:1113–1116

Huff GH, Bernhard JT (2006) Integration of packaged RF MEMS switches with radiation pattern reconfigurable square spiral microstrip antennas. IEEE Trans Antennas Propag 54(2):464–469

Huff GH, Feng J, Zhang S, Bernhard JT (2003) A novel radiation pattern and frequency reconfigurable single turn square spiral microstrip antenna. IEEE Microw Wirel Component Lett 13(2):57–59

ITU-R (2001) Recommendation, P.372-7, "Radio noise"

Johnson RC, Jasik H (eds) (1993) Antenna engineering handbook, 3rd edn. McGraw-Hill, New York, Smith GS, Chapter 5: Loop antennas. ISBN 1063-665X

Jung CW, Lee M, Li GP, Flaviis FD (2006) Reconfigurable scan-beam single-arm spiral antenna integrated with RF-MEMS switches. IEEE Trans Antennas Propag 54(2):455–463

Lee KF (1984) Principles of antenna theory. Wiley, Chichester

Lee C, Jung C (2015) Radiation-pattern-reconfigurable antenna using monopole-loop for fitbit flex wristband. IEEE Antennas Wirel Propag Lett 14:269–272

Li R, DeJean G, Laskar J, Tentzeris MM (2005) Investigations of circularly polarized loop antennas with a parasitic element for bandwidth enhancement. IEEE Trans Antennas Propag 53(12):3930–3939

Liberti JC, Rappaport TS (1999) Smart antennas for wireless communications: IS-95 and third generation CDMA applications. Prentice-Hall, Englewood Cliffs

Massey PJ (2001) New formulae for practical pager design. In: 11th international conference on antennas and propagation (ICAP), Manchester, vol 1, pp 265–268

Massey PJ (2009) Single tuned electrically small antennas. In: Loughborough antennas and propagation conference, 16–17 Nov, pp 497–500

Mehta A, Mirshekar-Syahkal D (2004) Spiral antenna with adaptive radiation pattern under electronic control. In: Antennas and Propagation Society international symposium, IEEE, vol 1, 20–25 June, pp 843–846

Mehta A, Mirshekar-Syahkal D (2007) Pattern steerable square loop antenna. Electron Lett 43(9):491–493

Mehta A, Mirshekar-Syahkal D, Nakano H (2006) Beam adaptive single arm rectangular spiral antenna with switches. IEE Proc Microw Antennas Propag 153(1):13–18

Morishita H, Hirasawa K, Nagao T (1998) Circularly polarised wire antenna with a dual rhombic loop. IEE Proc Microw Antennas Propag 145(3):219–224

Nakano H, Yoshida K, Yamauchi J (2013a) Radiation characteristics of a metaloop antenna. IEEE Antennas Wirel Propag Lett 12:861–863

Nakano H, Iitsuka Y, Yamauchi J (2013b) Loop-based circularly polarized grid array antenna with edge excitation. IEEE Trans Antennas Propag 61(8):4045–4053

DOI 10.1007/978-981-4560-75-7_26-1

Pal A, Mehta A, Mirshekar-Syahkal D, Massey P (2008) Short-circuited feed terminations on beam steering square loop antennas. Electron Lett 44(24):1389–1390

Pal A, Mehta A, Mirshekar-Syahkal D, Nakano H (2011) A square-loop antenna with 4-port feeding network generating semi-doughnut pattern for vehicular and wireless applications. IEEE Antennas Wirel Propag Lett 10:338–341

Pal A, Mehta A, Marhic ME (2013) Generating a pure circularly polarised axial beam from a pattern reconfigurable square loop antenna. IET Microw Antennas Propag 7(3):208–213

Pal A, Mehta A, Lewis R, Clow N (2014) Phased array system consisting of unit pattern reconfigurable square loop antennas. In: Antennas and Propagation Society international symposium (APSURSI), IEEE, pp 1658–1659, 6–11 Jul

Pozar DM (1986) Finite phased arrays of rectangular microstrip patches. IEEE Trans Antennas Propag AP-34(5):658–665

Roach TL, Huff GH, Bernhard JT (2007) On the applications for a radiation reconfigurable antenna. In: Second NASA/ESA conference on adaptive hardware and systems (AHS), Aug, pp 7–13

Schelkunoff SA (1939) A general radiation formula. In: Proceedings of the I.R.E., Oct 1939, pp 660–666

Snelling EC (1988) Soft ferrites – properties and applications, 2nd edn. Butterworth, London. ISBN 0-408-02760-6

van der Zaag PJ (1999) New views on the dissipation in soft magnetic ferrites. J Magn Magn Mater 196–197(1999):315–319

Wagstaff AJ and Merricks N (2003) Man-Made Noise Measurement Programme (AY4119) final report. Issue 2, Mass Consultants Limited, Sept 2003. This report was downloadable from www.ofcom.org.uk. Last accessed 30 Dec 2014. In particular see section 5.1

Wu JC, Chang CC, Chin TY, Huang SY, Chang SF (2010) Sidelobe level reduction in wide-angle scanning array system using pattern-reconfigurable antennas. Microwave Symposium Digest (MTT), IEEE MTT-S, 23–28 May, pp 1274–1277

Zhang Y, Wei K, Zhang Z, Li Y, Feng Z (2015) A compact dual-mode metamaterial-based loop antenna for pattern diversity. IEEE Antennas Propag Lett 14:394–397

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_29-1

Microstrip Patch Antennas

Kai Fong Lee[a]* and Kin-Fai Tong[b]
[a]Department of Electrical Engineering, The University of Mississippi, University, USA
[b]Department of Electronic and Electrical Engineering, University College London, London, UK

Abstract

The basic geometry of a microstrip patch antenna (MPA) consists of a metallic patch which is either printed on a grounded substrate or suspended above a ground plane. The antenna is usually fed either by a coaxial probe or a stripline. In the coaxial case, the center conductor is directly connected to the patch and the outer conductor to the ground. In the stripline case, energy is coupled to the patch in several ways: by direct connection, by proximity coupling, and by aperture coupling. The patch antenna idea appeared to be originated in the early 1950s, but there was little activity for almost two decades, mainly due to its inherent narrow bandwidth. It began to attract the serious attention of the antenna community in the 1970s, as antenna designers began to appreciate the advantages offered by this type of antennas, which include low profile, conformability to a shaped surface, ease of fabrication, and compatibility with integrated circuit technology. In the last three decades, extensive studies have been devoted to improving the bandwidth and other performance characteristics. This chapter begins with a brief description of the modeling techniques and basic characteristics of the MPA. Methods for broadbanding are then discussed, followed by dual- and multiband designs, size reduction techniques, circularly polarized patch antennas, and frequency-agile and polarization-agile designs. The chapter ends with some concluding remarks.

Keywords

Microstrip antennas; Wideband patch antennas; Patch antennas for circular polarization; Small-size patch antennas; Dual- and multiband patch antennas

Introduction

The idea of microstrip patch antennas arose from utilizing printed circuit technology not only for the circuit components and transmission lines but also for the radiating elements of an electronic system. The basic geometries are shown in Fig. 1. It consists of a metallic patch which is either printed on a grounded substrate or suspended above a ground plane and fed against the ground at an appropriate location. The patch shape can in principle be arbitrary. In practice, the rectangle, the circle, the equitriangle, and the annular ring are common shapes. Feeding is usually by a coaxial probe or a stripline. The latter can be directly connected to the patch, or proximity coupled to the patch, or coupled through an aperture. These are illustrated in Fig. 2 for the rectangular patch. Electromagnetic energy is first guided or coupled to the region under the patch, which acts like a resonant cavity with open circuits on the sides. Some of the energy leaks out of the cavity and radiates into space, resulting in an antenna.

The advantages and disadvantages of the four feeding methods are listed in Table 1. Of the four, the coaxial feed and the direct stripline feed are the most common.

*Email: leek@olemiss.edu

DOI 10.1007/978-981-4560-75-7_29-1

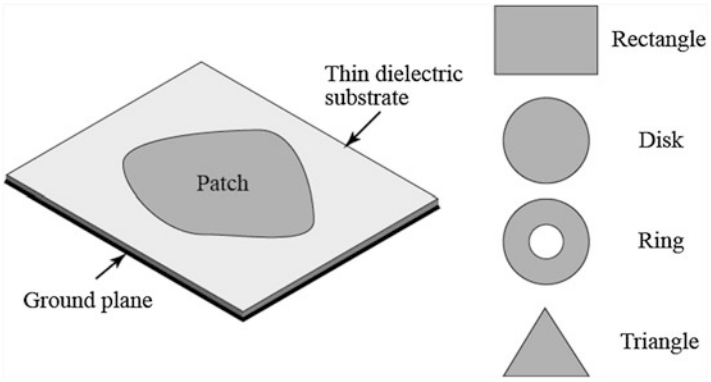


Fig. 1 The basic structure of the microstrip patch antenna

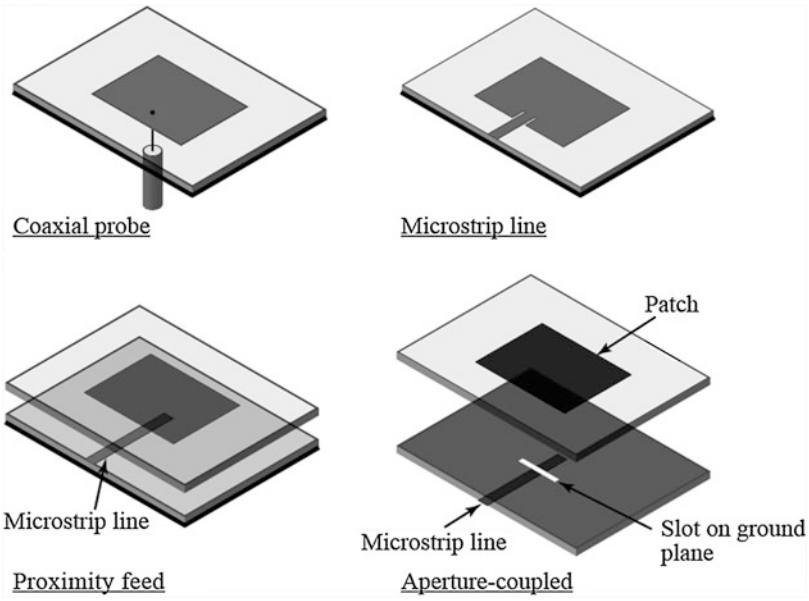


Fig. 2 Four common feeding methods of microstrip patch antenna

Table 1 The comparisons between the four common feeding methods for microstrip patch antenna

	Advantages	Disadvantages
Coaxial feed	Easy to match Low spurious radiation	Large inductance for thick substrate Soldering required
Microstrip line	Monolithic Easy to fabricate Easy to match by controlling insert position	Spurious radiation from feed line, especially for thick substrate when line width is significant
Proximity coupled	No direct contact between feed and patch Can have large effective thickness for patch substrate and much thinner feed substrate	Multilayer fabrication required
Aperture coupled	Use of two substrates avoids deleterious effect of a high dielectric constant substrate on the bandwidth and efficiency No direct contact between feed and patch avoiding large probe reactance or wide microstrip line No radiation from the feed and active devices since a ground plane separates them from the radiating patch	Multilayer fabrication required Higher back lobe radiation

DOI 10.1007/978-981-4560-75-7_29-1

The patch antenna idea appeared to be proposed by Deschamps in the early 1950s (Deschamps and Sichak 1953), but there was little activity for some two decades. It was not until the late 1970s that this type of antenna attracted serious attention of the antenna community, due to the recognition of the many advantages it offered. These include low profile, conformability to a shaped surface, ease of fabrication, and compatibility with integrated circuit technology, since active and passive circuit elements can be etched on the same substrate. Its main drawback is narrow bandwidth, typically less than 5 %. Research in the last several decades has resulted not only in patch antennas with double digit bandwidths but a host of techniques for designing dual- and multiband, dual- and circular polarization characteristics, size reduction, and reconfigurable designs. The MPA has found extensive applications in the wireless industry, both commercial and military, and it is probably not an exaggeration to say that this type of antennas has become the favorite of antenna designers. The literature on MPAs is vast, comprising thousands of research articles in archival journals. There were also several review articles, handbook chapters, and more than a dozen books, a partial list of which are given in the references [Bahl and Bhartia (1980); Chen and Chia (2005); Debatosh and Antar (2010); Garg (2000); Huang (2008); James and Hall (1981, 1989); Kumar and Ray (2003); Lee and Tong (2012); Lee and Chen (1997); Richards et al. (1981); Pozar (1992); Shafai (2007);Wong (2002); Zurcher and Gardiol (1995)].

In this chapter, a brief description of the modeling techniques and basic characteristics of the MPA is given in section "Modeling Techniques and Basic Characteristics." Section "Broadbanding Techniques" presents four major broadbanding techniques. Dual- and triple-band designs are discussed in section "Dual- and Multiband Designs." Methods for size reduction and circular polarization are the topics of sections "Methods of Reducing the Patch Size" and "Designs for Circular Polarization," respectively. Section "Frequency-Agile and Polarization-Agile Patch Antennas" presents examples of frequency-agile and polarization-agile patch antennas. Concluding remarks are given in section "Conclusion." To keep the chapter within the allotted length, dual-polarization designs and the subject of microstrip antenna array are not addressed in this chapter.

Modeling Techniques and Basic Characteristics

Brief Description of Modeling Techniques

There are two approaches to deducing the performance characteristics of microstrip patch antennas. One is to devise a physical model based on a number of simplifying assumptions; the other is to attempt to solve Maxwell's equations subject to the boundary conditions.

For the coaxial feed microstrip patch antenna shown in Fig. 1a, a physical model, known as the cavity model (Lo et al. 1979), has been developed to deduce the characteristics of the antenna. It is based on a number of simplifying assumptions which are valid for thin substrates. These assumptions enable the fields between the patch and the ground plane to be determined analytically for a number of patch shapes. From these, radiation and impedance characteristics of the basic MPA geometry can be calculated. Basic characteristics obtained from the cavity model are presented in section "Basic Characteristics."

More general methods which aim at solving the full set of Maxwell's equations without relying on the thin substrate assumption are known as full-wave methods. One such method treats the currents on the antenna structure to be the unknowns to be solved. By requiring the fields produced by the currents to satisfy the boundary and excitation conditions, integral equations are obtained for the unknown currents, which are solved by some numerical method, such as the moment method (Mosig and Gardiol 1985). Another approach of the full-wave method converts Maxwell's equations into difference equations, which, together with appropriate models for the boundary conditions, are solved numerically in the time domain (Reineix and Jecko 1989). This is known as the finite difference time domain (FDTD)

DOI 10.1007/978-981-4560-75-7_29-1

Table 2 A list of commercial electromagnetic simulation software

Software name	Theoretical model	Model dimension	Company
Ensemble (DesignerRF)	Moment method	2.5	Ansoft (ANSYS)
IE3D	Moment method	2.5	Zeland
Momentum	Moment method	2.5	Agilent
EM	Moment method	2.5	Sonnet
PiCasso	Moment method/genetic	2.5	EMAG
FEKO	Moment method	3D	EMSS
Microwave studio	FDTD	3D	CST
Fidelity	FDTD	3D	Zeland
HFSS	Finite element	3D	Ansoft (ANSYS)

method. Still another method, known as the finite element method, solves Maxwell's eq. in the form of the vector wave eq. by the Rayleigh-Ritz variational method (Salon and Chari 1999).

The cavity model provides physical insight, but works well only for the basic geometry of the MPA, namely, a conducting patch on a thin, grounded substrate, with the patch shape amenable to the method of separation of variables. The full-wave methods have the advantages of being able to handle thick substrates, various feeding methods, and thick complicated structures such as multilayers, multiple patches with or without parasitic elements, shorting walls, and shorting pins, as well as slots cut in patches. However, they require extensive computation time and provide little physical insight.

In the last two decades, many commercial simulation software based on full-wave methods are available. These software have played a key role in the rapid advancement of microstrip antenna research. They have become indispensable tools in the design of these antennas. Some of these are listed in Table 2. Many of the results to be presented in this chapter were based on IE3D and HFSS. Some were obtained using moment method and FDTD codes developed in-house.

Basic Characteristics

The basic characteristics of microstrip patch antennas can be obtained by applying the cavity model to a coaxially fed microstrip patch antenna. The basic assumption in the cavity model is that the thickness of the substrate is much smaller than the wavelength. The region under the patch can then be considered as a cavity bounded by electric walls on the top and bottom ($E_t = 0$) and, as a consequence of the thin substrate assumption, bounded by a magnetic wall ($H_t = 0$) on the sides. The fields under the patch are transverse magnetic (TM), with the electric field perpendicular to the patch. These fields are determined by solving the wave equation subject to the conditions on the boundary and the feed. By virtue of the equivalence principle, the fields outside the cavity can be calculated from the equivalent sources on the exit region of the cavity.

Using this approach, the performance characteristics of the probe-fed rectangular, circular, annular-ring, and equitriangular patch antennas were predicted theoretically and verified experimentally in the late 1970s and 1980s (Lee and Dahele 1989). While differing in detail, there are a number of similar features, irrespective of the shapes of the patch. These are delineated below for the rectangular patch shown in Fig. 3. The rectangular patch is probably the most commonly used microstrip patch antenna.

(A) The fields under the cavity are transverse magnetic, with the electric field in the z direction and independent of z. There are an infinite number of modes, each characterized by a pair of integers (m, n):

DOI 10.1007/978-981-4560-75-7_29-1

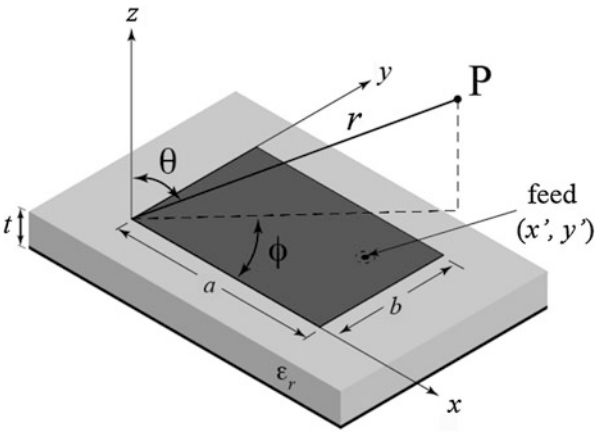


Fig. 3 Geometry of a rectangular patch antenna

$$E_z = E_0 \cos\frac{m\pi x}{a} \cos\frac{n\pi y}{b} \quad m, n = \text{integers} \tag{1}$$

(B) For the cavity bounded by electric walls on the top and a magnetic wall on the side, the resonant frequency of each mode is governed by the dimensions of the patch and the relative permittivity of the substrate ε_r. It is given by

$$f_{mn} = \frac{k_{mn} \cdot c}{2\pi\sqrt{\varepsilon_r}} \tag{2}$$

where

$$k_{mn} = \left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right]^{\frac{1}{2}} \tag{3}$$

c = velocity of light in free space.

(C) Because of fringing fields at the edge of the patch, the patch behaves as if it has a slightly larger dimension. Semiempirical factors are usually introduced in the cavity model-based design formulas to account for this effect, as well as the fact that the dielectric above the patch (usually air) is different from the dielectric under the patch. These factors vary from patch to patch.

For the rectangular patch, with a > b, a commonly used formula for the fundamental mode, accurate to within 3 % of measured values, is

$$f_r = \frac{c}{2(a+t)\sqrt{\varepsilon_e}} \tag{4}$$

where

$$\varepsilon_e = \frac{(\varepsilon_r + 1)}{2} + \frac{(\varepsilon_r - 1)}{2}\left[1 + \frac{10t}{b}\right]^{\frac{-1}{2}} \tag{5}$$

is the effective permittivity (Schneider 1969).

DOI 10.1007/978-981-4560-75-7_29-1

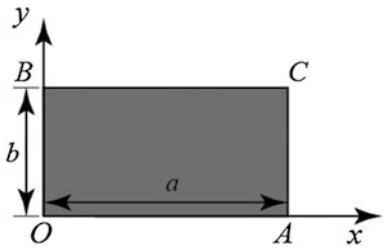


Fig. 4 Top view of the rectangular patch

(D) The equivalent sources at the exit region (the vertical side walls) are the surface magnetic current densities, related to the tangential electric fields in those locations. Referring to Fig. 4, they are:

Along the wall OA,

$$\vec{M}_s = -\hat{n} \times \vec{E} = +\hat{y} \times \hat{z} \times E_0 \cos\frac{m\pi x}{a} = +\hat{x}E_0\frac{m\pi x}{a}. \tag{6}$$

Along the wall AC,

$$\vec{M}_s = -\hat{n} \times \vec{E} = -\hat{x} \times \hat{z} \times E_0 \cos m\pi \cos\frac{n\pi y}{b} = +\hat{y}E_0 \cos m\pi \cos\frac{n\pi y}{b}. \tag{7}$$

Along the wall CB,

$$\vec{M}_s = -\hat{n} \times \vec{E} = -\hat{y} \times \hat{z} \times E_0 \cos\frac{m\pi x}{a} \cos n\pi = -\hat{x}E_0 \cos\frac{m\pi x}{a} \cos n\pi. \tag{8}$$

Along the wall BO,

$$\vec{M}_s = -\hat{n} \times \vec{E} = +\hat{x} \times \hat{z} \times E_0 \cos\frac{n\pi y}{b} = -\hat{y}E_0 \cos\frac{n\pi y}{b}. \tag{9}$$

The effect of the infinite ground plane is to double the strength of each $\vec{M}_s$ source. The electric field and magnetic surface current distributions on the side walls for the TM_{01} and TM_{10} modes are illustrated in Fig. 5. For the TM_{10} mode, the magnetic currents along b are constant and in phase, while those along a vary sinusoidally and are out of phase. For this reason, the b edge is known as the radiating edge since it contributes predominately to the radiation. The a edge is known as the non-radiating edge. Similarly, for the TM_{01} mode, the magnetic currents are constant and in phase along a and are out of phase and vary sinusoidally along b. The a edge is thus the radiating edge for the TM_{01} mode.

(E) To satisfy the boundary condition imposed by the feed, the fields under the patch are expressed as a summation of the various modes. The amplitudes and phases of these modes are determined by the excitation. The mode with resonant frequency equal to the excitation frequency will be at resonance and has the largest amplitude.

(F) Each resonant mode has its own characteristic radiation pattern. For the rectangular patch, the commonly used modes are TM_{10} and TM_{01}. However, the TM_{03} mode has also received some attention. These three modes all have broadside radiation patterns. The computed patterns for a = 1.5b and two values of ε_r are shown in Fig. 6. In the principal planes, the TM_{01} and TM_{03} modes have similar linear polarizations, while that of the TM_{10} mode is orthogonal to the other two.

DOI 10.1007/978-981-4560-75-7_29-1

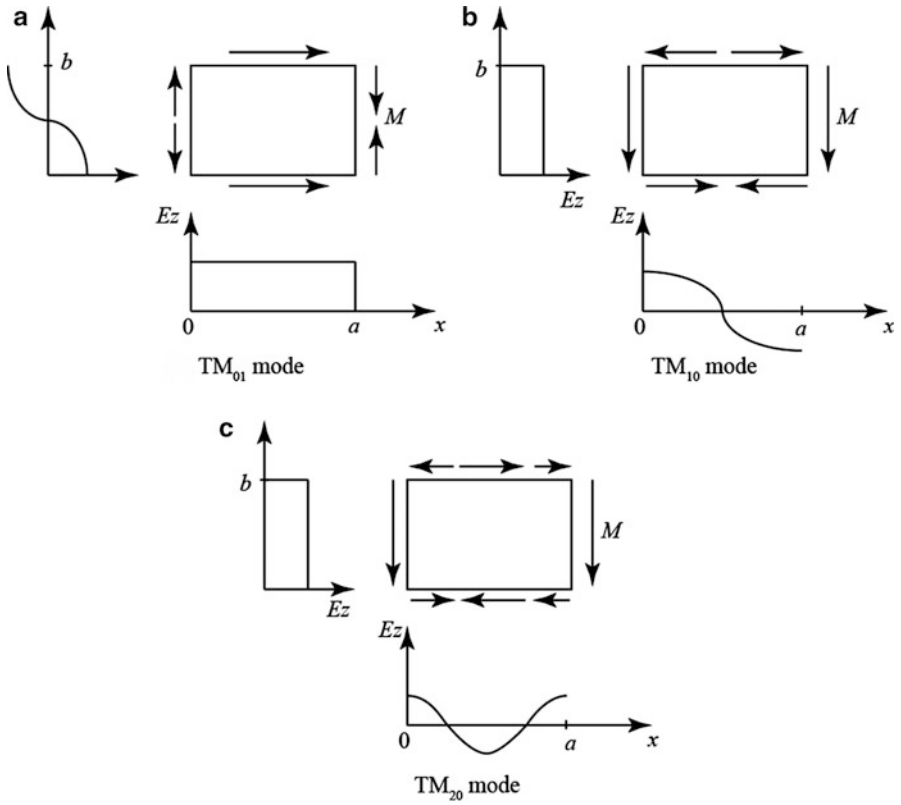


Fig. 5 Illustrations of the magnetic current density distribution on the side walls

The patterns do not appear to be sensitive to a/b or t. However, they change appreciably with ε_r. Typical half-power beamwidths of the TM_{10} and TM_{01} modes are of the order of 100° and the gains are typically 5 dBi. The patterns of most of the other modes have maxima off broadside. For example, those of the TM_{11} mode are illustrated in Fig. 6g.

Figure 7 shows the computed and measured radiation patterns of the TM_{10} and TM_{01} modes obtained by Lo et al. (1979) for a rectangular patch with a $= 11.43$ cm, b $= 7.6$ cm, $\varepsilon_r = 2.62$,and t $= 0.159$ cm. Both $\overrightarrow{E_\theta}$ and $\overrightarrow{E_\phi}$ were measured in each of the two cuts, $\phi = 0^{\circ}$ and $\phi = 90^{\circ}$. It was found in their experiment that, in each case, one component polarization was negligible when compared to the other and is not shown. The dominant component is called the co-polarization, and the component orthogonal to the dominant component is called the cross polarization. For thin substrates, cross-polarization fields are mainly due to the excitation of higher-order modes. A study of the dependence of the cross-polarized fields on aspect ratio (a/b), feed position, and relative permittivity can be found in Huynh et al. (1988).

For each mode, there are two orthogonal planes in the far-field region – one designated as E-plane and the other designated as H-plane. The far-zone electric field lies in the E-plane and the far-zone magnetic field lies in the H-plane. The patterns in these planes are referred to as the E- and H-plane patterns, respectively. For the TM_{01} mode, the contributions to the far fields are from the magnetic surface current densities on the side walls containing the radiating edges. Referring to Figs. 3 and 5, one can show from the directions of these magnetic currents that the E-plane is the y-z plane ($\phi = 90^{o}$) and the H-plane is the x-z plane ($\phi = 0^{o}$).

For the TM_{10} mode, again referring to Figs. 3 and 5, one can show that the E-plane is the x-z plane ($\phi = 0^{o}$) and the H-plane is the y-z plane ($\phi = 90^{o}$).

DOI 10.1007/978-981-4560-75-7_29-1

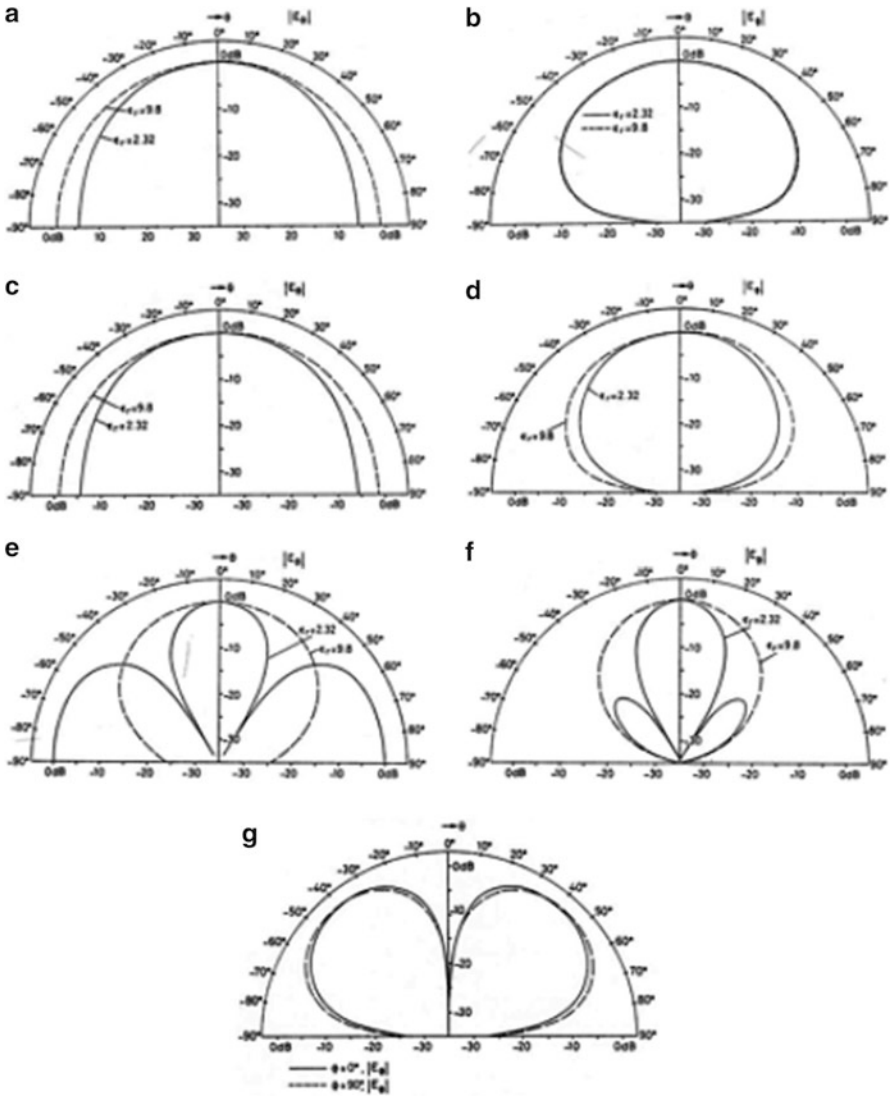


Fig. 6 Relative field patterns for a rectangular patch with a/b = 1.5, f_{nm} = 1 GHz, and (i) $\varepsilon_r = 2.32$, t = 0.318, 0.159, 0.0795 cm; (ii) $\varepsilon_r = 9.8$, t = 0.127, 0.0635, 0.0254 cm. (**a**) TM_{10}, $\phi = 0°$. (**b**) TM_{10}, $\phi = 90°$. (**c**) TM_{01}, $\phi = 90°$. (**d**) TM_{01}, $\phi = 0°$. (**e**) TM_{03}, $\phi = 90°$. (**f**) TM_{03}, $\phi = 0°$. (**g**) TM_{11}, $\varepsilon_r = 2.32$

With appropriate design, circular polarization can be achieved by utilizing two modes. This will be discussed in detail in section "Designs for Circular Polarization."

(G) At resonance, the input reactance is small for thin substrates, while the input resistance is largest when the feed is near the edge of the patch and decreases as the feed moves inside the edge. The decrease follows the square of a cosine function for the TM_{10} and TM_{01} modes of a coaxial-fed rectangular patch. Figure 8 shows the theoretical and measured values of the resonant resistances of the first two modes of a coaxial-fed rectangular patch.

(H) By choosing the feed location properly, the resonant resistance can be matched to the feed line resistance, while the use of thin substrates (thickness $t \leq 0.03\ \lambda_0$) will minimize the feed inductance at resonance, resulting in a very near unity voltage standing wave ratio (VSWR or SWR). As the frequency deviates from resonance, VSWR increases. For linear polarization, a common definition of impedance bandwidth is the range of frequencies for which VSWR is less than or equal to two, corresponding to 10 dB return loss or −10 dB for the reflection coefficient (scattering parameter S_{11}).

DOI 10.1007/978-981-4560-75-7_29-1

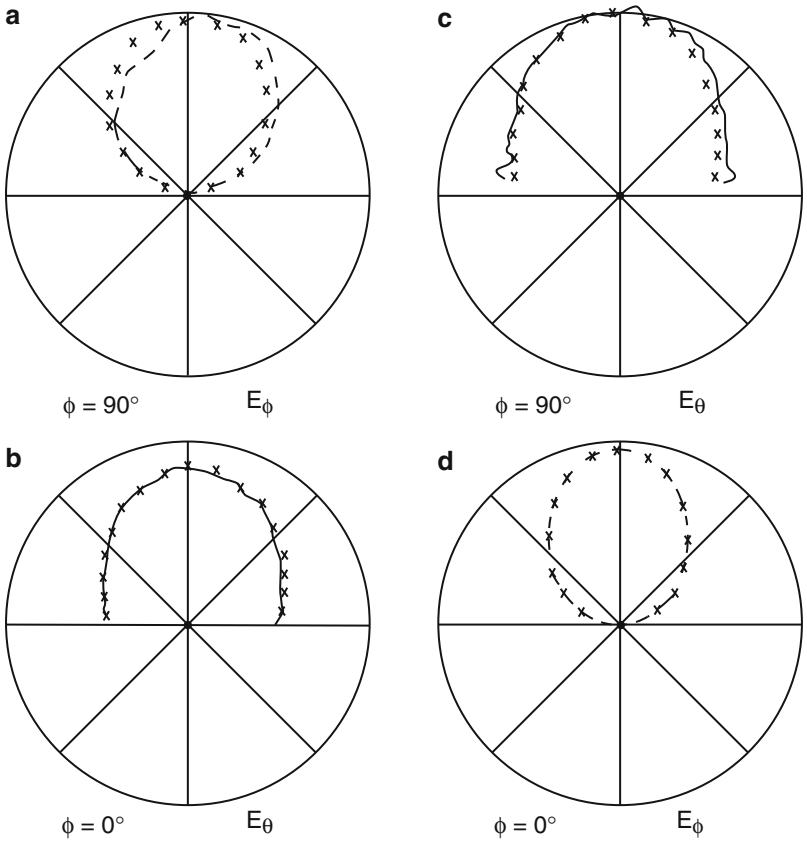


Fig. 7 Theoretical (x) and measured (*solid or dashed line*) radiation patterns in $\phi = 0°$ and $\phi = 90°$ planes of a rectangular patch antenna with a = 11.43 cm, b = 7.6 cm, $\varepsilon_r = 2.62$, t = 0.159 cm. (**a**) and (**b**) at resonant frequency 804 MHz of (1, 0) mode; (**c**) and (**d**) at resonant frequency 1,187 MHz of (0, 1) mode (From Lo et al. (1979). © 1979 I.E. Reprinted with permission)

This is usually also the antenna bandwidth, as the patterns are much less sensitive to frequency. For circular polarization, bandwidth is determined by both VSWR ≤ 2 and axial ratio $\leq$3 dB.

(I) The losses in the patch antenna comprise radiation, copper, dielectric, and surface-wave losses. For thin substrates, surface wave can be neglected. According to Wood (1981), it is required that $t/\lambda_o < 0.07$ for $\varepsilon_r = 2.3$ and $t/\lambda_o < 0.023$ for $\varepsilon_r = 9.8$, if the antenna is to launch no more than 25 % of the total radiated power as surface waves. The quality factor Q of a particular mode is determined by the ratio of the stored to loss energy and determines the impedance bandwidth of the antenna.

(J) In general, the impedance bandwidth is found to increase with substrate thickness t and inversely proportional to $\sqrt{\varepsilon_r}$. However, the use of low-permittivity substrates can lead to high levels of radiation from the feed lines, while for higher permittivities, an increase in substrate thickness can lead to decrease in efficiency due to surface-wave generation. Additionally, when the substrate thickness exceeds about 0.05 λ_o, where λ_o is a free-space wavelength, the antenna cannot be matched to the feed line. As a result, for the basic MPA geometry, the impedance bandwidth is limited to about 5 %.

DOI 10.1007/978-981-4560-75-7_29-1

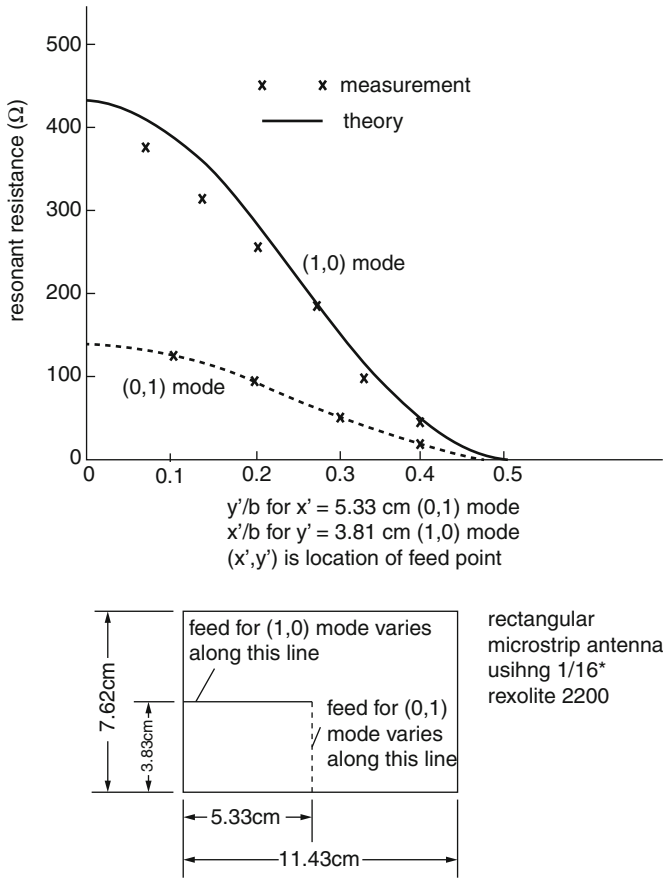


Fig. 8 Variation of resonant resistance with feed position of the TM_{10} and TM_{01} mode in a rectangular patch antenna with a = 11.43 cm, b = 7.62 cm, $\varepsilon_r = 2.62$, t = 0.159 cm (From Richards et al. (1981). © 1981 I.E. Reprinted with permission)

Broadbanding Techniques

General Remarks

The operating frequencies of some wireless communication systems are shown in Table 3.

While the bandwidths of several systems (e.g., GPS) are quite narrow, five of the systems shown in Table 3 require bandwidths in excess of 8 %. As discussed previously, the microstrip patch antenna in the basic form of a conducting patch in a grounded substrate is inherently narrow band and is not able to meet the requirements of these systems. While bandwidth can be increased by using lossy substrates, this is usually not desirable as efficiency will be reduced. In the last three decades, a number of techniques have been developed to broaden the bandwidths of microstrip patch antennas, without compromising efficiency. The various designs provide bandwidths in the range of 10–60 %. Note that a 20 % bandwidth will be able to cover the first two applications in Table 3, while a 30 % bandwidth will cover the third, fourth, and fifth applications combined.

The methods developed for efficient wideband patch antenna design are based on one or more of the following principles:

(A) By means of parasitic elements or slots, additional resonances are introduced so that, in conjunction with the main resonance, an overall broader band response is obtained.
(B) Thick substrates of low permittivity are used.
(C) A scheme is devised to reduce the mismatch problem associated with thick substrates.

DOI 10.1007/978-981-4560-75-7_29-1

Table 3 Frequencies of some wireless communication systems

System	Operating frequency	Overall bandwidth
Advanced mobile phone service (AMPS)	UL: 824–849 MHz DL: 869–894 MHz	70 MHz (8.1 %)
Global system for mobile communications (GSM)	UL: 880–915 MHz DL: 925–960 MHz	80 MHz (8.7 %)
Global positioning system (GPS)	L1: 1,575.42 MHz L2: 1,227.60 MHz L5: 1,176.45 MHz	P-code[a]: 20.46 MHz C/A code[b]: 2.046 MHz (max. 1.7 %)
Personal communications service (PCS)	UL: 1,710–1,785 MHz DL: 1,805–1880 MHz	170 MHz (9.5 %)
Universal mobile telecommunication systems (UMTS)	UL: 1,920–1,980 MHz DL: 2,110–170 MHz	250 MHz (12.2 %)
Wi-Fi IEEE 802.11(a, b, g, n) IEEE802.11 ac	(b, g, n) 2,400–2,497 MHz (a, n) 5,150–5,350 MHz (a, n) 5,725–5,825 MHz 5,170–5,330 MHz 5,490–5,710 MHz 5,735–5,835 MHz	(a, b, g) 20 MHz (n) 20, 40 MHz (max. 0.8 %) 20, 40, 80, 160 MHz (max. 3 %)
Wireless gigabit	56,160–64,800 MHz	8,640 MHz (14 %)
WiMAX	M2300T-01/02: 2,300–2,400 MHz M2500T-01: 2,496–2,690 MHz M3500T-02/03/05: 3,400–3,600 MHz	100 MHz 194 MHz 200 MHz (max. 5.7 %)
Bluetooth	2,400–2,483.5 MHz	83.5 MHz (3.4 %)

[a]Precision (P)
[b]Coarse-acquisition (C/A)

Several wideband designs based on the above principles will now be presented.

Stacked Patches

A popular method for increasing the bandwidth of microstrip patch antennas is to use parasitic elements to create dual or multiple resonances. If these resonances are suitably arranged, the resulting overall frequency response can be significantly broader than the case when only a single resonance is present. This is similar to the situation in coupled tuned circuits. There are two arrangements: the stacked geometry in which the fed and parasitic patches are on separate layers (Sabban 1983; Lee et al. 1987, 1995; Barlately et al. 1990) and the coplanar geometry in which all patches are on the same layer (Kumar and Gupta 1984). Thin substrates are used. The coplanar geometry increases the lateral area occupied by the patches and is not as popular as the stacked geometry and will not be discussed further.

One of the first journal papers on two-layer stacked patches is that by Lee et al. (1987), who studied the configuration shown in Fig. 9.

The substrate has relative permittivities, $\varepsilon_r = 2.17$ and thickness $t = 0.254$ mm. They are separated by a distance s by means of spacers. The dimensions of the patches are identical, with $a = 1.5$ cm and $b = 1.0$ cm. It is found that when the patches are excited at the TM_{01} mode, with resonant frequency approximately 10 GHz, the impedance bandwidth is about 13 % when the spacing is around 0.0508 cm ($0.017\ \lambda_0$). This is almost an order of magnitude larger than the case when there is no parasitic patch.

Subsequently, Lee et al. (1995) developed a full-wave moment method analysis and a computer program for multilayer microstrip antennas. Using this program, representative design guides for the configuration of Fig. 10, operating at the center frequency of 5 GHz, are shown in Table 4. In Table 4,

DOI 10.1007/978-981-4560-75-7_29-1

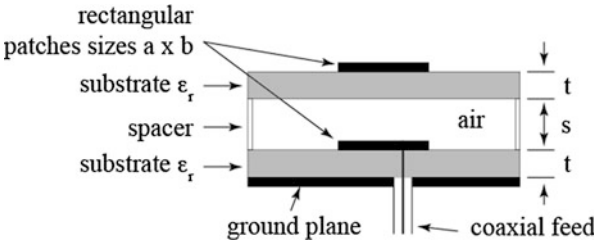


Fig. 9 Geometry of rectangular electromagnetically coupled stacked patch antenna

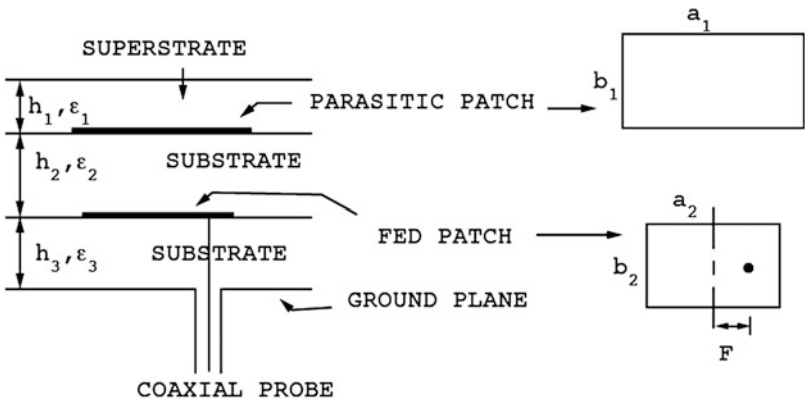


Fig. 10 Geometry of the stacked electromagnetically coupled microstrip antenna with superstrate

Table 4 Design examples for stacked electromagnetically coupled patch antennas at the center frequency of 5 GHz

	12 %	12 %	15 %
BW (VSWR $\leq$ 2)	Design 1	Design 2	Design 3
ε_1	1.0	2.2	1.0
ε_2	1.2	1.2	1.2
ε_3	2.2	2.2	2.2
h_1 (mm)	0.0	0.26	0.0
h_2 (mm)	3.580	3.630	3.500
h_3 (mm)	0.493	0.486	1.200
a_1 (cm)	2.296	2.198	2.300
b_1 (cm)	1.275	1.465	1.278
a_2 (cm)	2.000	2.027	2.000
b_2 (cm)	1.111	1.351	1.111
F (cm)	0.928	0.940	0.900

design 1 gives the parameters which achieve a bandwidth of 12 % for the case when there is no superstrate (dielectric cover). When a superstrate of thickness of 0.26 mm and relative permittivity of 2.2 is placed on top of the parasitic patch, the parameters which yield 12 % impedance bandwidth are given by design 2. Design 3 provides the antenna parameters which result in a bandwidth of 15 % when no superstrate is present. If the center frequency is changed, it is only necessary to scale the length parameters accordingly (patch dimensions, substrate and superstrate thickness, feed location).

DOI 10.1007/978-981-4560-75-7_29-1

The patterns of stacked patches are stable across the impedance bandwidth. Typical E- and H-plane half-power bandwidths are 76° and 86°, respectively. This is to be compared with 92° and 86° for the single patch. The gain of the stacked patches is about 6.0 dBi and that of the single patch is about 5.2 dBi.

It is obvious that, by extending the stacked patch geometry to multiple parasitic patches in multiple layers, additional resonances can be introduced, further broadening the bandwidth of the antenna. However, the antenna becomes very thick when there are more than two layers.

Aperture-Coupled Patches

General Remarks About Aperture-Coupled Feed

While the most popular methods for feeding a patch antenna are directly connecting a coaxial probe or stripline to the patch, a method which has a number of advantages is to couple energy from the stripline through an aperture (slot) in the ground plane, as shown in Fig. 1d. The open-ended microstrip line is located on a dielectric slab below the ground plane, while the patch antenna is formed on a separate dielectric slab above the ground plane. The two structures are electromagnetically coupled through an aperture in the ground plane between them. In the original paper by Pozar (1985), the aperture was in the form of a small circular hole. Subsequently, a more common shape of the aperture was in the form of a narrow rectangular slot.

As listed in Table 1, one advantage of this feeding method is that the feed network is isolated from the radiating element by the ground plane, which prevents spurious radiation. Another advantage is that active devices such as phase shifters and amplifiers can be fabricated in the feed substrate with high dielectric constant for size reduction, while the radiating patch can be mounted on low dielectric constant substrate in order to increase bandwidth and radiation efficiency.

The slot that couples energy from the stripline to the patch can be either resonant or nonresonant. The bandwidth achievable by nonresonant slot excitation is typically 6–7 % but can be as large as 10–13 % by utilizing thick substrates, since the problem of probe inductance is not applicable here. If the slot is resonant, it provides another resonance in addition to the patch resonance, resulting in an antenna with impedance bandwidth in excess of 20 %. However, the resonant slot has a strong back lobe radiation which reduces the gain of the antenna and causes interference problems in certain applications.

As with coaxially fed microstrip antennas, the impedance bandwidth of aperture-coupled patch antennas can be further enhanced by using stacked or coplanar elements.

In the next section, an example of a wideband aperture-coupled patch antenna utilizing a resonant slot is presented.

Example of a Wideband Aperture-Coupled Patch Antenna

By using a resonant slot and relatively thick foam substrate for the patch, Croq and Papiernik (1990) studied the aperture-coupled patch antenna shown in Fig. 11. Note that there was a dielectric cover (radome) protecting the patch.

The measured and computed impedances showed that, in the frequency range 4.85–6.1 GHz, the VSWR was less than 1.5, corresponding to a bandwidth of about 22 %.

The antenna gain was found to be about 8 dB for the entire bandwidth. The maximum back to front level was about −14 dB at the frequency of 5.6 GHz and was about −12 dB over the band. The strong back radiation is a major disadvantage of a resonant slot aperture-coupled patch antenna.

The Wideband U-Slot Patch Antenna

The stacked patch geometry with one fed patch and one parasitic patch can provide up to about 20 % bandwidth, while the aperture-coupled patch utilizing a resonant slot can yield up to about 25 %

DOI 10.1007/978-981-4560-75-7_29-1

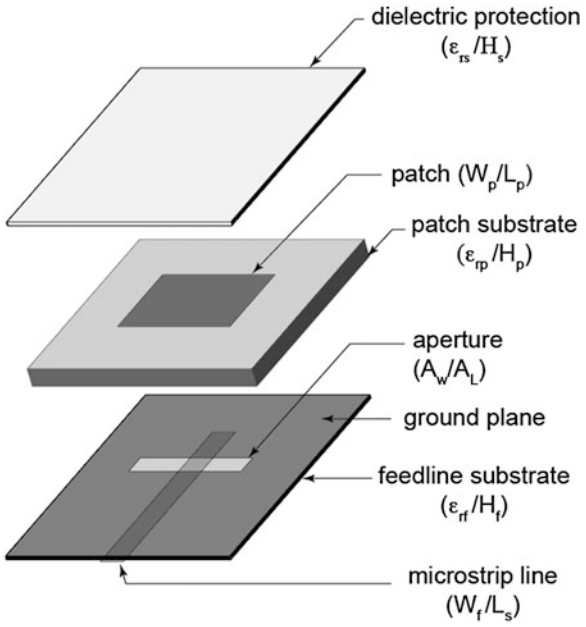


Fig. 11 Aperture-coupled patch antenna studied in the paper by Croq and Papiernik (1990). (**a**) Feed: $\varepsilon_{rf} = 2.2$; $tg_\delta = 0.001$; $H_f = 0.762$ mm; $W_f = 2.32$ mm; $L_s = 2.85$ mm. (**b**) Slot: $A_w = 0.8$ mm; $A_L = 15.4$ mm. (**c**) Square patch: $W_p = 17$ mm; $H_p = 5.5$ mm; $\varepsilon_{rp} = 1$. (**d**) Radome: $H_s = 1.6$ mm; $\varepsilon_{rs} = 2.2$; $tg_\delta = 0.001$

bandwidth. The former involves more than one layer and more than one patch, while the latter also involves more than one layer, as well as a more complicated feeding arrangement and strong back lobe radiation. A wideband design which is single layer and single patch is the U-slot patch antenna shown in Fig. 12, first introduced in 1995 by Huynh and Lee (1995). A number of studies by Lee and coworkers, as well as by others, followed (Lee et al. 1997, 2010; Tong et al. 2000; Clenet and Shafai 1999; Weigand et al. 2003). It was firmly established that the U-slot patch antenna can provide impedance bandwidths in excess of 30 % for an air substrate of thickness about 0.08 λ_0 and in excess of 20 % for material substrates of similar thickness.

Air Substrate

In the original study of Huynh and Lee (1995), the wide-bandwidth characteristics of the antenna were demonstrated experimentally. It was pointed out in their paper that the factors contributing to the wideband behavior were (1) the air substrate; (2) a relative thick substrate (about 0.08 λ_0); (3) the capacitance introduced by the U-slot, which countered the feed inductance; and (4) the additional resonance introduced by the U-slot, which combined with the patch resonance to produce a broadband response. Subsequently, Lee et al. (1997) produced a more comprehensive experimental and FDTD modeling of the air-substrate U-slot patch antenna. Results for the VSWR response of one such antenna are shown in Fig. 13. The impedance bandwidth is about 30 %. The measured patterns (not shown here) were stable across the band. The E-plane and H-plane beamwidths were about 70° and 65°, respectively. The gain of the antenna was around 7.5 dBi, about 2 dB higher than the traditional microstrip antenna. The cross polarization was negligible in the E-plane. It was about 12 dB below the co-polarization level in the H-plane at the center of the band, but it increased to about 8 dB at the outer edges of the band.

While the abovementioned studies, as well as others, have shown that more than 30 % impedance bandwidth can be obtained when an air-substrate thickness of about 0.08 λ_0 is used, it should be pointed out that some applications do not need such a wide bandwidth. For example, referring to Table 3, in Advanced Mobile Phone Service (AMPS), about an 8.1 % bandwidth is sufficient. Likewise, in Global

DOI 10.1007/978-981-4560-75-7_29-1

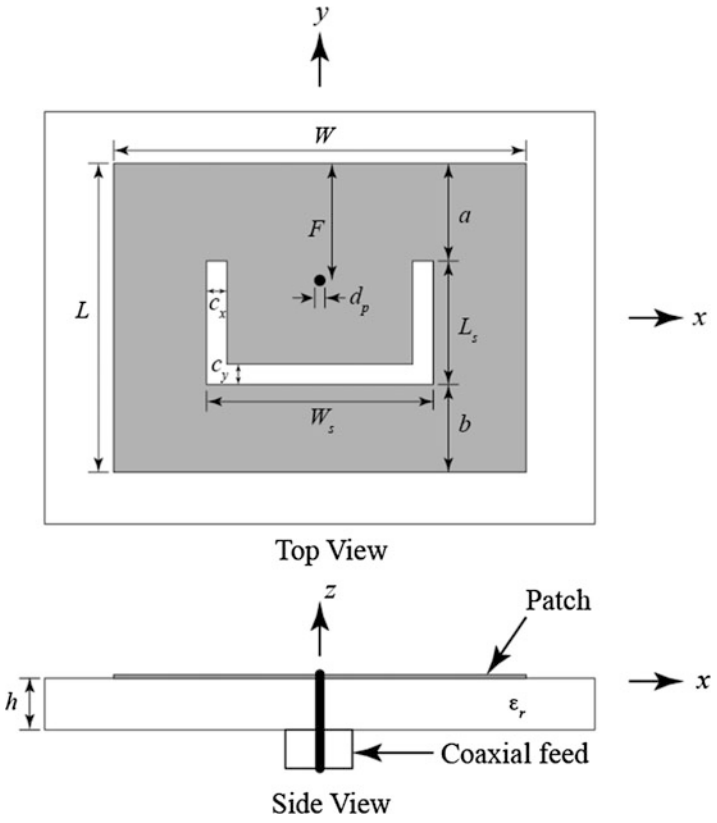


Fig. 12 Geometry of the rectangular U-slot patch antenna

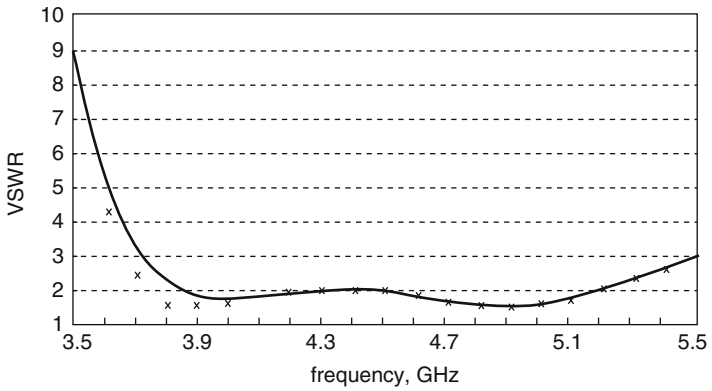


Fig. 13 VSWR of the U-slot patch antenna with dimensions: W = 36 mm, L = 26 mm, F = 13 mm, W_s = 12 mm, L_s = 20 mm, a = 2 mm, b = 4 mm, $c_x = c_y$ = 2 mm, and h = 5 mm. (x measured, — computed) (From Lee et al. (1997). © 1997 IET, Reprinted with permission)

System for Mobile Communications (GSM), only an 8.7 % bandwidth is needed. While such bandwidths cannot be realized by the traditional patch antenna (Lee and Luk 2010), it has been demonstrated that these can be realized by a U-slot patch antenna only 0.033 λ_0 thick, which has a 12 % bandwidth. The dimensions of the antenna used in this study are shown in Table 5, and the results are summarized in Table 6.

DOI 10.1007/978-981-4560-75-7_29-1

Table 5 Dimension of the U-slot patch antenna with different air-substrate thickness (unit: mm) (From Lee et al. (2010). © 2010 IEEE, Reprinted with permission)

h (mm)	W	L	W_s	L_s	a	b	F
1	35.5	26	12	10.7	0.1	0.6	15
2	35.5	26	11	12.3	0.6	0.8	15
3	35.5	26	11	14.2	0.9	1	15
4	35.5	26	11	16.2	1.8	1.6	15
5	35.5	26	11	20	3.8	2.1	15
6	35.5	26	11	21	4.3	2.1	15

Table 6 Simulated and measured impedance matching bandwidths of probe-fed U-slot patch antenna with different air-substrate thickness. The dimensions of the antennas are slightly different for the different thicknesses. In all cases, the inner radius of the coaxial feed is 0.4 mm (From Lee et al. (2010). © 2010 IEEE, Reprinted with permission)

	Simulation			Measurement	
h (mm)	Thickness in lambda	Matching freq/GHz	% BW	Matching freq/GHz	%BW
1	0.018	5.13–5.41	5.31		
2	0.033	4.73–5.23	10.04	4.44–5.01	12.1
3	0.048	4.45–5.2	15.54	4.19–4.95	16.6
4	0.063	4.2–5.28	22.78	4.05–5.1	23.0
5	0.078	3.98–5.32	28.81	3.8–4.94	26.1
6	0.089	3.8–5.14	29.97	3.74–4.98	28.4

Table 7 Dimensions of antenna in millimeters

ε_r	W	L	W_s	L_s	b	F	c_x	c_y	h
2.33	36.0	26.0	14.0	18.0	4.0	13.0	2.0	2.0	6.4

Table 8 Operating frequencies and bandwidth of the antenna in Table 7

	f_l(GHz)	f_o(GHz)	f_u(GHz)	BW (GHz)	%BW (%)
Computed	2.87	3.28	3.69	0.82	25.0
Measured	2.76	3.16	3.56	0.80	25.3

Material Substrate

Although the first series of investigations of the U-slot patch antenna used an air or foam substrate, subsequent investigations have confirmed that this wideband design can also be implemented with material substrates. As expected, the bandwidth of an antenna on a material substrate is smaller than an antenna on an air or foam substrate.

Tong et al. (2000) presented both experimental study and FDTD analyses of two antennas on a substrate with relative permittivity $\varepsilon_r = 2.33$. The dimensions of one of these antennas are shown in Table 7. The operating frequencies and bandwidths of this antenna are shown in Table 8. The 3 dB gain bandwidths were about the same as the impedance bandwidths, and the average gains of the antennas were about 7 dBi across the matching band.

DOI 10.1007/978-981-4560-75-7_29-1

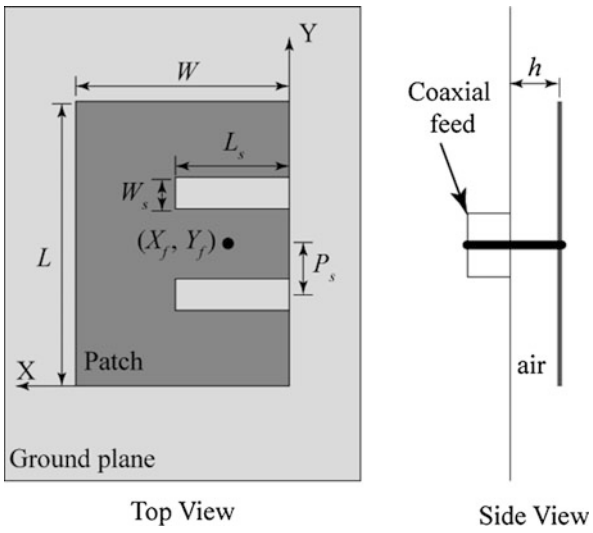


Fig. 14 Geometry of the E-patch antenna

Variations of the U-Slot Patch Antenna and the E-Patch

The use of the U-slot to achieve wideband performance is not limited to a rectangular patch. Wideband circular patch antennas with a U-shaped slot have been reported (Bhalla and Shafai 2002). The shape of the slot can also take on other forms, such as a circular slot and a V-slot (Rafi and Shafai 2004). By letting the width of the horizontal slot go to zero and extending the two vertical slots to the edge of the patch, an E-patch is obtained (Yang et al. 2001a). This geometry is shown in Fig. 14. As in the U-slot, the parallel slots provide an additional path for the currents, giving rise to a second resonance. The parallel slots can also introduce a capacitance which compensates for the probe inductance, thus enabling the use of relatively thick substrates. In Yang et al. (2001a), impedance bandwidths of about 30 % were obtained for E-patches operating at the center frequency of around 2.4 GHz, using air substrate of about 0.08 λ_0. The antenna parameters for one such antenna are listed below, in mm: L = 70, W = 30, h = 15, $X_f = 35$, $Y_f = 6$, $L_s = 40$, $W_s = 6$, and $P_s = 10$. Ground plane size = 14 × 21 cm.

The Wideband L-Probe Coupled Patch Antenna

Shortly after the publication of the first paper on the wideband U-slot patch antenna, another wideband single-layer single-patch antenna was introduced. This design achieves wideband operation using an L-shaped probe feeding method. The L-probe feed was first used by Nakano in the electromagnetically coupled curl antenna (Nakano et al. 1997). Its use in patch antennas was due to Luk et al. (1998a). The geometry is shown in Fig. 15. This design uses low-permittivity substrate (air or foam) of thickness about 0.1 free-space wavelength. The feed is a modified version of the coaxial probe. Instead of the center conductor extending vertically to the patch and connected to it, a portion of it is bent in the horizontal direction. The horizontal arm of the probe is approximately a quarter of a wavelength long. It provides a capacitance to counteract the inductance due to the vertical part. In the experiment of Luk et al. (1998a), the rectangular patch had a width $W_x = 30$ mm and a length $W_y = 25$ mm. The L-shaped probe, connected to the inner conductor of a 50 Ω SMA launcher, had a parallel arm $L_h = 10.5$ mm and a perpendicular arm $L_v = 4.95$ mm. It was excited in the TM_{01} mode of the patch antenna. The patch was supported by a foam layer of thickness H = 6.6 mm with a dielectric constant close to unity. Within the matching band, this thickness corresponded to 0.08–0.12 λ_0. The distance between the lower edge of the patch and the perpendicular arm of the L-probe was D = 2 mm. The probe radius R = 0.5 mm. The measured VSWR and gain as a function of frequency are shown in Fig. 16. The bandwidth (VSWR < 2)

DOI 10.1007/978-981-4560-75-7_29-1

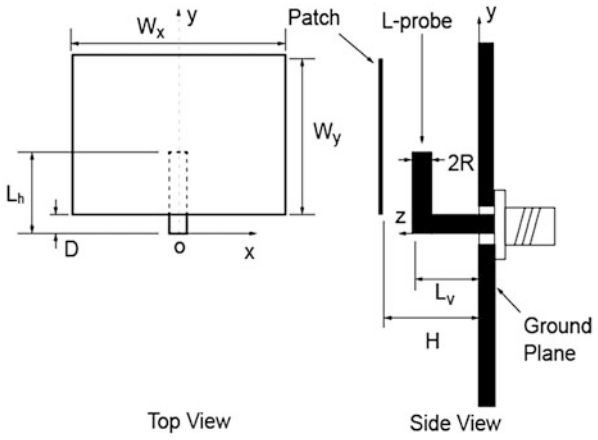


Fig. 15 Geometry of the patch antenna with an L-probe feed

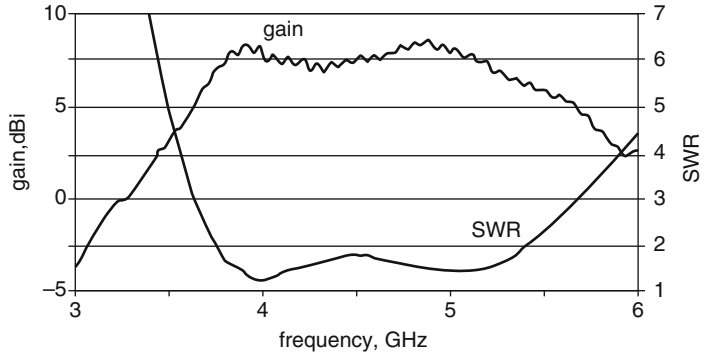


Fig. 16 Measured gain and SWR against frequency (From Luk et al. (1998a). © 1998 IET, Reprinted with permission)

was 36 % and the average gain was 7.5 dBi. The measured radiation patterns at 4.53 GHz are shown in Fig. 17. It is seen that the co-polarization patterns were symmetrical with respect to the broadside direction. The cross polarization was negligible in the E-plane (y-z plane). It was quite high in the H-plane (x-z plane) in the direction around $\theta = 30°$. Simulation results of this antenna using the software IE3D were in good agreement with experimental results.

Subsequent related studies included the use of twin L-probes to increase gain and reduce cross polarization (Mak et al. 2005) and the T-probe-fed patch antennas (Mak et al. 2000).

Dual- and Multiband Designs

Introductory Remarks

There are many applications in wireless communications that involve two or more distinct frequency bands. It is sometimes possible that a broadband microstrip antenna can cover the frequencies of interest. However, the disadvantage of using a broadband antenna is that it also receives non-desired frequencies unless some kind of filtering network is introduced to reject such frequencies. On the other hand, the advantage of a dual- or multiband design is that it focuses only on frequencies of interest and is thus more desirable. The frequency bands can be relatively close or relatively far apart. A variety of designs have been developed to provide the desired characteristics: by providing multi-resonances, by using multi-

DOI 10.1007/978-981-4560-75-7_29-1

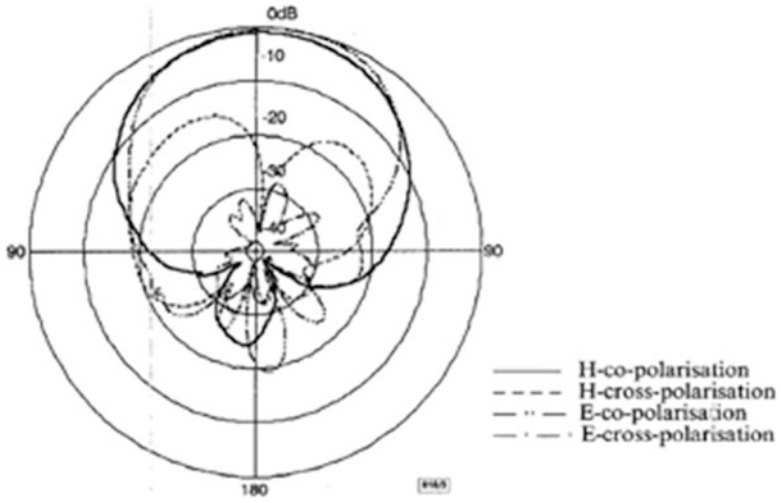


Fig. 17 Measured radiation pattern at 4.53 GHz (From Luk et al. (1998a). © 1998 IET, Reprinted with permission)

modes, or by introducing band notches in a broadband antenna. Long and Walton(1979) and Dahele et al. (1987) used stacked circular and stacked annular-ring patches for dual-band designs. This technique was extended to multibands by Anguera et al. (2003). Dual- or triple-band characteristics can also be realized by loading a rectangular patch with one or two stubs (Deshmukh and Ray 2010), by utilizing the TM_{01} and TM_{03} modes of a rectangular patch (Zhong and Lo 1983) or the TM_{10}, TM_{20}, and TM_{11} modes of the equilateral-triangular patch (Lee et al. 1988). Recently, it was found that the U-slot patch geometry, originally introduced for broadband operation, can be suitably modified for dual- and multiband operations (Lee et al. 2008, 2011). This method is discussed in detail in the next section.

The Use of U-Slot Patches

Most of the studies of the U-slot patch antenna were concerned with its broadband capabilities. However, recent studies have shown that it can be also used for dual- or multiband designs. There are two approaches, depending on whether the frequency ratios are larger than or less than about 1.5. The two cases will be referred to as large frequency ratio and small frequency ratio, respectively.

Large Frequency Ratios

For the case when the frequency ratio is larger than about 1.5, the approach to dual-band design is to use a coaxial feed and adjust the U-slot dimensions so that the patch resonance and the slot resonance do not merge to yield a broadband response. Figure 18 shows a U-slot patch antenna which operates at a band centered at 2.0 GHz and another band centered at 4.8 GHz. The patch dimension L (54 mm) determines the lower resonance, while the upper resonance is governed by the dimensions of the U-slot. The simulated reflection coefficient (S_{11}) versus frequency is shown in Fig. 19. The impedance bandwidths at the lower and upper bands are 3.5 % and 18.2 %, respectively. The radiation patterns are similar to those of the broadband U-slot patch antenna.

To achieve triple-band operation, two slots are necessary. While it is possible to use two U-slots, it was found that the second slot in the form of an H-slot offers better flexibility. Figure 20 shows a U-slot in combination with an H-slot to produce resonances at 1.94, 4.16, and 5.44 GHz. Figure 21 shows the reflection coefficient (S_{11}) response curve. The impedance bandwidths at the three bands are 2.6 % for the lower band, 9.8 % for the middle band, and 10.4 % for the upper band. The radiation patterns at the three bands are shown in Fig. 22. The co-polarization patterns are stable. The cross-polarization levels are 20 dB below the co-polarization at the lower band. At the center frequency of the middle band (4.2 GHz)

DOI 10.1007/978-981-4560-75-7_29-1

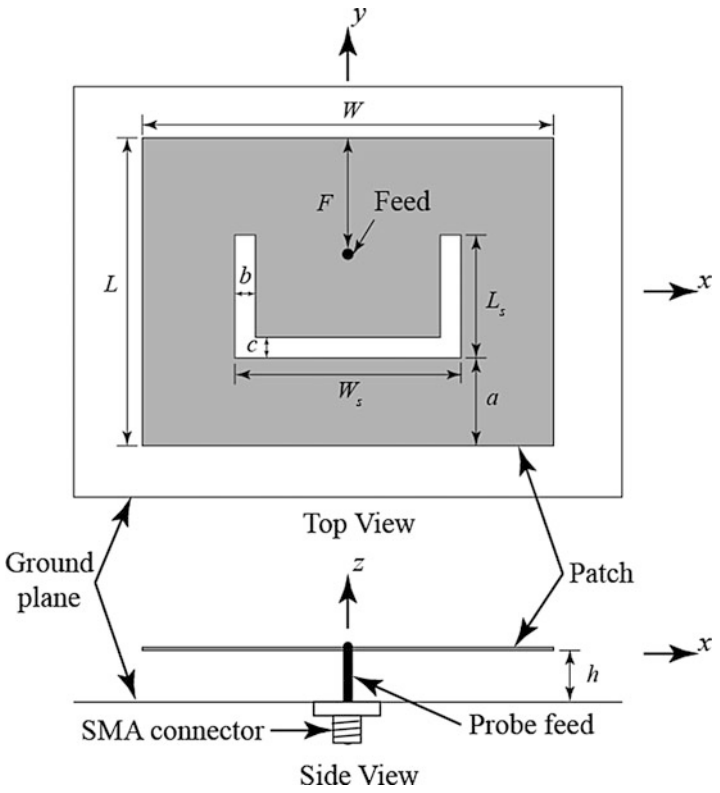


Fig. 18 Geometry of dual-band U-slot patch antenna

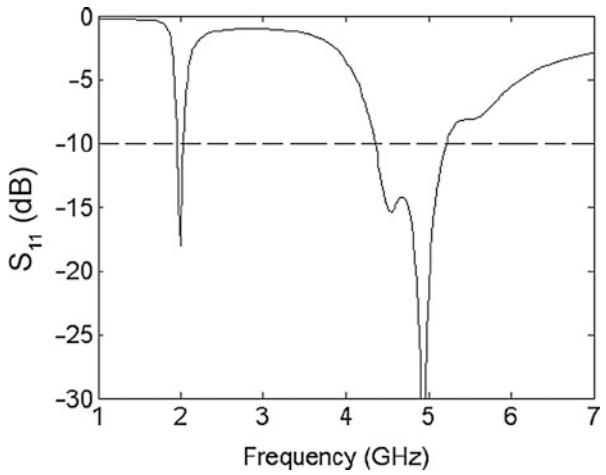


Fig. 19 Simulated reflection coefficient (S_{11}) of the dual-band U-slot patch antenna with dimensions W = 64 mm, L = 54 mm, F = 28.4 mm, W_s = 34 mm, L_s = 22 mm, a = 8.4 mm, b = 11 mm, and h = 6 mm

and high band (5.4 GHz), it becomes very high at oblique angles, as the antenna is electrically thick ($>0.08\ \lambda_o$) and radiation from the vertical coaxial feed becomes significant.

Rather than cutting a second slot in the patch, a triple-band antenna can also be realized by placing an open circuit stub on the radiating edge of a dual-band U-slot rectangular patch antenna. Quadruple-band response results if two stubs are used. An example is shown in Fig. 23a (ground plane not shown). The substrate is glass epoxy ($\varepsilon_r = 4.3$, h = 0.159 cm, and $\tan\delta = 0.02$). The measured VSWR of the triple-

DOI 10.1007/978-981-4560-75-7_29-1

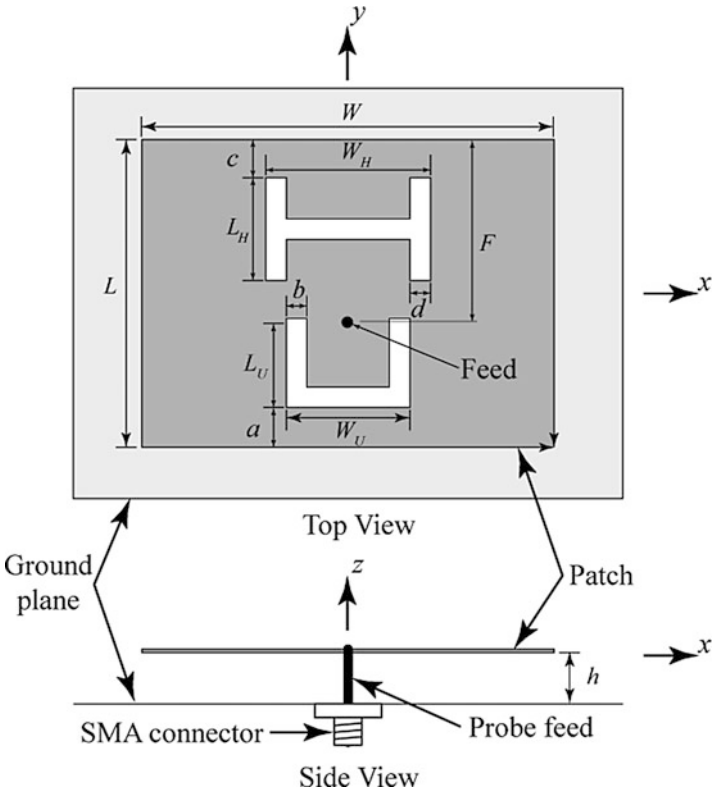


Fig. 20 Geometry of tri-band U-slot patch antenna

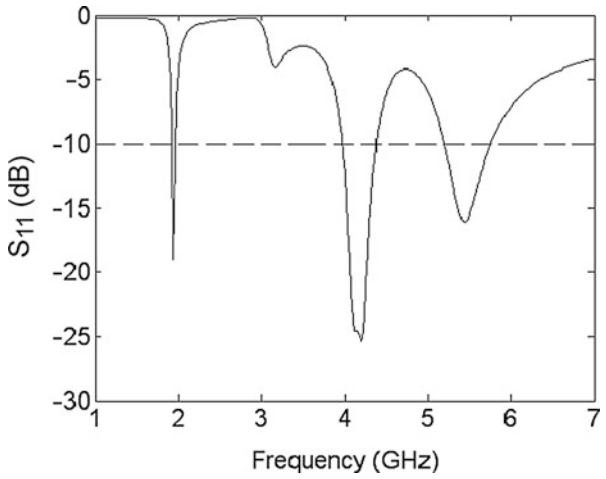


Fig. 21 Simulated reflection coefficient (S_{11}) of the tri-band U-slot patch antenna of Fig. 20 with dimensions W = 64 mm, L = 54 mm, F = 28.4 mm, W_U = 34 mm, L_U = 22 mm, W_H = 28 mm, L_H = 13.5 mm, a = 8.4 mm, b = 11 mm, c = 1.5 mm, d = 2 mm, and h = 6 mm

band antenna is shown in Fig. 24. The measured resonant frequencies agree well with the simulated values of 770, 952, and 1,100 MHz and corresponding bandwidths 16, 18, and 22 MHz. Finally, quadruple-band response can be obtained with a stub-loaded rectangular patch with two unequal-length half U-slots cut into it, as illustrated in Fig. 23b.

DOI 10.1007/978-981-4560-75-7_29-1

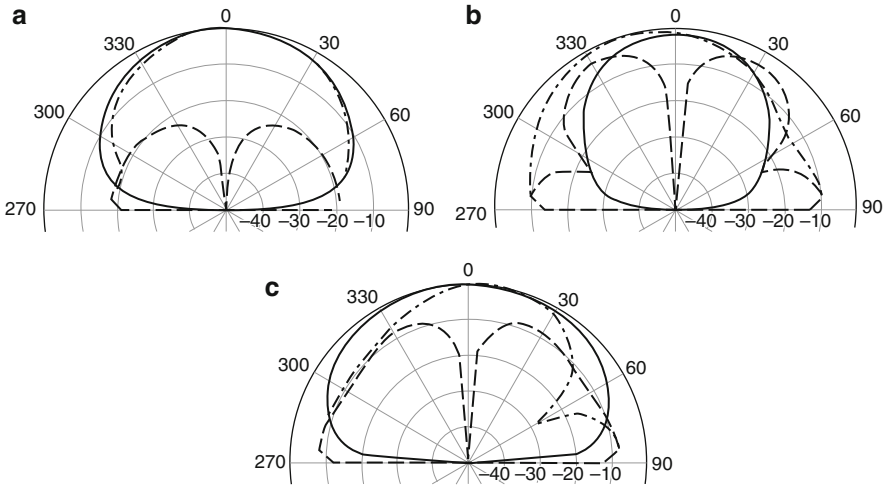


Fig. 22 Simulated radiation pattern of the tri-band U-slot antenna at (**a**) 1.94 GHz, (**b**) 4.16 GHz, and (**c**) 5.44 GHz ((—— H-Co, — — H-x, — · - E-Co)) (From Lee et al. (2010). © 2010 IEEE, Reprinted with permission)

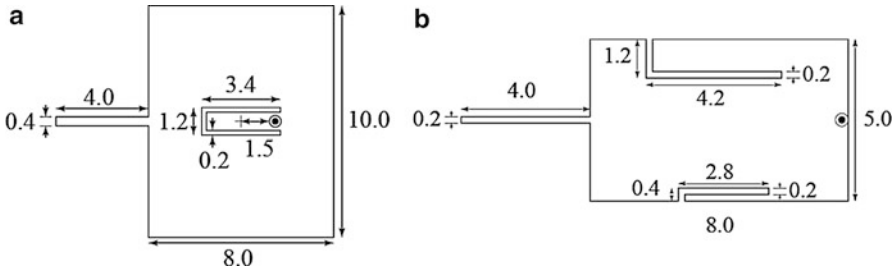


Fig. 23 Geometries of the stub-loaded U-slot patch antenna. (**a**) Triple band; (**b**) quadruple band (From Deshmukh and Ray (2010). © 2010 IEEE, Reprinted with permission)

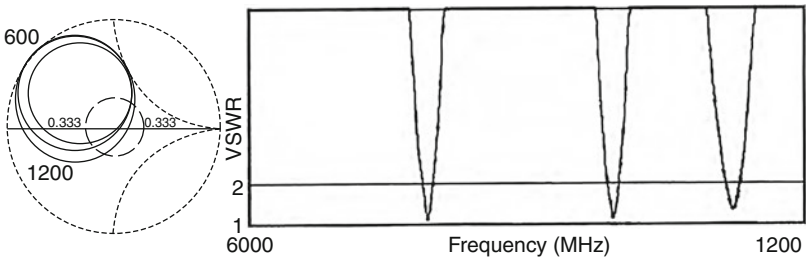


Fig. 24 VSWR response of the triple-band antenna in Fig. 23a (From Deshmukh and Ray (2010). © 2010 IEEE, Reprinted with permission)

Small Frequency Ratios

A simple approach to designing dual- and triple-band patch antenna with frequency ratios less than about 1.5 was introduced in Lee et al. (2008, 2011). The method starts with a broadband antenna. When a suitably located U-slot of the appropriate dimensions is cut in the patch, it introduces a band notch at a frequency within the original broadband. The broadband antenna is turned into a dual-band antenna. When a second U-slot is cut in the patch, two band notches can be introduced and a triple-band antenna results.

DOI 10.1007/978-981-4560-75-7_29-1

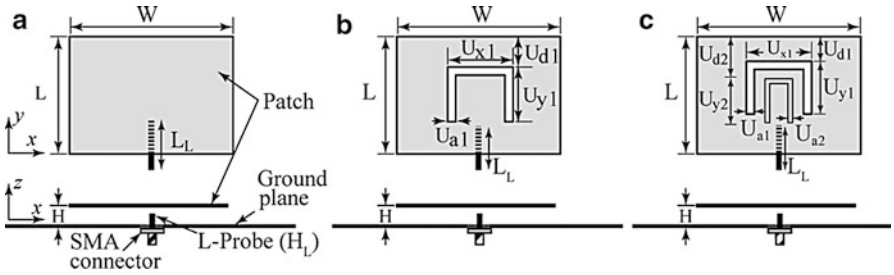


Fig. 25 Geometry of three antennas: (**a**) broadband, (**b**) dual band, and (**c**) triple band

Table 9 Dimensions of the proposed antennas (unit: mm) (From Lee et al. (2008). © 2008 IEEE, Reprinted with permission)

Antenna	W	L	H	H_L	L_L	U_{a1}	U_{d1}	U_{x1}	U_{y1}	U_{a2}	U_{d2}	U_{x2}	U_{y2}
a	22	18	5	3.5	8.5								
b	22	18	5	3.5	8.5	0.8	2	7.5	10.8				
c	22	18	5	3.5	8.5	0.8	2	7.5	10.8	0.8	3.5	4.5	10.8

The original broadband antennas can be an L-probe-fed patch, an M-probe-fed patch, coax-fed stacked patches, aperture-coupled stacked patches (Lee et al. 2011), or a U-slot patch (Mok et al. 2013). In what follows, the use of an L-probe-fed patch is presented in detail.

Figure 25a shows the wideband L-probe-fed rectangular patch. When a U-slot is cut in the patch (Fig. 25b), a notch is introduced within the matching band and a dual-band antenna results. With two U-slots (Fig. 25c), two notches are introduced and a triple-band antenna results.

In Lee et al. (2008), these antennas with the dimensions shown in Table 9 were fabricated. Air substrate was used. Simulation results, using IE3D software, were obtained, as well as measured results. The simulated and measured return losses are shown in Fig. 26, and the impedance bandwidths ($S_{11} < -10$ dB) are summarized in Table 10. Reasonable agreement was obtained between the simulated and measured results. The discrepancies between the results were due to the fabrication tolerance of the fabricated antenna prototypes. Figures 27, 28, and 29 show the measured radiation patterns for the three antennas. The patterns are broadside and the polarization is linear.

For antenna b, the U-slot introduced a notch at around 5.5 GHz. The relationship between the band notch frequency and the total U-slot length is shown in Table 11. The total length of the slot is approximately half wavelength for the two proposed U-slot(s) antennas. The resulting dual-band antenna has a frequency ratio $f_2/f_1 = 1.30$, where f_2 and f_1 are the center frequencies of the upper and lower bands, respectively. For antenna c, the two U-slots introduced a notch at around 5.5 GHz and another notch at around 6.2 GHz. The resulting triple-band antenna has frequency ratios $f_3/f_1 = 1.34$ and $f_2/f_1 = 1.16$, where f_3, f_2, and f_1 are the center frequencies of the upper, middle, and lower frequencies. It should be noted that the frequency ratios are determined by the bandwidth of antenna a, which was 30 % for this case. If one starts with an antenna with broader bandwidth, e.g., 40 %, the frequency ratios of the dual- and triple-band antenna will be larger.

A patch antenna with quadruple-band characteristics can be realized by cutting three half-wavelength long U-slots on the patch. Another method is to use stacked patches in multiple layers (Anguera et al. 2003).

DOI 10.1007/978-981-4560-75-7_29-1

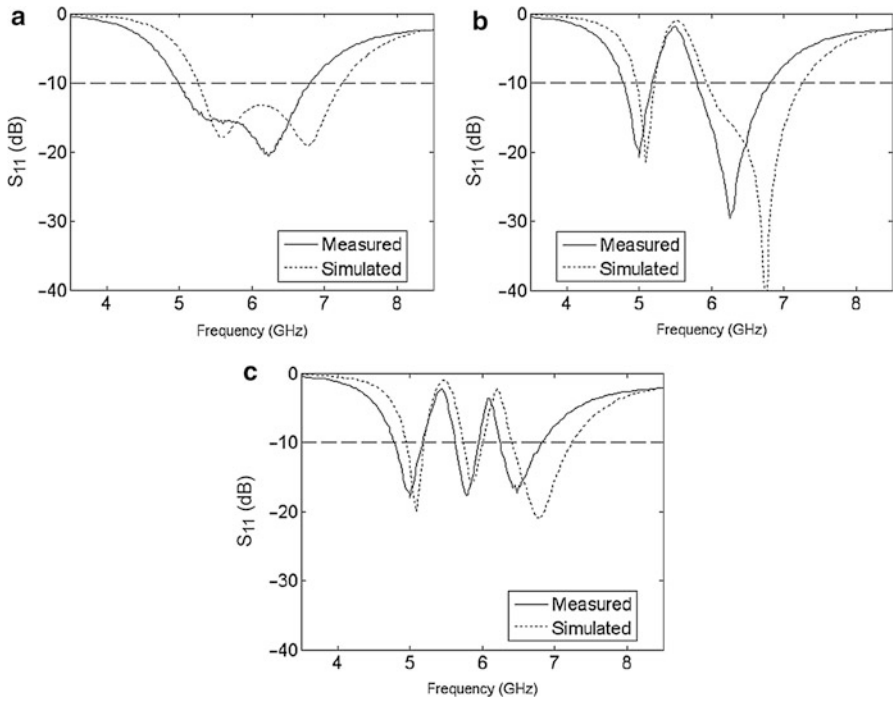


Fig. 26 Simulated and measured reflection coefficient (S_{11}) of the three antennas: (**a**) broadband, (**b**) dual band, and (**c**) triple band

Table 10 Simulated and measured impedance bandwidths (From Lee et al. (2008) © 2008 IEEE, Reprinted with permission)

Antenna	Simulation/GHz	Measurement/GHz
a	5.26–7.25 (31.8 %)	5.00–6.80 (30.5 %)
b	4.97–5.22 (4.9 %), 5.94–7.26 (20 %)	4.80–5.18 (7.6 %), 5.80–6.80(15.9 %)
c	4.95–5.20 (4.9 %), 5.74–6.00 (4.4 %), 6.41–7.24 (12.2 %)	4.80–5.18 (7.6 %), 5.63–5.95 (5.5 %), 6.25–6.83 (8.9 %)

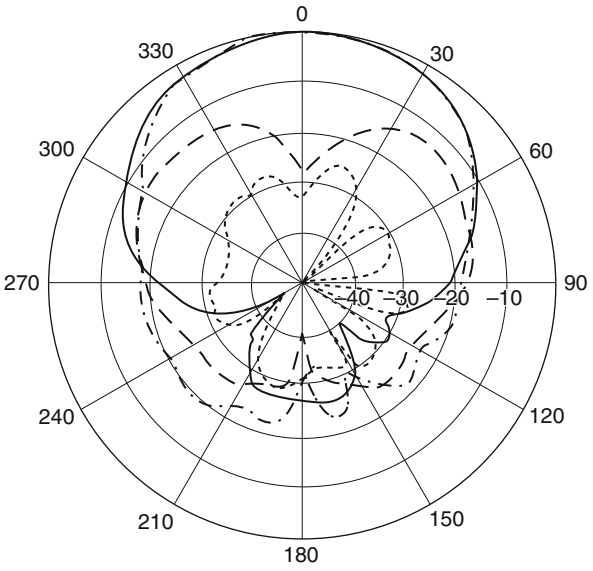


Fig. 27 Measured radiation patterns of the broadband antenna of Table 9a at 5.9 GHz ((——— H-Co, — — H-x, — · — E-Co, ····· E-x)) (From Lee et al. (2008). © 2008 IEEE, Reprinted with permission)

DOI 10.1007/978-981-4560-75-7_29-1

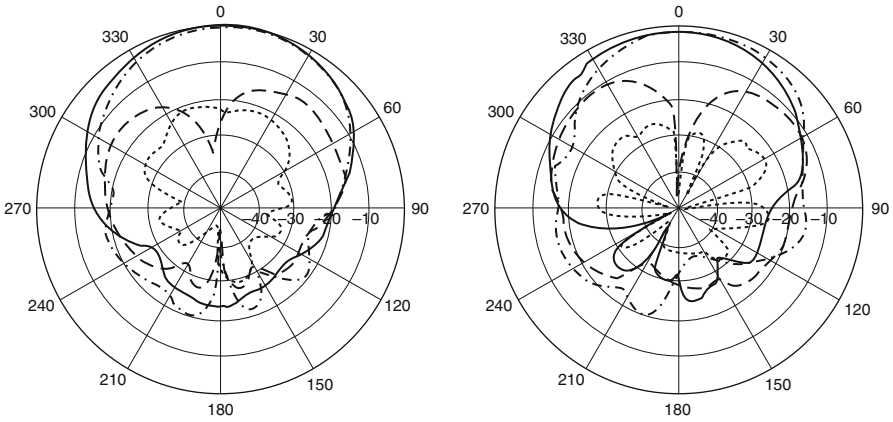


Fig. 28 Measured radiation patterns of the dual-band antenna of Table 9b: (**a**) 5.0 GHz and (**b**) 6.3 GHz ((——— H-Co, — — H-x, — · - E-Co, ····· E-x)) (From Lee et al. (2008). © 2008 IEEE, Reprinted with permission)

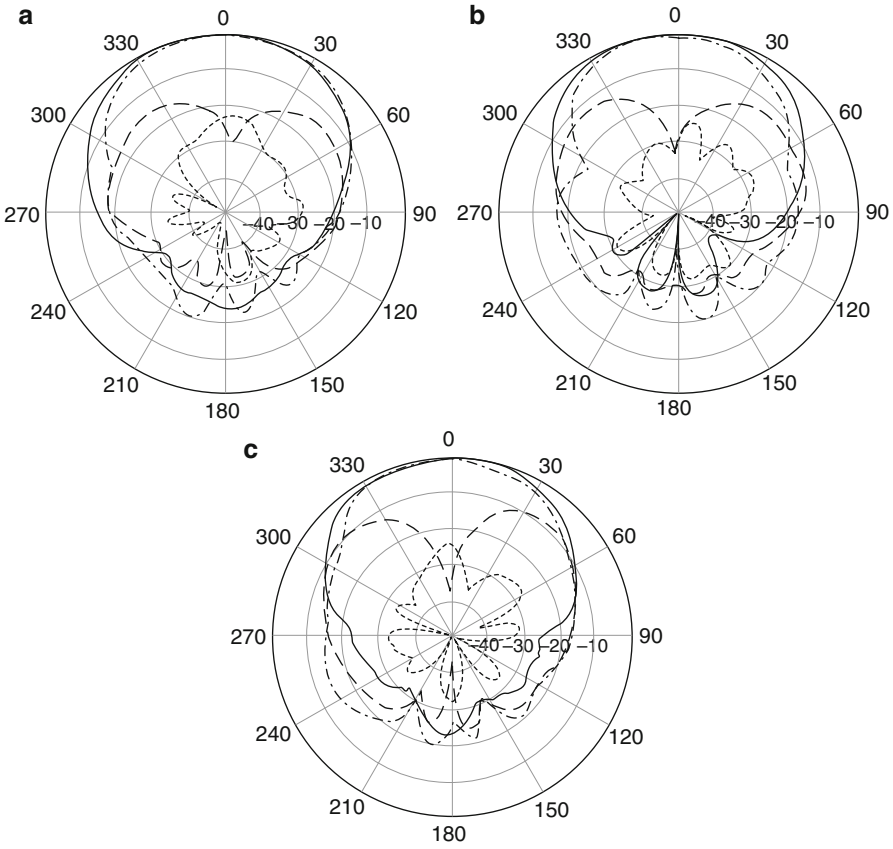


Fig. 29 Measured radiation patterns of the triple-band antenna of Table 9c: (**a**) 5.0 GHz, (**b**) 5.8 GHz, and (**c**) 6.55 GHz ((——— H-Co, — — H-x, — · - E-Co, ····· E-x)) (From Lee et al. (2008). © 2008 IEEE, Reprinted with permission)

Table 11 Relationship between the band notch frequency and the total U-slot length (From Lee et al. (2010). © 2010 IEEE, Reprinted with permission)

Antenna	Freq. of band notch		Total U-slot length	
b	5.5 GHz		27.5 mm (0.504 λ_o)	
c	5.5 GHz	6.2 GHz	27.5 mm (0.504 λ_o)	24.5 mm (0.506 λ_o)

DOI 10.1007/978-981-4560-75-7_29-1

Methods of Reducing the Patch Size

General Remarks

In many applications, it is desirable for the dimensions of the patch to be a small fraction of the free-space wavelength. The resonant length of the microstrip patch antenna is approximately λ/2, where λ is the wavelength in the dielectric substrate. It follows that the size of the patch can be reduced by using a substrate material with high dielectric constant. However, the resulting patch antenna will have narrow impedance bandwidth. This motivated the search for other size reduction methods.

By placing a shorting wall along the null in the electric field across the center of the patch, the resonant length can be reduced by a factor of two (Pinhas and Shtrikman 1988; Chair et al. 1999; Lee et al. 2000). The area occupied by the patch will be reduced by a factor of four, if the aspect ratio is kept the same. Another technique to reduce the resonant length is to add a shorting pin in close proximity to the feed (Waterhouse et al. 1998). The shorting pin is capacitively coupled to the resonant circuit of the patch, effectively increasing the permittivity of the substrate. It has been shown that a suitably placed shorting pin can reduce the resonant length of a circular patch by a factor of three and the area of the patch by a factor of nine. Broadbanding techniques such as stacked patches, U-slot patch, and L-probe feed can be applied to obtain small-size wideband patch antennas (Shackelford et al. 2003). All these methods result in radiation patterns with high cross polarization. This may not be a disadvantage in indoor mobile communication applications. A low cross-polarization design is that of the folded patch, which, however, is thicker and more difficult to fabricate (Luk et al. 1998b).

In what follows, the use of shorting wall, shorting pin, and the U-slot patch with shorting pin is discussed in detail.

The Use of Shorting Wall: Quarter-Wave Patch

Chair et al. (1999) presented experimental results of the quarter-wave patch shown in Fig. 30. The substrate between the patch and the ground plane is foam, with thickness h and relative permittivity 1.08. The sides of the patch are a = b = 3.06 cm long, with one side shorted. The patch is fed by a coaxial probe, with the feed point at x = 0 and y = d, where d is the distance between the feed point and the open edge and is adjusted for best match.

Resonant frequency, pattern, and bandwidth (VSWR < 2) measurements were performed for several thicknesses h, from 2 to 7 mm. Resonant frequencies ranged from 2.19 to 2.46 GHz, while the bandwidths ranged from 3.59 % for h = 2 mm (0.017 λ_o) to 17.66 % for h = 7 mm (0.058 λ_o). The corresponding bandwidths for a half-wave regular patch on the same thickness and the same width but double in length were 1.57 % and 5.55 %, respectively. These results were confirmed qualitatively by simulation software. That the quarter-wave shorted patch has a wider bandwidth than the half-wave patch is due to its smaller volume and therefore less stored energy, leading to a smaller Q and larger bandwidth. This conclusion was found to hold only for air/foam substrate and not for material substrate which supports surface waves.

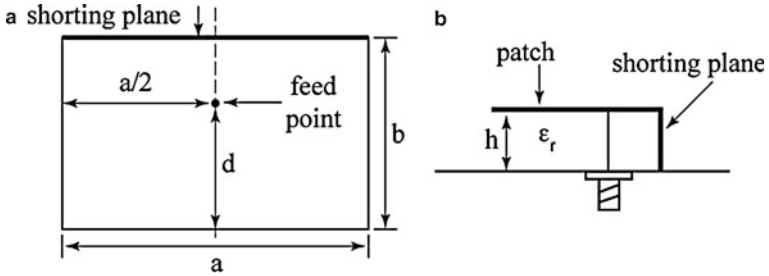


Fig. 30 Geometry of the shorted patch. (**a**) Top view; (**b**) side view

DOI 10.1007/978-981-4560-75-7_29-1

Table 12 Measured gains of the shorted square patch of Fig. 31

h (mm)	f_o (GHz)	Gain (dBi) at broadside	Maximum gain and direction
3	2.06	2.5	2.5 at 0°
5	2.17	2.5	3.5 at 30°
7	2.46	0.2	2.2 at 45°

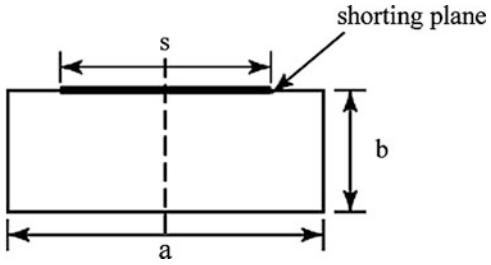


Fig. 31 Geometry of partially shorted patch

In Lee et al. (2000), it was found that shorted patches on substrates with $\varepsilon_r = 2.32$ and $\varepsilon_r = 4.0$ had smaller bandwidths than their half-wave counterparts. This was attributed to the fact that the surface-wave loss was considerably more in a half-wave patch than a quarter-wave patch, leading to a larger bandwidth for the former.

The measured patterns by Chair et al. (1999) showed large cross polarization in the E-plane. They also showed that, depending on the thickness, the maximum radiation could occur off broadside. The measured gains for h = 3, 5, and 7 mm at the resonant frequencies in the broadside direction as well as in the maximum direction are summarized in Table 12. It is seen that typical values of the maximum gain were in the range 2–3.5 dBi. This is about half that of the regular half-wave patch.

Maximum gain direction is measured from the perpendicular direction in the E-plane (perpendicular to the shorting plane).

Partially Shorted Patch and Planar Inverted-F Antenna

Figure 31 shows the geometry in which the shorting wall, instead of extending fully across the width of the patch a, has a width s, where $s \leq a$.

It was shown in Hirasawa and Haneishi (1992) that the use of a partially shorted wall had the effect of reducing the resonant frequency of the antenna. Lee et al. (2000) showed that this was accomplished at the expense of bandwidth. Their calculated results using IE3D simulation software for an antenna with a = 3.8 cm, b = 2.5 cm, h = 3.2 cm, and $\varepsilon_r = 1.0$ showed that as s/a decreased from 1.0 to 0.1, the resonant frequency decreased from 2.69 to 1.61 GHz, representing a 60 % reduction in frequency or size. However, the bandwidth was reduced from 7.4 % for s/a = 1.0 to 3.7 % for s/a = 0.1.

The partially shorted patch in the form shown in Fig. 32 is known as the planar inverted-F antenna (PIFA), because the side view looks like an inverted F. The width of the shorting wall w is approximately 0.2 L_1, while the dimensions of L_1 and L_2 are on the order of 1/8 λ_o.

The Use of Shorting Pin

Another technique for reducing the patch size, very similar to the inverted-F method, is to use a shorting pin (Waterhouse et al. 1998). This is illustrated in Fig. 33.

The shorting pin causes the fields underneath the patch to bounce back and forth. The field starts to radiate once the bouncing distance reaches half wavelength. As a result of the multiple bounces, the

DOI 10.1007/978-981-4560-75-7_29-1

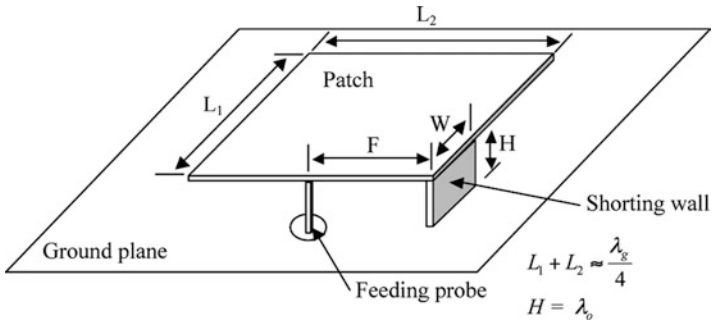


Fig. 32 Size reduction by using an inverted-F patch

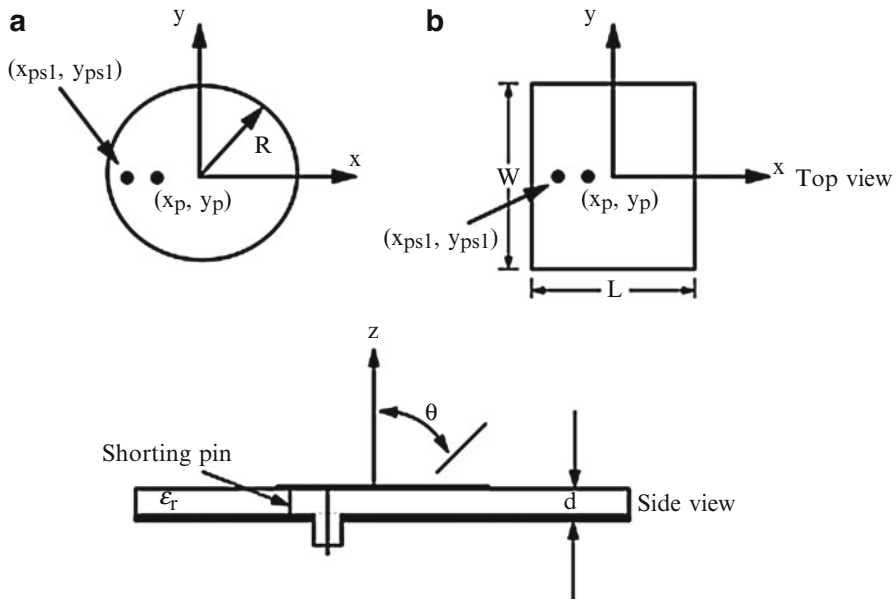


Fig. 33 Circular (**a**) and rectangular (**b**) patches with shorting pin

physical size of the patch is reduced. Since the bounces are non-unidirectional, the fields can radiate out from almost all edges of the patch, resulting in high cross polarization. However, for certain applications such as cellular phone communication in a multipath environment, high cross-polarized fields are not a concern.

If the shorting pin is close to the feed, the resonant circuit of the patch is capacitively coupled to the pin. This is equivalent to increasing the permittivity of the substrate, which further contributes to reduction in frequency or size of the patch (measured in wavelength).

The solid curve of Fig. 34 shows the simulation results for return loss for the circular patch shown in Fig. 34a with $x_p = 6.2$ mm, $x_{ps} = 8.3$ mm, and $y_p = y_{ps} = 0$, on a foam substrate with $\varepsilon_r = 1.07$ and $t = 10$ mm. IE3D simulation software was used. The radius of the patch is reduced by a factor of 3 and the area by a factor of 9 when compared to the case of no shorting pin. The thickness of the foam substrate is 0.06 λ_o, and the impedance bandwidth is about 6.3 %. The simulated radiation patterns at 1.9 GHz are shown in Fig. 35. As noted earlier, the cross polarization of this type of antenna is very high. The simulation results are consistent with the experiments reported in Waterhouse et al. (1998).

The bandwidth can be slightly improved by using multiple shorting pins. A circular patch with two and three shorting pins is shown in Fig. 36. The reflection coefficient (S_{11}) for these cases is shown in the

DOI 10.1007/978-981-4560-75-7_29-1

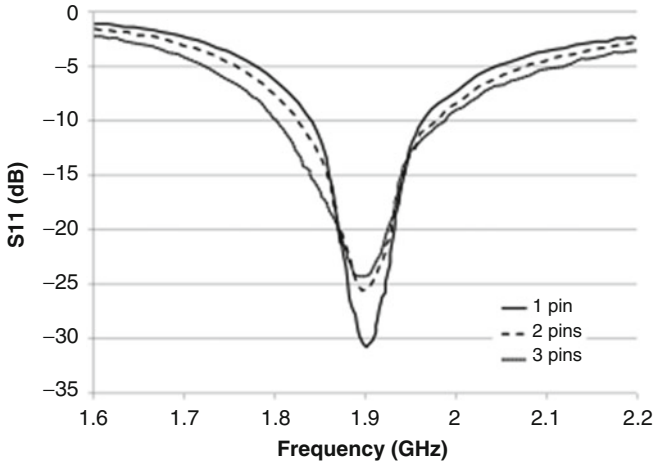


Fig. 34 Simulated return loss of the miniature patch antenna with different numbers of shorting pins

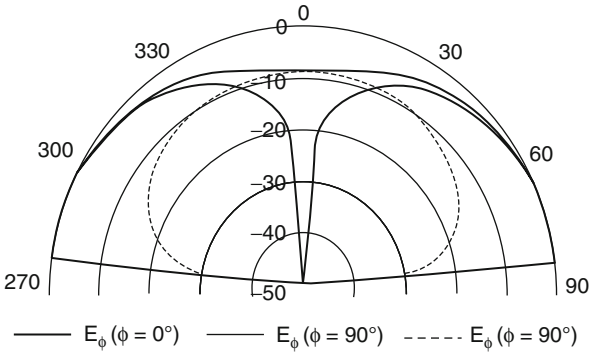


Fig. 35 Simulated radiation pattern of the miniature patch antenna with one shorting pin at 1.9 GHz (10 dB/div)

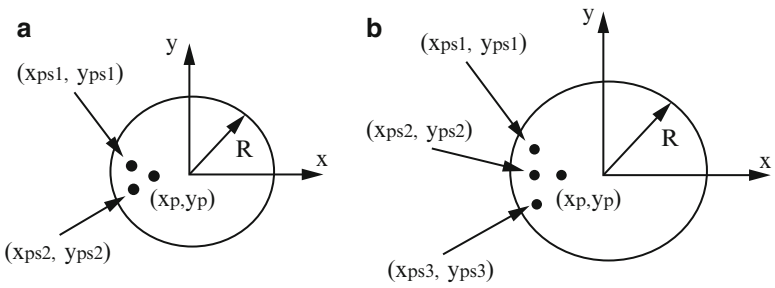


Fig. 36 Circular patch with (**a**) 2 and (**b**) 3 shorting pins (units in mm and not to scale) (**a**) R = 13.2, $x_p = 4.95$, $x_{ps1} = x_{ps2} = 11.08$, $y_{ps1} = 1.95$, $y_{ps2} = -1.95$; (**b**)R = 15.4, $x_p = 2.35$, $x_{ps1} = 13.3$, $x_{ps2} = x_{ps3} = 14.1$, $y_{ps1} = 0$, $y_{ps2} = 5.13$, $y_{ps3} = -5.13$

broken curves of Fig. 34. The impedance bandwidths are 7.9 % for the patch with two pins and 10 % for the patch with three pins.

Wideband U-Slot Patch Antenna with Shorting Pin

The shorting pin technique can be applied to the wideband U-slot patch discussed in section "Aperture-Coupled Patches." An example of the U-slot patch with shorting pin is shown in Fig. 37a

DOI 10.1007/978-981-4560-75-7_29-1

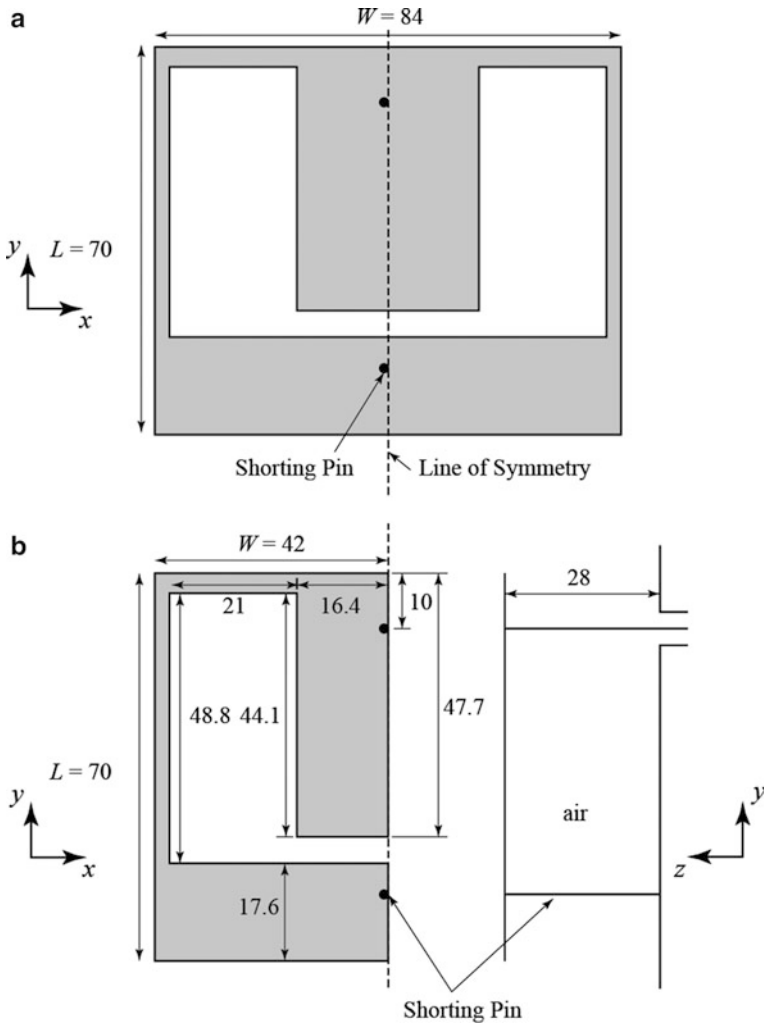


Fig. 37 (**a**) U-slot patch with shorting pin; (**b**) half U-slot patch with shorting pin

(Shackelford et al. 2001). The U-slot patch has a length L = 70 mm (0.2 λ_o) and a width W = 84 mm (0.24 λ_o) where λ_o is the free-space wavelength at the center frequency 0.86 GHz of the U-slot patch. The patch is supported by the shorting pin and the probe wire in air. The radii of the shorting pin and the probe feed are 4.65 and 2 mm, respectively. The shorting pin and the probe feed are located at the non-radiating edge of the U-slot patch, which is placed on a 1 × 1 λ_o ground plane. Figure 38 shows the measured and simulated SWR of this antenna. The matching frequency range is from 0.75 to 0.97 GHz with an impedance bandwidth of 25.6 %. The area of the U-slot patch is only 0.055 λ_o. Figure 39 shows the simulated x-z and y-z plane radiation patterns at 0.85 GHz.

Studies of the U-slot patch with shorting pin on material substrates have been reported in Shackelford et al. (2003). It was also possible to further reduce the size of the patch by removing half of the structure along the plane of symmetry (Chair et al. 2005). Such an antenna is shown in Fig. 38b and its VSWR response is also shown in Fig. 39. The corresponding bandwidth is 20 %.

Discussion on Finite Ground Plane Size Effect

The papers from which the above examples were drawn, as well as most of the papers in the literature on patch size reduction techniques, did not address the ground plane size effects. Based on the recent study for the shorted patch on foam substrate using CST Microwave Studio simulation software, it was found

DOI 10.1007/978-981-4560-75-7_29-1

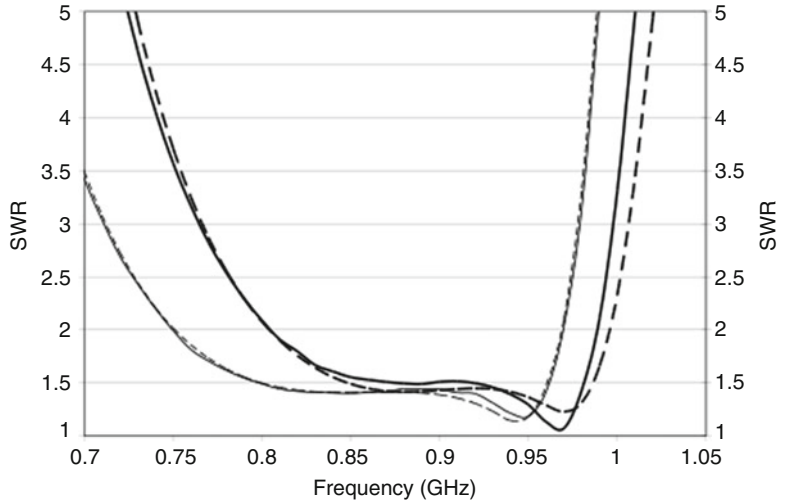


Fig. 38 Measured and simulated SWR of the U-slot patch antennas with shorting pin. ——Half U-slot (measured). ····· Half U-slot (simulated). —— Full U-slot (measured). ----- Full U-slot (simulated) (From Chair et al. (2005). © 2005 IEEE, Reprinted with permission)

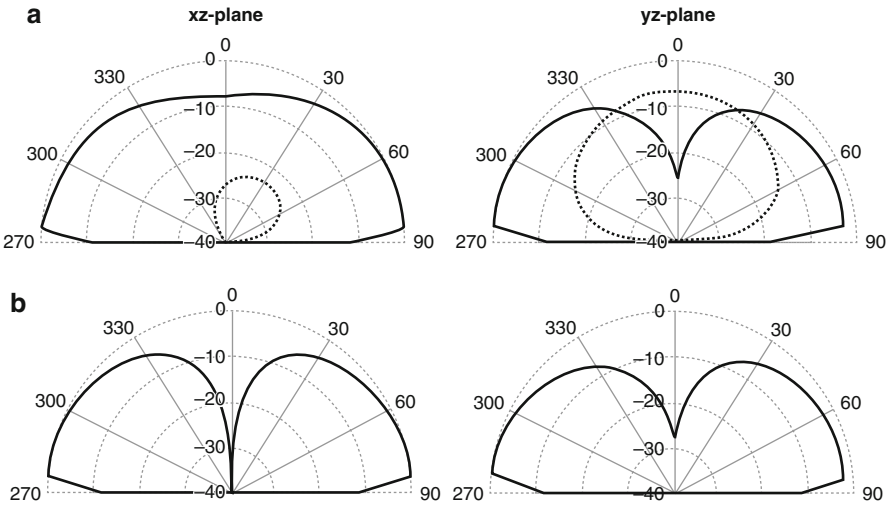


Fig. 39 Simulated radiation patterns of the U-slot patch antenna with shorting pin at 0.9 GHz (10 dB/div). (**a**) Half U-slot. (**b**) Full U-slot E_θ ——— E_ϕ ------

that, when compared to the infinite ground case, the center resonant frequency, bandwidth, and gain will deviate less than 5 % and 10 % and 1.5 dB, respectively, until the ground plane is reduced to smaller than $0.5\times\ 0.5\ \lambda_o$ (Tong et al. 2011).

In general, the ground plane size effects have to be considered with care, as they are dependent on the particular antenna and the locations of the antenna and the feed (Best 2009).

DOI 10.1007/978-981-4560-75-7_29-1

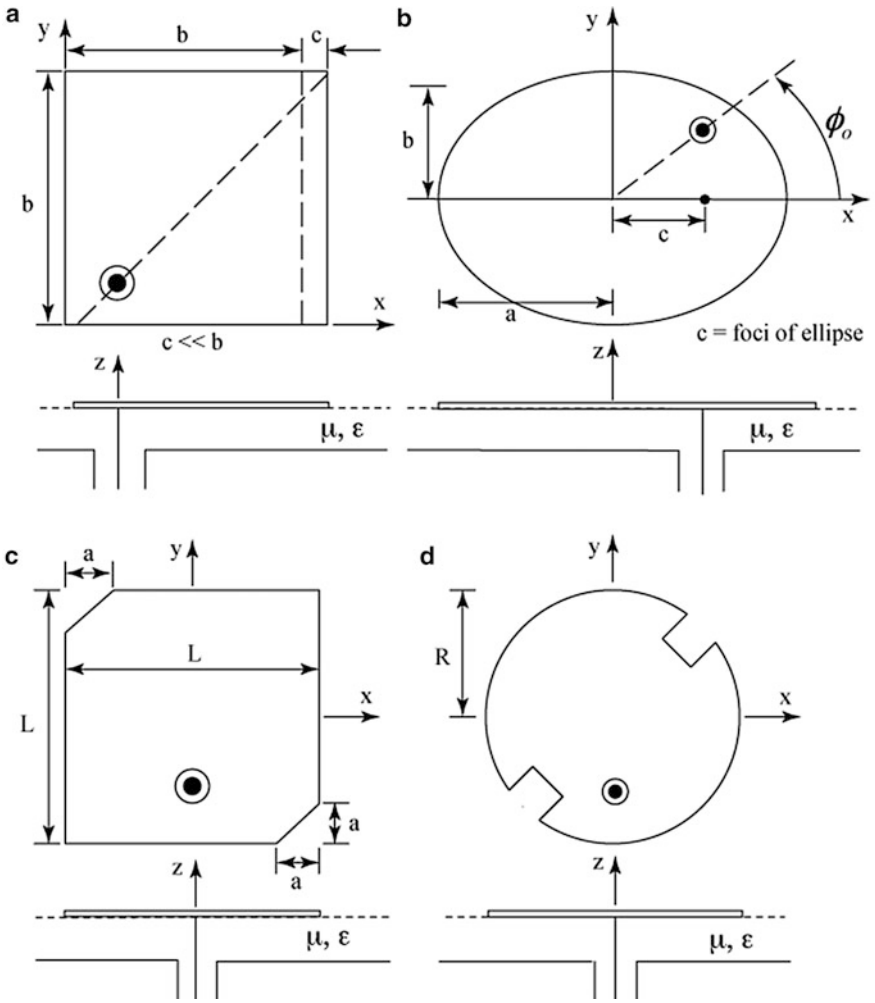


Fig. 40 Single-feed circularly polarized patches. (**a**) Almost square patch. (**b**) Almost circular (elliptical patch). (**c**) Square patch with truncated corners. (**d**) Circular patch with indentations

Designs for Circular Polarization

Basic Principles of Circularly Polarized Patch Antennas

The designs described in the previous sections are for linear polarization. There are many circumstances, such as multipath/fading environments and in communication with space vehicles above the earth's ionosphere, in which it is more reliable to use circular polarization. Circularly polarized patch antennas can be broadly classified into three types – single feed, dual feed, and sequentially rotated feed. Figure 41 shows four single-feed designs.

Figure 40a and b shows an almost square patch and an almost circular (elliptical) patch. Figure 40c and d shows a square patch with truncated corners and a circular patch with indentations. Although coaxial feeds are shown in Fig. 40, the patches can also be fed by a microstrip line or by coupling through an aperture.

At the appropriate frequency, all four configurations will generate two orthogonal linearly polarized waves which are in phase quadrature in the broadside direction. This design is simple, but suffers from very narrow axial ratio bandwidth for thin substrates. By using the U-slot or L-probe for relatively thick substrates or the use of stacked patch techniques, the axial ratio bandwidth can be increased to over 10 %.

The basic principle of the single-feed CP patch antenna is that a perturbation of the dimensions be introduced such that, by feeding the patch at the appropriate location, two modes with orthogonal

DOI 10.1007/978-981-4560-75-7_29-1

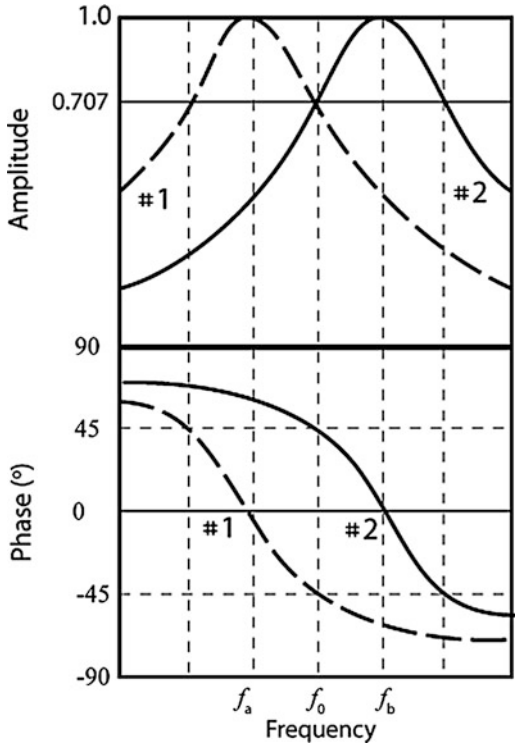


Fig. 41 The amplitudes and phases of two orthogonal modes

polarizations are generated with resonant frequencies which are slightly different. In the almost square patch and the almost circular patch, the perturbation is in the form of one of the dimensions being slightly different from the other. In the square patch with truncated corners and the circular patch with indentations, the perturbation is in the form of two slightly different diagonal lengths.

In the broadside direction, the amplitudes and phases of the radiated electric fields of the two orthogonal modes generated are illustrated in Fig. 41. The resonant frequencies of these modes are f_a and f_b, respectively.

At the frequency f_o, approximately midway between f_a and f_b, the amplitudes of the two modes are equal and the phases differ by 90°, resulting in circular polarization. This condition for perfect CP occurs only at f_o and in the broadside direction. As the frequency deviates from f_o and/or the direction is off broadside, elliptical polarization results.

The antenna is only useful for the range of frequencies within which both the return loss is less than 10 dB (RLBW) and the axial ratio is less than 3 dB (ARBW). This overlapping region will be referred to as the CP bandwidth.

Broadbanding of Single-Feed Circularly Polarized Patch Antennas

As with the case of the RLBW, the ARBW is found to increase as the substrate thickness increases. However, this thickness is very limited if the ARBW is to overlap with the RLBW. For the single-feed patch antennas of Fig. 40, achievable CP bandwidth is less than 1 %. For thicker substrates, the use of the U-slot patch, the E-patch, stacked patches, as well as the L-probe feeding techniques is found to be effective not only in increasing the RLBW and the ARBW but also to enable them to overlap. The truncated square patch with U-slot, the CP modified U-slot patch, and the CP modified E-patches will now be discussed in detail.

DOI 10.1007/978-981-4560-75-7_29-1

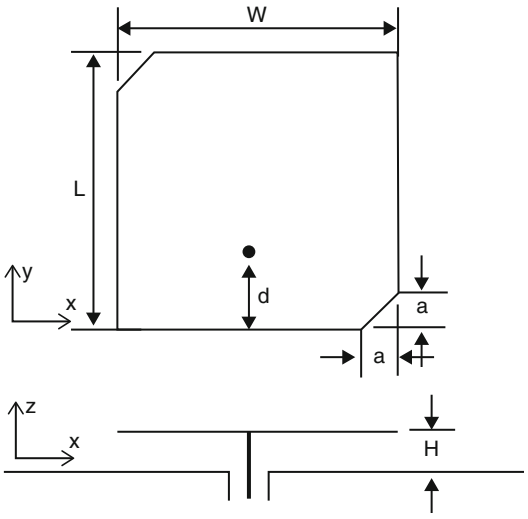


Fig. 42 Geometry of single probe feed square patch antenna with truncated corners

Table 13 Parameters of antenna as shown in Fig. 43. Units of a and d are in mm

Case	Substrate thickness (H)		a	d
1	1 mm	0.016 λ_o	3.3	8.2
2	1.5 mm	0.024 λ_o	4.5	7.2
3	2 mm	0.032 λ_o	4.9	5.5
4	3 mm	0.046 λ_o	5.9	5.1
5	4 mm	0.06 λ_o	6.9	4.3

The Truncated Square Patch with U-Slot

A circularly polarized antenna can be designed using a square patch with truncated corners. One of the requirements for the CP antenna is that the axial ratio bandwidth (ARBW) should be within the return loss bandwidth (RLBW) or, equivalently, the impedance bandwidth. This overlapping bandwidth is very narrow for the square patch with truncated corners. Typically, it is about 0.8 %, for an air-substrate thickness of around 0.02 λ_o. Yang et al. (2008) have shown that, while increasing the substrate thickness can increase axial ratio bandwidth and the impedance bandwidth individually, this increase is very limited if the ARBW is to be within the RLBW. On the other hand, by using a U-slot patch with truncated corners and using substrates of thickness around 0.1 λ_o, the axial ratio bandwidth can be as wide as 6.1 % and still stay within the impedance bandwidth. The antenna geometry studied by Yang et al. (2008) is shown in Fig. 42.

The patch has dimensions 28.6 mm (W) $\times$ 28.6 mm (L) and the ground plane is 100 $\times$ 100 mm. It is fed by a coaxial probe and the substrate is air. Using the IE3D simulation software, the RLBW and ARBW are studied for five thicknesses. For each case, the truncation parameter a and the feed position d are adjusted for best performance and are shown in Table 13. The simulation results are shown in Table 14.

It is seen that, as the substrate thickness increases from 1 mm (0.016 λ_o) to 3 mm (0.046 λ_o), the RLBW increases from 4.0 % to 7.4 % while the ARBW increases from 0.82 % to 3.1 %. When the thickness increases further to 4 mm (0.06 λ_o), the RLBW drops to 4.0 % while the ARBW continues to increase to 3.8 %. The important point to note is that, beginning with case 2, the two frequency bands do not overlap. Thus, only in case 1 the ARBW lies within the RLBW, and the useful frequency band is determined by the ARBW, from 4.90 to 4.94 GHz, corresponding to 0.82 %.

DOI 10.1007/978-981-4560-75-7_29-1

Table 14 Simulation results of the different cases as shown in Table 13

			Simulation (GHz)	
Case	Substrate thickness (H)		RLBW	ARBW
1	1 mm	0.016 λ_o	4.86–5.06 (4.0 %)	4.90–4.94 (0.82 %)
2	1.5 mm	0.024 λ_o	4.88–5.08 (4.0 %)	4.79–4.86 (1.45 %)
3	2 mm	0.032 λ_o	4.85–5.11 (5.2 %)	4.68–4.78 (2.1 %)
4	3 mm	0.046 λ_o	4.81–5.18 (7.4 %)	4.52–4.66 (3.1 %)
5	4 mm	0.06 λ_o	4.89–5.09 (4.0 %)	4.40–4.57 (3.8 %)

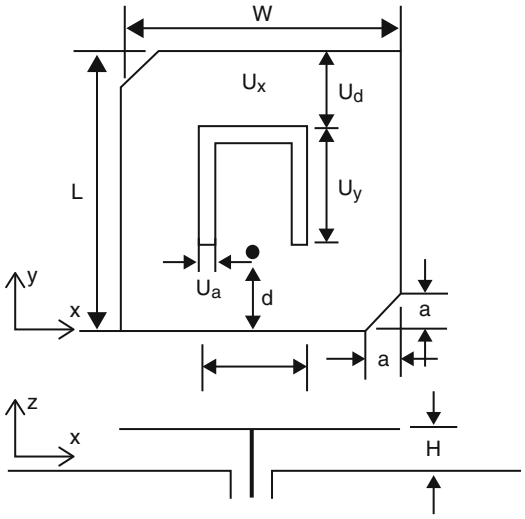


Fig. 43 Geometry of the single-feed U-slot square patch antenna with truncated corners

Table 15 Parameters of the antenna as shown in Fig. 44. Units for a, d, U_a, U_d, U_x, and U_y are in mm

Case	Substrate thickness (H)		a	d	U_a	U_d	U_x	U_y
6	4 mm	0.05 λ_o	5.7	12.6	1	9.8	12	14
7	6 mm	0.08 λ_o	7.7	9.6	1	9.8	12	14
8	7.5 mm	0.1 λ_o	8.2	5.6	1	9.8	11	14

Figure 43 shows the geometry when a U-slot is added to the square patch with truncated corners. Three thicknesses are studied. The dimensions of the antenna are shown in Table 15, and the simulation and measurement results are shown in Table 16.

The following is observed:

1. When the U-slot is added on the patch, the resonant frequency is lowered compared to the case without the U-slot.
2. By adjusting the dimensions of the U-slot, the RL and AR bandwidths can be tuned to overlap.
3. The measured overlapping bandwidth is 5.23 % for H = 6 mm (0.08 λ_o). This is to be compared to 0.82 % achievable for the case of no U-slot. This is achieved, of course, at the expense of increasing the thickness of the antenna.

DOI 10.1007/978-981-4560-75-7_29-1

Table 16 Results of the different cases as shown in Table 15

			Simulation (GHz)		Measurement (GHz)	
Case	Substrate thickness (H)		RLBW	ARBW	RLBW	ARBW
6	4 mm	0.05 λ_o	3.83–4.18 (8.7 %)	3.96–4.05 (2.2 %)	–	–
7	6 mm	0.08 λ_o	3.73–4.2 (11.9 %)	3.96–4.12 (4.0 %)	3.66–4.16 (12.8 %)	3.91–4.12 (5.23 %)
8	7.5 mm	0.1 λ_o	3.84–4.08 (6.1 %)	3.84–4.09 (6.3 %)	3.88–4.08 (5.0 %)	3.82–4.05 (5.84 %)

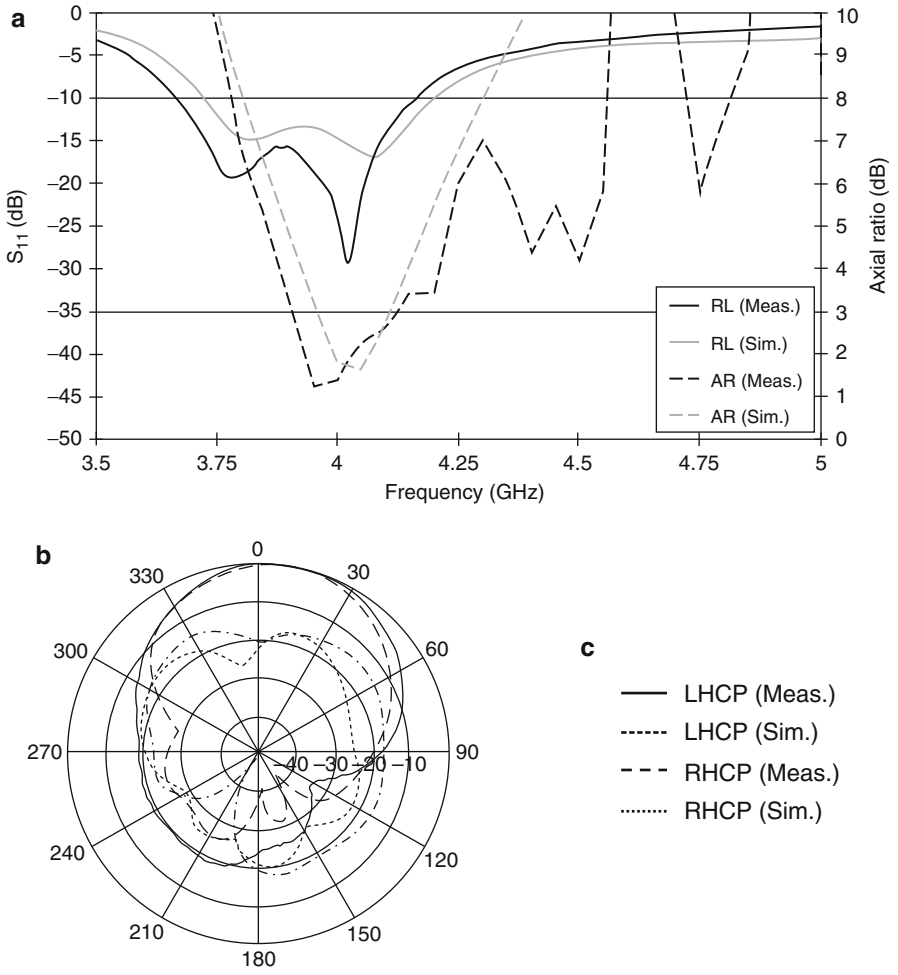


Fig. 44 Simulated and measured performance of case 7, radiation patterns are at 3.95 GHz

The simulation and measurement results for S_{11}, axial ratio, and radiation patterns at one frequency for the cases of H = 6 mm and H = 7.5 mm are shown in Figs. 44 and 45.

The CP Modified U-Slot Patch and the CP Modified E-Patch

Instead of truncating the corners of a square patch, circular polarization can also be generated by the use of an asymmetric U-slot (unequal arms) etched on a square patch (Tong and Wong 2007) or a modified E-patch in which the two parallel slots are unequal (Khidre et al. 2010). The geometries of these are shown in Figs. 46 and 47, respectively. By using relatively thick substrates (0.08 λ_o), the CP bandwidths achievable are comparable to the truncated U-slot square patch.

DOI 10.1007/978-981-4560-75-7_29-1

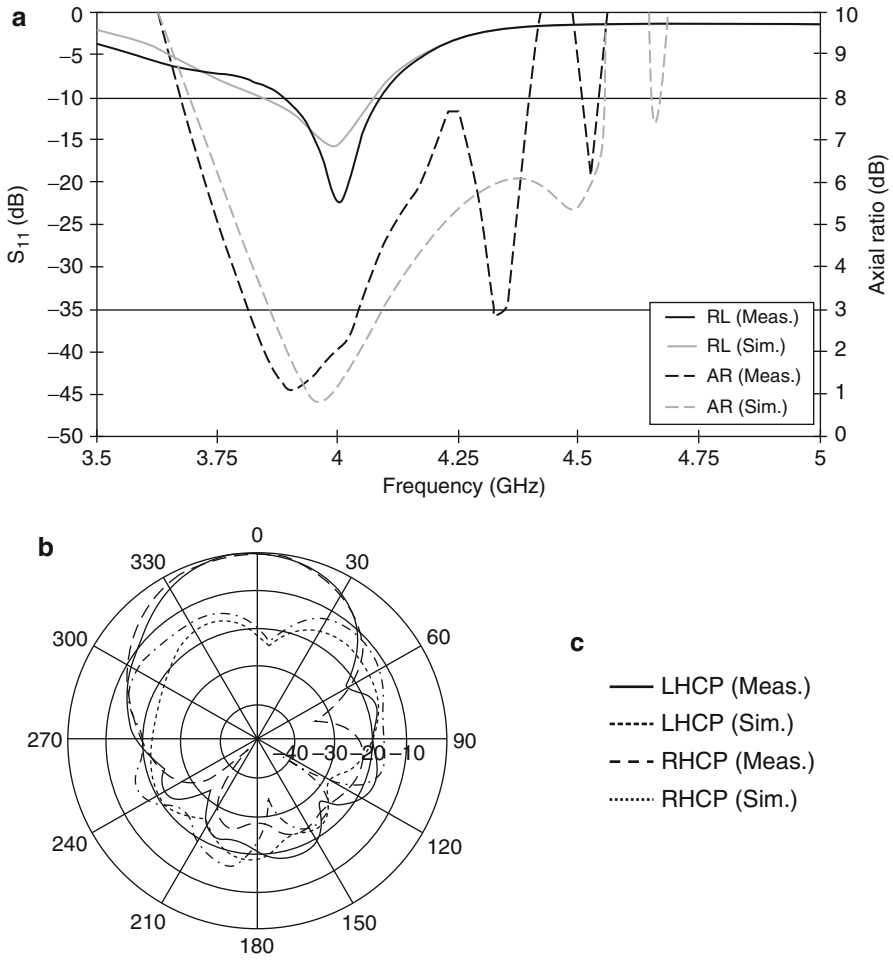


Fig. 45 Simulated and measured performance of case 8, radiation patterns are at 4 GHz

In Khidre et al. (2010), a study was made of the three designs using both air substrate and a substrate with $\varepsilon_r = 2.2$.

The optimized dimensions of the designs to work at IEEE 802.1 lb/g band (2.4–2.5 GHz) are tabulated in Tables 17, 18, and 19. The simulated S_{11} and axial ratios of the three air-substrate antennas are shown in Figs. 48 and 49, respectively. From the results, the modified E-patch exhibits 10.1 % impedance bandwidth (2.35–2.6 GHz), while unequal arms U-slot has 9.25 % (2.37–2.6 GHz) and truncated corner U-slot has 9.65 % (2.27–2.5 GHz). It is clear that all ARBW are aligned with the corresponding RLBW, resulting in CP bandwidths of 6.5 % (2.38–2.54 GHz) for the modified E-patch, 4 % (2.39–2.49 GHz) for the unequal arms U-slot, and 4.5 % (2.39–2.5 GHz) for the truncated corner U-slot.

The simulated S_{11} and axial ratios of the material substrate antennas are shown in Figs. 50 and 51, respectively. From the results, the modified E-patch exhibits 10.6 % impedance bandwidth (2.39–2.66 GHz), while unequal arms U-slot has 6.9 % (2.38–2.65 GHz) and truncated corner U-slot has 13.1 % (2.41–2.75 GHz). Again, all AR bandwidths are aligned with the corresponding S_{11} bandwidth, resulting in CP bandwidths of 3.6 % (2.41–2.5 GHz) for the modified E-patch, 2.8 % (2.36–2.55 GHz) for the unequal arms U-slot, and 3.3 % (2.39–2.5 GHz) for the truncated corner U-slot.

The CP characteristics of the three designs were verified experimentally in Yang et al. (2008), Tong and Wong (2007), and Khidre et al. (2010).

DOI 10.1007/978-981-4560-75-7_29-1

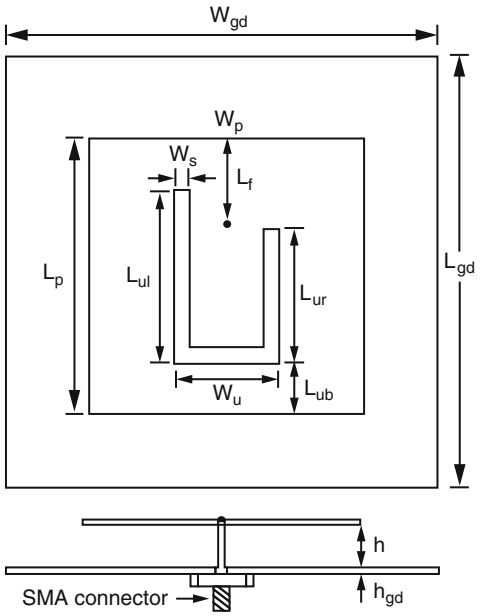


Fig. 46 Geometry of CP unequal arms U-slot patch antenna

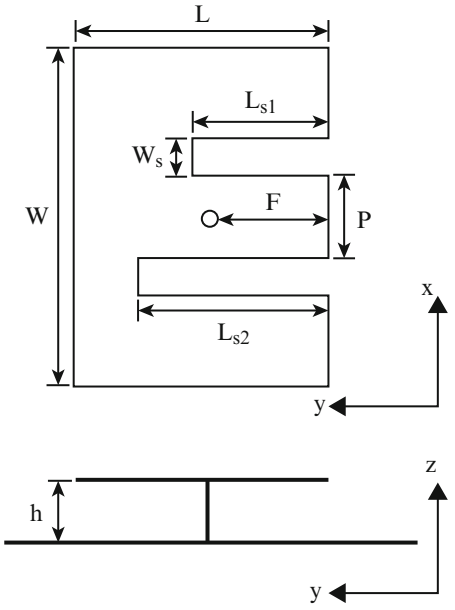


Fig. 47 Geometry of the CP modified E-patch antenna

Table 17 Dimensions of CP modified E-shaped patch in mm (From Khidre et al. (2010). © 2010 IEEE, Reprinted with permission)

ε_r	H	W	L	Ws	Ls1	Ls2	P	F
1	10	77	47.5	7	19	44.5	14	17
2.2	6.7	63	33.5	4	27	6	20	10

DOI 10.1007/978-981-4560-75-7_29-1

Table 18 Dimensions of CP modified U-slot patch in mm (From Khidre et al. (2010). © 2010 IEEE, Reprinted with permission)

ε_r	H	L_p	W_p	L_{ul}	L_{ur}	L_{ub}	W_u	W_s	L_f
1	10	43.7	43.7	27.3	19.8	10.3	16.9	2.3	12.5
2.2	6.7	32.7	32.7	20	13.5	8.9	11.4	1.5	12

Table 19 Dimensions of CP modified U-slot patch in mm (From Khidre et al. (2010). © 2010 IEEE, Reprinted with permission)

ε_r	H	L	W	a	d	U_a	U_d	U_x	U_y
1	10	48.2	48.2	12.2	10	2.1	19.8	23	19
2.2	6.7	36	36	7.7	6.5	1.2	19.6	15.5	11.5

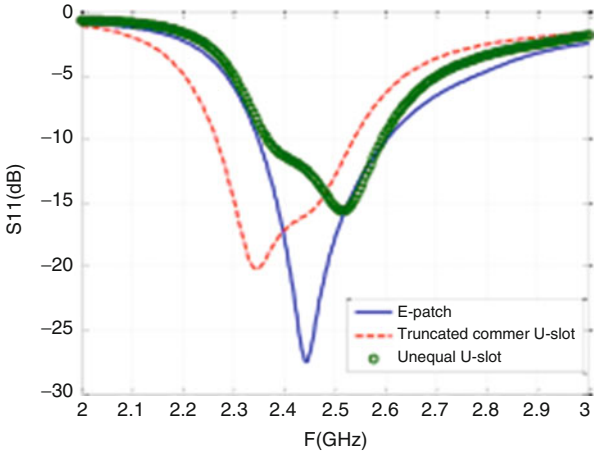


Fig. 48 S_{11} for E-shaped, unequal arms U-slot and truncated corner U-slot patch antennas. Air substrate (From Khidre et al. (2010). © 2011 IEEE, Reprinted with permission)

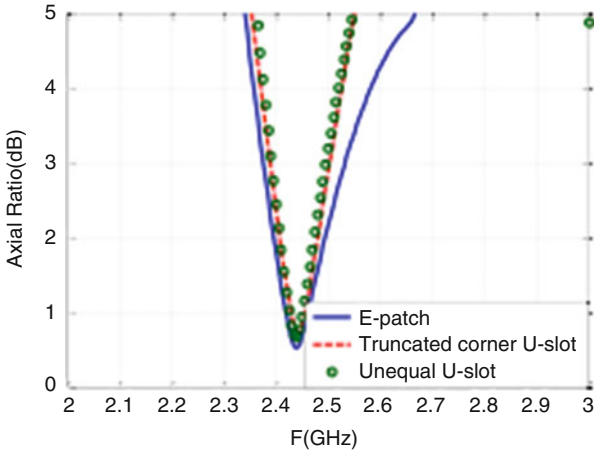


Fig. 49 Axial ratio for E-shaped, unequal arms U-slot, and truncated corner U-slot patch antennas. Air substrate (From Khidre et al. (2010). © 2011 IEEE, Reprinted with permission)

DOI 10.1007/978-981-4560-75-7_29-1

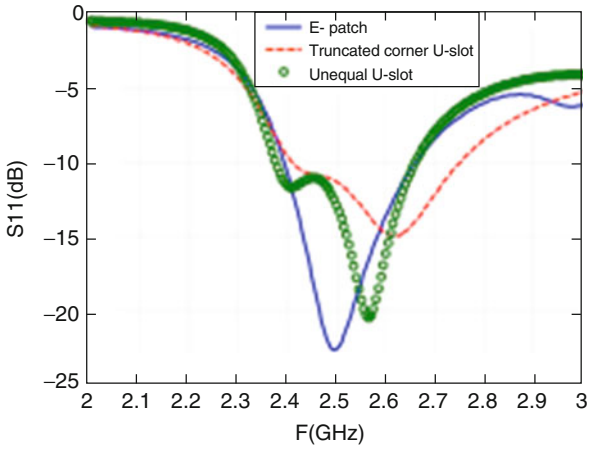


Fig. 50 S_{11} for E-shaped, unequal arms U-slot, and truncated corner U-slot antennas. $\varepsilon_r = 2.2$ (From Khidre et al. (2010). © 2011 IEEE, Reprinted with permission)

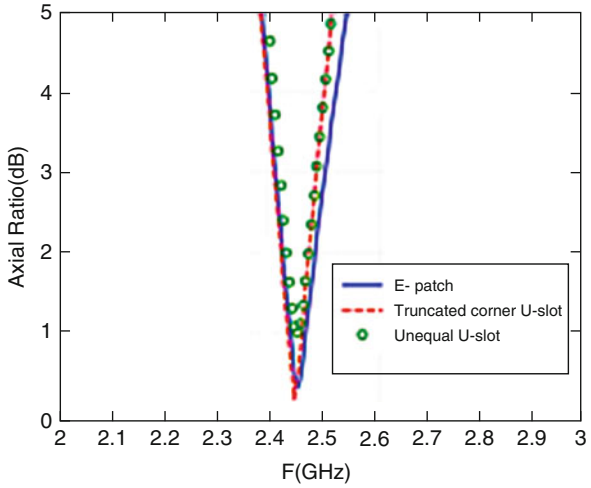


Fig. 51 Axial ratio for E-shaped, unequal arms U-slot, and truncated corner U-slot antennas. $\varepsilon_r = 2.2$ (From Khidre et al. (2010). © 2011 IEEE, Reprinted with permission)

Designs for Wider CP Bandwidth

In Yang et al. (2008), a truncated square patch fed by an L-probe with an air substrate attained a usable bandwidth of 16 %. In Chung and Mohan (2003), truncated corner stacked square patches achieved a usable bandwidth of 17 %. Both antennas are relatively thick.

For thin substrates ($<0.03\ \lambda_0$), good CP bandwidth can be obtained by the use of dual feeds to provide the 90° phase difference for the two orthogonal modes, as illustrated in Fig. 52. This feeding method can attain about 10 % axial ratio bandwidth while keeping the substrate thin, although at the expense of more complexity in the feed circuitry (Hall and Dahele 1997).

The dual-feed design can be considered a special case of multi-feed design. It was shown that sequentially feeding a patch at suitable multiple locations with appropriate phasing leads to the cancelation of the cross circularly polarized components, resulting in improved axial ratio bandwidth (Hall and Dahele 1997).

As in the case of linear polarization, it is desirable in many situations to reduce the size of CP patch antennas, and a number of methods have been developed. Because of page limitations, they are not

DOI 10.1007/978-981-4560-75-7_29-1

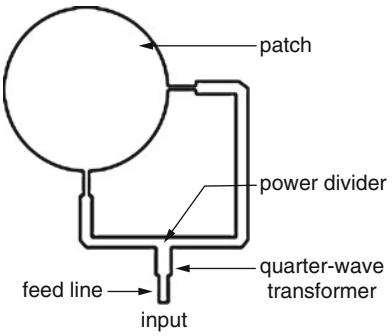


Fig. 52 Dual-feed circularly polarized patch antenna

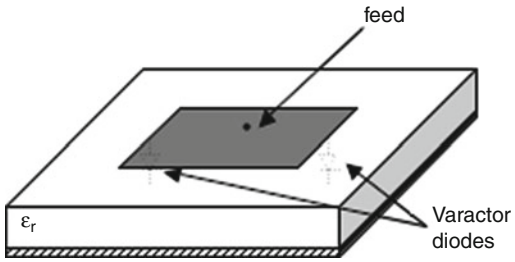


Fig. 53 Illustrating the use of varactor diodes for tuning

described here. The reader is referred to the article on "Small Antennas in Wireless Communications" (Wong et al. 2012).

Frequency-Agile and Polarization-Agile Patch Antennas

In some applications, it is desirable to be able to tune the operating frequency of the patch antenna without having to fabricate a new antenna. In other applications, the ability to alter the polarization of the antenna will enable more efficient use of the allocated frequencies. The methods for designing frequency-agile patch antennas and polarization-agile patch antennas include tuning using varactor diodes, shorting pins, adjustable air gap, patch antenna with switchable slots, and U-slot and E-patches with switchable slots. Several of these will be described in this section.

Frequency-Agile Microstrip Patch Antennas

Tuning Using Varactor Diodes

For a given set of patch dimensions, the resonant frequency is primarily governed by the value of the relative permittivity ε_r of the substrate. If some means is available to alter ε_r, the resonant frequency will change. One method of achieving this is to introduce varactor diodes between the patch and the ground plane, as shown in Fig. 53. The diodes are provided with a bias voltage, which controls the varactor capacitance and hence the effective permittivity of the substrate. Bhartia and Bahl (1982) performed an experiment on this method and the results are shown in Fig. 54. The resonant frequency (f_r) of the lowest mode of the rectangular patch increases with the bias voltage, which changes the diode capacitance. It is seen that, in this experiment, a tuning range of some 20 % was achieved with a 10 V bias. The range

DOI 10.1007/978-981-4560-75-7_29-1

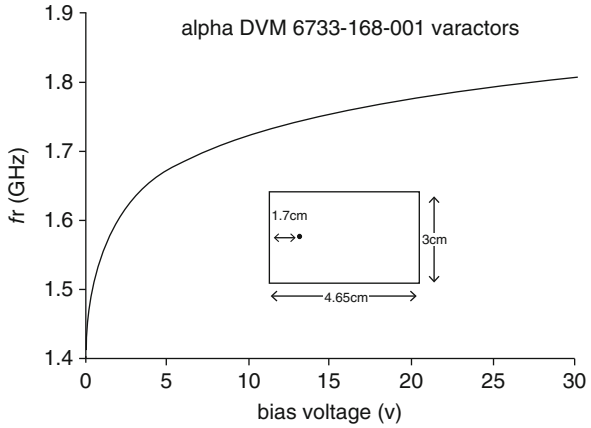


Fig. 54 Resonant frequency versus bias voltage for a varactor-loaded rectangular patch antenna

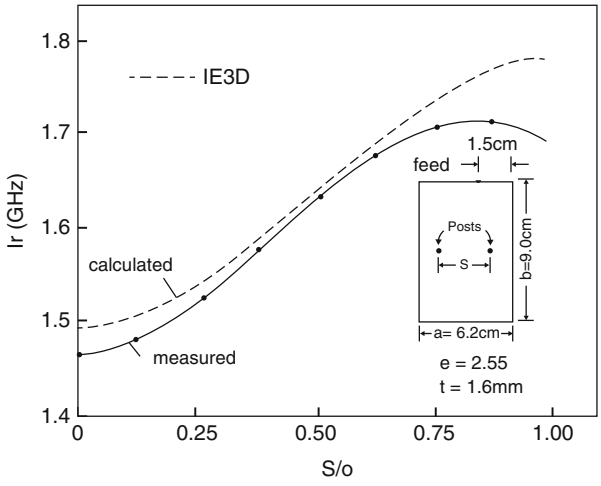


Fig. 55 Resonant frequency versus separation of posts for 6.2 × 9.0 cm rectangular patch antenna with $\varepsilon_r = 2.55$, t = 1.6 mm

increased to about 30 % with a 30 V bias. Note that the curve of frequency versus bias voltage is not a linear one. Similar results were obtained in Waterhouse and Shuley (1994) and Kalialakis and Hall (2007).

Tuning Using Shorting Posts (Pins)

The value of ε_r can also be changed by introducing shorting posts (pins) at various points between the patch and the ground plane. These shorting posts present an inductance and therefore affect the effective permittivity of the substrate. The method appeared to be first introduced by Schaubert et al. (1981), who performed an experiment on a rectangular patch using two pins, as shown in Fig. 55.

The solid curve in Fig. 55 was measured results, which shows that the resonant frequency is dependent on the separation of the two posts. A tuning range of some 18 % is obtained as the separation varies between 0 mm and the whole width of the patch.

It is not possible to analyze the rectangular patch with shorting pins using the cavity model. Full-wave analysis is necessary and a number of commercial software can be used to obtain simulation results.

The results using the Zeland IE3D simulation software are shown in the dashed curve in Fig. 55 for resonant frequency and in Fig. 56 for S_{11}.

DOI 10.1007/978-981-4560-75-7_29-1

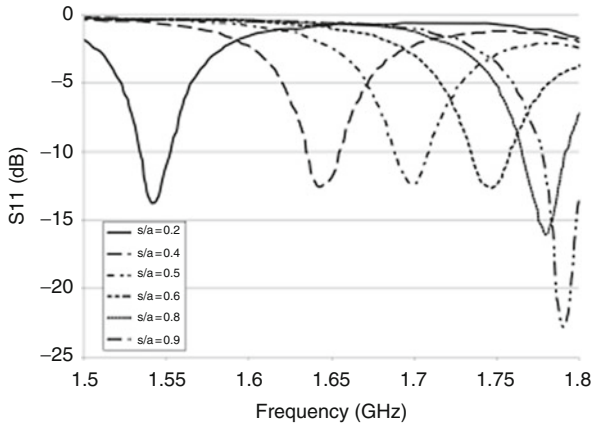


Fig. 56 Simulated reflection coefficient (S_{11}) versus frequency for six values of pin spacing

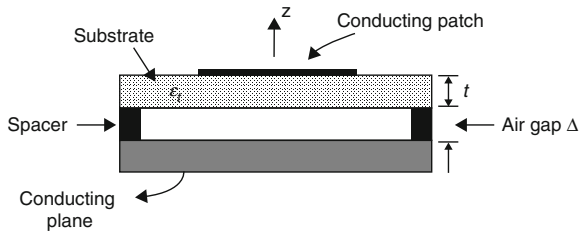


Fig. 57 Geometry of a microstrip patch antenna with an air gap

Tuning by Means of an Adjustable Air Gap

By introducing an air gap between the substrate and the ground plane in a microstrip patch antenna, the effective permittivity of the cavity will change. This can be used to tune the resonant frequency (Dahele and Lee 1985).

The geometry of a microstrip patch antenna with an air gap is shown in Fig. 57. The feed is not shown and will be discussed shortly.

Consider the cavity under the conducting patch. It is made of two layers: a substrate of thickness t and an air region of thickness Δ. Compared to the case with no air gap, the effective permittivity of the cavity is smaller. As a result, the resonant frequencies of the various modes will increase. Since the effective permittivity decreases as Δ increases, tending toward the free-space value of unity as $\Delta \to \infty$, it follows that the resonant frequencies can be tuned by adjusting the air gap with Δ. As a by-product, the bandwidth will also increase partly due to the increase in the height of the dielectric medium and partly because the effective permittivity is smaller. It is also to be noted that substrate and etching tolerances can be compensated by adjusting Δ.

A heuristic derivation of the effective permittivity and the resonant frequency of the two-layer cavity can be obtained by considering the capacitance of a capacitor with two dielectric layers, as shown in Fig. 58.

The overall capacitance is the two capacitors in series:

$$C = \frac{\varepsilon_1 \varepsilon_2 A}{\varepsilon_1 d_2 + \varepsilon_2 d_1} = \frac{\varepsilon_{\text{eff}} A}{d_1 + d_2} \tag{10}$$

where

DOI 10.1007/978-981-4560-75-7_29-1

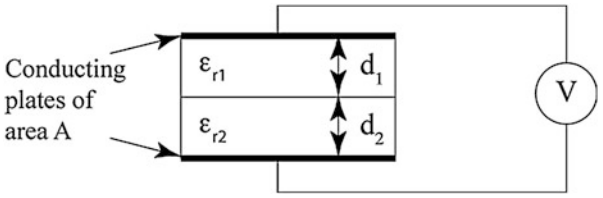


Fig. 58 A capacitor consisting of two dielectric layers

$$\varepsilon_{\text{eff}} = \frac{\varepsilon_1 \varepsilon_2 (d_1 + d_2)}{\varepsilon_1 d_2 + \varepsilon_2 d_1}. \tag{11}$$

Applying this result to the geometry of Fig. 58,

$$\varepsilon_{\text{eff}} = \frac{\varepsilon \varepsilon_0 (t + \Delta)}{\varepsilon_r \varepsilon_0 \Delta + \varepsilon_0 t} = \frac{\varepsilon (t + \Delta)}{(t + \Delta \varepsilon_r)} \tag{12}$$

Since $\varepsilon_r > 1$,

$$\varepsilon_{\text{eff}} < \varepsilon \qquad \text{when} \quad \Delta > 0$$

$$\varepsilon_{\text{eff}} = \varepsilon \qquad \text{when} \quad \Delta = 0$$

If the effective permittivity is used in the formula for the resonant frequency, we obtain, for the TM_{nm} mode,

$$f_{nm}(\Delta) = f(0) \sqrt{\frac{\varepsilon}{\varepsilon_{\text{eff}}}} \tag{13}$$

where $f_{\text{nm}}(0)$ is the resonant frequency when there is no air gap.

The formula for f_{nm} applies to any patch shape. As the air gap width increases, ε_{eff} decreases and the resonant frequency increases. The dependence of f_{nm} on Δ, however, is not a linear one.

Rectangular, circular, and annular-ring patches with air gaps fed coaxially had been studied by Lee and coworkers by means of the cavity model as well by full-wave analysis (Dahele and Lee 1985). The results were validated by an experiment. Typical tuning range is from 0 % to 20 %.

The disadvantage of applying the adjustable air gap method to coaxially fed patches is that, every time the air gap width is changed, the coaxial probe has to be de-soldered and resoldered to the patch. This disadvantage can be avoided if the patch is fed using coupling through an aperture or stripline, which was demonstrated in Mao et al. (2011).

Discussion Compared to the methods of using varactor diodes, or shorting pins, or switchable slot (Yang and Rahmat-Samii 2001b), the air gap method has the following advantages:

1. It can be applied to patches of any shape.
2. There is no need to add components or etch a slot in the patch.
3. The method is particularly attractive for an array consisting of many elements. If the elements are fed by apertures or strip lines, the resonant frequencies of all the elements, and therefore of the array, can be tuned by a single adjustment of the air gap width.

DOI 10.1007/978-981-4560-75-7_29-1

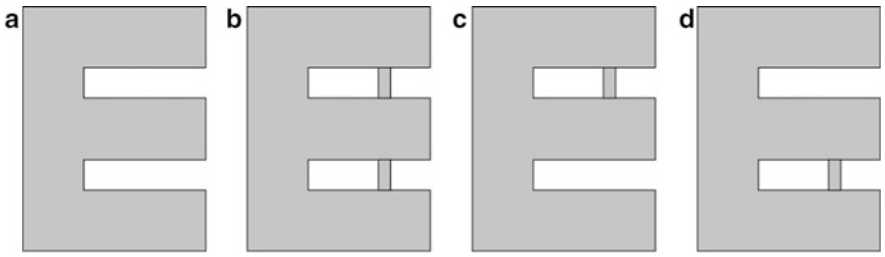


Fig. 59 Switching states of the reconfigurable E-shaped patch antenna: (**a**) LP state 1; (**b**) LP state 2; (**c**) LHCP state 3; and (**d**) RHCP state 4

Table 20 Antenna possible configurations (From Khidre et al. (2013). © 2013 IEEE, Reprinted with permission)

State	Switch 1	Switch 2	Frequency	Polarization
1	OFF	OFF	f_L	LP
2	ON	ON	f_H	LP
3	ON	OFF	f	LHCP
4	ON	OFF	f	RHCP

The disadvantage is that the air gap has to be changed mechanically. However, a MEMS-based electrostatically tunable circular patch antenna has been developed (Jackson and Ramadoss 2007).

Polarization-Agile E-Shaped Patch Antenna

Polarization-agile antennas are antennas whose polarization characteristics can be altered. They are also called polarization reconfigurable antennas and have received increasing attention in the last decade. Polarization agility is a desirable feature for wireless applications because it can double the system capacity through frequency reuse with two orthogonal polarizations without the need of integrating several individual antennas, thus rendering the wireless devices lighter and more compact. Reconfigurable circularly polarized antennas also provide a powerful modulation scheme for microwave tagging systems. They also have potential applications in multiple-input multiple-output (MIMO) systems. Polarization switching could be accomplished with the aid of switching devices such as PIN diodes and RF-MEMs.

Using switchable slots, Yang and Rahmat-Samii (2002) demonstrated a reconfigurable patch antenna for circular polarization diversity. The CP bandwidth obtained was 3 %. By using PIN diodes in U-slot patches, Qin et al. (2010) built two antenna prototypes; the first enables switching between linear and circular polarization, and the second switches between the two circular polarizations, with a CP bandwidth of 2.8 %. Khidre et al. (2013) designed and built CP polarization-agile patch antennas incorporating PIN diodes in an E-shaped patch and obtained CP bandwidth of 7 %. The operation principle, design, and performance of the polarization-agile antenna studied by Khidre et al. (2013) are presented below.

Operation Principle

When an E-shaped patch antenna is loaded with two switches on the slots as shown in Fig. 59, the antenna will have four possible states listed in Table 20. At state 1, the resonant frequency of the original E-shaped patch would be f_L and radiate linear polarized (LP) fields (Yang et al. 2001a). At state 2, the two switches are ON which allows the surface currents to pass through them, making the current path around the slots become shorter. Thus, the electrical length is smaller and the resonant frequency f_H would be higher ($f_H > f_L$). Due to the symmetry of the antenna structure, at state 2, the polarization would also be linear.

DOI 10.1007/978-981-4560-75-7_29-1

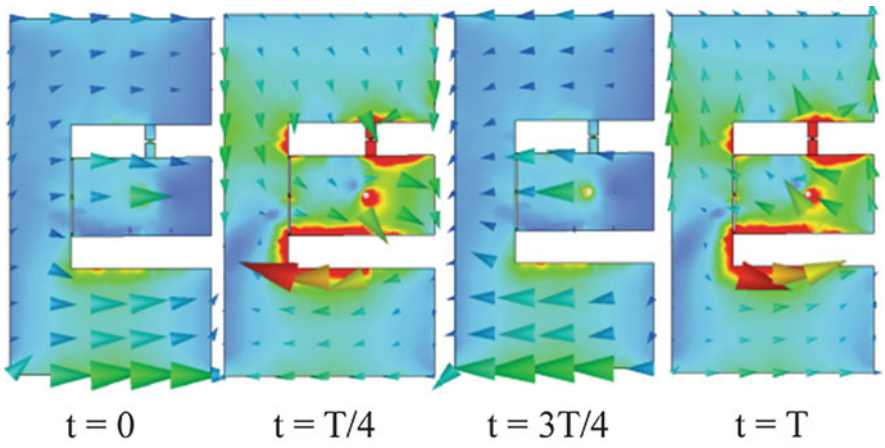


Fig. 60 Current distribution on the E-shaped patch antenna in state 3 during one-time cycle *T* (From Khidre et al. (2013). © 2013 IEEE, Reprinted with permission)

Consequently, going from state 1 to state 2 or vice versa is a frequency reconfigurable E-shaped patch antenna with linear polarization.

Recently, it has been shown that CP fields could be excited by E-shaped patch antenna with a wide effective bandwidth (9.27 %) (Khidre et al. 2010). Letting one of its slots shorter than the other would introduce asymmetry and perturb the field beneath the patch, thus producing CP fields. Therefore, an E-shaped patch antenna at state 3 effectively has the upper slot shorter than the lower one, and at an appropriate switch position on the upper slot, LHCP could be attained at resonant frequency f ($f_L < f < f_H$). In a similar fashion, state 4 gives RHCP. According to the above mechanism, wideband polarization reconfigurable E-shaped patch antenna is realized. The current distribution of an E-shaped patch antenna in state 3 is shown in Fig. 60, where the current rotates clockwise which implies LHCP.

Antenna Design

The reconfigurable E-shaped patch antenna geometry with labels is shown in Fig. 61. It is printed on a thin sheet of RT/duroid 5880 dielectric substrate with $\varepsilon_r = 2.2$ and thickness t = 0.787 mm. The substrate is mounted 10 mm above the ground plane by a coaxial single-feed pin. Although the antenna is built over air substrate for wide bandwidth, the thin dielectric substrate facilitates mounting the microwave components. The dimensions of the antenna are obtained through the procedure described in Khidre et al. (2010).

Two RF PIN diodes are inserted on the slots of the E-shaped patch antenna. A narrow slit of width $s = 0.5$ mm is incorporated on the E-shaped patch surface to avoid DC short across the diodes' terminals. Therefore, the E-shaped patch is divided into two parts: inner and outer one. Three DC block capacitors are inserted on the narrow slit to maintain RF continuity. The outer part of the E-shaped patch is DC grounded through a narrow shorted λ/4 transmission line, and the short circuit is realized by a vertical via as shown in Fig. 62b. This ensures high impedance at the edge of the E-shaped patch and hence keeps the RF current unperturbed on the E-shaped patch surface. The ground plane is common for both the DC and the RF signals. Its size is 200 × 100 mm. The RF and DC control signals are supplied to the antenna via the coaxial pin feed and are superimposed together through a tee bias for RF/DC isolation. Thus, the D2 diode would be ON and the D1 diode would be OFF when the inner part of the E-shaped patch is positively charged, resulting in RHCP. For LHCP, the terminals of the DC source have to be reversed; hence, the D1 diode would be ON and the D2 diode would be OFF.

DOI 10.1007/978-981-4560-75-7_29-1

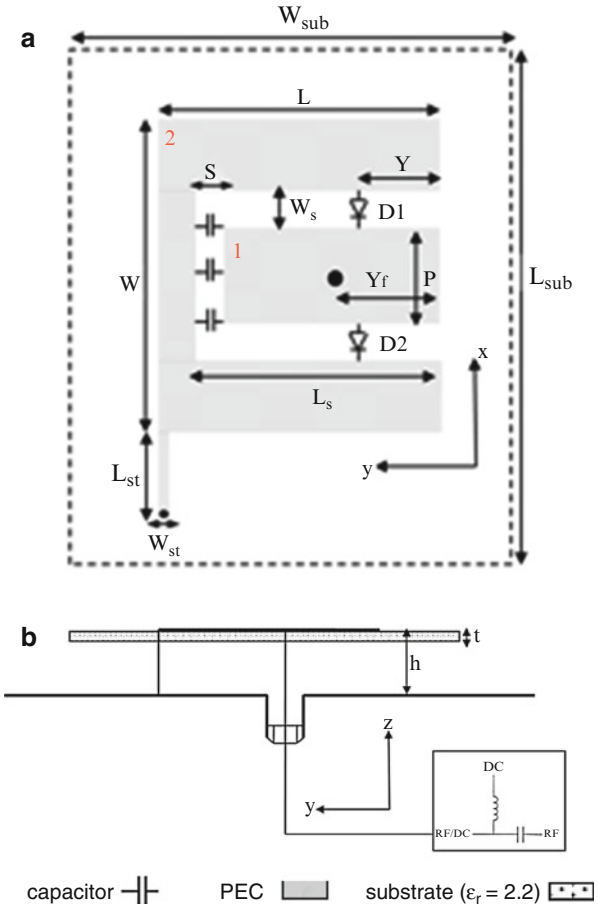


Fig. 61 Geometry of a single-feed reconfigurable E-shaped patch antenna with integrated DC biasing circuit: (**a**) top view; (**b**) side view: L_{sub} =140 mm, W_{sub} =80 mm, L = 43 mm, W = 77 mm, L_s = 30 mm, W_s = 7 mm, P = 17 mm, Y_f =14 mm, L_{st} =28 mm, W_{st} = 0.3 mm, S = 0.5 mm, h = 10 mm, t = 0.787 mm (From Khidre et al. (2013). © 2013 IEEE, Reprinted with permission)

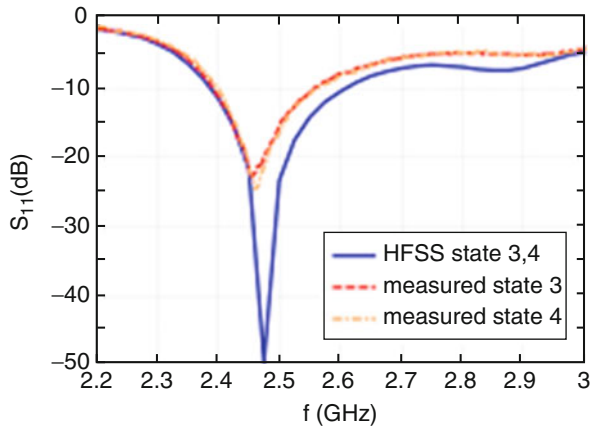


Fig. 62 Simulated and measured S_{11} for the proposed antenna in states 3 and 4 (From Khidre et al. (2013). © 2013 IEEE, Reprinted with permission)

DOI 10.1007/978-981-4560-75-7_29-1

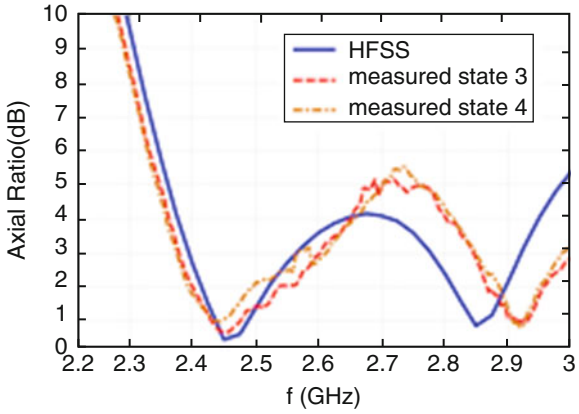


Fig. 63 Simulated and measured axial ratio for the proposed antenna in states 3 and 4 (From Khidre et al. (2013). © 2013 IEEE, Reprinted with permission)

Table 21 Antenna bandwidth simulated versus measured results at states 3 and 4 (From Khidre et al. (2013). © 2013 IEEE, Reprinted with permission)

Parameter	Simulation	Measured
S_{11} (*<–10 dB*)	2.39–2.6 GHz (8.4 %)	2.4–2.575 GHz (7 %)
Axial ratio(*<3 dB*)	2.4–2.6 GHz (8 %)	2.38–2.6 GHz (8.8 %)

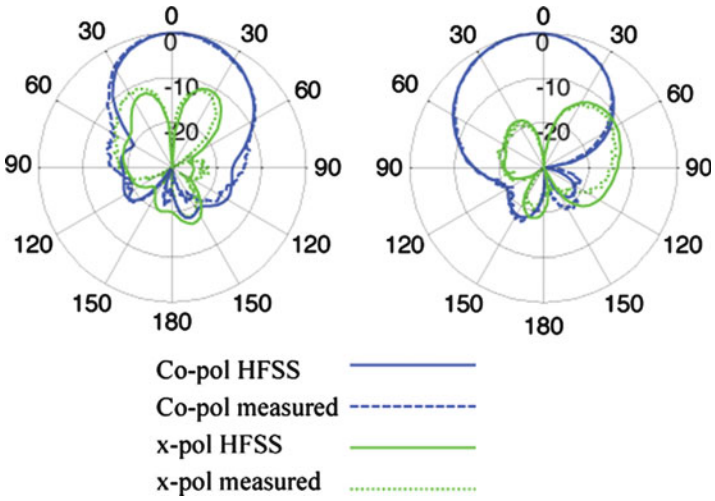


Fig. 64 Simulated and measured radiation pattern of the proposed antenna at 2.45 GHz: (**a**) x-z plane at state 3; (**b**) y-z plane at state 3; (**c**) x-z plane at state 4; and (**d**) y-z plane at state 4 (From Khidre et al. (2013). © 2013 IEEE, Reprinted with permission)

Antenna Performance

The simulated versus measured results of S_{11} and axial ratio in both switching states 3 and 4 are shown in Figs. 62 and 63, respectively. Good agreement between simulated and measured results is observed. Due to the structure symmetry, S_{11} and axial ratio are maintained along switching between the two states, which is an advantage of this design. This can be observed from Fig. 62, where S_{11} in state 3 is almost the

DOI 10.1007/978-981-4560-75-7_29-1

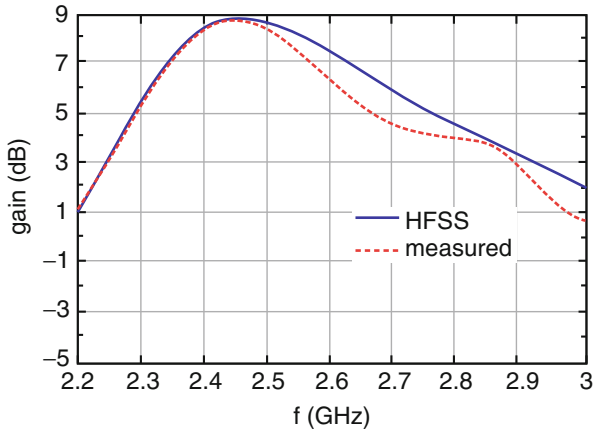


Fig. 65 Simulated and measured gain of the proposed antenna in either state 3 or state 4 (From Khidre et al. (2013). © 2013 IEEE, Reprinted with permission)

same as in state 4. Similarly, in Fig. 63 the axial ratio is kept the same in both switching states. The achieved impedance and axial ratio bandwidths calculated from simulation in comparison to measurements are tabulated in Table 21. From Table 21, the measured effective overlapped bandwidth (−10 dB S_{11} and 3 dB axial ratio) is 7 % (2.4–2.575 GHz), which is wideband for single-feed polarization-agile microstrip antenna designs.

Antenna radiation pattern at 2.45 GHz is shown in Fig. 64, where good agreement between the simulated and measured results is observed. The radiation pattern in y-z plane remains the same after switching, whereas in x-z plane, the radiation pattern is mirrored due to the symmetry of the structure. Experimental results along with the simulated one for the antenna gain versus frequency are shown in Fig. 65. Again, because of the symmetry, the broadside gain is the same in both modes of the antenna. To avoid redundancy, only RHCP mode is shown. Maximum realized gain is 8.7 dBi at 2.45 GHz according to both simulation and measurement, whereas the 3 dB gain bandwidth is 2.31–2.63 GHz (13 %) from measurement which covers the S_{11} and axial ratio bandwidths shown in Table 21.

Conclusion

This chapter began with descriptions of the basic geometry of the microstrip patch antenna, feeding methods, operation principles, and performance characteristics. The chapter then proceeded to present its development into a sophisticated and versatile antenna which is used in numerous military, space, commercial, and medical applications.

While the advances of the last three decades have been impressive, a number of challenges remain. For example, in size reduction, most of the methods developed focused on reducing the patch size, while useful guidelines on the effects of ground plane size on antenna characteristics have not been adequately developed. In broadbanding, the efficient methods developed all result in increased volume, which somewhat mitigate the low-profile advantage of microstrip patch antennas. The search for wideband, efficient, and low-profile patch antennas remains an area of active research.

Other than the basic geometry, there is a lack of design guides for most of the structures discussed in this chapter, such as the wideband U-slot patch, the L-probe-fed patch, or the patch with shorting pin. Antenna

DOI 10.1007/978-981-4560-75-7_29-1

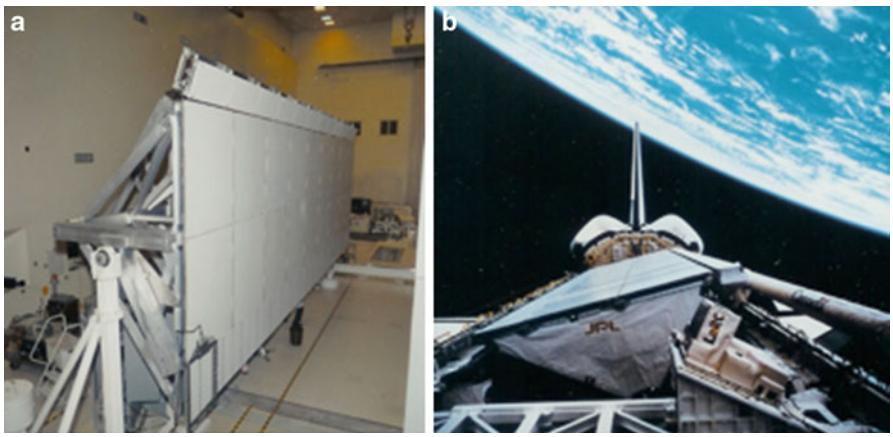


Fig. 66 The picture of a shuttle imaging radar antenna consisting of an array of microstrip patch antennas, (**a**) in the laboratory and (**b**) during flight (Courtesy of Jet Propulsion Laboratory, NASA)

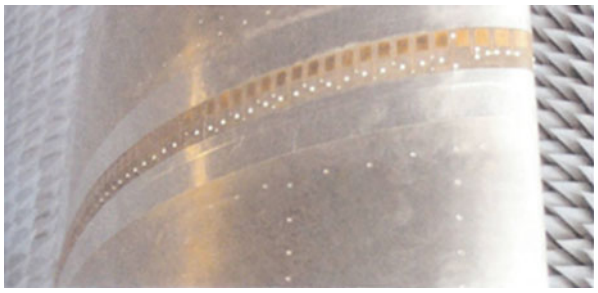


Fig. 67 An array of microstrip antennas are made conformal to a wing shape surface (Courtesy of Air Force Research Lab./ Antenna Technology Branch, Hanscom AFB, USA)

Fig. 68 A patch antenna in a cellular phone

designers have to rely on trial and error using a simulation software. It will be very helpful to develop useful formulas/guidelines/procedures which enable the designer to arrive at the first cut parameters, without relying on trial and error.

It is instructive to end this chapter by showing the photos and a sketch of several microstrip patch antennas used in space, military, commercial, and medical applications.

Figure 66 shows an imaging radar antenna used in the Space Shuttle, consisting of an array of microstrip patch antennas, in the laboratory (a) and during flight (b). Figure 67 is a picture of an array of microstrip antennas which are made conformal to a wing shape surface such as the fuselage of an airplane or rocket. Figure 68 shows a patch antenna in a cellular phone. A base station antenna array

DOI 10.1007/978-981-4560-75-7_29-1

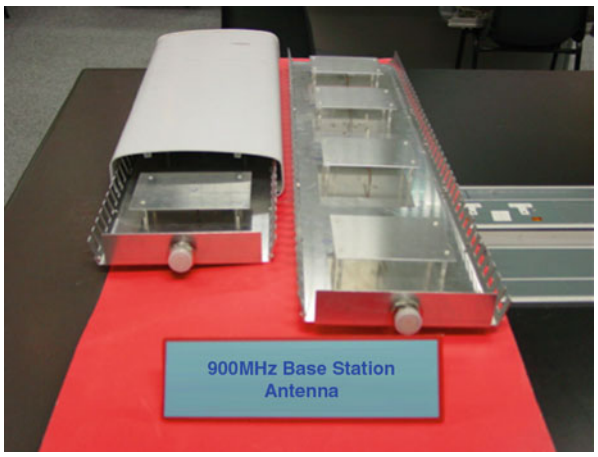


Fig. 69 A base station antenna array consisting of L-probe-fed patches (Courtesy of the Department of Electronic Engineering, City University of Hong Kong)

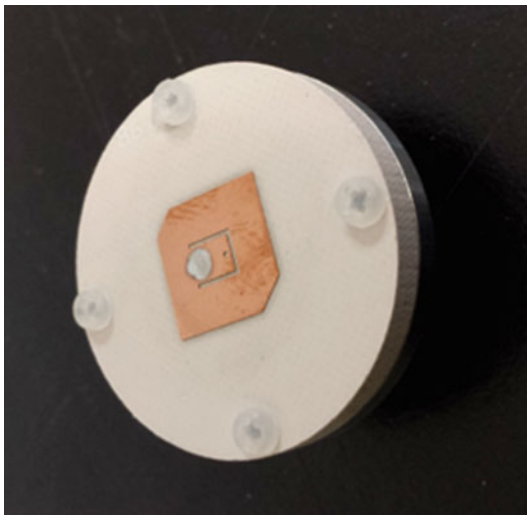


Fig. 70 A GPS antenna in the form of a U-slot square patch with truncated corners (Courtesy of the Department of Electronic Engineering, City University of Hong Kong)

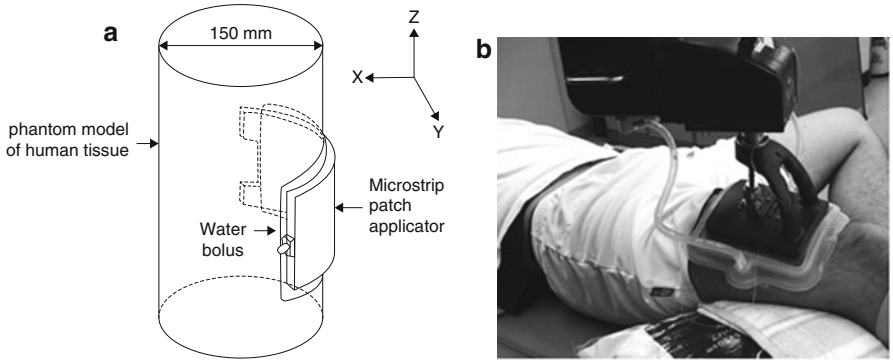


Fig. 71 (**a**) Flexible microstrip patch mounted on a curved surface for hyperthermia medical application (From Kobayashi et al. (1989). © 1989 IEEE, Reprinted with permission). (**b**) The applicator for the hyperthermia system positioned on a human thigh (From Ichinoseki-Sekine et al. (2007). © 2007, British Journal of Sports Medicine, Reprinted with Permission)

DOI 10.1007/978-981-4560-75-7_29-1

consisting of L-probe-fed patches is shown in Fig. 69. Figure 70 is a circularly polarized antenna for GPS applications in the form of a U-slot square patch with truncated corners. Finally, a sketch of a flexible microstrip patch applicator conformable to a shaped surface for microwave hyperthermia is given in Fig. 71a (Kobayashi et al. 1989). A photography of this type of applicator (Ichinoseki-Sekine et al. 2007) in a clinical environment is shown in Fig. 71b, where the patch antenna is used to impart localized microwave energy at 434 MHz into the human body to destroy malignant tumors.

Cross-References

- ▶ Antennas in Hand-Held Devices
- ▶ Broadband and Multiband Planar Antennas
- ▶ Circularly Polarized Antennas
- ▶ Low-Profile Antennas
- ▶ Phased Arrays
- ▶ Reconfigurable Antenna Arrays for Wireless Communications
- ▶ Small Antennas (PIFA/PILA/Loading Antenna/etc.)

References

Anguera J, Font G, Puente C, Borja C, Soler J (2003) Multifrequency microstrip patch antenna using multiple stacked elements. IEEE Microwave Wireless Compon Lett 13(3):123–124

Bahl I, Bhartia P (1980) Microstrip antennas. Artech House, Dedham

Barlately L, Mosig JR, Sphicopoulos T (1990) Analysis of stacked microstrip patches with a mixed potential integral equation. IEEE Trans Antennas Propag 38(5):608–615

Best SR (2009) The significance of ground-plane size and antenna location in establishing the performance of ground-plane-dependent antennas. IEEE Antennas Propag Mag 51(6):29–43

Bhalla R, Shafai L (2002) Broadband patch antenna with a circular arc shaped slot. In: IEEE Antennas and Propagation Society international symposium (IEEE Cat. No.02CH37313), IEEE, pp 394–397, San Antonio, Texas

Bhartia P, Bahl I (1982) A frequency agile microstrip antenna. In: 1982 Antennas and Propagation Society international symposium. Institute of Electrical and Electronics Engineers, pp 304–307, Albuquerque, New Mexico

Chair R, Lee KF, Luk KM (1999) Bandwidth and cross-polarization characteristics of quarter-wave shorted patch antennas. Microw Opt Technol Lett 22(2):101–103

Chair R, Mak C-L, Kishk AA (2005) Miniature wide-band half U-slot and half E-shaped patch antennas. IEEE Trans Antennas Propag 53(8):2645–2652

Chen ZN, Chia MYW (2005) Broadband planar antennas: design and applications. Wiley-Chichester

Chung KL, Mohan AS (2003) A systematic design method to obtain broadband characteristics for singly-fed electromagnetically coupled patch antennas for circular polarization. IEEE Trans Antennas Propag 51(12):3239–3248

Clenet M, Shafai L (1999) Multiple resonances and polarisation of U-slot patch antenna. Electron Lett 35(2):101–103

Croq F, Papiernik A (1990) Large bandwidth aperture-coupled microstrip antenna. Electron Lett 26(16):1293–1294

Dahele JS, Lee KF (1985) Theory and experiment on microstrip antennas with airgaps. IEE Proc H Microwaves Antennas Propag 132(7):455–460
Dahele J, Lee K, Wong D (1987) Dual-frequency stacked annular-ring microstrip antenna. IEEE Trans Antennas Propag 35(11):1281–1285
Debatosh G, Antar YMM (2010) Microstrip and printed antennas: new trends, techniques and applications. Wiley, Hoboken
Deschamps GA, Sichak W (1953) Microstrip microwave antennas. In: Third USAF symposium on antennas, Monticello, Illinois
Deshmukh AA, Ray KP (2010) Multi-band configurations of stub-loaded slotted rectangular microstrip antennas. IEEE Antennas Propag Mag 52(1):89–103
Garg R et al (2000) Microstrip antenna design handbook. Artech House, Boston
Hall PS, Dahele JS (1997) Dual-band circularly polarized microstrip antenna. In: Lee KF, Chen W (eds) Advances in microstrip and printed antennas. Wiley Interscience, New York, pp 163–217
Hirasawa K, Haneishi M (1992) Analysis, design, and measurement of small and low-profile antennas. Artech House Publishers, Boston
Huang J (2008) Microstrip antennas: analysis, design, and application. In: Balanis CA (ed) Modern antenna handbook. Wiley, Hoboken
Huynh T, Lee KF (1995) Single-layer single-patch wideband microstrip antenna. Electron Lett 31(16):1310–1312
Huynh T, Lee KF, Lee R (1988) Crosspolarisation characteristics of rectangular patch antennas. Electron Lett 24(8):463–464
Ichinoseki-Sekine N et al (2007) Changes in muscle temperature induced by 434 MHz microwave hyperthermia. Br J Sports Med 41(7):425–429
Jackson R, Ramadoss R (2007) A MEMS-based electrostatically tunable circular microstrip patch antenna. J Micromech Microeng 17(1):1–8
James JR, Hall PS (eds) (1989) Handbook of microstrip antennas. Peregrinus, London
James JR, Hall PS, Wood C (1981) Microstrip antenna theory and design. Peregrinus, London
Kalialakis C, Hall PS (2007) Analysis and experiment on harmonic radiation and frequency tuning of varactor-loaded microstrip antennas. IET Microwaves Antennas Propag 1(2):527–535
Khidre A, Lee KF, Yang F, Elsherbeni A (2010) Wideband circularly polarized E-shaped patch antenna for wireless applications. IEEE Antennas Propag Mag 52(5):219–229
Khidre A, Lee KF, Elsherbeni A, Yang F (2013) Circular polarization reconfigurable wideband E- shaped patch antenna for wireless applications. IEEE Trans Antennas Propag 61(2):960–964
Kobayashi H, Nikawa Y, Okada F, Mori S (1989) Flexible microstrip patch applicator for hyperthermia. In: Digest on Antennas and Propagation Society international symposium, IEEE, pp 536–539, San Jose, CA
Kumar G, Gupta K (1984) Broad-band microstrip antennas using additional resonators gap-coupled to the radiating edges. IEEE Trans Antennas Propag 32(12):1375–1379
Kumar G, Ray KP (2003) Broadband microstrip antennas. Artech House Publishers, Boston
Lee KF, Chen W (1997) Advances in microstrip and printed antennas. Wiley Interscience, New York
Lee KF, Dahele JS (1989) Characteristics of microstrip patch antennas and some methods of improving frequency agility and bandwidth. In: James JR, Hall PS (eds) Handbook of microstrip antennas. Peregrinus, London, pp 111–214
Lee KF, Luk KM (2010) Microstrip patch antennas. Imperial College Press, London
Lee K, Tong K (2012) Microstrip patch antennas – basic characteristics and some recent advances. Proc IEEE 100(7):2169–2180

Lee RQ, Lee K, Bobinchak J (1987) Characteristics of a two-layer electromagnetically coupled rectangular patch antenna. Electron Lett 23(20):1070–1072
Lee KF, Luk KM, Dahele JS (1988) Characteristics of the equilateral triangular patch antenna. IEEE Trans Antennas Propag 36(11):1510–1518
Lee K-F, Chen W, Lee RQ (1995) Studies of stacked electromagnetically coupled patch antennas. Microw Opt Technol Lett 8(4):212–215
Lee KF, Luk KM, Tong KF, Shum SM, Huynh T, Lee RQ (1997) Experimental and simulation studies of the coaxially fed U-slot rectangular patch antenna. IEE Proc Microwaves Antennas Propag 144(5):354–358
Lee KF, Guo YX, Hawkins JA, Chair R, Luk KM (2000) Theory and experiment on microstrip patch antennas with shorting walls. IEE Proc Microwaves Antennas Propag 147(6):521–525
Lee KF, Yang SLS, Kishk AA (2008) Dual- and multiband U-slot patch antennas. IEEE Antennas Wirel Propag Lett 7:645–647
Lee KF, Yang SL, Kishk AA, Luk KM (2010) The versatile U-slot patch antenna. IEEE Antennas Propag Mag 52(1):71–88
Lee KF, Luk KM, Mak KM, Yang SLS (2011) On the use of U-slots in the design of dual-and triple-band patch antennas. IEEE Antennas Propag Mag 53(3):60–74
Lo Y, Solomon D, Richards W (1979) Theory and experiment on microstrip antennas. IEEE Trans Antennas Propag 27(2):137–145
Long S, Walton M (1979) A dual-frequency stacked circular-disc antenna. IEEE Trans Antennas Propag 27(2):270–273
Luk KM, Mak CL et al (1998a) Broadband microstrip patch antenna. Electron Lett 34(15):1442–1443
Luk KM, Chair R, Lee KF (1998b) Small rectangular patch antenna. Electron Lett 34(25):2366–2367
Mak C, Lee K, Luk K (2000) Broadband patch antenna with a T-shaped probe. IEE Proc Microwaves Antennas Propag 147(2):73–76
Mak C-L, Wong H, Luk K-M (2005) High-gain and wide-band single-layer patch antenna for wireless communications. IEEE Trans Veh Technol 54(1):33–40
Mao Y, Padooru Y, Lee KF, Elsherbeni A, Yang F (2011) Air gap tuning of patch antenna resonance. In: 2011 I.E. international symposium on antennas and propagation (APSURSI), IEEE, pp 3088–3090, Spokane, Washington
Mok WC, Wong SH, Luk KM, Lee KF (2013) Single-layer single-patch dual-band and triple-band patch antennas. IEEE Trans Antennas Propag 61(8):4341–4344
Mosig J, Gardiol F (1985) General integral equation formulation for microstrip antennas and scatterers. IEE Proc H Microwaves Antennas Propag 132(7):424–432
Nakano H, Yamazaki M, Yamauchi J (1997) Electromagnetically coupled curl antenna. Electron Lett 33(12):1003–1004
Pinhas S, Shtrikman S (1988) Comparison between computed and measured bandwidth of quarter-wave microstrip radiators. IEEE Trans Antennas Propag 36(11):1615–1616
Pozar DM (1985) Microstrip antenna aperture-coupled to a microstripline. Electron Lett 21(2):49–50
Pozar DM (1992) Microstrip antennas. Proc IEEE 80(1):79–91
Qin P-Y et al (2010) Polarization reconfigurable U-slot patch antenna. IEEE Trans Antennas Propag 58(10):3383–3388
Rafi G, Shafai L (2004) Broadband microstrip patch antenna with V-slot. IEE Proc Microwaves Antennas Propag 151(5):435–440
Reineix A, Jecko B (1989) Analysis of microstrip patch antennas using finite difference time domain method. IEEE Trans Antennas Propag 37(11):1361–1369

DOI 10.1007/978-981-4560-75-7_29-1

Richards WF, Lo YT, Harrison DD (1981) An improved theory for microstrip antennas and applications. IEEE Trans Antennas Propag 29(1):38–46

Sabban A (1983) A new broadband stacked two-layer microstrip antenna. In: 1983 Antennas and Propagation Society international symposium, Institute of Electrical and Electronics Engineers, pp 63–66, Houston, Texas

Salon S, Chari MVK (1999) Numerical methods in electromagnetism. Academic, San Diego

Schaubert D et al (1981) Microstrip antennas with frequency agility and polarization diversity. IEEE Trans Antennas Propag 29(1):118–123

Schneider MV (1969) Microstrip lines for microwave integrated circuits. Bell Syst Tech J 48(5):1421–1444

Shackelford AK et al (2001) U-slot patch antenna with shorting pin. Electron Lett 37(12):729–730

Shackelford AK, Lee KF, Luk KM (2003) Design of small-size wide-bandwidth microstrip-patch antennas. IEEE Antennas Propag Mag 45(1):75–83

Shafai L (2007) Wideband microstrip antennas. In: Volakis JL (ed) Antenna engineering handbook. McGraw Hill, New York

Tong K-F, Wong T-P (2007) Circularly polarized U-slot antenna. IEEE Trans Antennas Propag 55(8):2382–2385

Tong KF et al (2000) A broad-band U-slot rectangular patch antenna on a microwave substrate. IEEE Trans Antennas Propag 48(6):954–960

Tong KF, Lee KF, Luk KM (2011) On the effect of ground plane size to wideband shorting-wall probe-fed patch antennas. In: 2011 IEEE-APS topical conference on antennas and propagation in wireless communications, IEEE, pp 486–486, Torino, Italy

Waterhouse RB, Shuley NV (1994) Full characterisation of varactor-loaded, probe-fed, rectangular, microstrip patch antennas. IEE Proc Microwaves Antennas Propag 141(5):367–373

Waterhouse RB, Targonski SD, Kokotoff DM (1998) Design and performance of small printed antennas. IEEE Trans Antennas Propag 46(11):1629–1633

Weigand S et al (2003) Analysis and design of broad-band single-layer rectangular u-slot microstrip patch antennas. IEEE Trans Antennas Propag 51(3):457–468

Wong KL (2002) Compact and broadband microstrip antennas. Wiley Interscience, New York

Wong H, Luk KM, Chan CH, Xue Q, So KK, Lai HW (2012) Small antennas in wireless communications. IEEE Proc 100(7):2109–2121

Wood C (1981) Analysis of microstrip circular patch antennas. IEE Proc H Microwaves Opt Antennas 128(2):69–76

Yang F, Zhang, X-X, Rahmat-Samii (2001a). Wide-band E-shaped patch antennas for wireless communications. IEEE Trans Antennas Propag 49(7):1094–1100

Yang F, Rahmat-Samii Y (2001b) Patch antenna with switchable slot (PASS): dual-frequency operation. Microw Opt Technol Lett 31(3):165–168

Yang F, Rahmat-Samii Y (2002) A reconfigurable patch antenna using switchable slots for circular polarization diversity. IEEE Microwave Wireless Compon Lett 12(3):96–98

Yang S et al (2008) Design and study of wideband single feed circularly polarized microstrip antennas. Prog Electromagn Res 80:45–61

Zhong SS, Lo YT (1983) Single-element rectangular microstrip antenna for dual-frequency operation. Electron Lett 19(8):298–300

Zurcher J, Gardiol F (1995) Broadband patch antenna. Artech House Publishers, Boston

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_30-1

Reflector Antennas

Trevor S. Bird*
Antengenuity, Eastwood, NSW, Australia

Abstract

Reflector antennas are in widespread use in communications and radar applications in the twenty-first century. They are seen on towers for point-to-point telecommunication links, on houses for pay tv and in news items of spacecraft travelling to distant galaxies. This chapter provides an introduction to reflector antenna fundamentals for design and analysis. The history of the focusing properties of reflectors extends back to ancient times, but some of the more intricate properties such as defocusing have not been realised until recent times. Some basic reflector geometries are considered initially through a review of geometric optics. The basic design principles of a reflector are established from a rigorous point of view through a study of the paraboloid geometry, its radiation pattern and focal region fields. Practical reflectors have imperfections due to surface errors or misalignments, and how these impact the reflector radiation pattern is outlined. The means of feeding a parabolic reflector is outlined through a description of the dipole, waveguide and horn feeds. Some other reflector configurations are discussed including the offset parabolic reflector, the symmetrical and offset Cassegrain geometries and the spheroid. An introduction to reflector shaping is given through a description of two techniques. The chapter concludes with a description of three typical reflector applications in satellite communications, weather radar and radio astronomy.

Keywords

Reflector; Paraboloid; Cassegrain; Spheroid; Shaping; Aperture blocking; Aberrations; Offset parabolic; Feeds; Antenna; Shaped beams; Dual-polarized radar; Radioastronomy

Introduction

Reflector antennas are ubiquitous in communications and radar applications in the twenty-first century. They are seen on towers for point-to-point telecommunication links, on houses for pay tv and in news items of spacecraft travelling to distant galaxies. For many in society, they have become symbolic of progress and advancement in technology. Whilst the design and operation of reflectors is of crucial interest to many, for some the concepts are considered well established and lack interest for further study. Indeed, many of the basic concepts are known, but like many areas there are detailed topics of vital importance that require detailed knowledge for future designs to meet the ever-increasing performance criteria. The purpose of this chapter is to provide an introduction to the design principles of reflector antennas. In this approach, important techniques and concepts are introduced to assist the reader with basic reflector design principles and also to assist in suggesting improvements in design approaches. It could be part of a first lecture course on advanced antennas.

*Email: tsbird@optusnet.com.au
*Email: tsbird@ieee.org

DOI 10.1007/978-981-4560-75-7_30-1

In transmission, a reflector antenna concentrates signals received from another antenna, called a feed, into a narrow beam of radiation in the far field. Close by, and in front of the reflector, the resulting aperture field is reasonably uniform. This field falls away rapidly normal to the reflector edge. Depending on the feeding arrangement, the beam can be scanned, or it can be broadened to cover a defined two-dimensional region, such as on the earth's surface from space. In reception, an incoming signal is directed by the reflector to the feed and the receiving electronics. As a result of the principle of reciprocity (Balanis 1982), the means of operation is the same whether in transmit or receive operation.

In the next section, a short history is given of reflector antennas up to recent times. This is followed by a description of some basic reflector geometries by means of geometric optics. The standard methods used in the analysis of reflector antennas are then reviewed as background. This is followed in the section "Introduction to Reflector Antennas" by basic design principles of a reflector from a rigorous point of view through study of the paraboloid. The impacts of reflector imperfections such as surface errors and misalignments on the reflector radiation pattern are outlined. The means of feeding a reflector is described where simple feed antennas and techniques are also discussed in the section "Introduction to Reflector Antennas." Other reflector configurations, such as the offset parabolic reflector and the Cassegrain geometry, are discussed later in the section as well as an introduction to reflector shaping. The chapter concludes with a description of three typical reflector applications, namely, satellite communications, dual-polarized weather radar and radio astronomy.

Background

Brief History of Reflector Antennas

The history of reflector antennas commences with basic configurations taken from optical astronomy such as the parabolic reflector and the Newtonian and Cassegrain geometries. The parabola has been well known from ancient times for its ability to focus rays from the sun. For example, it has been claimed that Archimedes in the third century BC, who had made a study of paraboloidal reflectors, used them to set the Roman fleet alight during the Siege of Syracuse. In the early experiments conducted by Heinrich Hertz in the late 1800s to verify the predictions of James Clerk Maxwell, a cylindrical parabolic reflector was used with his spark gap generator at an estimated operating frequency of about 450 MHz. The reflector was made of zinc sheet metal supported by a wooden frame and had a spark-gap excited dipole along its focus as shown in Fig. 1. The aperture size was 2 m high by 1.2 m wide, with a focal length of 0.12 m. Some of these antennas still exist today (see Fig. 1). Other early workers such as Bose employed plane reflectors in 1895 as part of their experiments on the electrical properties of materials. Following these early experiments, reflectors appear to have been little used until the 1920s when UHF and microwave sources became more readily available and experiments were conducted on communications and radar. During the 1930s, Marconi used a parabolic reflector on his ship *Electra* for investigations of UHF transmission. In March 1931, a 1.67 GHz microwave relay telephone link, called Micro-Ray, was demonstrated across the English Channel between Calais, France, and Dover, UK, using 10 ft (3 m) parabolic dishes (see Fig. 2a). As a result of this demonstration, a permanent link was established in 1934 between Lympne, France, and St Inglevert, UK. Also in the 1930s, Bell Technical Laboratories (BTL), USA, worked on a range of equipment, including reflectors operating at frequencies up to 500 MHz, for potential radio relay links operating at these frequencies. These advances would have to wait until efficient sources became available and higher frequencies were required. In 1937, a 9 m dish was built by pioneering radio astronomer Grote Reber in his backyard, where he undertook a sky survey that resulted ultimately in the new field of radio astronomy. About this time, in 1938, early airborne radar experiments were conducted in the UK using the 30 cm wavelength and a 2 ft (60 cm) square cylindrical parabola. The invention of the resonant magnetron

DOI 10.1007/978-981-4560-75-7_30-1

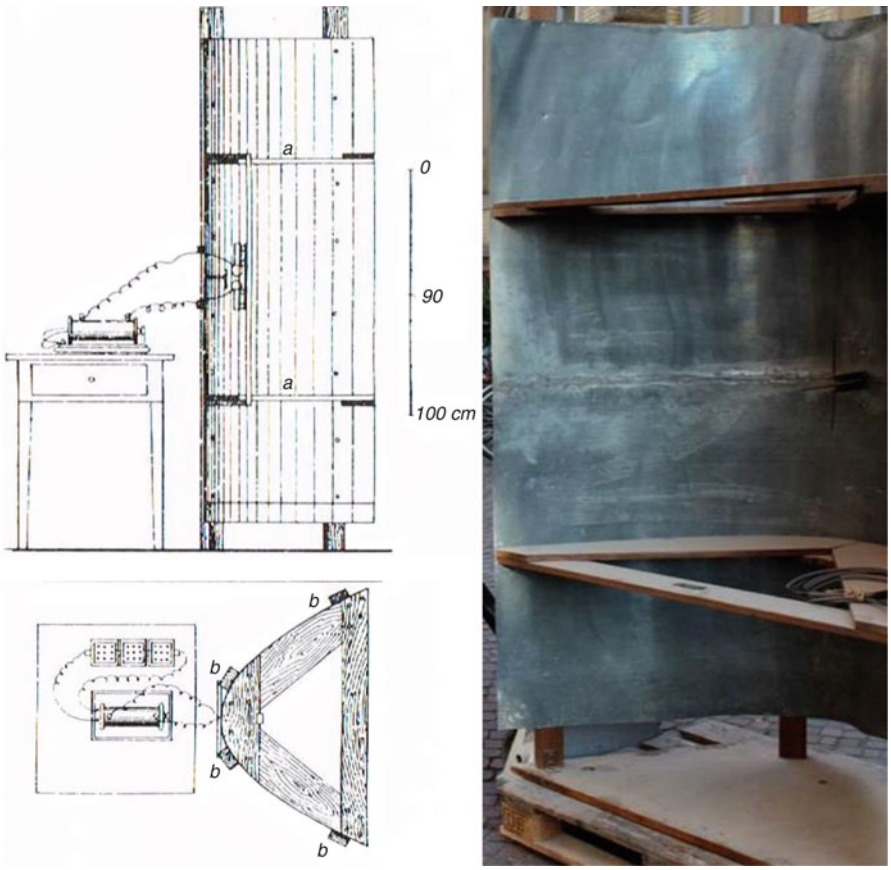


Fig. 1 The Hertz cylindrical parabolic reflector or mirror used for experiments on 'electric radiation' at Karlsruhe, Germany (Hertz 1892)

by Boot and Randall in 1940 made centimetre-wave radar technically feasible, as well as communications. The former application resulted in the evolution of air interception radar and the use of dishes in aircraft from 1941 onwards in Britain and the USA. Other defence equipment employing reflectors were developed at this time (e.g., gun-laying radar). Theoretical aspects of reflectors were not neglected. The development of techniques beyond geometric optics in occurred in several institutes in the 1940s and particularly the MIT Radiation Laboratory in the USA, where technical reports and notes were disseminated after 1945 in textbooks. In particular, the book by Samuel Silver (1985) on microwave antennas contained a detailed description of reflector antenna design that was the definitive statement on the topic for upwards of 30 years and is still a very major reference. Thus, during the early 1940s, radar provided a great incentive for reflector antenna research, and this resulted in many new developments such as improved feeding techniques and shaped-beam antennas, in which the reflector is shaped differently in two orthogonal planes to produce a beam of a particular shape. The consolidation of the concepts in the 1950s resulted in further applications such as terrestrial links, and also advances in radio astronomy saw the construction of three large parabolic reflectors over 60 m in diameter at Jodrell Bank, UK, Parkes, Australia, and Greenbank, USA. The introduction of geostationary satellites for global communications in the 1960s witnessed the wide use of many types of dish antennas for communication networks, which conveyed telephone calls and television programmes across continents. One of the first parabolic reflector antennas used for commercial satellite communications was constructed in 1962 at Goonhilly Downs, Cornwall, England, to communicate with the Telstar satellite. The stabilized satellites that then became available were able to point a narrow beam continuously towards the earth (with a despun platform, for

DOI 10.1007/978-981-4560-75-7_30-1

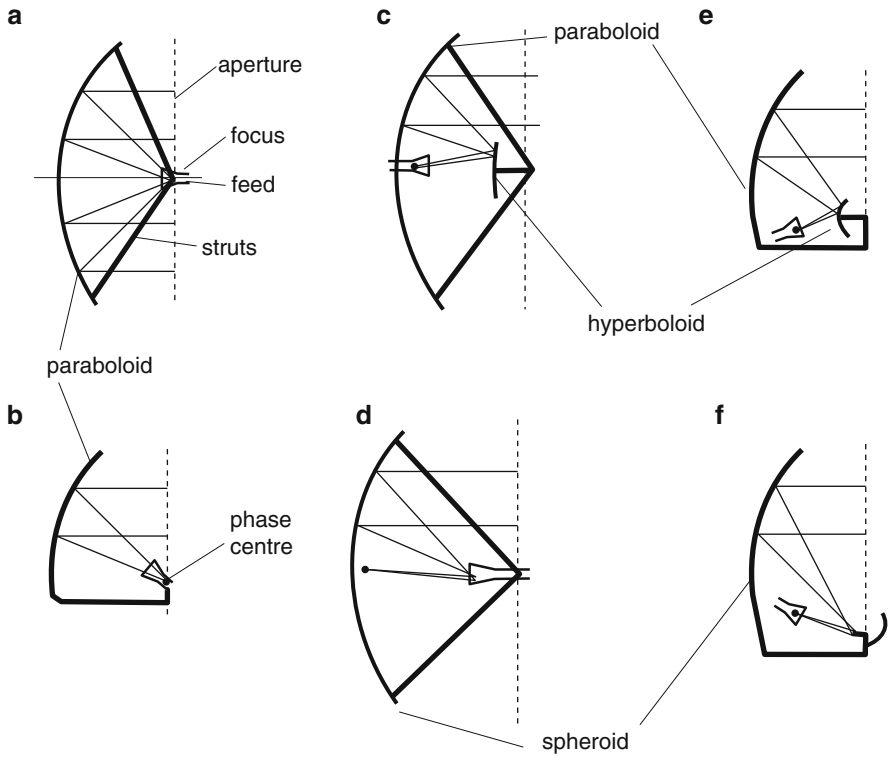


Fig. 2 Reflector geometries; (**a**) symmetrical paraboloid; (**b**) offset parabolic reflector; (**c**) symmetrical Cassegrain; (**d**) spherical reflector; (**e**) offset Cassegrain; and (**f**) offset Gregorian

example) with the result that high-gain reflectors were introduced on-board satellites. On the ground, with the increased signal, smaller and cheaper reflectors (e.g., 30 m) could be used. Many countries around the world consolidated the dishes for this purpose at a central location, now referred to as teleport. The first reflector with a dual-reflector Cassegrain configuration (see Fig. 2e) was installed in Japan in 1963. These were assisted by improvements in understanding of reflector design and associated design techniques. The widespread use of computers in the 1960s and the advent in the 1970s of computer design tools as well as the further development of techniques such as the geometrical theory of diffraction led to improved design accuracy and reflector performance. Some of these techniques will be described in this chapter. There are two major reviews of developments in reflector antenna design and application that contain valuable information, one in the 1980s (Rusch 1984) and another at the end of the 1990s (Bird and James 1999), both of which are worth consulting for further details of reflectors.

Basic Reflector Geometries

The basic reflector geometries are shown in Fig. 2. Most configurations shown arose in optical astronomy and are designed to maximize the signal in one direction only, which is determined by the amplitude and phase distribution across the aperture. The paraboloid in Fig. 2a is the most commonly occurring type of reflector configuration. An offset version is shown in Fig. 2b. Another reflector shape that is used is the spherical reflector Fig. 2d, which does not have a single point focus but a focal line along its axis. The use of a subreflector with a parabola in a Cassegrain configuration as shown in Fig. 2c, e can have enhanced performance such as increased antenna gain because feed lines can be short, which also assists to reduce the antenna noise temperature. A disadvantage of any axisymmetric reflector antenna system is the blockage created by the feed or the subreflector and any associated strut supports. These obstructions have a deleterious effect on the antenna gain and, often more importantly, on sidelobe levels. Blockage

DOI 10.1007/978-981-4560-75-7_30-1

effects can be avoided by means of the offset-fed reflector configurations shown in Fig. 2b, e, f. Although the lack of symmetry creates a number of design problems, these antennas are capable of performance which is superior to that of their axisymmetric counterparts. A difficulty of single offset reflectors Fig. 2b is the high cross-polarization due to the offset feed. However, this can be largely overcome in Fig. 2e, f by means of a subreflector, which, along with the feed offset angle, can be adjusted to cancel out the cross-polarization due to the offset feeding. A spherical reflector equivalent is the Gregorian configuration which has a concave subreflector illustrated in Fig. 2f. In reception, the secondary reflector, which is located axially outside of the primary focus, forms a properly oriented final focus by refocussing the diverging wave emanating from the primary reflector. Furthermore, if the subreflector is shaped, the primary and subreflector combination can be arranged to have a point focus. There are other well-known reflector configurations that are not shown here but which are most suitable for some applications such as the Schwartzfeld system for beam steering.

Geometric Optics and Ray Tracing Description of Reflectors

The laws of reflection at a perfect conductor require the incident and reflected ray directions, $\mathbf{s_i}$ and $\mathbf{s_r}$, to satisfy the following:

$$(\mathbf{s_i} - \mathbf{s_r}) \times \hat{n} = 0 \tag{1a}$$

$$(\mathbf{s_i} + \mathbf{s_r}) \cdot \hat{n} = 0 \tag{1b}$$

where $\hat{n}$ is the normal to the reflector surface, Σ, so that

$$\mathbf{s_r} = \mathbf{s_i} - 2\hat{n}(\hat{n} \cdot \mathbf{s_i}). \tag{1c}$$

Thus, Eq. 1 indicates $\mathbf{s_i}$ and $\mathbf{s_r}$ lie in the same plane and also the angle of incidence relative to the normal equals the angle of reflection after taking into account that one angle is 360° minus the angle for the other ray.

The boundary condition at the surface requires the total net tangential electric field to be zero and for continuity of the normal components of the electric field. Referring to Fig. 1.2.1, if $\mathbf{E_i}$. is the incident electric field and $\mathbf{E_r}$ is the reflected field, these require

$$(\mathbf{E_i} + \mathbf{E_r}) \times \hat{n} = 0 \tag{2a}$$

and

$$(\mathbf{E_i} - \mathbf{E_r}) \cdot \hat{n} = 0. \tag{2b}$$

Taking the cross-product of Eq. 2a and $\hat{n}$ results in $(\mathbf{E_i} + \mathbf{E_r}) - \hat{n}[(\mathbf{E_i} + \mathbf{E_r}) \cdot \hat{n}] = 0$.

Utilizing Eq. 2b, it is found that

$$\mathbf{E_r} = 2\hat{n}(\hat{n} \cdot \mathbf{E_i}) - \mathbf{E_i}. \tag{3}$$

Equation 3 gives the reflected electric field at the surface, Σ, in terms of the incident field. The field reflected from a paraboloid is a plane wave, and extra phase shift must be included to account for the plane wave path s_r (Fig. 3) from the reflector to the aperture. Since the total path length is $2f$ and the ray path length from the focus to the reflector is ρ,

DOI 10.1007/978-981-4560-75-7_30-1

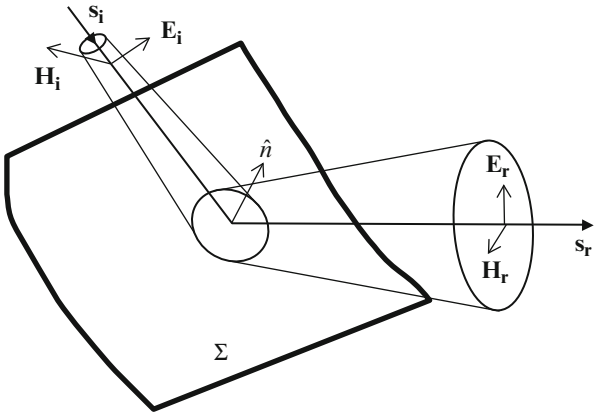


Fig. 3 Reflection at a conducting surface Σ

$$s_{\mathbf{r}} = 2f - \rho.$$

Therefore, the approximate aperture field in the plane at $z = 0$ due to an arbitrary incident field is

$$\mathbf{E}_a = \left[2\hat{n}\left(\hat{n} \cdot \mathbf{E}_f\right) - \mathbf{E}_f\right] e^{-jk(2f-\rho)} \tag{4a}$$

$$\mathbf{H}_a = \frac{1}{\eta_o}\hat{z} \times \mathbf{E}_a \tag{4b}$$

Suppose the feed antenna radiates an incident electric field at the reflector in the following form:

$$\mathbf{E_i} = \left(\hat{\psi}' F_\psi(\psi', \xi') + \hat{\xi}' F_\xi(\psi', \xi')\right) \exp(-jk\rho')/\rho' \tag{5}$$

where the primed coordinates (ρ', ψ', ξ') are coordinates relative to the feed phase centre. Many feed types such as waveguides and horns can be represented by the feed functions

$$F_\psi(\psi', \xi') = A(\psi') \cos \xi' \text{and } F_\xi(\psi', \xi') = -B(\psi') \sin \xi'. \tag{6}$$

If $A(\psi) = B(\psi)$, Eq. 5 becomes

$$\mathbf{E_i} = A(\psi')\left(\hat{\psi}' \cos \xi' - \hat{\xi}' \sin \xi'\right)\Big) \exp(-jk\rho')/\rho' = \hat{x}' A(\psi') \exp(-jk\rho')/\rho'$$

indicating the incident field is linearly polarized and the radiation pattern is axisymmetric.

Suppose the feed phase centre is coincident with the focus of the reflector. For illustrative purposes, assume the reflector is a perfect paraboloid, in which case $\rho = 2f/(1 + \cos\psi)$ is the distance from the focus to the surface of the paraboloid. By elementary methods, the normal to the surface can be shown to be given by

DOI 10.1007/978-981-4560-75-7_30-1

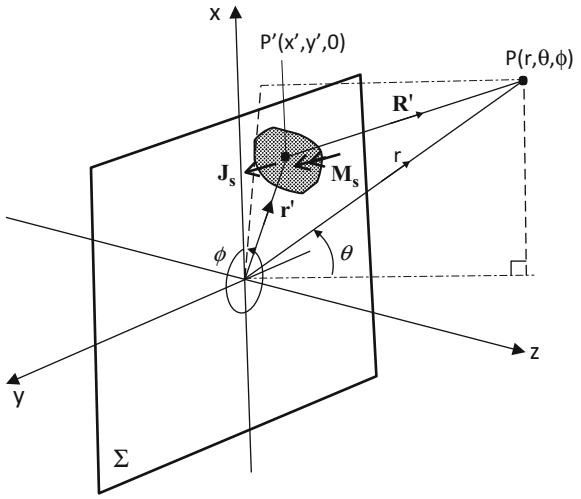


Fig. 4 Calculation of radiated fields at observation point P from aperture currents at source point P′

$$\hat{n} = \hat{\rho}\cos\frac{\psi}{2} + \hat{\psi}\sin\frac{\psi}{2}.$$

Substituting this and Eq. 5 into Eq. 4a results in electric fields in the aperture plane as

$$\mathbf{E}_a = -\left[\hat{\rho}F_\psi \sin\psi + \hat{\psi}F_\psi \cos\psi + \hat{\xi}F_\xi\right](1 + \cos\psi)\frac{\exp(-jk2f)}{2f}.$$

When expressed in rectangular coordinates, the field is

$$\mathbf{E}_a = -\left[\hat{x}\left(F_\psi \cos\xi - F_\xi \sin\xi\right) + \hat{y}\left(F_\psi \sin\xi + F_\xi \cos\xi\right)\right](1 + \cos\psi)\frac{\exp(-jk2f)}{2f}. \tag{7}$$

Introducing Eq. 25 into Eq. 26, the aperture field reduces to

$$\mathbf{E}_a = -\left[\hat{x}\left(A(\psi)\cos^2\xi + B(\psi)\sin^2\xi\right) + \hat{y}\sin 2\xi\left(\frac{A(\psi) - B(\psi)}{2}\right)\right](1 + \cos\psi)\frac{\exp(-jk2f)}{2f} \tag{8}$$

Further, for a feed with an axisymmetric pattern, the aperture field is also linearly polarized:

$$\mathbf{E}_a = -\ \hat{x}A(\psi)(1 + \cos\psi)\frac{\exp(-jk2f)}{2f} \tag{9}$$

Aperture Fields and Fourier Transforms

It will be shown that fields radiated by a planar antenna from a surface on which currents are specified can be determined by means of aperture integration. In general, the antenna can be represented by both electric and magnetic currents on its aperture plane. Suppose these currents are denoted by $\mathbf{J_s}$ and $\mathbf{M_s}$ on a surface Σ as shown in Fig. 4. These currents can be estimated or calculated from either approximate or rigorous analysis methods. In the far field of the surface Σ on which the currents are located, the radiated fields are given by

DOI 10.1007/978-981-4560-75-7_30-1

$$\mathbf{E} \approx \frac{jk}{4\pi}\frac{e^{-jkr}}{r}\hat{r} \times \int_{\Sigma}\left[\mathbf{M_s} - \eta \mathbf{J_s} \times \hat{R}\right]\exp(jk\hat{r}\cdot\mathbf{r}')dS' \tag{10a}$$

and

$$\mathbf{H} = \frac{1}{\eta_o}\hat{r} \times \mathbf{E} \tag{10b}$$

where $\hat{r}$ is a unit vector in the radial direction, $\hat{R}$ is a unit vector from the source point (in primed coordinates) to the far-field point as shown in Fig. 4. $\eta_o = \sqrt{\mu_o/\varepsilon_o}$ is the free-space wave impedance, where μ_o and ε_o are the permeability and permittivity of free space, respectively. Equation 10b implies that the field radiated by the aperture is a wave with a spherical wave front. In the far-field region, the field is polarized tangential to the radiation sphere since there are no radial components (i.e., $E_r = 0 = H_r$). For a plane aperture coincident with the x-y plane, the surface currents are given by

$$\mathbf{J_s} = \hat{n} \times \mathbf{H_a} \text{ and } \mathbf{M_s} = -\hat{n} \times \mathbf{E_a} \tag{11}$$

where ($\mathbf{E_a}$, $\mathbf{H_a}$) are the total fields on the aperture. From Eq. 9a, the radiated field is

$$\mathbf{E} \approx -\frac{jk}{4\pi}\frac{e^{-jkr}}{r}\hat{r} \times \int_{A}[\hat{z} \times \mathbf{E_a} + \eta(\hat{z} \times \mathbf{H_a}) \times \hat{r}]\exp(jk\hat{r}\cdot\mathbf{r}')dS'. \tag{12}$$

Using the standard vector relationships between rectangular and spherical coordinate systems, namely,

$$\hat{x} = \hat{r}\sin\theta\cos\phi + \hat{\theta}\cos\theta\cos\phi - \hat{\phi}\sin\phi$$

$$\hat{y} = \hat{r}\sin\theta\sin\phi + \hat{\theta}\cos\theta\sin\phi + \hat{\phi}\sin\phi$$

$$\hat{z} = \hat{r}\cos\theta - \hat{\theta}\sin\theta$$

from which Eq. 31 can be expressed in spherical components as

$$E_r = 0 \tag{13a}$$

$$E_\theta \approx \frac{jk}{4\pi}\frac{e^{-jkr}}{r}\left[\left(N_x\cos\phi + N_y\sin\phi\right) - \eta_o\cos\theta\left(L_x\sin\phi + L_y\cos\phi\right)\right] \tag{13b}$$

$$E_\phi \approx \frac{jk}{4\pi}\frac{e^{-jkr}}{r}\left[\cos\theta\left(-N_x\sin\phi + N_y\cos\phi\right) + \eta_o\left(L_x\cos\phi + L_y\sin\phi\right)\right] \tag{13c}$$

where N_x, N_y, L_x and L_y are the components of the vectors

$$\mathbf{N}(u,v) = \int_{A}\mathbf{E_a}(x',y')\exp(j2\pi(ux' + vy'))dx'dy' \tag{14a}$$

DOI 10.1007/978-981-4560-75-7_30-1

$$\mathbf{L}(u,v) = \int_A \mathbf{H_a}(x',y')\exp(j2\pi(ux' + vy'))dx'dy'. \tag{14b}$$

The functions

$$u = \frac{1}{\lambda}\sin\theta\cos\phi \tag{15a}$$

and

$$v = \frac{1}{\lambda}\sin\theta\sin\phi \tag{15b}$$

are variables in the u-v plane, which is often preferred over the standard spherical angles for presenting the far fields in two dimensions. It can be seen from Eq. 14 that the components of **N** and **L** are two-dimensional Fourier transforms of the aperture field components between the x-y and u-v planes. When the magnetic field in the aperture is related to the electric field through Eq. 4b, **L** is related to **N** as follows:

$$\mathbf{L} = \frac{1}{\eta_o}\hat{z} \times \mathbf{N}.$$

Consequently, Eqs. 13b and 13c simplify to

$$E_\theta \approx \frac{jk}{4\pi}\frac{e^{-jkr}}{r}(1 + \cos\theta)\left(N_x\cos\phi + N_y\sin\phi\right) \tag{16a}$$

and

$$E_\phi \approx \frac{jk}{4\pi}\frac{e^{-jkr}}{r}(1 + \cos\theta)\left(-N_x\sin\phi + N_y\cos\phi\right) \tag{16b}$$

Through Eqs. 16 or 13, it is apparent that the far-zone fields are related to the Fourier transforms of the aperture field distributions. Some relationships for simple aperture distributions are illustrated in Fig. 5.

Physical Optics Method

A powerful approach to analysing radiation from reflectors involves estimating the currents on the actual reflector surfaces and by means of Maxwell's equations transforming these into the far field. The field expressions given in the previous section are modified to suit the change in summation of the elementary currents – for a reflector, either the main or a subreflector, which are usually large in terms of wavelengths. A good approximation is to assume that at any point on the reflector, the current that is induced is the same as if for an infinite plane conductor. This principle is called physical optics (PO). Relevant to the previous equations, on a conducting surface there are no magnetic currents.

If the feed radiates a magnetic field $\mathbf{H_f}$ at the reflector surface, the surface current is given by

$$\mathbf{J_s} = 2\hat{n} \times \mathbf{H_f}|_{reflector\Sigma} \tag{17}$$

DOI 10.1007/978-981-4560-75-7_30-1

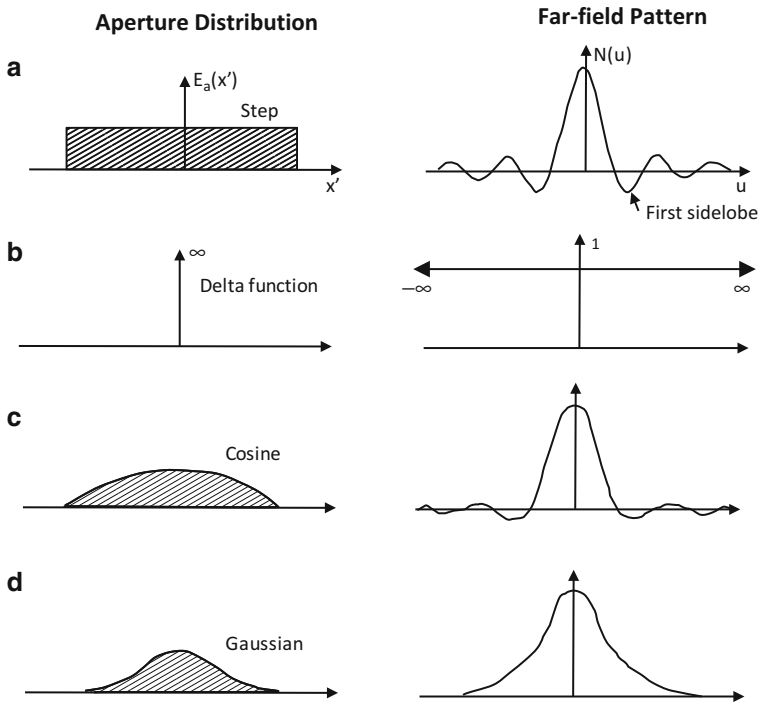


Fig. 5 Fourier transform relationships for simple aperture distributions and the resulting transforms. (**a**) Unit step; (**b**) delta function; **c** cosine; and Gaussian distributions.

where $\hat{n}$ is the unit outward normal to the reflector surface Σ. The factor of 2 appears in Eq. 17 because the total magnetic field at a plane conductor is twice the incident field.

Equation 17 may be used in Eq. 10 to calculate the radiated fields. There is no magnetic surface current on the surface Σ, and as a result, the radiated electric field can be rewritten as

$$\mathbf{E}(r,\theta,\phi) = -\frac{jk\eta_o}{4\pi}\frac{e^{-jkr}}{r}[\mathbf{F}(\theta,\phi) - \hat{r}(\mathbf{F}(\theta,\phi)\cdot\hat{r})] \tag{18}$$

where

$$\mathbf{F}(\theta,\phi) = \int_{\Sigma} \mathbf{J_s}\exp(jk\,\hat{r}\cdot\mathbf{r}')dS'. \tag{19}$$

The primed coordinates in this instance refer to the source components on the reflector surface. The purpose of the second term inside the square brackets of Eq. 18 is to cancel out a radial vector component produced by the first term because there is no radial component in the far-field region.

Introduction to Reflector Antennas

With some fundamentals in hand and awareness of some basic reflector geometries in Fig. 2, some of the basic properties of the radiation from reflector antennas will now be investigated. Before commencing this, there are some standard terms that require definition.

DOI 10.1007/978-981-4560-75-7_30-1

Elementary Terminology

Feed

A basic component of a reflector is the feed antenna. The aim is to illuminate the reflector to give the desired energy distribution because as the Fourier transform relations show, if the illumination is highly tapered, the antenna beam is broader than when the illumination is uniform. In addition, the amplitude of the first sidelobe is higher for a uniform distribution compared with a tapered one. When the phase has a gradient across the aperture, the beam will be steered at some angle from boresight.

The most common type of feed is a waveguide or horn although the dipole is sometimes preferred at frequencies below about 1 GHz. Without some type of reflector or cup at the rear of the dipole, a significant amount of power is radiated as a backlobe, which combines with the radiation of the reflector to fill in the sidelobe nulls and also raises the levels of the odd-numbered sidelobe levels. The horn has low levels of rear radiation and can have a low input reflection coefficient over a broad range of frequencies.

Beamwidth and Radiation Patterns

The beamwidth of a reflector antenna is very strongly dependent upon the feed illumination over its surface. Based on Fourier relationships, a circular aperture of diameter D that is uniformly illuminated has a half-power beamwidth (HPBW) approximately equal to $1.02/(D/\lambda)$. With a $\left(1-(2r/D)^2\right)^p$ aperture distribution where r is the distance from the aperture centre, the HPBW increases for $p \geq 1$ as $\approx \left[1.27^{(p+1)/2}\right](\lambda/D)$. The radiation patterns are usually presented in the form of pattern cuts through the far-field sphere for the radiation from a reflector that has two planes of symmetry. For a reflector producing a shaped beam over a surface, it is usual to make a two-dimensional plot of the radiation. For a reflector excited by a feed with radiation given by Eq. 25, the radiation patterns in the two principal planes corresponding to the planes parallel to the electric field polarization, the E-plane, and also the pattern cut orthogonal to it, the H-plane, are of interest. Also of interest is the cross-polarized radiation in the

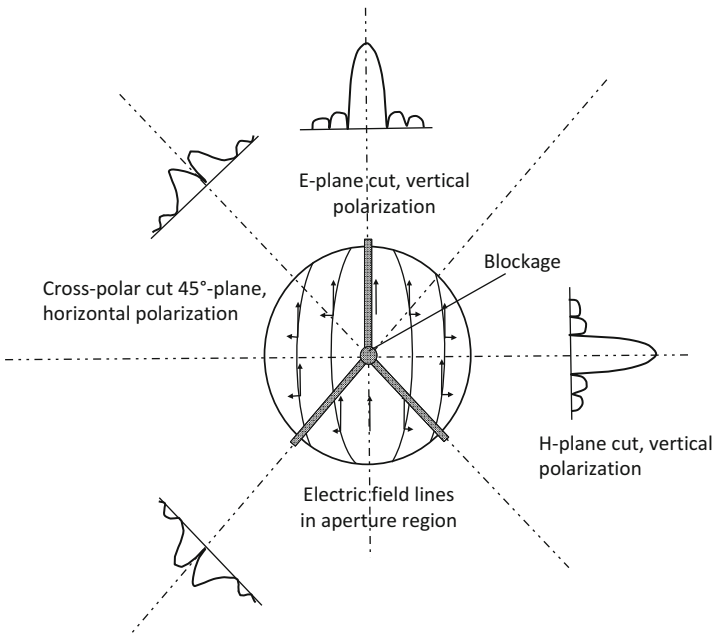


Fig. 6 Relationship of aperture field to radiation pattern cuts.

DOI 10.1007/978-981-4560-75-7_30-1

45-degree planes (see Fig. 6). How these patterns are defined shall be described in more detail later in this section.

Reflector Antenna Gain and Efficiency

Antenna gain is defined as the ratio of the power received at (θ,ϕ) if this same power were radiated isotropically as compared with the total input power. Thus, when P_r is the power density radiated at (θ,ϕ), and P_T is the total input power; the gain at (θ,ϕ) is defined as

$$G(\theta,\phi) = \frac{4\pi r^2 P_r}{P_T}. \tag{20a}$$

Furthermore, the power density in the far field is given by $P_r = |\mathbf{E}\cdot\mathbf{E}|/2\eta_o$ where the electric field is $\mathbf{E}(r,\theta,\phi) = \left(\hat{\theta}E_\theta(\theta,\phi) + \hat{\phi}E_\phi(\theta,\phi)\right)\exp(-jkr)/r$. Also in the far field, the magnetic field is given by $\mathbf{H}(\theta,\phi) = (\hat{r}\times\mathbf{E})/\eta_o$. Therefore, Eq. 20a reduces to

$$G(\theta,\phi) = \frac{2\pi\left(|E_\theta|^2 + |E_\phi|^2\right)}{\eta_o P_T}. \tag{20b}$$

Maximum gain occurs at directions close to the axis either in the boresight direction or in a steered direction at some angle not far from it.

Under conditions of uniform aperture and phase across the aperture, a reflector of diameter D has a maximum gain of

$$G_o = \left(\frac{\pi D}{\lambda}\right)^2 \tag{21}$$

The reflector illumination is usually tapered in order to keep sidelobes to acceptably low levels. As a result, the gain is less than that given by Eq. 21. To include illumination tapering, the maximum gain is usually defined as

$$G_{\max} = \eta_a G_o \tag{22}$$

where η_a is the aperture efficiency as compared with uniform illumination.

Equation 22 neglects losses in the reflector system due to feed spillover, mismatch, reflector surface errors, etc. These power losses can be accommodated in Eq. 20 by introducing efficiency factors for each loss mechanism. Initially, consider one of these losses called feed spillover. The power density radiated by a feed is

$$P_f = \frac{1}{2\eta}|\mathbf{E_f}|^2.$$

which results in a total radiated power

$$P_T = \int_0^{2\pi} d\xi \int_0^{\pi} P_f \rho \sin\psi d\psi. \tag{23}$$

DOI 10.1007/978-981-4560-75-7_30-1

Not all the power in Eq. 23, however, is intercepted by the reflector. Some of it falls outside the reflector causing power loss. This power loss is called spillover. The power collected by a reflector subtending an angle ψ_c is

$$P_c = \frac{1}{2\eta}\int_0^{2\pi} d\xi \int_0^{\psi_c} |\mathbf{E_f}|^2 \rho \sin\psi d\psi. \tag{24}$$

By conservation of power, the power lost due to spillover is

$$P_s = P_T - P_c = \frac{1}{2\eta}\int_0^{2\pi} d\xi \int_{\psi_c}^{\pi} |\mathbf{E_f}|^2 \rho \sin\psi d\psi.$$

The spillover efficiency is defined as the ratio of the power that is collected by the reflector and the total radiated power ie.

$$\eta_s = \frac{P_c}{P_T} = 1 - \frac{P_s}{P_T}. \tag{25}$$

By means of Eqs. 23 and 24, this is expressed as

$$\eta_s = \frac{\int_0^{2\pi} d\xi \int_0^{\psi_c} |\mathbf{E_f}|^2 \rho \sin\psi d\psi}{\int_0^{2\pi} d\xi \int_0^{\pi} |\mathbf{E_f}|^2 \rho \sin\psi d\psi}. \tag{26}$$

Ideally, η_s should be close to 1; typically, it has a value around 80–90 %. The maximum antenna gain given by Eq. 22 now becomes

$$G_{\max} = \eta_a \eta_s \left(\frac{\pi D}{\lambda}\right)^2 = \eta_a \eta_s G_o. \tag{27}$$

In the same way, efficiency factors may also be defined for other losses. Let

$\eta_f =$ feed mismatch efficiency
$\eta_c =$ conductor efficiency
$\eta_r =$ reflector surface roughness efficiency

These efficiencies may be included in the gain function in much the same way as described for spillover. To include all such losses, it is convenient to define an overall efficiency factor as the product of all efficiencies as

$$\eta_T = \eta_a \eta_s \eta_f \eta_c \eta_r. \tag{28}$$

Finally, the maximum antenna gain is expressed as

$$G_{\max} = \eta_T G_o. \tag{29}$$

DOI 10.1007/978-981-4560-75-7_30-1

Gain will be discussed in connection with specific reflector geometries in the following sections.

Edge Taper and Edge Illumination

As the feed illuminates the reflector from some distance, depending on the type of reflector, this illumination can vary with angle. For example, for a paraboloid, the distance from the focus to the reflector edge is greater than to the centre of the reflector. As has been discussed, the radiation pattern is closely related to the aperture illumination and in particular the tapering. It is, therefore, common practice to refer to the level of the field at the edge of the reflector as a rule of thumb for the antenna beamwidth and sidelobe levels. Two different terms are used, often interchangeably.

1. Edge illumination (i.e., illumination at the edge) is the ratio of the field strength radiated by the feed in the direction of the edge and its level at the reflector vertex when both are measured on the surface of a sphere (of radius given by the distance from the focus to the vertex).
2. Edge taper is the ratio of the feed field strength at the actual reflector edge and the corresponding level at the reflector vertex. The difference between edge illumination and edge taper is the freespace loss due to the additional distance δ from the sphere of radius f to the reflector rim. If E is the edge illumination, by definition edge taper, T, is

$$T = E/L_e$$

where $L_e = (f+\delta)/f$ is the edge taper loss. For a paraboloid, $\delta = f/(4f/D)^2$ and, therefore, $L_e = 1 + \left[1/(4f/D)^2\right]$. In dB, the relationship is

$$\text{Edge taper (dB)} = \text{Edge illumination(dB)} - \text{L}_\text{e}\text{(dB)}. \tag{30}$$

For example, a paraboloid with $f/D = 0.35$ is illuminated by a feed with an edge illumination of -10 *dB;* the edge taper is -13.58 *dB*.

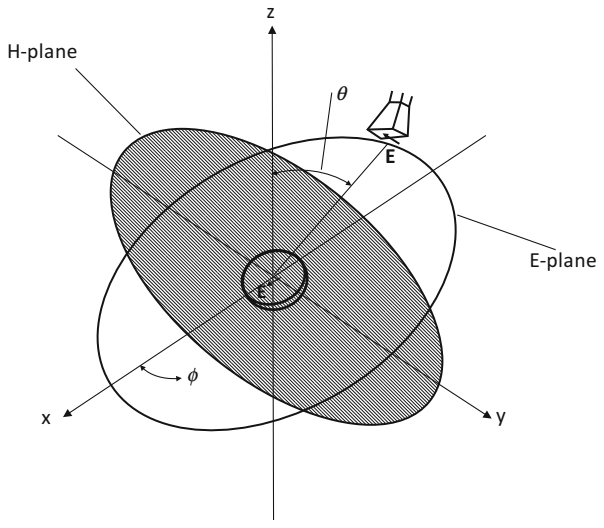


Fig. 7 Principal plane radiation pattern cuts

DOI 10.1007/978-981-4560-75-7_30-1

Co- and Cross-Polarized Radiation Patterns

The far fields of an antenna are tangential to the surface of a sphere centred on the antenna since $E_r = 0$ as illustrated in Fig. 7. Thus, the remaining field components on the surface are aligned in the $\hat{\theta}$ and $\hat{\varphi}$ directions. The field is linearly polarized if the components, E_θ and E_ϕ, are everywhere in constant phase difference in time. When the field components are $\pm 90^\circ$ out of phase in time, the field is circularly polarized. The field is elliptically polarized for an arbitrary phase difference.

The polarization generally varies over the surface of the far-field sphere, and as a consequence, it is important to control the polarization in the main directions required for the application. The principal plane cuts in the E- and H-directions, shown in Fig. 7, usually provide sufficient information about a reflector and/or feed that have two planes of symmetry. Furthermore, it is useful to understand the field components relative to the chosen field polarization. To do this, consider a field given by **E**. Let a unit vector in the direction of the reference polarization, or the co-polarized component, be **p,** and let a unit vector in the orthogonal or cross-polarized direction be **q**. As a result,

$E_p = \mathbf{E} \cdot \mathbf{p}$ – co-polarxized component of the electric field

$E_q = \mathbf{E} \cdot \mathbf{q}$ – cross-polarized component of the electric field.

With this definition of the **p** and **q** vectors, the two components of **E** on the surface are expressed as

$$\begin{bmatrix} E_p(\theta) \\ E_q(\theta) \end{bmatrix} = \begin{bmatrix} \cos(\phi - \phi_o) & \sin(\phi - \phi_o) \\ \sin(\phi - \phi_o) & -\cos(\phi - \phi_o) \end{bmatrix} \cdot \begin{bmatrix} E_\theta(\theta, \phi) \\ E_\phi(\theta, \phi) \end{bmatrix} \tag{31}$$

where ϕ_o is the reference direction. For example, if the reference polarization is oriented at $\phi_o = 0^\circ$, the E-plane corresponds to the $\phi = 0^\circ$ plane where

$$E_p(\theta) = E_\theta(\theta, 0) \text{and} E_q = -E_\phi(\theta, 0) \tag{32a}$$

while the orthogonal H-plane corresponds to the $\phi = 90^\circ$ plane where

$$E_p(\theta) = E_\phi\left(\theta, \frac{\pi}{2}\right) \text{and} E_q = E_\theta\left(\theta, \frac{\pi}{2}\right). \tag{32b}$$

The components of the field in the $\phi = 45^\circ$ plane relative to the reference are

$$E_p(\theta) = \frac{1}{\sqrt{2}}\left[E_\theta\left(\theta, \frac{\pi}{4}\right) + E_\phi\left(\theta, \frac{\pi}{4}\right)\right] \tag{33a}$$

$$E_q(\theta) = \frac{1}{\sqrt{2}}\left[E_\theta\left(\theta, \frac{\pi}{4}\right) - E_\phi\left(\theta, \frac{\pi}{4}\right)\right]. \tag{33b}$$

The E-plane pattern is measured or calculated by aligning the reference or test antenna with the electric field and then rotating the antenna under test (AUT) in that plane whilst measuring the received signal. The H-plane pattern is found by rotating the AUT in the plane orthogonal to the reference but with the AUT aligned with the reference. Similarly, the co-polar pattern in the 45 ° plane is obtained by aligning the reference parallel to the electric field while the AUT is rotated in the 45 ° plane. The cross-polarization in

DOI 10.1007/978-981-4560-75-7_30-1

an arbitrary $\phi -$ plane is found by initially rotating the AUT to coincide with the polarization of the reference. Then the reference is rotated by 90 °, and the pattern is then measured in the $\phi -$ plane.

Paraboloidal Reflector

Radiation from a Paraboloid

The aperture field predicted by geometric optics is given by Eq. 8. The field radiated by this distribution is obtained by substituting this field into Eq. 12 and carrying out the Fourier transform. This can be done numerically using quadrature methods such as Simpson's rule or by means of a two-dimensional fast Fourier transform. In making use of the latter, only the fields on the aperture should be used as input. The rest of the matrix should be filled with zeros. If a grid point is outside the aperture by less than a quarter of a wavelength, a non-zero estimate could be used which has a value that is half the value on the field on the rim of the aperture. If the feed functions F_ψ and F_ξ are integrable functions, the integral may be evaluated in closed form from Eqs. 16 and 14a. The transform in this case is

$$\mathbf{N}(\theta, \phi) = \int_0^{2\pi} d\xi \int_0^{D/2} \mathbf{E_a}(t, \xi) \exp(jwt \cos(\phi - \xi))\, t\, dt. \tag{34}$$

$w = k \sin\theta, t = \rho \sin\psi$ and $\rho = 2f/(1 + \cos\psi) = f \sec^2\psi/2$, for a paraboloid.

Consider initially the aperture field given by Eq. 9. The multiplier corresponds to an illumination function $F(\psi) = A(\psi)(1 + \cos\psi)$. Suppose also $A(\psi)$ has an inverse taper so that the aperture illumination is uniform, i.e., $F(\psi) = 1$. Equation 34 gives

$$N_x(\theta, \phi) = -\frac{\exp(-jk2f)}{2f} \int_0^{2\pi} d\xi \int_0^{D/2} \exp(jwt \cos(\phi - \xi))\, t\, dt.$$

Making use of the Bessel identity

$$\int_0^{2\pi} \begin{matrix}\cos \\ \sin\end{matrix} p\phi' \Bigg\} e^{jz\cos(\phi - \phi')} d\phi' = 2\pi j^p J_p(z) \Bigg\{ \begin{matrix}\cos \\ \sin\end{matrix} p\phi \tag{35}$$

where $J_p(z)$ is an ordinary Bessel function of order p with argument z, the transform simplifies to

$$N_x(\theta, \phi) = -\pi \frac{\exp(-jk2f)}{f} \int_0^{D/2} J_0(wt) t\, dt.$$

Since $\int_0^1 J_0(at) = J_1(a)/a$ then

$$N_x(\theta, \phi) = -\pi \left(\frac{D}{2}\right)^2 \frac{\exp(-jk2f)}{f} \frac{J_1(X)}{X} \tag{36}$$

where $X = \frac{\pi D}{\lambda} \sin\theta$. A plot of the electric field radiated by the paraboloid with uniform illumination is shown in Fig. 8. The first sidelobe level occurs at −17.6 dB relative to the peak, and the beamwidth is $1.02\lambda/D$.

DOI 10.1007/978-981-4560-75-7_30-1

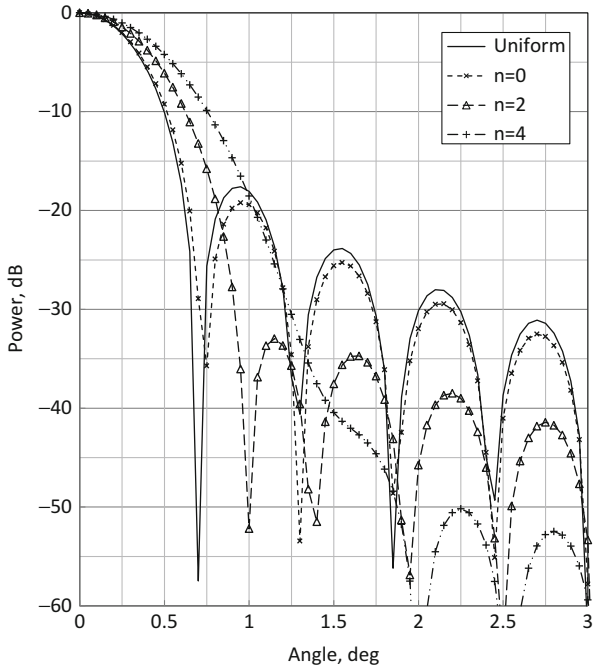


Fig. 8 Radiation patterns of a paraboloid with $D = 100\ \lambda$ and $f/D = 0.4$ with a feed with a symmetric pattern. Order of feed taper is n. Also shown is the pattern for uniform aperture illumination

A tapered illumination is considered next. Suppose $A(\psi) = \cos^n\psi = B(\psi)$ where n can have arbitrary values. Large positive values of n correspond to higher edge taper. Negative values correspond to inverse tapers. Following from Eq. 9

$$N_x(\theta,\phi) = -\pi\frac{\exp(-jk2f)}{f}\int_0^{D/2} J_0(wt)\cos^n\psi(1+\cos\psi)t\,dt$$

where $t = \rho\sin\psi = 2f\tan(\psi/2)$. Therefore, $dt = f\sec^2(\psi/2)d\psi$. Let $\psi_c = 2\tan^{-1}(D/4f)$. Then

$$N_x(\theta,\phi) = -2\pi f\exp(-jk2f)\int_0^{\psi_c} J_0(2f\,w\tan[\psi/2])\sin\psi\cos^n\psi\sec^2(\psi/2)d\psi. \tag{37}$$

Also, making a further substitution, $u = \tan(\psi/2)$ where $du = (1/2)\sec^2(\psi/2)\,d\psi$, and also $\sin\psi = 2u/(1+u^2)$ and $\cos\psi = (1-u^2)/(1+u^2)$. Therefore,

$$N_x(\theta,\phi) = -8\pi f\exp(-jk2f)\int_0^{D/4f} J_0(2f\,w\,u)\left(\frac{1-u^2}{1+u^2}\right)^n\frac{u\,du}{(1+u^2)}. \tag{38}$$

The integral can be evaluated by replacing the zero-order Bessel function with its series and then integrating term by term. However, it is usually more convenient to evaluate the integral numerically. Plots of the far field obtained from Eqs. 38 and 36 are shown in Fig. 8 for a 100 λ diameter dish with $f/D = 0.4$ for a feed with tapers defined by $n = 0$, 2 and 5. These results are compared with uniform aperture illumination of a circular aperture. As the taper is increased, the beamwidth increases, and the sidelobes reduce. The first sidelobe level for the first two cases $n = 0$ and 2 is, respectively, -19.2 dB and -34.0 dB. For $n = 4$, the widened main beam and the first sidelobe have merged. The radiation patterns of a paraboloid with specific feed antenna types will be discussed in the next section.

DOI 10.1007/978-981-4560-75-7_30-1

A paraboloidal reflector and a feed with an axisymmetric pattern radiates power at boresight direction given by

$$P_r = \frac{2}{\eta}\left(\frac{kf}{r}\right)^2 \left| \int_0^{\psi_c} A(\psi) \tan\frac{\psi}{2} d\psi \right|^2 \tag{39}$$

The substitution $t = \rho \cos\psi$ was made in Eq. 23 to obtain this equation. The power collected by the reflector from the feed is

$$P_c = \frac{\pi}{\eta} \int_0^{\psi_c} |A(\psi)|^2 \sin\psi d\psi. \tag{40}$$

Let the maximum gain be $G_{\max} = \eta_a \eta_s G_o$. That is,

$$G_{\max} = \eta_s \left(\frac{\pi D}{\lambda}\right)^2 2\cot^2\frac{\psi}{2} \frac{\left| \int_0^{\psi_c} A(\psi) \tan\frac{\psi}{2} d\psi \right|^2}{\int_0^{\psi_c} |A(\psi)|^2 \sin\psi \, d\psi} \tag{41}$$

The aperture efficiency is given by

$$\eta_a = 2\cot^2\frac{\psi}{2} \frac{\left| \int_0^{\psi_c} A(\psi) \tan\frac{\psi}{2} d\psi \right|^{12}}{\int_0^{\psi_c} |A(\psi)|^2 \sin\psi d\psi} \tag{42}$$

Blockage due to the feed and its supporting struts can be included in an approximate manner by altering the limits of integration. For example, feed blockage of diameter a is introduced in Eq. 34 by changing the lower limit of integration. Thus,

$$\begin{aligned} \mathbf{N}(\theta,\phi) &= \int_0^{2\pi} d\xi \int_{a/2}^{D/2} \mathbf{E_a}(t,\xi) \exp(jwt\cos(\phi-\xi)) t\, dt \\ &= \int_0^{2\pi} d\xi \left\{ \int_0^{D/2} - \int_0^{a/2} dt\, t \right\} \mathbf{E_a}(t,\xi) \exp(jwt\cos(\phi-\xi)). \end{aligned} \tag{43}$$

This shows that the vector transform is the difference of two transforms. The domain of one transform is over the full diameter, and the other domain is only over the blocked area. Therefore, for a uniform illumination,

$$N_x(\theta,\phi) = -\pi \frac{\exp(-jk2f)}{f} \left[\left(\frac{D}{2}\right)^2 \frac{J_1(X)}{X} - \left(\frac{a}{2}\right)^2 \frac{J_1(X')}{X'} \right] \tag{44}$$

where $X' = \frac{\pi a}{\lambda} \sin\theta$. The field radiated by a 100 λ diameter dish with $f/D = 0.4$ and central blockages of 5 λ and 10 λ are shown in Fig. 9. Also shown is the effect of blockage on a symmetric tapered feed with

DOI 10.1007/978-981-4560-75-7_30-1

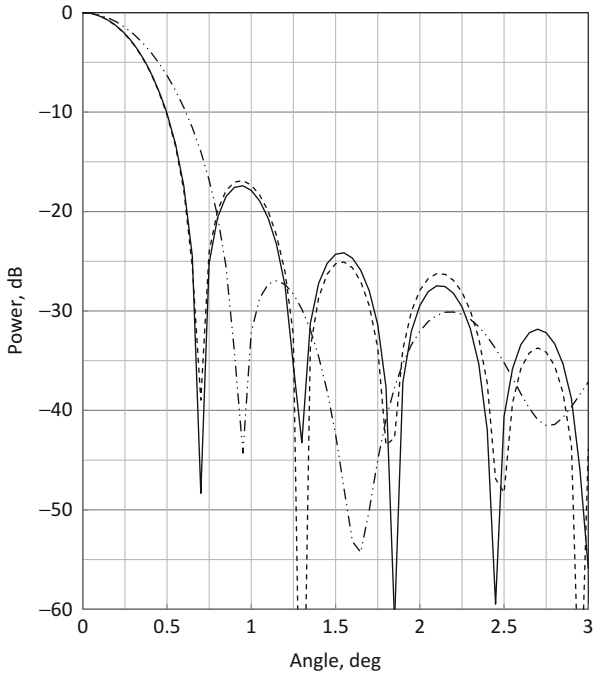


Fig. 9 Radiation patterns of a $D = 100\ \lambda$ paraboloid with central blockage. The reflector has $f/D = 0.4$ and the aperture is uniformly illuminated. *Solid line*: blockage diameter $a = 5\ \lambda$ and uniform illumination; *short dashed line*: blockage diameter $a = 10\ \lambda$; *long dashed line*: blockage diameter $a = 5\ \lambda$ with tapered feed $n = 2$

$n = 2$. Comparing this figure with Fig. 8, it is seen that central blockage has had a major impact on the radiation pattern of the reflector with a tapered feed, indicating that low sidelobes can be badly affected by blockage.

Struts can be handled in an approximate manner using a similar technique. Suppose that a feed is supported by three struts that are separated by 120° apart. These struts can be approximated by angular segments of width 2Δ. Then,

$$\begin{aligned}\mathbf{N}(\theta,\phi) &= \left\{\sum_{m=1}^{3}\int_{2(m-1)\pi/3+\Delta}^{2m\pi/3-\Delta} d\xi\right\}\int_0^{D/2}\mathbf{E_a}(t,\xi)\exp[jwt\cos(\phi-\xi)]t\,dt\\ &= \int_0^{2\pi} d\xi\int_0^{D/2}\mathbf{E_a}(t,\xi)\exp[jwt\cos(\phi-\xi)]t\,dt\\ &\quad-\left\{\sum_{m=1}^{3}\int_{2m\pi/3-\Delta}^{2m\pi/3+\Delta} d\xi\right\}\int_0^{D/2}\mathbf{E_a}(t,\xi)\exp(jwt\cos(\phi-\xi))t\,dt.\end{aligned}\tag{45}$$

With an axisymmetric illumination function, the integral over ξ can be expressed in closed form as

$$\begin{aligned}I &= \left\{\sum_{m=1}^{3}\int_{2m\pi/3-\Delta}^{2m\pi/3+\Delta} d\xi\right\}\int_0^{D/2}\mathbf{E_a}(t,\xi)\exp(jwt\cos(\phi-\xi))t\,dt\\ &= 2\Delta\sum_{m=1}^{3}\sum_{k,\,l=-\infty}^{\infty} j^k\exp\left[\frac{2m\pi}{3}(l-k)\right]S[\Delta(l-k)]\int_0^{D/2}\mathbf{E_a}(t,\xi)J_k(wt\cos\phi)J_l(wt\sin\phi)t\,dt\end{aligned}\tag{46}$$

DOI 10.1007/978-981-4560-75-7_30-1

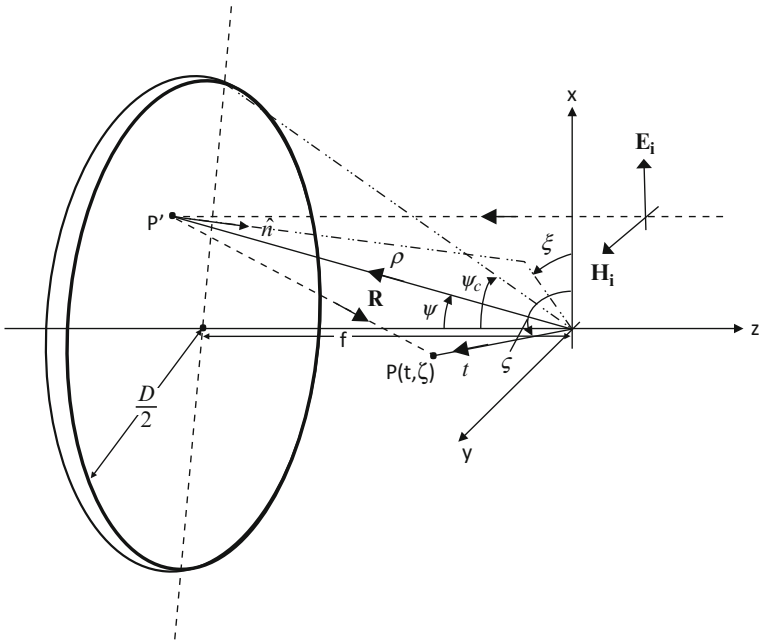


Fig. 10 Focal region of a paraboloid

where $S(x) = \sin x/x$. In most cases, it is preferable to evaluate the integrals in Eq. 43 directly by numerical means rather than by Eq. 45.

Focal Region Fields of a Paraboloid

The fields in the vicinity of the focus of a reflector are of particular importance as they indicate the most appropriate excitation from the feed for a desirable excitation and hence highest efficiency. The focal fields can be calculated from geometric optics or from physical optics. In both cases, the reflector is illuminated by a plane wave, and the field arriving at the focus is determined by each method. The PO method will be used here. An incident plane wave is assumed to propagate towards the reflector parallel to the z-axis as illustrated in Fig. 10. The incident fields are

$$\mathbf{E_i} = \hat{x} E_o \exp(jkz) \tag{47}$$

and

$$\mathbf{H_i} = \frac{1}{\eta_o} \hat{z} \times \mathbf{E_i}.$$

An electric current induced on the paraboloid is

$$\mathbf{J_s} \approx \hat{n} \times \mathbf{H_i} = \frac{2E_o}{\eta_o} (\hat{n} \times \hat{y}) \exp(jkz) \Big|_{\text{reflector } \Sigma}. \tag{48}$$

Introducing the normal to the paraboloid, in rectangular co-ordinates the surface current is

$$\mathbf{J_s} = \frac{2E_o}{\eta_o} e^{jkz} \left(\hat{x} \cos \frac{\psi}{2} + \hat{z} \sin \frac{\psi}{2} \cos \xi \right). \tag{49}$$

DOI 10.1007/978-981-4560-75-7_30-1

The fields in the focal region (at $z = 0$) are obtained by means of the PO expression Eq. 10 where the surface of integration is the reflector and in this instance there are no magnetic currents.

Although the focal region is not always in the far field of the reflector, approximations to the scattered field included in Eq. 10a are applicable. As shown in Fig. 10, the vector $\mathbf{R} = \mathbf{t} - \boldsymbol{\rho}$ is a vector from a point P′ on the reflector to P in the focal region. For a large reflector, in the vicinity of the focus, consider distances $|\mathbf{t}| << |\boldsymbol{\rho}|$. Further, let $R = |\mathbf{R}| \approx |\boldsymbol{\rho}| = \hat{\rho} \cdot \mathbf{t}$ in the phase function of the integrand of Eq. 10a and $\hat{R} \approx \hat{\rho}$ in the amplitude function of the integrand. Thus, Eq. 10a can be expressed as

$$\mathbf{E}_F(t,\zeta) \approx -\frac{jk\eta_o}{4\pi}\int_\Sigma [\mathbf{J_s} - (\hat{\rho}\cdot\mathbf{J_s})\hat{\rho}]\frac{e^{-jk\rho}}{\rho}\exp(jk\hat{\rho}\cdot\mathbf{t})dS'. \tag{50}$$

For a paraboloid, the surface element is given by $dS' = \rho^2 \sec(\psi/2)\sin\psi d\psi d\xi$.

The amplitude function of Eq. 50 is expanded as follows:

$$[\mathbf{J_s} - \hat{\rho}(\mathbf{J_s}\cdot\hat{\rho})]\rho\, d\psi d\xi = -\frac{2fE_o}{\eta_o}e^{jkz}\hat{x}\left(1 - \tan^2\frac{\psi}{2}\cos 2\xi\right)\sin\psi - \hat{y}\sin\psi\tan^2\frac{\psi}{2}\sin 2\xi + \hat{z}2\sin\psi\tan\frac{\psi}{2}\cos\xi. \tag{51}$$

The phase function inside the integrand is given by

$$\begin{aligned}\exp[jk(-\rho + z + \hat{\rho}\cdot\mathbf{t})] &\approx \exp[jk(-\rho(1+\cos\psi) + (x_F\cos\xi + y_F\sin\xi)\sin\psi)] \\ &= \exp[jk(-2f + t\sin\psi\cos(\zeta - \xi))]\end{aligned} \tag{52}$$

where, as shown in Fig. 10, $x_F = t\cos\zeta$ and $y_F = t\sin\zeta$ are rectangular coordinates in the focal plane on which (t,ζ) are the polar coordinates. With these approximations and substitutions, the components of Eq. 50 become

$$E_{Fx} = -\frac{jkfE_o}{2\pi}e^{-jk2f}\int_0^{2\pi}d\xi\int_0^{\psi_c}\left(1 - \tan^2\frac{\psi}{2}\cos 2\xi\right)\sin\psi\exp[jkt\sin\psi\cos(\zeta - \xi)]d\psi \tag{53a}$$

$$E_{Fy} = \frac{jkfE_o}{2\pi}e^{-jk2f}\int_0^{2\pi}d\xi\int_0^{\psi_c}\sin\psi\tan^2\frac{\psi}{2}\sin 2\xi\exp[jkt\sin\psi\cos(\zeta - \xi)]d\psi \tag{53b}$$

$$E_{Fz} = -\frac{jkfE_o}{\pi}e^{-jk2f}\int_0^{2\pi}d\xi\int_0^{\psi_c}\sin\psi\tan\frac{\psi}{2}\cos\xi\exp[jkt\sin\psi\cos(\zeta - \xi)]d\psi. \tag{53c}$$

The integration with respect to ξ can be completed by means of the integral identity Eq. 35, which allows Eq. 53 to be simplified to

$$E_{Fx}(t,\zeta) = \Omega_0(t) + \Omega_2(t)\cos 2\zeta \tag{54a}$$

$$E_{Fy}(t,\zeta) = \Omega_2(t)\sin 2\zeta \tag{54b}$$

$$E_{Fz}(t,\zeta) = -2j\Omega_1(t)\cos\zeta \tag{54c}$$

wherein

DOI 10.1007/978-981-4560-75-7_30-1

$$\Omega_n(t) = \kappa \int_0^{\psi_c} J_n(kt\,\sin\psi)\tan^n\frac{\psi}{2}\,\sin\psi d\psi \tag{55}$$

and

$$\kappa = -jkf\,E_o e^{-jk2f}.$$

For a long–focal length paraboloid, Eq. 54 can be simplified further by virtue of the small angle to the rim, ψ_c, leading to

$$\Omega_0(t) \approx 2\kappa\,\sin^2\frac{\psi_c}{2}\left[2\frac{J_1(U)}{U}\right] \tag{56a}$$

$$\Omega_1(t) \approx 2\kappa\,\sin^2\frac{\psi_c}{2}\left[\psi_c\frac{J_2(U)}{U}\right] \tag{56b}$$

$$\Omega_2(t) \approx 0 \tag{56c}$$

where $U = kt\sin\psi_c$. Thus, the transverse component of the focal field is

$$E_{Fx}(t,\zeta) \approx -j2kE_o e^{-jk2f}\left[2\frac{J_1(kt\sin\psi_c)}{(kt\sin\psi_c)}\right].$$

The field consists of a series of high- and low-intensity bands known as Airy rings. The latter correspond to the zeros of the $J_1(U)$ Bessel function. The first two of these occur at $U = 3.85$ and 7.02. The peaks occur at the maxima of the function $2J_1(U)/U$. The first two peaks occur at $U = 0$ and $U = 5.14$. A contour plot of the electric field focal region given by Eq. 53 for a paraboloid with diameter $D = 100\ \lambda$ and $f/D = 0.35$ is shown in Fig. 11. The low-intensity rings are clearly evident and are close to the approximations of Eq. 52 ($t = 0.66\ \lambda$ and $1.21\ \lambda$) even though the focal length is relatively short. It can be seen that close to optimum excitation could be achieved with a feed with a uniform aperture distribution and a radius α approximately equal to $0.66\ \lambda$.

Imperfect Reflectors

Up to this point, it has been assumed that the reflector surface is a smooth function with no surface imperfections or misalignments in the structure. In practice, due to manufacturing tolerances and physical variations, a reflector system has some imperfections. In this section, two types of imperfections are considered. The first is due to surface errors, and the second is related to various misalignments that result in aberrations in the radiated pattern.

Surface Errors

Random surface errors, if small, can be considered to add a phase function to the aperture field. Let the aperture field with phase error be approximated by

$$\mathbf{E_a}' = \mathbf{E_a}e^{-j\alpha} \tag{57}$$

DOI 10.1007/978-981-4560-75-7_30-1

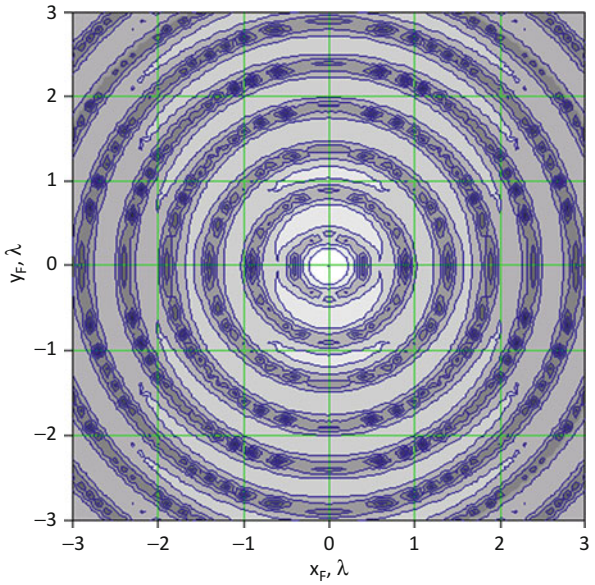


Fig. 11 Focal region field of a paraboloid reflector with diameter $D = 100\ \lambda$ and $f/D = 0.35$

where α is a small random function in the surface profile, and $\mathbf{E_a}$ is the aperture field with no reflector surface errors. Since $\alpha \ll 1$, let $\exp(-j\alpha) \approx (1 - \alpha^2/2) + j\alpha$. For simplicity, consider a paraboloid with a feed having an axisymmetric pattern. With the aperture distribution given by Eq. 57, the aperture efficiency is modified to

$$\eta_a{}' = \frac{1}{2\pi^2}\cot^2\frac{\psi}{2}\frac{\left|\int_0^{2\pi} d\xi \int_0^{\psi_c}\left(1 - \frac{\alpha^2}{2} + j\alpha\right)A(\psi)\tan\frac{\psi}{2}d\psi\right|^2}{\int_0^{\psi_c}|A(\psi)|^2\sin\psi d\psi} \approx \eta_a\left(1 - \overline{\alpha}^2 + (\overline{\alpha})^2\right) \tag{58}$$

where η_a is the aperture efficiency with no surface errors. The quantities $\overline{\alpha}^2$ and $(\overline{\alpha})^2$ are the mean-square phase error and average phase error weighted by the aperture illumination function $A(\psi)\tan(\psi/2)$. The phase reference plane can be chosen arbitrarily so that latter quantity $(\overline{\alpha})^2 = 0$. Therefore,

$$\eta_a{}' \approx \eta_a\left(1 - \overline{\alpha}^2\right). \tag{59}$$

Another more sophisticated model of surface errors has been developed (Ruze 1966), which is valid for large surface errors that are Gaussian distributed. If the errors are completely uncorrelated in small regions of the aperture of dimensions much less than D, the aperture efficiency is

$$\eta_a{}' \approx \eta_a \exp\left(-\overline{\delta}^2\right) \tag{60}$$

where $\overline{\delta}^2$ is the mean square error of the Gaussian distribution. For small errors, Eq. 60 is almost identical to Eq. 59 with the exception that $\overline{\alpha}^2$ is a weighted quantity. A useful indicator for practical

DOI 10.1007/978-981-4560-75-7_30-1

reflectors is the rms surface error ε, which is equivalent to the rms error of a sinusoid of period $2\pi/k = \lambda$, is given by

$$\varepsilon = \frac{\sqrt{\overline{\delta^2}}}{2k} = \frac{\lambda}{4\pi}\sqrt{\overline{\delta^2}}. \tag{61}$$

As an example, a reflector was designed to have an aperture efficiency of 70 % at 30 GHz. When the reflector was fabricated, the rms surface error was found to be 50 μm. Therefore in practice, the effective aperture efficiency is about 66 % and the gain 0.28 dB below what was expected. As a rule of thumb, a surface tolerance as high as $\lambda/30$ may be acceptable in many applications. This results in a gain loss of up to 1.8 dB. In other applications, this loss may not be acceptable, and a smaller surface error, typically $\lambda/50$, may be required. This error corresponds to a gain loss of about 1 dB.

Aberrations

The radiation pattern of an aperture antenna is sensitively dependent upon the phase distribution of the aperture field. Variation from an ideal uniform phase distribution can cause aberrations. These aberrations can also occur inadvertently during antenna design and manufacture. Sometimes they arise intentionally when, for example, beam shaping or for beam steering. Due to aberrations, the ideal aperture field $\mathbf{E_a}$ can be considered to have a phase distribution superimposed on it; thus,

$$\mathbf{E_a}\exp(j\Phi(x',y')) \tag{62}$$

where for a circular aperture, the polar coordinates $x' = t\cos\xi$ and $y' = t\sin\xi$ allow the aberration function to be expressed as

$$\Phi(x',y') = \Phi(t,\xi) = \sum_{m,\,n=0}^{\infty} \Delta_{mn} t^n \cos m\xi. \tag{63}$$

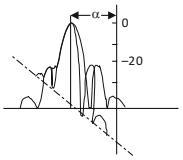
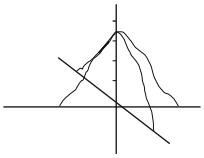
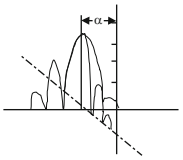
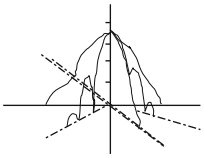
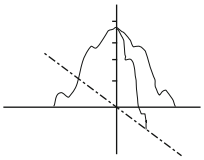
From Eq. 63, it is possible to see the effect of each term by letting all coefficients be zero apart from the desired one. The primary aberrations are illustrated in Table 1. They are referred to as linear ($n = 1$, $m = 0$), quadratic ($n = 2$, $m = 0$), coma ($n = 3$, $m = 1$), astigmatic ($n = 2$, $m = 2$) and spherical ($n = 4$, $m = 0$) aberration. Linear aberration in Table 2 shifts the direction of the main beam by an amount $\alpha = \sin^{-1}(\Delta_{11}/k)$ without changing the structure of the beam. Quadratic phase error causes a reduction in antenna gain and increases both the beamwidth and sidelobe level. Another effect is that the nulls in the pattern are filled in. Cubic phase error, or coma, causes the beam to shift an angle $\alpha = \arcsin(2\Delta_{31}a^2/3k)$ and also reduces gain. In addition, the pattern is asymmetrical in the plane containing the shifted beam and the central axis (the plane of 'scan'). The sidelobes closest to the central axis are lower than the sidelobes without coma, and those in the direction of scan are higher than without coma. The effect of astigmatism is similar to a quadratic phase error. When astigmatism and quadratic phase error occur together, the width of the main beam and the sidelobes are different in the two principal planes. Finally in Table 1, spherical aberration produces a symmetrical distortion of the radiation pattern with an effect similar to quadratic phase error.

The gain of an aperture affected by aberration, G_a, is approximately given by

$$G_a = \frac{1}{1+\kappa\Delta_e^2}G \tag{64}$$

DOI 10.1007/978-981-4560-75-7_30-1

Table 1 Aberrations on a circular aperture of unit radius

Aberration	Order n, m	Aperture phase	Radiation pattern
Linear	1,0	$\Phi = \Delta_{11} t$	 $\alpha = \sin^{-1}(\Delta_{11}/K)$
Quadratic	2,0	$\Phi = \Delta_{20} t^2$	
Coma	1,3	$\Phi = \Delta_{31} t^3 \cos\xi$	 $\alpha = \tan^{-1}(2\Delta_{21}/3K)$
Astigmatism	2,2	$\Phi = \Delta_{22} t^2 \cos 2\xi$	
Spherical	4,0	$\Phi = \Delta_{40} t^4$	

where G is the gain without aberration, Δ_e is the phase error in radians at the edge of the aperture and K is a constant the value of which depends on the type of aberration and equals 0 for linear, 1/12 for quadratic, 1/72 for coma, 1/6 for astigmatism and 4/45 for spherical aberration.

In any practical antenna, all types of aberrations can occur, some to a greater extent than others. For example, a method commonly used to scan the beam of a reflector antenna is to displace the feed laterally from the reflector axis in order to produce a linear phase shift across the aperture. As well as the desired linear aberration (beam shift), coma is strongly represented in the radiation pattern. Astigmatism is also produced, but it is of lesser importance than coma for small lateral shifts. However, when the feed is moved from the focus in the axial direction, either towards or away from the reflector vertex, quadratic and spherical aberrations occur.

DOI 10.1007/978-981-4560-75-7_30-1

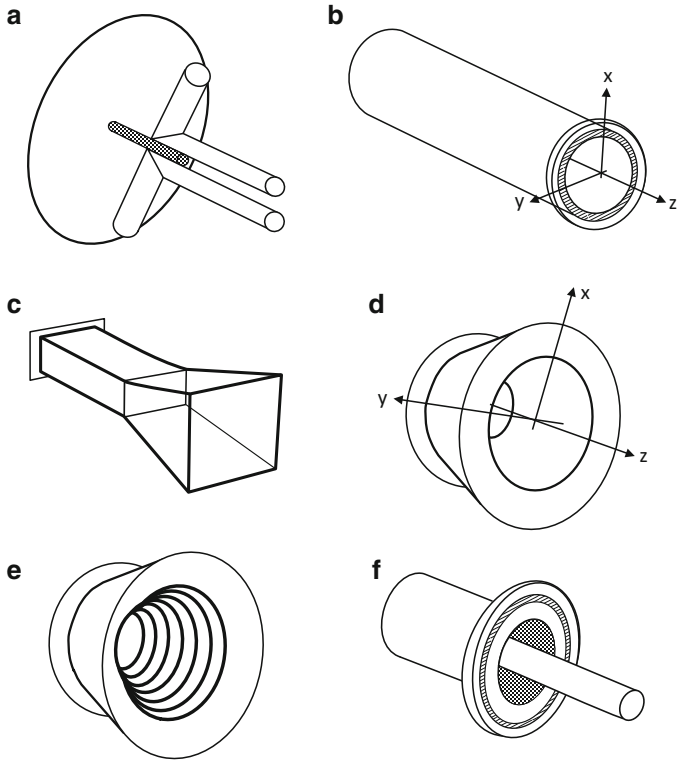


Fig. 12 Feeds for reflectors. (**a**) Dipole with reflector; (**b**) circular waveguide; (**c**) pyramidal horn; (**d**) conical horn; (**e**) corrugated horn; and (**f**) cup feed

Feeds for Reflectors

The feed antenna for a reflector is a vital part of the overall system operation and performance, and therefore, its design requires careful attention. It should have low input reflection over the desired frequency band, efficiently illuminate the reflector and have low spillover and in some applications it should have low cross-polarization. The latter is important in applications such as communications, radar and polarimetry that use orthogonally polarized signals for different bands. The types of feed vary significantly as shown in Fig. 12. As has been mentioned previously, the dipole feed was the first feed that was used, and it still is a common solution where bandwidth may not be significant and at frequencies where weight and size preclude other methods. As shown in Fig. 12a, the dipole can be fitted with a small reflector to improve the directionality of its beam. An option as a feed for a front-fed reflector instead of a dipole is the cup feed as illustrated in Fig. 12f, which can have a bandwidth of about 20 %. The advantage of the dipole or the cup feed is that they are self-supporting so that struts are not required, which improves pattern symmetry and does not reduce gain. A waveguide feed, which is either a circular waveguide as in Fig. 12b or a rectangular horn as shown in Fig. 12c, are suitable in some situations. Aperture slots or tuning rings are often included to improve the radiation pattern of the feed. For the efficient illumination of a reflector system with a subreflector, a feed with a narrow beam is usually required. Hence in dual-reflector applications, a tapered horn is frequently employed. These come in the form of a smooth wall conical horn as in Fig. 4.1d, an optimized profile or one with corrugations as shown in Fig. 4.1e.

Dipole

One of the simplest feeds to construct and connect to an amplifier or power source is the dipole. Consider the dipole shown in Fig. 12a minus the reflector. The electric field radiated by a half-wave dipole oriented parallel to the x-direction is

DOI 10.1007/978-981-4560-75-7_30-1

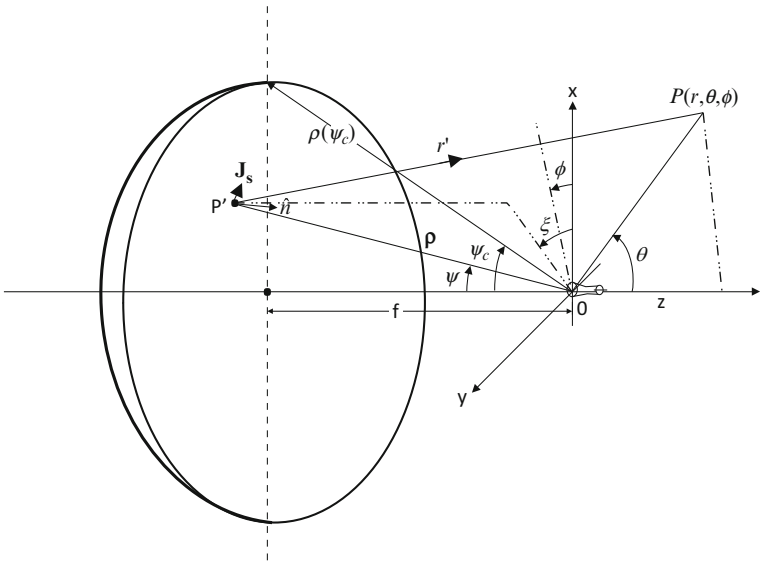


Fig. 13 Geometry of reflector and feed in transmit mode

$$\mathbf{E}_f = E_o \frac{e^{-jkr}}{r} A(\theta, \phi)\left[\hat{r}\ \sin\theta\cos\phi + \hat{\theta}\cos\theta\cos\phi - \hat{\phi}\sin\phi\right] \tag{65a}$$

$$\mathbf{H}_f = \frac{1}{\eta_o}\hat{r} \times \mathbf{E}_f \tag{65b}$$

where E_o is a constant and

$$A(\theta, \phi) = \frac{\cos\left(\frac{\pi}{2}\sin\theta\right)}{\cos\theta}$$

is the pattern function for a half-wave dipole. The pattern function is kept general in the present example for extension to an array of dipoles as a feed in a later chapter of this book ("▶ Applications of Phased Array Feeders for Reflector Antennas." In the E-plane ($\phi = 0$ or π) and H-plane ($\phi = \pm\,\pi/2$), the patterns are different.

The radiation from a paraboloid with a dipole feed can be found from PO or the aperture field method. The former approach will be used here. The geometry of the reflector and feed is shown in Fig. 13. When the reflector is illuminated by the feed, surface currents are created, and these can be approximated by PO as described above. To do this, Eq. 65b is substituted into Eq. 17, and from Eq. 19, it is found that

$$\mathbf{F}(\theta, \phi) = -\frac{2}{\eta_o}\int_0^{2\pi}\int_0^{\psi_c} \hat{n} \times \left(\hat{\rho} \times \mathbf{E}_f\right)\exp(jk\ \hat{r}\cdot\boldsymbol{\rho})\rho^2 \sec\left(\frac{\psi}{2}\right)\sin\psi\, d\psi\, d\xi \tag{66}$$

where $\hat{r}\cdot\boldsymbol{\rho} = \rho(\sin\theta\sin\psi\cos(\phi - \xi) - \cos\theta\cos\psi)$. Now,

DOI 10.1007/978-981-4560-75-7_30-1

$$\begin{aligned}\hat{n}\times\left(\hat{\rho}\times\mathbf{E}_f\right) &= \left(\hat{\rho}\cos\frac{\psi}{2}+\hat{\psi}\sin\frac{\psi}{2}\right)\times\left(\hat{\xi}\cos\psi\cos\xi+\hat{\psi}\sin\xi\right)E_o\frac{e^{-jk\rho}}{\rho}A(\psi,\xi)\\ &= E_o\frac{e^{-jk\rho}}{\rho}A(\psi,\xi)\left(\hat{\rho}\sin\frac{\psi}{2}\cos\psi\cos\xi-\hat{\psi}\cos\frac{\psi}{2}\cos\psi\cos\xi+\hat{\xi}\cos\frac{\psi}{2}\sin\xi\right).\end{aligned} \tag{67}$$

In order to express Eq. 66 in the global coordinate system, the vector components of Eq. 67 must be transformed by means of

$$\begin{aligned}\hat{\rho} = \hat{r}(\sin\theta\sin\psi\cos(\phi-\xi)-\cos\theta\cos\varphi)+\hat{\theta}(\cos\theta\sin\psi\cos(\phi-\xi)+\sin\theta\cos\psi)\\ -\hat{\phi}\sin\psi\sin(\phi-\xi)\end{aligned} \tag{68a}$$

$$\begin{aligned}\hat{\psi} = \hat{r}(\sin\theta\cos\psi\cos(\phi-\xi)-\cos\theta\cos\varphi)+\hat{\theta}(\cos\theta\cos\psi\cos(\phi-\xi)-\sin\theta\sin\psi)\\ -\hat{\phi}\cos\psi\sin(\phi-\xi)\end{aligned} \tag{68b}$$

and

$$\hat{\xi} = -\hat{r}\sin\theta\sin(\phi-\xi)+\hat{\theta}\cos\theta\sin(\phi-\xi)+\hat{\phi}\cos(\phi-\xi). \tag{68c}$$

As a result, the vector triple product becomes

$$\hat{n}\times\left(\hat{\rho}\times\mathbf{E}_f\right) = E_o\frac{e^{-jk\rho}}{\rho}A(\theta,\phi)\left(-\hat{r}\sin\frac{\psi}{2}\cos\psi\cos\xi+\hat{\theta}\cos\frac{\psi}{2}\cos\psi\cos\xi-\hat{\phi}\cos\frac{\psi}{2}\sin\xi\right). \tag{69}$$

The radial components of the electric field in Eq. 18 cancel leaving the following far-field components

$$E_\theta = \frac{jkfE_o}{\pi}\frac{e^{-jkr}}{r}B(\theta,\phi) \tag{70a}$$

and

$$E_\phi = \frac{jkfE_o}{\pi}\frac{e^{-jkr}}{r}C(\theta,\phi) \tag{70b}$$

where

$$\begin{aligned}B(\theta,\phi) = \int_0^{2\pi}d\xi\cos\xi\int_0^{\psi_c}d\psi A(\psi,\xi)\exp[jk\rho(\sin\theta\sin\psi\cos(\phi-\xi)-(1+\cos\theta\cos\psi))]\times\\ \left[\cos\theta(\cos\psi\cos\xi\cos(\phi-\xi)-\sin\xi\sin(\phi-\xi))-\sin\theta\tan\frac{\psi}{2}\cos\phi\cos\xi\right]\tan\frac{\psi}{2}\end{aligned} \tag{71a}$$

$$\begin{aligned}C(\theta,\phi) = \int_0^{2\pi}d\xi\sin\xi\int_0^{\psi_c}d\psi A(\psi,\xi)\exp[jk\rho(\sin\theta\sin\psi\cos(\phi-\xi)-(1+\cos\theta\cos\psi))]\times\\ [\cos\psi\cos\xi\sin(\phi-\xi)+\sin\xi\cos(\phi-\xi)]\tan\frac{\psi}{2}.\end{aligned} \tag{71b}$$

DOI 10.1007/978-981-4560-75-7_30-1

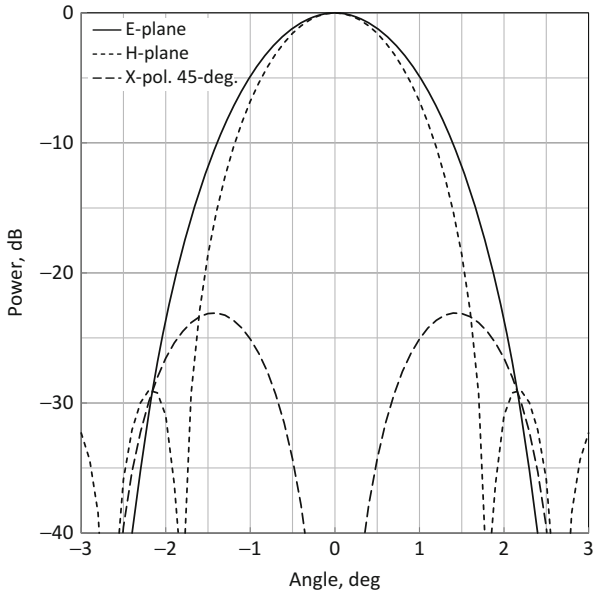


Fig. 14 Principal radiation patterns of a paraboloid reflector with a half-wavelength dipole feed at an operating frequency of 10 GHz. Reflector diameter $D = 1.5\ m$ and $f/D = 0.35$ i.e., $\psi_c = 71.075°$. *Solid line*: E-plane; *short-dashed line*: H-plane; *long-dashed line*: X-polar 45°-plane

The radiation patterns of a reflector with a diameter 50 λ and $f/D = 0.35$ that is fed with a half-wave dipole have been computed, and these are shown in Fig. 14. The E-plane field corresponds to $E_\theta(\theta, 0)$ and the H-plane pattern to $E_\phi(\theta, \pi/2)$, which are given by the functions $B(\theta, 0)$ and $C(\theta, \pi/2)$, respectively. The computed gain is 41 dBi, which gives an aperture efficiency of 51.1 %. The cross-polar level in the 45°-plane is relatively high and is due to deformation of the polarization vector from the far-field sphere to the parabolic surface as well as variation on the surface of the feed field pattern from one principal plane to the other. When the feed pattern is axisymmetric, the latter variation is not present.

Waveguides and Horns

A range of waveguide feeds have been devised for prime focus reflector and also have been flared or profiled into horns for more directive feeding of dual-reflector systems. A basic type of waveguide feed for a paraboloid is a circular waveguide, of which there are many variations devised often through modifying the flange at the aperture plane as illustrated in Fig. 12b.

The reflector geometry is as shown in Fig. 13, and PO will be used once again to determine the far fields of the reflector. Consider a circular waveguide shown in Fig. 12b with a radius α and a smooth flange. Here the electric field radiated by this circular aperture is assumed to have its principal polarization parallel to the x-axis. Therefore, the electric field in the far zone is given by

$$\mathbf{E}_f = \frac{jkaE_o}{2}\frac{e^{-jkr}}{r}\left[\hat{\theta}A(\theta)\cos\phi - \hat{\phi}B(\theta)\sin\phi\right] \tag{72}$$

where

$$A(\theta) = \left(1 + \frac{\beta}{k}\cos\theta\right)\frac{J_1(k_c a)}{k_c}\frac{J_1(wa)}{wa} \tag{73a}$$

and

$$B(\theta) = \left(\frac{\beta}{k} + \cos\theta\right) J_1(k_c a) \frac{k_c J_1'(wa)}{k_c^2 - w^2} \tag{73b}$$

are the feed pattern functions in which $w = k\sin\theta$, $\beta = \sqrt{k^2 - k_c^2}$ is the TE_{11} mode propagation constant and $k_c a = 1.84118$ is the cut-off wavenumber. The associated magnetic field is given by Eq. 4.1b.

The surface current on the paraboloid due to feed radiation given by Eq. 72 is

$$\begin{aligned}
\mathbf{J}_s &= \frac{2}{\eta_o}\hat{n} \times \left(\hat{\rho} \times \mathbf{E}_f\right) = \frac{jkaE_o}{\eta_o}\frac{e^{-jk\rho}}{\rho}\left(-\hat{\rho}\cos\frac{\psi}{2} + \hat{\psi}\sin\frac{\psi}{2}\right) \times \left(\hat{\xi}A(\psi)\cos\xi - \hat{\psi}B(\psi)\sin\xi\right) \\
&= \frac{jkaE_o}{\eta_o}\frac{e^{-jk\rho}}{\rho}\left(\hat{\rho}\,A(\psi)\sin\frac{\psi}{2}\cos\xi + \hat{\psi}A(\psi)\cos\frac{\psi}{2}\cos\xi - \hat{\xi}B(\psi)\cos\frac{\psi}{2}\sin\xi\right) \\
&= \frac{jkaE_o}{\eta_o}\frac{e^{-jk\rho}}{\rho}\cos\frac{\psi}{2}\left\{\hat{r}\left[A(\psi)\cos\xi\cos(\phi-\xi)\left(\sin\theta - \cos\theta\tan\frac{\psi}{2}\right) + B(\psi)\sin\theta\sin\xi\sin(\phi-\xi)\right]\right. \\
&+ \hat{\theta}\left[A(\psi)\cos\xi\cos(\phi-\xi)\left(\cos\theta - \sin\theta\tan\frac{\psi}{2}\right) - B(\psi)\cos\theta\sin\xi\sin(\phi-\xi)\right] \\
&\left. + \hat{\phi}\left[-A(\psi)\cos\xi\sin(\phi-\xi) + B(\psi)\sin\xi\cos(\phi-\xi)\right]\right\}.
\end{aligned}$$

The radial components in the integrand cancel out, and therefore, the radiated electric field is given by

$$\begin{aligned}
E_\theta(r,\theta,\phi) = \frac{k^2 af E_o}{2\pi}\frac{e^{-jkr}}{r}\int_0^{2\pi} d\xi \int_0^{\psi_c} d\psi \exp[jk\rho(\sin\theta\sin\psi\cos(\phi-\xi) - (1+\cos\theta\cos\psi))] \times \\
\left[A(\psi)\cos\xi\cos(\phi-\xi)\left(\cos\theta - \sin\theta\tan\frac{\psi}{2}\right) - B(\psi)\cos\theta\sin\xi\sin(\phi-\xi)\right]\tan\frac{\psi}{2}
\end{aligned} \tag{74a}$$

and

$$\begin{aligned}
E_\phi(r,\theta,\phi) = -\frac{k^2 af E_o}{2\pi}\frac{e^{-jkr}}{r}\int_0^{2\pi} d\xi \int_0^{\psi_c} d\psi \exp[jk\rho(\sin\theta\sin\psi\cos(\phi-\xi) - \\
(1+\cos\theta\cos\psi))][A(\psi)\cos\xi\sin(\phi-\xi) + B(\psi)\sin\xi\cos(\phi-\xi)]\tan\frac{\psi}{2}.
\end{aligned} \tag{74b}$$

The maximum gain of paraboloid with a circular waveguide feed can be calculated by means of Eqs. 20b and 74. It may be shown that the total power radiated by a circular waveguide operating in the TE_{11} mode is

$$P_T = \frac{|E_o|^2 \pi a^2}{4\eta_o} J_1^2(k_c a)\left(1 - \frac{1}{(k_c a)^2}\right).$$

The maximum gain then follows from Eq. 20b where the field components of the electric field are obtained from Eq. 74 that are evaluated in the boresight direction at (0°,0°).

As an example of a prime-focus reflector and circular feed, consider a parabolic reflector of diameter $D = 100\ \lambda$ and $f/D = 0.35$ excited by a circular waveguide with a radius of $a = 0.66\ \lambda$, which corresponds

DOI 10.1007/978-981-4560-75-7_30-1

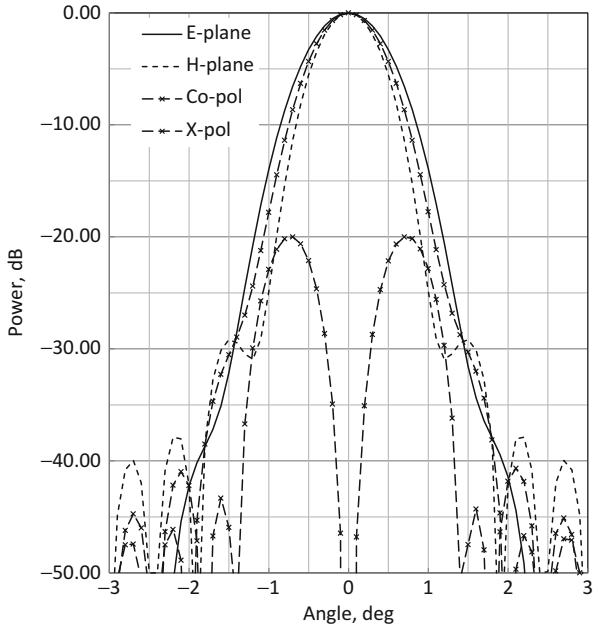


Fig. 15 Radiation patterns of paraboloid with a circular waveguide feed. Reflector dimensions $D = 100\ \lambda$ and $f/D = 0.35$. Circular waveguide feed radius is $a = 0.66\ \lambda$, *Solid plot*: E-plane; *dashed plot*: H-plane; *long dash* – x-: co-polar in the 45° plane; and *short dash* – x-: cross-polar in the 45° plane

to the radius of the first Airy ring in the focal region. The principal plane radiation patterns and co- and cross-polarized patterns are shown in Fig. 15. The computed gain was 48.1 dBi where the aperture efficiency is 65.8 %. This corresponds close to the peak gain for this reflector. Variations in the waveguide diameter either smaller or larger produce a reduction in gain.

Other waveguide feed types can also be studied by means of Eq. 74 as long as the feed radiation can be represented by variations of Eq. 72. For example, a corrugated waveguide operating in the HE_{11} mode at the balanced hybrid–hybrid condition the feed pattern functions Eq. 73 are replaced by

$$A(\theta) = \frac{2\pi k_\rho a}{\left(k_\rho^2 - w^2\right)}\left(1 + \frac{\beta}{k}\cos\theta\right)J_1(k_c a)J_0(wa) \tag{75a}$$

$$B(\theta) = \frac{2\pi k_\rho a}{\left(k_\rho^2 - w^2\right)}\left(\frac{\beta}{k} + \cos\theta\right)J_1(k_c a)J_0(wa) \tag{75b}$$

where $k_\rho a = 2.4048$ and $w = k\sin\theta$. The radiation pattern of the corrugated waveguide is axisymmetric about the boresight direction except for the Huygen's factors. However, in usual operation, $\beta \approx k$. This enables some simplifications to be made to the field radiated by the reflector in Eq. 74. The co- and cross-polarized radiation patterns of the previous reflector of diameter $D = 100\ \lambda$ with a corrugated waveguide of radius $a = \lambda$ are shown in Fig. 16. It is observed that the cross-polarization level is significantly lower than in the case of the circular waveguide feed with a peak of −43 dB and the principal plane patterns are almost identical. Even though the feed cross-polarization is low, there is still some cross-polarization with the reflector. This is due to diffraction that occurs at the reflector rim, which produces a slight depolarization of the radiated field.

DOI 10.1007/978-981-4560-75-7_30-1

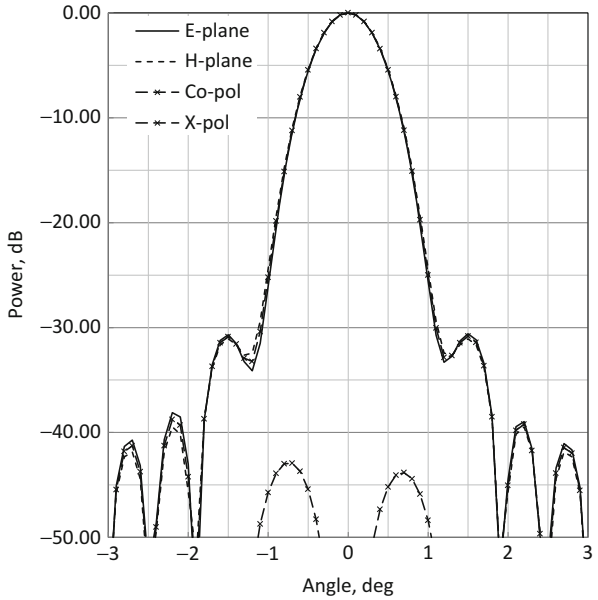


Fig. 16 Radiation patterns of a paraboloid with a corrugated waveguide feed. Reflector dimensions $D = 100\ \lambda$ and $f/D = 0.35$. Waveguide feed radius is $a = 0.66\ \lambda$, *Solid plot*: E-plane; *dashed plot*: H-plane; *long dash* – x-: co-polar in the 45° plane; and *short dash* – x-: cross-polar pattern in the 45° plane

Linear flared horns can also be modelled approximately with Eqs. 72 and 74 provided Eq. 73 can be modified by adding a quadratic phase factor to account for the expansion of the wavefront. For example, a conical horn has an approximate radiated field that is given by the fields radiated by a circular horn multiplied by the quadratic phase factor

$$\exp\left(\frac{k}{2}\frac{\rho'^2}{L}\right)$$

where $\rho' = \sqrt{x'^2 + y'^2}$ is the radial distance to the source point and L is the distance from the horn apex to the aperture. By introducing this factor into the field of the TE_{11} mode of circular waveguide, the far-field of the conical horn is obtained as

$$\begin{Bmatrix} E_\theta \\ E_\phi \end{Bmatrix}(r,\theta,\phi) = \pm\frac{jkE_o}{4}\frac{e^{-jkr}}{r}\ (Q_0(k_c,\theta) \mp Q_2(k_c,\theta))(1+\cos\theta)\ \begin{Bmatrix} \cos\phi \\ \sin\phi \end{Bmatrix} \tag{76}$$

where

$$Q_m(\xi,\theta) = \int_0^a J_m(\xi\rho')J_m(k\rho'\sin\theta)\exp\left(-jk\rho'^2/L\right)\rho' d\rho' \tag{77}$$

and $k_c a = 1.841184$. Equation 77 is sufficiently accurate for most applications providing the cone semi-angle, θ_o, is less than about 30°.

It is assumed in Eq. 76 that $\beta \approx k$. The half-power beamwidth of a conical horn can be shown to be $HPBW \approx 0.6 * 180\lambda/(\pi a)$, where a is the radius at the aperture. Thus, if the feed radius at the aperture is $a = 1\ \lambda$ the $HPBW \approx 34^\circ$. Suppose conical horn is used as a feed in a paraboloid with $D = 50\ \lambda$ and $f/$

DOI 10.1007/978-981-4560-75-7_30-1

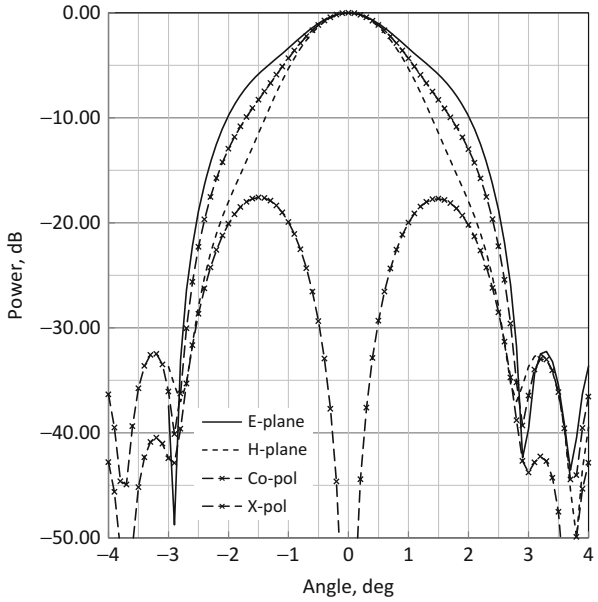


Fig. 17 Radiation patterns of paraboloid with a conical horn feed. Reflector dimensions $D = 50\ \lambda$ and $f/D = 0.5$. Waveguide feed radius is $\alpha = 1\ \lambda$, *Solid plot*: E-plane; *dashed plot*: H-plane; *long dash* – x-: co-polar in the 45° plane; and *short dash* – x-: cross-polar pattern in the 45° plane

$D = 0.5$. The co-polar and cross-polar patterns of this reflector with the conical horn shown in Fig. 17. Obviously the cross-polar level is quite high and this frequently the case with smooth wall conical horns. The cross-polar peak is reduced as in Fig. 16 with a corrugated conical horn. Similarly, a quadratic phase factor can be applied to a corrugated waveguide to model the radiated field of a corrugated horn. The result is

$$\begin{Bmatrix} E_\theta \\ E_\phi \end{Bmatrix}(r,\theta,\phi) = \pm\frac{jkE_o}{2}\frac{e^{-jkr}}{r}(1+\cos\theta)Q_0(k_\rho)\begin{Bmatrix} \cos \\ \sin \end{Bmatrix}\phi \tag{78}$$

where Q_0 is given by Eq. 77.

Advanced Feeds

The feeds illustrated in Fig. 12 represent canonical feeds although extensions shown in Fig. 18 are now in general use. These feeds provide several improvements such as increased efficiency, pattern symmetry, low cross-polarization, increased bandwidth, well-defined phase centre and low input mismatch, which increased the performance of suitable reflector systems. A short summary of these feeds is given in this section. Further details on them as well as other feeds can be found in the references (Olver et al. 1994; Bird and Love 2007).

A simple extension of the circular waveguide in Fig. 12b is to corrugate the output flange as shown in Fig. 18a. This enables the patterns for pipes of smaller diameter, and therefore wider beamwidth, to be axisymmetric, which is useful for deep dishes (i.e., f/D < 0.35). In addition, the number of corrugations may be reduced so that there is only one (as shown in Fig. 18c) or two ring slots and these can be tuned to give low cross-polarization at several narrow frequency bands. The depth of the slot varies from about 0.2 to 0.375 λ depending on the aperture diameter and the distance the slot(s) needs to be placed for the chosen edge illumination. Another option as a feed for deep reflectors is to use a coaxial waveguide.

DOI 10.1007/978-981-4560-75-7_30-1

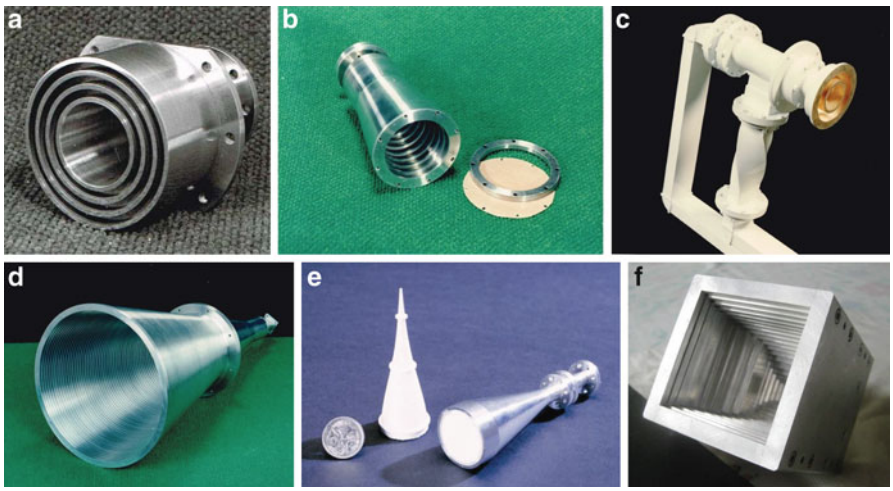


Fig. 18 Advanced feeds for reflector antennas. (**a**) Circular waveguide with corrugated flange; (**b**) corrugated horn for an offset parabolic reflector; (**c**) dual-polarized circular waveguide with ring slot flange; (**d**) wideband corrugated horn for a symmetrical Cassegrain reflector; (**e**) dielectric-lined horn with profiled dielectric insert; and (**f**) profiled rectangular horn feed (Pictures (**a**) to (**e**) courtesy CSIRO Australia)

A coaxial feed works best when the inner-to-outer conductor radii are close to 0.3. When this ratio is larger, the H-plane pattern becomes narrow and the E-plane pattern broadens. A flange with ring slots can also be used to adjust the pattern for coaxial feeds.

The corrugated horns illustrated in Fig. 18b, d continue to be an effective feed option for long–focal length reflectors (i.e., f/D >0.5) and for dual reflectors. Its excellent characteristics such as beam symmetry, bandwidth, low cross-polarization and good input reflection coefficient make it an attractive option for many applications. Fabrication can still be a problem as it is important to have accurate corrugations, which can be obtained with numerical machining.

Dielectric-lined horns as shown in Fig. 18e are another option in place of corrugated horns. The dielectric constant can be profiled transversely or have a central core of material with dielectric constant of about 1.2 and a small air gap between the core and the metal wall. When properly designed, a dielectric-lined horn can have very large bandwidths sometimes in excess of 30:1. There are practical issues around the use of dielectrics such as moisture trapping and outgassing in a vacuum. They have to be suitable for the operating environment either on the ground or in space as well as structurally stable. Nonetheless, their wide bandwidth makes them an attractive possibility.

The performance of standard rectangular and circular horns can be improved with a number of steps along their profile to generate higher-order modes. This has been used for many years to obtain incremental improvements. This process has been extended recently by introducing many steps to profile the horn from a desired input waveguide diameter to an output diameter that is able to achieve the desired beamwidth, input match and cross-polar maximum. A rectangular horn designed by this method is shown in Fig. 18f. The profile is created by computer optimization in much the same way described in a later section ("Reflector Shaping") for profiling reflectors. By this means, a horn's performance can be tailored to suit the intended application including a desired bandwidth, edge illumination and input reflection coefficient.

The feed geometries referred to above enable high-performance reflector systems to operate reliably. The references should be consulted (Olver et al. 1994; Bird and Love 2007) for further details as well as information of other feed types.

DOI 10.1007/978-981-4560-75-7_30-1

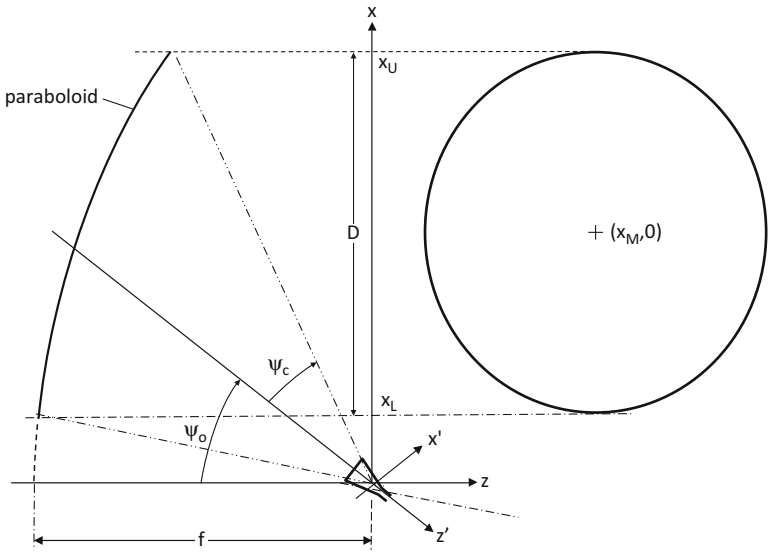


Fig. 19 The offset parabolic reflector geometry and its rotated co-ordinate system {X′} for the feed along with the projected aperture

Other Reflector Geometries

While the paraboloid is the most common type of reflector in everyday use, there are other reflector geometries that are used. In this section, the basic properties of the offset parabolic reflector are described as are the properties of the classical symmetrical Cassegrain and its offset equivalent. The radiation from the prime-focus fed spherical reflector is described, and focal fields of this reflector are studied. Finally, some basic techniques for designing shaped reflectors will be outlined.

Offset-Fed Parabolic Reflector

Blockage by the feed and feed support struts in a paraboloidal reflector reduces the gain and increases the sidelobe level. This can be overcome with the offset-fed configuration as illustrated in Fig. 19. The advantage of offset reflecting systems is that the offset parabolic antenna is formed by rotating the feed through an angle as shown in Fig. 19, which illuminates only part of the paraboloid and allows the feed to be supported by a mount below the reflector. The feed and reflector are contained within a cone of half-angle ψ_c with the feed at its apex and the reflector rim lying on its surface. The projection of the rim onto the x-y plane is a circle of diameter

$$D = \frac{4f \sin \psi_c}{\cos \psi_o + \cos \psi_c} \tag{79}$$

which is centred at (x_m, 0) where

$$x_m = \frac{2f \sin \psi_o}{\cos \psi_o + \cos \psi_c}. \tag{80}$$

f is the focal length of the original paraboloid and ψ_o is the rotation angle ("offset angle") of the primed coordinate system about the focus. The cone angle ψ_c lies between the rotated axis of the feed and the rim of the offset reflector. The angles are given by

DOI 10.1007/978-981-4560-75-7_30-1

$$\psi_{o,c} = \tan^{-1}\left(\frac{x_m + D/2}{2f}\right) \pm \tan^{-1}\left(\frac{x_m - D/2}{2f}\right).$$

Feed blockage is avoided if the largest feed upper extremity in the positive x-direction is less than the clearance, x_L, between the lower rim and the z-axis, where

$$x_L = 2f \tan\left[\left(\frac{\psi_o - \psi_c}{2}\right)\right] \tag{81}$$

From the focus along the offset axis, the reflector rim lies on the projected elliptical cone that has major and minor axes of length

$$a = \frac{D}{2\sin\gamma};\ b = \frac{D}{2} \tag{82}$$

where $\gamma = \tan^{-1}\frac{2f}{x_m}$. The centre of the ellipse is (x_m,0,z_m), where $x_m$is given by Eq. 80 and

$$z_m = f\left[\frac{\sin^2\psi_o + \sin^2\psi_c}{(\cos\psi_o + \cos\psi_c)^2} - 1\right]. \tag{83}$$

With respect to the rotated axes, the spherical polar co-ordinates in this system are (ρ, ψ, ξ) and a point on the paraboloid has coordinates

$$x' = \rho\sin\psi\cos\xi; y' = \rho\sin\psi\sin\xi; z' = -\rho\cos\psi \tag{84}$$

where

$$\rho = \frac{2f}{1 - \cos\xi\sin\psi\sin\psi_o + \cos\psi\cos\psi_o} \tag{85}$$

The relation between the two coordinate systems is

$$x = x'\cos\psi_o - z'\sin\psi_o; y = y'; z = x'\sin\psi_o + z'\cos\psi_o. \tag{86}$$

The normal is expressed

$$\hat{n} = -\sqrt{\frac{\rho}{4f}}\{-\hat{x}(\cos\xi\sin\psi\cos\psi_o + \cos\psi\sin\psi_o) - \hat{y}\sin\xi\sin\psi + \hat{z}(1 - \cos\xi\sin\psi\sin\psi_o + \cos\psi\cos\psi_o)\}.$$

A geometric optics approximation to the aperture field is obtained by the approach for symmetrical reflectors leading up to Eq. 27. Clearly, feed rotation does not alter the paraboloid's basic properties, viz. the distance from the focus to the aperture plane at $z = 0$ to the reflector and back to the aperture is $2f$; also there is no z-component of the aperture field because the output is a plane wave. These properties allow the aperture field expressions to be simplified. For the incident field from the feed expressed by Eq. 25, the rectangular components of the aperture field in the projected aperture are given by Eq. 47:

DOI 10.1007/978-981-4560-75-7_30-1

$$E_{ax} = g_o\left[a_1 F_\psi(\psi,\xi) - b_1 F_\xi(\psi,\xi)\right] \tag{87a}$$

$$E_{ay} = g_o\left[b_1 F_\psi(\psi,\xi) + a_1 F_\xi(\psi,\xi)\right] \tag{87b}$$

where

$$a_1 = \cos\xi(1 + \cos\psi\cos\psi_o) - \sin\psi\sin\psi_o,$$

$$b_1 = \sin\xi(\cos\psi + \cos\psi_o)$$

and

$$g_o = -\exp(-2fk)/2f.$$

The far fields are obtained from Eq. 87 by applying Eq. 16 in the aperture plane. To do this, define polar co-ordinates $(t,\ \zeta)$ centred on $(x_m,\ 0)$ in the aperture such that $x = x_m + t\cos\zeta$, $y = t\sin\zeta$ and $z = \left(x^2 + y^2 - 4f^2\right)/4f$. The resulting integrals are evaluated on the projected circle, Fig. 19. To obtain the angle co-ordinates relative to the rotated feed coordinates, the inverse of Eq. 86 are used to express $\psi = \text{asin}(z'/\rho)$ and $\xi = \text{atan}(y'/x')$ where $\rho = \sqrt{x^2 + y^2 + z^2}$. When the feed pattern is axisymmetric (i.e., $F_\psi(\psi,\xi) = A(\psi)\cos\xi$ and $F_\xi(\psi,\xi) = -A(\psi)\sin\xi$ the aperture fields will be symmetric about the vertical (x-) axis. However, compared with the symmetrical paraboloid there is cross-polarization. This is maximum in the plane of asymmetry ($\xi = 90°$ and 270°). Cross-polarization results because the feed has been rotated and the field illuminating the reflector is no longer symmetrically located. In general, a feed with asymmetric patterns has maximum cross-polarization between the 90 and 45° planes, depending on the level of feed cross-polarization. Special feeds have been designed to introduce low levels of cross-polarization to cancel out the cross-polarization of the offset paraboloid (Rudge and Adatia 1975).

As described above, the radiation field is obtained by substituting Eq. 87 into Eq. 16 and evaluating the required transform either numerically or by means of the FFT. An example of the radiation patterns possible for an offset parabolic reflector is shown in Fig. 20. The E-, H-, co- and cross-polar patterns in the 90°-plane are shown for an offset reflector with $\psi_{o=44°}$, $\psi_c = 30°$ and $D = 100\ \lambda$. The reflector is fed with a rectangular waveguide having aperture dimensions $1.57\lambda \times 2.14\ \lambda$. Its E-plane pattern is obviously coincident with the x-z plane. The dimensions of the waveguide and the operating frequency result in edge illuminations in both the E- and H-planes of about −12 dB. The computed gain is 37.67 dBi, and the aperture efficiency is 72 %. The peak cross-polar level is −23.6 dB below the co-polar maximum in the $\phi = 90°$-plane. The results given in Fig. 20 were computed from Eq. 87 by numerical integration and they agree well with experimental values (Rudge 1975).

The feed can be modified from Eq. 5.9 to one with pattern axial symmetry given by the Gaussian function

$$A(\psi) = \exp\left(-\alpha\psi^2\right).$$

where the coefficient α is chosen to give an edge illumination of −12 dB. Thus, $\alpha = -EdB/\left(20\psi_c^2\log_{10}e\right)$ wherein EdB is the edge illumination $EdB = -12$. With this axisymmetric feed, the offset antenna gain has an increased gain of 38.0 dBi and an efficiency of 77.9 %. The radiation patterns are similar to those shown in Fig. 5.2 except that first sidelobe is lower at −29.6 dB below the peak and the cross-polar maximum is slightly lower at −24.1 dB. The peak cross polarization occurs near to the 45° plane for an asymmetric feed and is at a higher level than in the axisymmetric case, which occurs

DOI 10.1007/978-981-4560-75-7_30-1

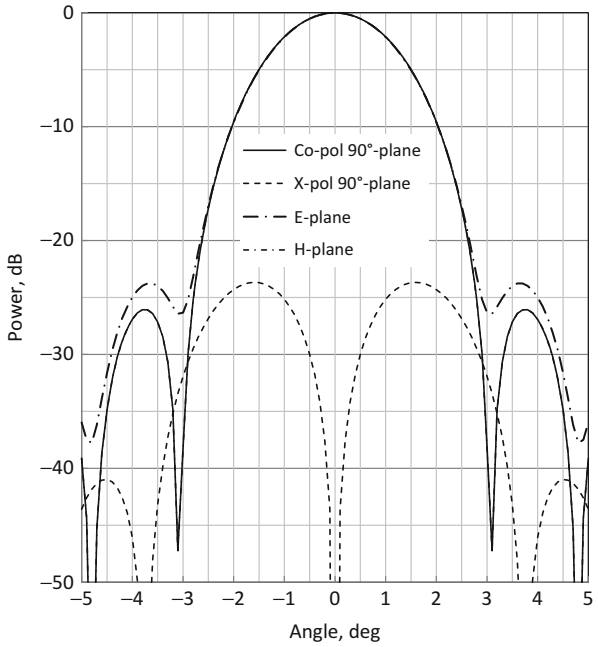


Fig. 20 Co- and cross-polarized radiation patterns of an offset reflector antenna with $\psi_o = 44°$, $\psi_o = 30°$ and $D = 100\ \lambda$ and that is fed with a rectangular waveguide of aperture dimensions $a = 1.57\ \lambda$ and $b = 2.14\ \lambda$

in the 90° plane. However, the maximum level of cross-polarization is approximately the same for both feeds.

An estimate of the HPBW of an offset parabolic reflector in the E-plane is approximately

$$HPBW\ \ (\text{deg.}) = (0.9\,E_{dB}\ \ +\ \ 58)\ \ (\lambda/D) \tag{88}$$

where E_{dB} is the edge illumination in dB. Thus, in the previous example, where the edge illumination is −12 dB, Eq. 5.10 predicts a HPBW of 2.4°, which agrees closely with the HPBW of the pattern in Fig. 20.

In the design of offset parabolic reflectors, the useful concept for establishing 'ball-park' figures and designing the co-polar pattern is that of the 'effective paraboloid.' The offset parabolic reflector is assumed identical to a symmetrical paraboloid of the same diameter but with an effective focal length

$$f_{eff} = \frac{2f}{1 + \cos\psi_o} \tag{89}$$

For the effective paraboloid, the focal length to diameter ratio is f_{eff}/D.

As an example, consider the design of an offset paraboloid to produce a beam in the far-field at the elevation and azimuth angles θ_b and ϕ_b. Assume θ_b is close to reflector boresight and consider an incoming ray from (θ_b, ϕ_b). In the effective paraboloid, the reflected ray makes an angle β with respect to the axis, where

$$\beta \approx \theta_b \left(1 + \frac{1}{32\left(f_{eff}/D\right)^2}\right) \tag{90}$$

DOI 10.1007/978-981-4560-75-7_30-1

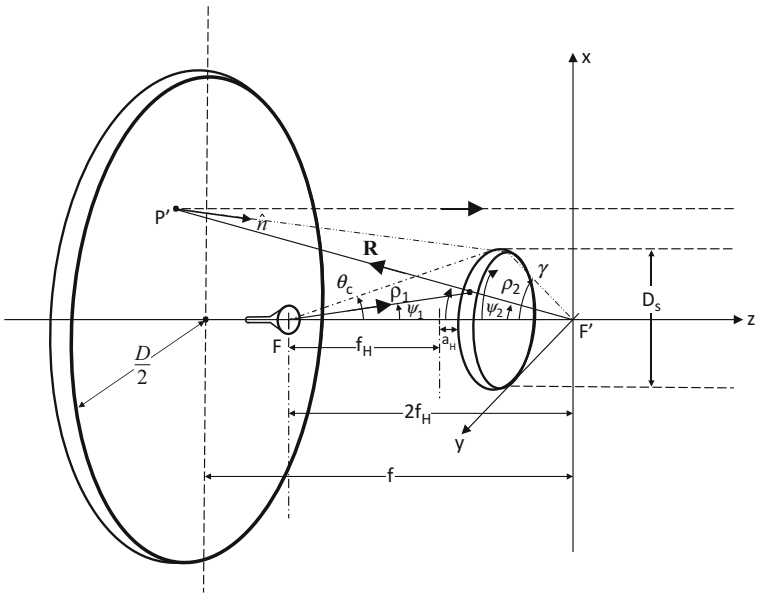


Fig. 21 Symmetrical Cassegrain antenna with hyperboloidal subreflector

which comes from the small angle estimate for a paraboloid. The position for the location of feed in the focal plane is estimated as

$$x_{oeff} = f_{eff} \sin\beta \cos\phi_b y_{oeff} = f_{eff} \sin\beta \sin\phi_b. \tag{91}$$

The accuracy of the estimates given by Eq. 91 is usually quite good, but they become less accurate as the angle from boresight increases due to the original estimate Eq. 90.

Cassegrain Antennas

The Cassegrain antenna as illustrated in Fig. 21 had its origins in an optical telescope that was devised in the seventeenth century. In its classical form, it consists of a paraboloidal main reflector and a smaller hyperboloidal subreflector. The hyperboloid has real and virtual foci, labelled F and F′ in Fig. 21, and it is symmetrical about the axis FF′. From the focus F the distance to the surface is given by

$$\rho_1 = \frac{-e\beta}{1 - e\cos\psi_1} \tag{92}$$

where $\beta = f_H(1 - 1/e^2)$ and $e = f_H/a_H$ is the hyperboloid' s eccentricity. Alternatively, from the virtual focus the surface is given by

$$\rho_2 = \frac{e\beta}{1 + e\cos\psi_2}. \tag{93a}$$

The relation between the angles from the axis to the reflector is

$$\tan\frac{\psi_2}{2} = M\tan\frac{\psi_1}{2} \tag{93b}$$

DOI 10.1007/978-981-4560-75-7_30-1

where $M = \frac{e+1}{e-1}$ is the magnification factor. The cone angle from the focus to the rim is given by

$$\theta_c = \tan^{-1}(-1/\alpha) + \left(\frac{\alpha/e}{\sqrt{\alpha^2+1}}\right) \tag{93c}$$

where $\alpha = 2\beta/D_s$. D_s is the diameter of the subreflector. The relation between the cone angles from the foci is

$$\cot\theta_c + \cot\gamma = f_H/D_s \tag{93d}$$

In the Cassegrain configuration, F′ of the hyperboloid is coincident with the paraboloid's focus, while the feed is located at F. When a source of spherical waves is located at the focus, the hyperboloid creates a reflected wave that for the paraboloid appears also to emanate from a source located at F′. The main advantages of the Cassegrain over a single reflector antenna are that the feed can be located close to the main reflector and the receiving/transmitting electronics and, also, feed spillover is directed toward the cold sky not towards a hot earth. In the classical symmetrical Cassegrain (Fig. 21) blockage still occurs due to the subreflector and the subreflector supports.

However, there is sufficient flexibility in the dual reflector system to shape the profiles of both reflectors in such a way that subreflector blockage is reduced. Reflector shaping applied to both surfaces allows the aperture illumination to be selected to enhance performance. This makes the symmetrical shaped Cassegrain superior in most respects to front-fed reflectors and is the reason for their widespread use in large earth stations.

Blockage can also be avoided by adopting an offset Cassegrain configuration, which is illustrated in Fig. 22. Feed spillover is an important contributor to the far-out sidelobes of all types of Cassegrain antennas, and to minimise this contribution, the feed sidelobes in particular should be small. Usually, subreflector edge illumination needs to be ≤ -16 dB to keep the spillover contributions to the sidelobes of the Cassegrain to an acceptable level.

The properties of Cassegrain antennas can be analysed by means of the techniques described in the section "Background." Ray tracing by geometric optics (GO) can be used at both reflectors to determine the aperture field. Although this approach is not accurate for small subreflectors (diameter less than about 10 λ) because of diffraction at the rim. GO is sufficiently accurate for most purposes when the diameter of the reflectors is large ($>10\ \lambda$). Considerable improvement in accuracy results for smaller reflectors when diffraction is included through methods such as the geometrical theory of diffraction (GTD). Accurate results are also possible by applying the physical optics approximation at both reflectors or by combining the techniques of GTD and physical optics.

Application of GO to the offset Cassegrain configuration shown in Fig. 22 enables an approximation to be found for the aperture fields. A feed with an axisymmetric pattern function $F(\psi)$, which is also polarized parallel to the plane of symmetry (x-z plane), results in aperture fields

$$E_{ax}(\psi',\zeta') = g_o(\psi',\zeta')[A\sin\psi'\cos\zeta' + B\left(\sin^2\zeta' + \cos\psi'\cos^2\zeta'\right) + C(1+\cos\psi')] \tag{94a}$$

$$E_{ay}(\psi',\zeta') = -g_o(\psi',\zeta')\sin\zeta'[A\sin\psi' + B(\cos\psi' - 1)\cos\zeta'] \tag{94b}$$

where

DOI 10.1007/978-981-4560-75-7_30-1

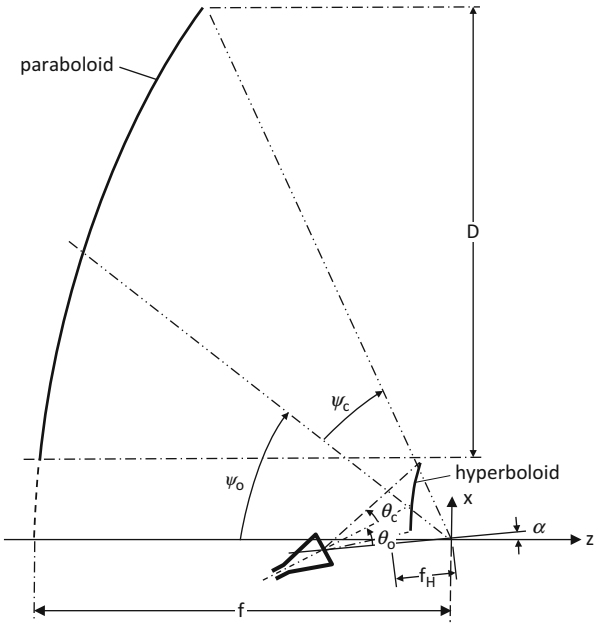


Fig. 22 Offset Cassegrain antenna

$$A = L\sin\alpha\cos\theta_o - \sin\theta_o(K + \cos\alpha)$$

$$B = L\sin\alpha\sin\theta_o + \cos\theta_o(K + \cos\alpha) - C$$

$$g_o(\psi', \zeta') = \frac{F(\psi')}{2fL}\exp\left[-2jk\left(f + \frac{f_H}{e}\right)\right]/(B\cos\psi' + A\sin\psi'\cos\zeta' + C(1 + \cos\psi'))$$

$C = 1 + K\cos\ \alpha$, $L = (e^2 - 1)/(1 + e^2)$ and $K = -2e^2/(1 + e^2)$. Also (Ψ', ζ') are the elevation and azimuth angles of a spherical coordinate system at the feed phase centre. Equation 5.16 are in an identical form to the single offset reflector except for the latter, $A = \sin\theta_0$, $B = 1 - \cos\theta_0$, $C = -1$ and D=exp$(-2jkf)/2f$. The aperture fields for a symmetrical Cassegrain are obtained from Eq. 94 by letting $\phi_0 = 0$, $\theta_0 = 0$ and $\alpha = 0$.

The offset Cassegrain can have low cross-polarization compared with a single offset by the appropriate selection of the feed and reflector offset angles. GO predicts zero cross-polarization, but due to subreflector diffraction cross-polarization has a small but finite level. Zero cross-polarization is predicted by Eq. 94 when the feed and hyperboloid offset angles satisfy the condition

$$\tan\frac{\theta_0}{2} = M\tan\frac{\alpha}{2}. \tag{95}$$

where M is the magnification factor defined for the Cassegrain (Eq. 93b).

Under this condition, the reflector pattern is axisymmetric also. When the feed pattern is not axisymmetric, Eq. 95 is the condition for zero cross-polarization in the plane of asymmetry (y-z plane). Furthermore, in that case the principal component of the aperture field is elliptical and maximum cross-polarization occurs between the 45 and 90° planes. The aperture field of an offset antenna with a feed with an axi-symmetric pattern that does not satisfy Eq. 95 is shown in Fig. 23. In this example, $D = 150\ \lambda$, $\psi_o = 53.553°$, $\psi_c = 20.511°$, $f = 160.77\ \lambda$, $\theta_o = 28°$, $\theta_c = 11$, e = 2.4575, $f_H = 41.054\ \lambda$ and $\alpha = 6°$ The feed pattern is a Gaussian function, which results in a subreflector edge illumination of −10 dB. As might

DOI 10.1007/978-981-4560-75-7_30-1

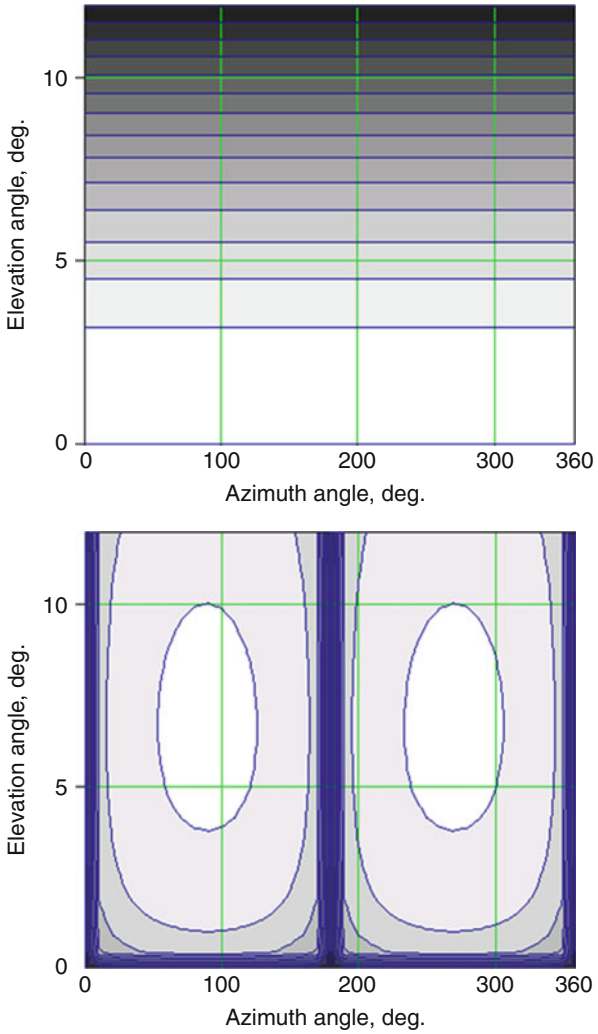


Fig. 23 Principal and cross-polar field components obtained by geometric optics in the aperture plane of an offset Cassegrain antenna defined by $D = 150\ \lambda$, $\psi_o = 53.553$, $\psi_c = 20.511°$, $\theta_o = 28°$, $\theta_c = 11°$, $e = 2.4575$, $f_H = 41.054\ \lambda$, $\alpha = 6°$. Feed has axisymmetric pattern with edge illumination of 10 dB. The electric field is polarized parallel to plane of symmetry

be expected, the co-polar contours are almost invariant with azimuth angle and peak cross-polarization occurs in 90 and 270° planes.

The radiation patterns can be obtained from Eq. 16 and the aperture fields in Eq. 94 by a series of coordinate changes from the aperture plane through to the local co-ordinate system of the feed. The geometric optics formulation described above does not take into account the diffraction from the subreflector, which contributes particularly to the cross-polarization. For example, in the plane of asymmetry (i.e., $\phi = \pm 90°$), a rigorous analysis of the antenna described in Fig. 23 has been shown that the cross-polar peak level at the angle α predicted by Eq. 95 varies from about -45 dB for a 10 λ diameter subreflector to about -30 dB for one with a subreflector 5 λ in diameter.

Spherical Reflectors

The spherical reflector is employed in many different applications, from satellite communications to scanning beams for radar tracking. The latter is achieved quite simply by rotating the feed at the chosen focal point as the surface appears identical over a limited angular range. However, the complication is that

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_30-1

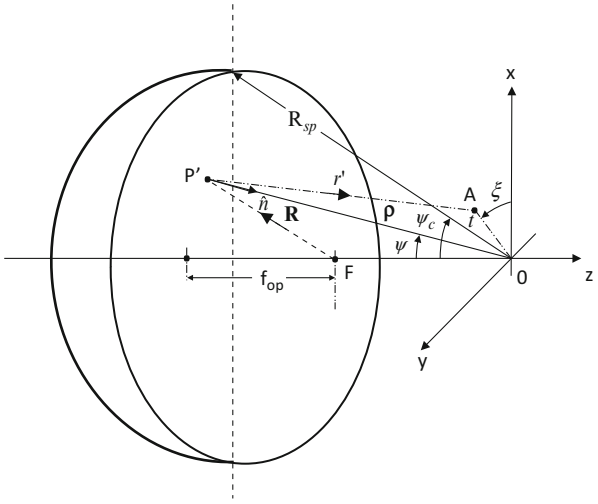


Fig. 24 Geometry of spherical reflector

without correction, a conventional feed horn provides poor illumination of a spheroid compared with a paraboloid. The classical location for a point source of a spherical reflector of radius R_{sp} is $f_{op} = R_{sp}/2$. However, marginal rays do not intersect this classical point but do so elsewhere along the paraxial region, i.e., close to the z-axis shown in Fig. 24. As a result, the optimum location focal point F can be closer to the reflector than the classical location (Ashmead and Pippard 1946). It has been shown that, to keep the phase error in the aperture to within $\pm\lambda/16$, the diameter of the aperture should not exceed $D = 256\lambda(f/D)^3$. Three approaches are commonly used for feeding a spherical reflector. These are an array of dipoles or waveguides, a correcting concave reflector known as a Gregorian-corrected configuration, and a line source in which a travelling wave feed radiates a field along the axis towards the reflector. To understand the spherical reflector further, the radiation from this reflector analysed by means of geometric optics. Through this analysis the optimum location can be estimated for a single feed. A waveguide feed is then used to excite a spherical reflector and the radiation pattern obtained. Finally, the focal region field of a spherical reflector is discussed.

The geometry of a spherical reflector is shown in Fig. 24. The distance from the focal point to the aperture plane at $z = 0$ is given by

$$s = FP' + P'A = \sqrt{t^2 + \left[\sqrt{\rho^2 - t^2} - \left(\rho - f_{op}\right)\right]^2} + \sqrt{\rho^2 - t^2} \tag{96}$$

where $\rho = R_{sp}$ the radius of the sphere, t is the radial distance from the z-axis and f_{op} is the distance from the vertex to the focal point at F. The path difference between a paraxial ray and a non-axial ray is given by

$$\Delta = R_{sp} + f_{op} - s$$

In wavelengths, this difference is

DOI 10.1007/978-981-4560-75-7_30-1

$$\frac{\Delta}{\lambda} = \frac{R_{sp}}{\lambda}\left(1 + \frac{f_{op}}{R_{sp}} - \sqrt{\left(\frac{t}{R_{sp}}\right)^2 + \left[\sqrt{1 - \left(\frac{t}{R_{sp}}\right)^2} - \left(1 - \frac{f_{op}}{R_{sp}}\right)\right]^2} - \sqrt{1 - \left(\frac{t}{R_{sp}}\right)^2}\right). \tag{97}$$

The total phase error over a prescribed aperture is least when the phase error at the aperture edge is zero. If the aperture radius is D/2, which describes a cone of angle $\psi_c = \sin^{-1}\left(\mathrm{D}/2\mathrm{R}_{\mathrm{sp}}\right)$ at O, the optimum focal length when $\Delta/\lambda = 0$ is

$$f_{op} = \frac{1}{4}\left(R_{sp} + \sqrt{R_{sp}^2 - \left(\frac{D}{2}\right)^2}\right). \tag{98}$$

Whilst this is the optimum location for a uniform illumination, it is not necessarily the most suitable location for a tapered pattern when the feed needs to move further away from the reflector vertex. Correspondingly there is a given aperture size for a total phase error. Thus, the phase error tolerance limits the aperture size. It has been determined that the maximum allowable total phase error $(\Delta/\lambda)_{\max}$ for a given aperture diameter D is

$$\left(\frac{D}{R_{sp}}\right)^4_{\max} = 75\pi\frac{(\Delta/\lambda)_{\max}}{(R_{sp}/\lambda)}. \tag{99}$$

For example, a spherical reflector with a radius of $R_{sp} = 1.52\ m$ is required to produce an aperture field with maximum total phase error of $(\Delta)_{\max} = \lambda/16$ at 11.2 GHz. The maximum possible aperture diameter predicted by Eq. 99 is $(D)_{\max} = 1.08m$.

The aperture field of the spherical reflector can be obtained from Eq. 4, where $\hat{n} = -\hat{r}$. It is assumed the feed radiates an incident electric field in the form given by Eq. 6, which results in

$$\mathbf{E_i} = \left(\hat{\psi}F_\psi(\psi,\xi) + \hat{\xi}F_\xi(\psi,\xi)\right)\exp(-jkR)/R$$

where, as shown in Fig. 24, R is the distance from the focus to the reflector surface. Therefore the aperture field is given by

$$\mathbf{E}_a = -\left(\hat{\psi}F_\psi(\psi,\xi) + \hat{\xi}F_\xi(\psi,\xi)\right)\exp(-jks)/R \tag{100}$$

where $s = R + r'$ is the total distance, $r' = \sqrt{R_{sp}^2 - t^2}$ and

$$R = \sqrt{t^2 + \left[\sqrt{R_{sp}^2 - t^2} - \left(R_{sp} - f_{op}\right)\right]^2}. \tag{101}$$

To calculate the radiated field from a spherical reflector, Eq. 100 is substituted in Eqs. 16. An asymmetrical feed is assumed so that $F_\psi(\psi,\xi) = A(\psi)\cos\xi$ and $F_\xi(\psi,\xi) = -B(\psi)\sin\xi$. The integral over ξ in this case can be completed in closed form by means of Eq. 35. The resulting transform functions are

DOI 10.1007/978-981-4560-75-7_30-1

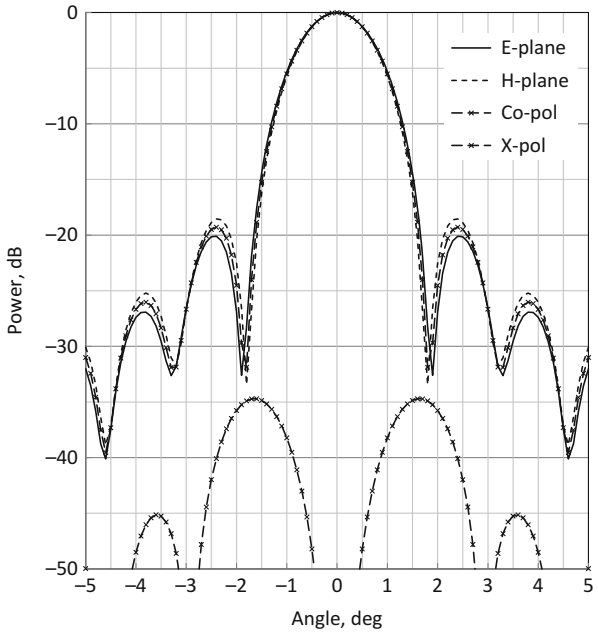


Fig. 25 Radiation patterns of a spherical reflector with a circular waveguide feed at an operating frequency of 11.2 GHz. The diameter of the spheroid is 305.8 cm, aperture diameter 108 cm, focal length 73.7 cm and $\psi_c = 20.89°$. Aperture radius of the circular waveguide feed is 1.87 cm

$$N_x(\theta,\phi) = -\pi\int_0^{D/2} dt\ \ t\left[A(\psi)\cos\psi(J_0(w) - J_2(w)\cos 2\phi) + B(\psi)\left(J_0(w) + J_2(w)\cos 2\phi\right)\right]\exp(-jks)/R \tag{102a}$$

$$N_y(\theta,\phi) = \pi\sin 2\phi\int_0^{D/2} dt\, tJ_2(w)[A(\psi)\cos\psi - B(\psi)]\exp(-jks)/R. \tag{102b}$$

where $w = kt\sin\theta$. It is convenient to integrate over the angle ψ in Eqs. 102 in the range $0 < \psi < \psi_c$. Substituting for $t = R_{sp}\sin\psi$ and

$$N_x(\theta,\phi) = -\frac{\pi R_{sp}^2}{2}\int_0^{\psi_c} d\psi\sin 2\psi[A(\psi)\cos\psi(J_0(w) - J_2(w)\cos 2\phi) + B(\psi)(J_0(w) + J_2(w)\cos 2\phi)\Big]\exp(-jks)/R \tag{103a}$$

$$N_y(\theta,\phi) = \frac{\pi R_{sp}^2}{2}\sin 2\phi\int_0^{\psi_c} d\psi\sin\psi J_2(w)[A(\psi)\cos\psi - B(\psi)]\exp(-jks)/R \tag{103b}$$

where $s = R_{sp}\cos\psi + R$ and $R = \sqrt{R_{sp}^2 + \left(R_{sp} - f_{op}\right)^2 - 2\cos\psi R_{sp}\left(R_{sp} - f_{op}\right)}$. Note that when $\psi = 0, R = f_{op}$, It can be seen from Eq. 103 that when the feed is axisymmetric, i.e., $A(\psi) = B(\psi)$, the contribution of the terms $J_2(w)\cos 2\phi$ and $J_2(w)\sin 2\phi$ to the far field will be small. These terms contribute to the cross-polarization so, therefore, for low cross-polar applications, an axisymmetric feed is preferred.

DOI 10.1007/978-981-4560-75-7_30-1

As an example, Fig. 25 shows the radiation patterns obtained with a 4.05 m diameter spherical reflector that is fed with a circular waveguide of radius 0.7 λ. The aperture diameter for a total phase error is given by Eq. 99 is $D = 108.7$ cm. The focal distance has been computed from Eq. 5.20 and also 74.7 cm. The edge illumination at $\psi_c = 20.89°$ in the E- and H-planes are -4.1 and -2.1 dB respectively. Good pattern symmetry is achieved and the first sidelobe levels lie near the -20 dB level. The cross-polar level is seen to have a peak of -35 dB relative to the co-polar peak. The computed peak gain is 39.76 dBi. The same reflector with a square-aperture horn feed with a sidelength of 4.9 cm is reported to have a measured gain of 39.4 dBi and comparable sidelobe levels (Li 1959) to those predicted in Fig. 25.

Although not mentioned earlier, wide-angle scanning is possible with a spherical reflector when a suitable feed – usually with a broad main lobe – is rotated at the focal distance. The scan angle width of the 4.05 m reflector reported above (Li 1959) was found to be 70°, which is typical for a spherical reflector.

The focal region fields of a spherical reflector can be obtained in the same way as described for the paraboloidal reflector in the section "Paraboloidal Reflector." The results can be used to design a feed that best matches these fields to obtain a high efficiency. The electric field in the axial region of a spherical reflector at a point (t, ξ, z) has been derived (Thomas et al. 1969) and is expressed

$$E_\rho(t,\xi,z) = G_\rho(t)\sin\xi \tag{104a}$$

$$E_\xi(t,\xi,z) = G_\xi(t)\cos\xi \tag{104b}$$

$$E_z(t,\xi,z) = G_z(t)\sin\xi \tag{104c}$$

where

$$G_\rho(t) = -jkR_{sp}E_o\sin^2\psi_c[A(t)+B(t)]$$

$$G_\xi(t) = -jkR_{sp}E_o\sin^2\psi_c[A(t)-B(t)]$$

$$G_z(t) = -2kR_{sp}E_o\sin^2\psi_c\,C(t)$$

$$A(t) = \frac{1}{2}\csc^2\psi_c\int_0^{\psi_c}\kappa(\psi)(1+\Gamma)J_0(wt)\exp(-j\phi)d\psi$$

$$B(t) = \frac{1}{2}\csc^2\psi_c\int_0^{\psi_c}\kappa(\psi)(1-\Gamma)J_0(wt)\exp(-j\phi)d\psi$$

$$C(t) = \frac{1}{2}\csc^2\psi_c\int_0^{\psi_c}\kappa(\psi)\frac{\sin\psi}{\zeta}J_1(wt)\exp(-j\phi)d\psi$$

in which $\kappa(\psi) = \frac{\sin\psi}{\zeta}\left(1-\frac{z}{R_{sp}}\cos\psi\right)$, $\zeta = \sqrt{1+\left(\frac{z}{R_{sp}}\right)-2\frac{z}{R_{sp}}\cos\psi}$, $w = \frac{k\sin\psi}{\zeta}$, $\Gamma = \frac{\zeta\cos\psi}{\left(1-\left(z/R_{sp}\right)\cos\psi\right.}$, and $\phi(\psi) = kR_{sp}(\cos\psi+\zeta)$. A contour plot of the focal fields that are predicted by Eq. 104 is shown in Fig. 26 for a spheroid of radius 1.52 m at a distance $z = 78.7$ cm from the vertex, which corresponds to the point F in Fig. 23 where $f_{op} = 73.7$ cm. The operating frequency is 11.2 GHz. The plot clearly shows the Airy rings and that a waveguide of radius about 1.8 cm would be a good starting point for feeding the reflector. This may mean exciting a number of waveguide modes to create the precise Airy rings. For example, Thomas et al. (1969) have shown that an efficiency >80 % is possible for

DOI 10.1007/978-981-4560-75-7_30-1

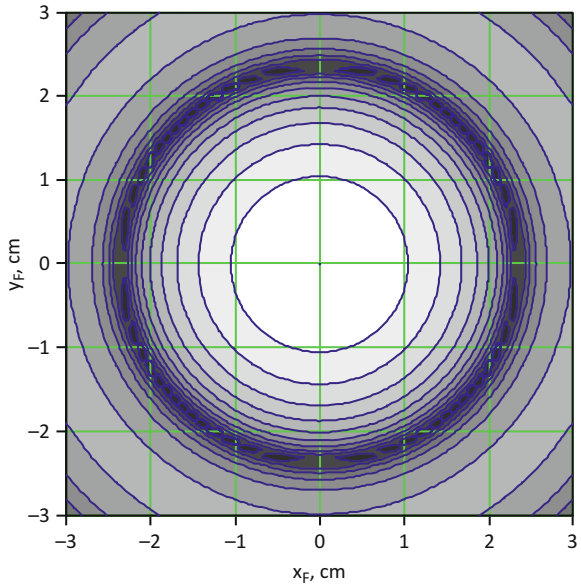


Fig. 26 Focal region fields at a distance $z = 78.7$ cm from the origin of a spherical reflector of radius 1.52 m. The operating frequency is 11.2 GHz

a spherical reflector with $R_{sp} = 400\ \lambda$ and $\psi_c = 20°$ for $z = 0.515R_{sp}$ with a feed having $ka > 15$, where α is the aperture radius, providing the correct combination of modes has been excited in the aperture.

A subreflector can be designed to completely correct for the spherical aberration (Holt and Bouche 1964) or numerically using the methods of the section "Reflector Shaping." which can also correct other deficiencies such as subreflector mismatch.

Reflector Shaping

All reflectors considered so far have been given in a closed form either as a parabola or a sphere. In this section, the objective is to achieve a desired radiation pattern through the shaping of the reflector surface. Two approaches are described. The first is a classical approach based on geometrical optics (Silver 1946). The second and more recent approach is based on numerical optimization. This latter approach is potentially more flexible because it allows the possibility of inclusion of diffraction effects, the mounting structure as well as limitations of the feed antenna. Accurate analysis methods can also be included such as physical optics, providing their implementation is fast enough for use with a standard optimization methods.

Reflector Synthesis by Geometric Optics

Reflector shaping or reflector synthesis by geometric optics methods was originally developed in the 1940s (Silver 1946) for a single reflector and has been extended to two reflectors (Galindo 1964) and later to more reflectors by continued application of the GO.

Consider the axisymmetric reflector shown in Fig. 27 for which the profile is required. The z_1-axis is taken as the axis of rotation of the reflector. The focus of the reflector is located at F a distance f from the reflector vertex, and also the reflector has a maximum dimension in the vertical direction specified by $x_{1\max}$. Let $\rho_1(\theta_1)$ be the radial distance from F to a point on the reflector at an elevation angle of θ_1 where clearly $\rho_1(0) = f$. An incident ray from the feed at F undergoes reflection and exits at an angle θ_2 to the z_1-

DOI 10.1007/978-981-4560-75-7_30-1

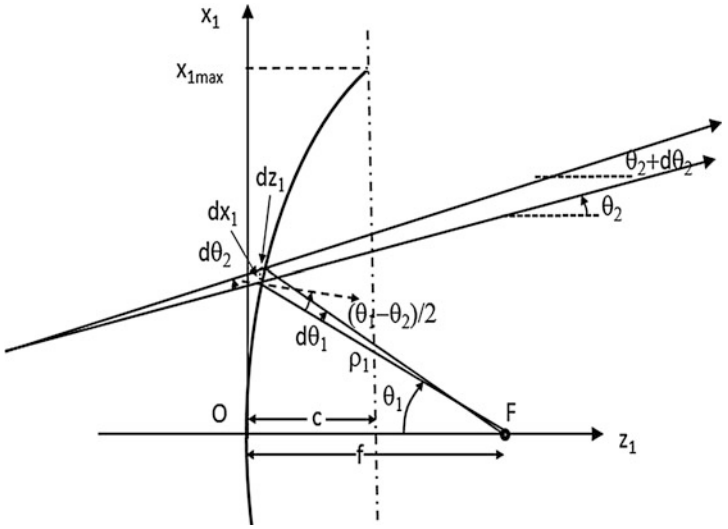


Fig. 27 Ray paths at a single reflector fed from the focus F

axis. By means of the second law of reflection, the angle between the normal to the curve and the incident and reflected rays is $(\theta_1 - \theta_2)/2$ as can be seen in Fig. 27. Application of Snell's law results in the following differential equation

$$\frac{dx_1}{dz_1} = \frac{1}{\rho_1}\frac{\partial \rho_1}{\partial \theta_1} = \tan\left(\frac{\theta_1 - \theta_2}{2}\right). \tag{105}$$

Integrating both sides of Eq. 105 with respect to θ_1 results in

$$\int_0^{\theta_1} \frac{1}{\rho_1}\frac{\partial \rho_1}{\partial \theta_1} d\theta_1 = \int_0^{\theta_1} \tan\left(\frac{\theta_1 - \theta_2}{2}\right) d\theta_1$$

and, therefore,

$$\rho_1(\theta_1) = f \exp\left[\int_0^{\theta_1} \tan\left(\frac{\theta_1 - \theta_2}{2}\right) d\theta_1\right]. \tag{106}$$

The solution of ρ_1 as a function of θ_1 through Eq. 106 gives the profile of the reflector.

As an example, if $\theta_2 = 0$ the exit ray is parallel to the z_1-axis. Setting this in Eq. 5.27 results in $\rho_1(\theta_1) = f \exp(2\ln|\sec(\theta_1/2)|)$, which simplifies to $\rho_1(\theta_1) = 2f/(1 + \cos\theta_1)$, the equation of a paraboloid.

In general, the angles θ_1 and θ_2 should be chosen to allow the energy to disperse into the secondary radiation pattern. This is done by ensuring that conservation of energy occurs from the feed to the aperture field. Incrementally, this energy is contained in the angular wedges $d\theta_1$ and $d\theta_2$. Thus, if $P(\theta_1)d\theta_1$ is the energy incident from the primary feed located at F, where $P(\theta_1)$ is the feed pattern function, and $I(\theta_2)d\theta_2$ is the reflected power, where $I(\theta_1)$ is the desired illumination function, then, for conservation of energy, it is required that

$P(\theta_1)d\theta_1 = I(\theta_2)d\theta_2$.

Therefore, over all angles this becomes

DOI 10.1007/978-981-4560-75-7_30-1

$$\int_0^{\theta_1} P(\theta_1)\sin\theta_1\,d\theta_1 = K\int_{\theta_{2\min}}^{\theta_2} I(\theta_2)\sin\theta_2\,d\theta_2 = K\int_{x_{2\min}}^{x_1} I(x_2')x_2'dx_2' \tag{107}$$

where $\theta_2 = \tan^{-1}\left(\frac{x_2-x_1}{z_2-z_1}\right)$, $\theta_{2\min}$ and $\theta_{2\max}$ are the minimum and maximum angles over which $I(\theta_2)$ is known or specified. $\theta_{2\max}$ can be chosen from almost zero through to several beamwidths. The constant K can be evaluated from Eq. 107 at the upper limits i.e., $\theta_1 = \theta_{1\max}$ and $\theta_2 = \theta_{2\max}$. This results in

$$K = \frac{\int_0^{\theta_{1\max}} P(\theta_1)\sin\theta_1\,d\theta_1}{\int_{\theta_{2\min}}^{\theta_{2\max}} I(\theta_2)\sin\theta_2\,d\theta_2}. \tag{108}$$

Equation 108 is substituted back into Eq. 107 to give

$$\frac{\int_0^{\theta_1} P(\theta_1)\sin\theta_1\,d\theta_1}{\int_0^{\theta_{1\max}} P(\theta_1)\sin\theta_1\,d\theta_1} = \frac{\int_{\theta_{2\min}}^{\theta_2} I(\theta_2)\sin\theta_2\,d\theta_2}{\int_{\theta_{2\min}}^{\theta_{2\max}} I(\theta_2)\sin\theta_2\,d\theta_2} \tag{109}$$

Equation 109 is a relationship between θ_2 and θ_1 that can be used in conjunction with Eq. 105 to determine $\rho_1(\theta_1)$. Alternatively, both $P(\theta_1)$ and $I(\theta_2)$ could be specified from measured data and a relationship obtained in order to solve Eq. 105.

A feed pattern that is quite useful for representing many practical feeds is $P(\theta_1) = \cos^n\theta$ where n is the index of the power pattern. In that case Eq. 107 becomes

$$\frac{\int_0^{\theta_1} \cos^{n+1}\theta_1\,\sin\theta_1\,d\theta_1}{\int_0^{\theta_{1\max}} \cos^{n+1}\theta_1\,\sin\theta_1\,d\theta_1} = \frac{\int_{\theta_{2\min}}^{\theta_2} I(\theta_2)\sin\theta_2\,d\theta_2}{\int_{\theta_{2\min}}^{\theta_{2\max}} I(\theta_2)\sin\theta_2\,d\theta_2}.$$

That is:

DOI 10.1007/978-981-4560-75-7_30-1

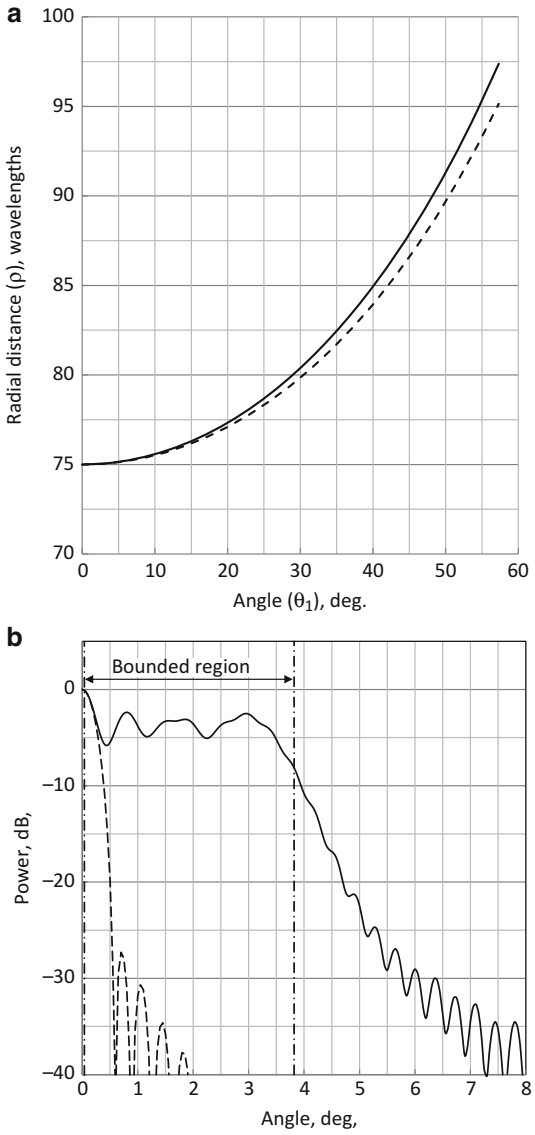


Fig. 28 Shaped reflector design for uniform illumination over the range $\lambda/10D \leq \theta \leq 10\ \lambda/D$(indicated by vertical *dashed lines*). $D = 150\ \lambda$ and $f/D = 0.5$ feed pattern index $n = 1.5$. (**a**) Reflector profile radial distance versus angle; (**b**) radiation pattern for synthesized reflector. Key – synthesized (*full curve*) and parabola (*dashed*)

$$\frac{(1 - \cos^{n+1}\theta_1)}{(1 - \cos^{n+1}\theta_{1\max})} = \frac{\int_{\theta_{2\min}}^{\theta_2} I(\theta_2)\sin\theta_2 d\theta_2}{\int_{\theta_{2\min}}^{\theta_{2\max}} I(\theta_2)\sin\theta_2 d\theta_2} \tag{110}$$

Continuing, suppose a uniform illumination is required over an angular range, i.e., $I(\theta_2) = 1$ for $\theta_{2\min} \leq \theta_2 \leq \theta_{2\max}$. Carrying out the integrations in Eq. 110 and reorganizing, the result is

DOI 10.1007/978-981-4560-75-7_30-1

$$\theta_2(\theta_1) = a\cos\left[\cos\theta_{2\min} + (\cos\theta_{2\max} - \cos\theta_{2\min})\frac{(1 - \cos^{n+1}\theta_1)}{(1 - \cos^{n+1}\theta_{1\max})}\right]. \tag{111}$$

This relation can be used in conjunction with Eq. 105 to obtain the reflector profile $\rho_1(\theta_1)$.

As an example, consider the design of the profile of reflector with a diameter of 150 λ and focal length $f = 75$ λ. A feed with $n = 1.5$ is chosen to give an edge illumination of about −7 dB. Suppose a uniform illumination function is required over the angular range of $\lambda/10 \leq \theta \leq 10\lambda/D$, i.e., in this case $0.030° \leq \theta \leq 3.8°$. The profile that is predicted by Eq. 106 along with Eq. 51 is shown in Fig. 28a. The resulting radiation pattern calculated from physical optics is shown in Fig. 28b. The average relative power level is −3.9 dB between the bounds indicated in Fig. 28b, which corresponds to an average gain of 28.7 dBi over the angular range. By comparison the profile and radiation pattern of a parabolic reflector with the same diameter and $f/D = 0.5$ is also shown in Fig. 28a, b produces a maximum gain on boresight of 52.5 dBi. The two profiles diverge as the edge approached and when $\theta_{2\max}$ is reduced the synthesized profile approaches the parabolic profile.

The approach described above is readily extended to two reflectors through the Galindo method (Galindo 1964). It was found that if an upper reflector limit equivalent to $x_{1\max}$ is chosen to be +1 (an arbitrary scale), a Cassegrain-type solution results. On the other hand, if $x_{1\max}$ is chosen as −1, a Gregorian-type reflector geometry is created, i.e., a spheroid with a concave reflector.

Reflector Synthesis by Numerical Optimization

The geometric optics approach described in the previous section has limitations as the number of physical constraints and requirements increase. The synthesis problem can be broadened to include accurate methods and more variables to be optimized through the use of numerical methods. There are several ways of doing this, both directly and indirectly. There are several direct techniques for reflector shaping that include effects such as diffraction, blockage and accurate feed modelling. One such approach is the method of successive projections. An indirect approach is to use standard numerical methods. A numerical representation of the reflector surface is used in concert with accurate radiation and feed models to meet the various system requirements. In this section a reflector synthesis approach is described which uses standard optimization methods.

A synthesis of a reflector surface by optimization is described in the flowchart shown in Fig. 29. The procedure uses a PO transmit-mode radiation-pattern analysis in conjunction with a standard optimization method. An important requirement for efficient optimization is the choice of representation of the surface and also the analysis methods so that each iteration can be quickly undertaken. One approach for representing the surface that has proved effective is with the basis-spline or B-spline functions. The degree of the spline function can be selected as required although, in many applications, a third-order basis spline has proved sufficient in both accuracy and computation time. The reflector surface $z = f(x, y)$ is represented by the B-spline series, the coefficients of which become the optimization variable in order to meet the performance requirements. A surface given in terms of B-splines of the same order in both dimensions is

$$f(x,y) = \sum_{i=0}^{m}\sum_{j=0}^{n}\alpha_{ij}B_{i,k}(x)B_{j,k}(y);\ 2 \leq k \leq m, n+1 \tag{112}$$

where $B_{i,k}(x)$ and $B_{j,k}(y)$ are standard B-spline basis functions of order k with $m + 1$ and $n + 1$ control points, respectively, and α_{ij} are the control point coefficients. A B-spline in a variable x is a piecewise polynomial function of degree $k - 1$ that is defined over a range $t_1 \leq x \prec t_m$ where $m = k + 1$. The points

DOI 10.1007/978-981-4560-75-7_30-1

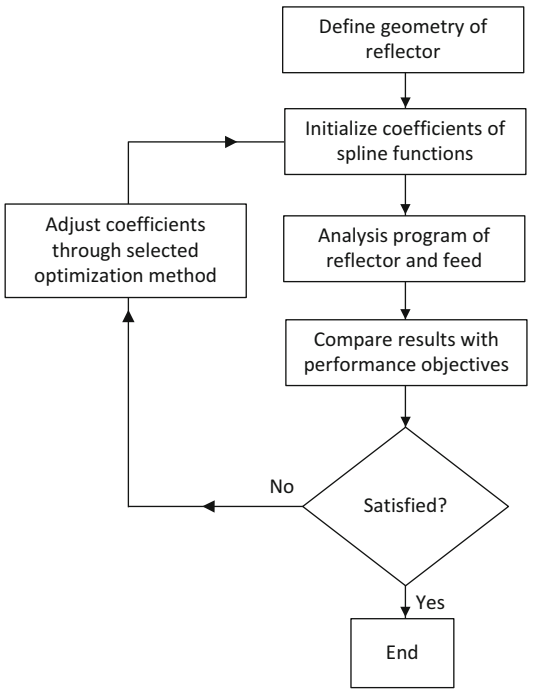


Fig. 29 Reflector shaping through numerical optimization.

where $x = t$ are known are called knots or break points, and these are arranged in ascending order. The number of knots is the minimum for the degree of the B-spline, which has non-zero values only in the range between the first and last knot. Each piece of the function is a polynomial of degree $k - 1$ between and including adjacent knots. The surface given by Eq. 5.33 has a set of $(m + 1)(n + 1)$ control points, which is in common with most other interpolation methods, except that the surface does not generally pass through the central control points. Expressions for the polynomial pieces are easily generated by means of available recursion formulae such as, for example, the Cox- de Boor formula for a k-spline which is given by

$$B_{i,k}(x) = \frac{(x - t_i)B_{i,k-1}(x)}{(t_{i+k-1} - t_i)} + \frac{(t_{i+k} - x)B_{i+1,k-1}(x)}{(t_{i+k} - t_{i+1})}$$

where t_i is the knot value at I and $t_i < t_{i+1}$.

The control point coefficients α_{ij} of Eq. 112 are the unknowns in the reflector synthesis, and in Fig. 28 they are determined by the optimization. If there is more than one reflector, all surfaces are expressed similarly to Eq. 112 and, therefore, the remaining coefficients could be included in the optimization as well.

Returning now to Fig. 29, after the coefficients of Eq. 112 have been initialised, an analysis method is used to determine the performance of the antenna with the initial surface. There are some standard reflector profiles that can be selected as a starting point for the optimization. The computed performance with the initial surface is compared with the performance objectives. These are usually included in a single objective functional for which a minimum is required. If the functional is not a minimum, the coefficients of the surface are adjusted further by the optimization method. The antenna is analysed once again with the revised coefficients and once again the performance is compared with the specifications through the functional. This process is repeated until a reflector shape is found where the computed performance meets the system requirements or a limit on the iterations is reached. The reflector-shaping procedure

DOI 10.1007/978-981-4560-75-7_30-1

could use a PO transmit-mode radiation-pattern analysis based on numerical integration of the surface current, within a search for an optimum reflector shape via variation of the coefficients, α_{ij}. A positive definite functional to be minimized by the optimizer could be sum of least-squares or a higher index between the actual and desired specifications. In addition, a combination of search methods can be useful. For instance, a genetic search algorithm has been found effective for arriving into the vicinity of a minimum of a functional and a gradient search such as a quasi-Newton search method can be usefully applied to reach the minimum in fewer functional evaluations.

Some Reflector Applications

As has been mentioned previously, reflectors occur in many areas of daily living, from communications, the weather and in understanding the wider universe such as through radioastronomy. Three examples of applications of reflector antennas in each of these areas are outlined in the following section.

Satellite Communications

Reflector antennas in satellite communications are used both on the ground and in space. The size and type of reflector vary significantly depending on the distance to the satellite and the level of available power. A satellite in geostationary orbit is located at a distance of 36,000 km from the earth while those in orbit about the earth may only be 800 km away. An earth station at a teleport of major earth station facility may have to transmit and receive to both types of satellite. The size of the reflector used in the home or at a business premises may be only 20–30 wavelengths in diameter while the earth stations used in network hubs and in satellite control may be over a hundred times larger. In this discussion the emphasis is on the latter antennas. In space, the reflector size is often determined by the size of the launch vehicle. A fully-formed shaped reflector at launch may have only a maximum diameter of 2 m whereas an unfurlable parabolic dish can be 30 m diameter or larger.

Antennas for major earth stations for international communications via lNTELSAT satellites are typically up to 32 m in diameter. They usually consist of dual shaped reflectors in an axi-symmetrical configuration. A recent trend has been towards greater use of 18 m or smaller diameter antennas for major earth stations as the power and sensitivity of the space borne antennas has increased. The major bands used by INTELSAT are the 6/4 GHz bands, where the 6 GHz band is for the uplink (transmitting) and 4 GHz for the downlink (receiving). The bandwidth in each link direction is 580 MHz, and both senses of circular polarization are used to double the communication capacity. Where a transmit beam is involved a number of additional specifications have to be met. These usually include the following:

1. The G/T on receive is important but on transmit the effective isotropic radiated power level (EIRP) in the direction of the satellite is specified depending on the type of service.
2. In both transmit and receive, the isolation in decibels (dB) between orthogonal polarizations is important. For a dual-polarized system a value of approximately ≤ -30 dB is typically specified for angles within 1 dB of the main beam.
3. The peak level of the radiation pattern sidelobes is specified for all angles greater than from boresight. This is specified as a function of $P(\theta)$ given in dBi. Most telecommunications administrations use the CCIR sidelobe recommendations and these depend on the diameter of the main reflector in wavelengths (D/λ) as follows:

(a) $D/\lambda > 100$

DOI 10.1007/978-981-4560-75-7_30-1

Fig. 30 An 11 m X-band earth station antenna that was designed for low profile and high isolation so as to meet the CCIR sidelobe specifications (Bird et al. 1995)

$$P(\theta) \leq \begin{cases} 32 - 25\log\theta\,\text{dBi}, & 1^\circ < \theta \leq 48^\circ \\ -10\,\text{dBi}, & \theta > 48^\circ \end{cases} \tag{113}$$

although in some applications the tighter specification 29-25log(theta) is used out to 48 degrees.

(b) $D/\lambda < 100$

$$P(\theta) \leq \max\begin{cases} 52 - 10\log(D/\lambda) - 25\log\theta\,\text{dBi}, & \theta > 1\text{st sidelobe} \\ -10\,\text{dBi} \end{cases}. \tag{114}$$

The latter requirement (i.e., Eq. 114) recognizes the fact that it is more difficult to achieve low sidelobe levels with small antennas.

As an example, the design of an axisymmetric dual-shaped reflector antenna with a diameter of <18 m is outlined. Figure 30 shows this dual-reflector antenna, where the feed-horn, polarizer-diplexer unit (feed-system), and low-noise receiver (LNA) are situated in the cone housing at the vertex. The antenna operates at X-band and consists of an 11 m diameter main reflector and a 1.2 m diameter subreflector. The reflector surfaces were shaped to maximize the antenna G/T and to meet the 29–25logθ CCIR sidelobe recommendation (Bird et al. 1995). The smaller reflector size than usual was specified in order to meet important environmental considerations. The antenna transmits right-hand circularly polarized signals in the frequency band 7.9–8.4 GHz and receives left-hand polarized signals in the 7.25–7.75 GHz band. The proximity of the two bands, combined with an upper limit on the antenna gain due to environmental factors, placed a stringent requirement on the isolation requirements of the feed system. In order to

DOI 10.1007/978-981-4560-75-7_30-1

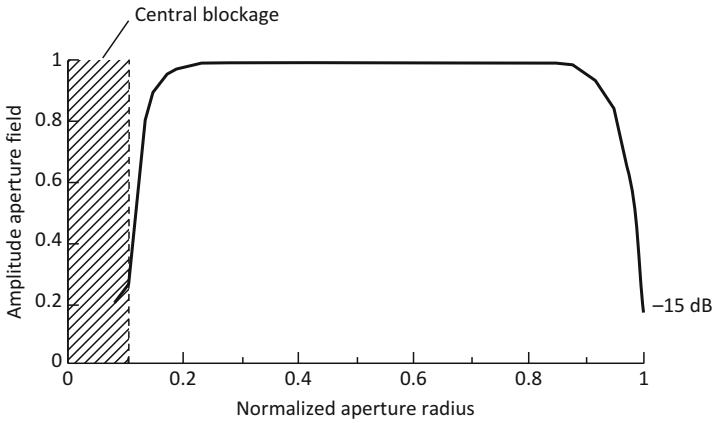


Fig. 31 Optimum aperture distribution for maximum G/T and low sidelobe levels

maximize the G/T, the transmit/receive filters had to have low insertion loss. This limits the order of the filters, and consequently the out-of-band performance. To achieve the desired high transmit/receive isolation, the feed system and the reflector had to have low reflections across the band. Low reflection from the reflector system was achieved by including a shaped matching cone at the centre of the subreflector (Wood 1980). The cone blends in smoothly with the remainder of the shaped subreflector and extends only over that the centre of the subreflector, which was already blocked by the feed. In this step, it is important to combine the feed mismatch and sidelobe designs together as the sidelobes have a large influence on the overall design.

Other improvements to existing earth stations can bring important advantages. This may be necessary to ensure that the expensive earth–station infrastructure meets any extension of the standards to permit access to any new satellites. Improvements to the input match of the feed to cover a wider bandwidth can have a significant impact. Also, as long as the new feed geometry fits the original feed enclosure, there will be minimum down-time for the earth station when the feed is replaced. Also, the optics of the antenna could be modified, for example, by replacing the usual uniform illumination over the full circle with other illuminations such as that shown in Fig. 31. By optimizing the taper at the upper edge of the illumination the spillover of energy is reduced past the edge of the main reflector and also of energy scattered by diffraction at the reflector rim. In addition, the reduction of energy in the central blocked region also minimizes the signal scattered by the secondary reflector. These changes can be included in an existing earth station by machining of the new subreflector and some re-setting of the surface panels so that any new sidelobe levels can be achieved.

Antennas on the satellite orbiting the earth operate in a very different environment and have objectives which usually differ significantly from those of earth station antennas. The communication antennas produce one or more beams, each beam defining a coverage region or “footprint” on the earth’s surface. Signals are either transmitted to or received from these footprints. The type of beams generated may be a single pencil, or a spot, beam over a major city, or a contoured beam designed to follow the boundary of a region or country, as illustrated in Fig. 32. An extreme case of this is a simple circular beam of angular width 17.4° to cover the entire earth from the geostationary orbit.

Apart from making the best use of the available satellite transmitter power, a contoured-beam antenna reduces interference to adjacent coverage regions and makes possible a more uniform coverage in the desired region. The isolation achieved between neighboring regions allows a common frequency band to be assigned, thus conserving the available frequency spectrum.

Ideally, a single shaped beam antenna produces the desired beams, both on transmit and receive. This may not be possible, and as a result transmit and receive functions are separated, which can be an

DOI 10.1007/978-981-4560-75-7_30-1

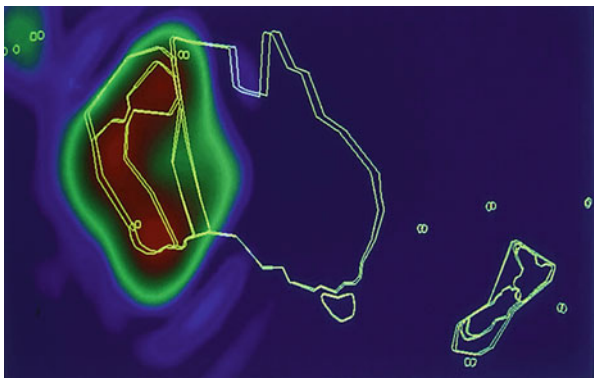


Fig. 32 Contoured beam design for Western Australia with a low-level illumination of the Cocos Islands

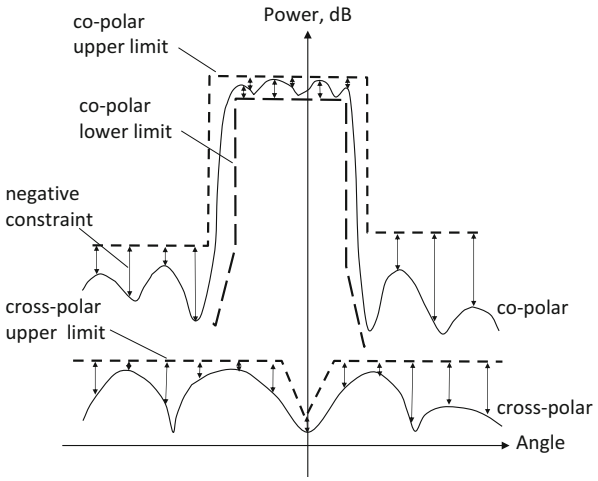


Fig. 33 Outline of constraints on a pattern for optimization through reflector shaping

advantage for isolation between bands. Currently, for each frequency band, a pair of antennas is used to generate all desired spot and area coverage beams, in two orthogonal polarizations. The type of antenna most commonly used for producing spot and contoured beams is a shaped offset reflector, which enables the feed and any beamforming network to be located close to the body of the spacecraft.

The procedure used to design shaped beams by numerical optimization is similar, irrespective of whether an array of feeds or reflector shaping is used. At first it is necessary to define maximum and minimum power levels at selected points inside the coverage region, and maximum levels outside of it, as depicted in Fig. 33. These levels define an envelope within which the radiated power must be contained.

It is important to know the relationship between the radiated power density in the defined area and the geometry of the reflector. This information is obtained by analysing the shaped reflector as described in the chapter on "► Applications of Phased Array Feeders for Reflector Antennas". The reflector surface or the feed excitation coefficients are adjusted by the optimizer so that the constraints on the power pattern could be satisfied. This is carried out by means of computer optimization. It may be impossible to determine all the unknowns automatically and some parameters may need to be selected on the basis of experience.

DOI 10.1007/978-981-4560-75-7_30-1

Fig. 34 A paraboloid reflector with high performance dual – polarized feed for a radar application (Hayman et al. 1998)

Table 2 Summary of dual-polarized meteorological radar reflector

Antenna type	Prime focus fed paraboloid
Diameter (D)	5.2 m
Focal ratio (f/D)	0.34
Beamwidth	0.9°
Gain	45.5 dBi
Sidelobes	<-21 dB
Cross-polarization	<-32 dB
Frequency	5,600–5,650 MHz

Dual-Polarized Weather Radar

Another important application for reflectors is in weather radar. The radar beams usually are required to have low-level sidelobes to avoid the clutter from signals scattered from the ground. In applications involving dual-polarization, high isolation is required between the polarization of two orthogonally polarized channels. The latter may be addressed with a shaped reflector but in some instances it can be addressed through a new feed system with high isolation. As an example of the latter, there has been an increased use of dual polarization in meteorological radars. The additional data provided by dual polarization has improved understanding of clouds and precipitation. For instance, droplet size can be estimated using the difference between the vertical and horizontal reflectivities, which can be assessed from radar returns of dual-polarized signals. Larger droplets are more oblate and will have a greater differential reflectivity than smaller droplets.

One approach that is sometimes adopted is the upgrade of an existing single-channel Doppler radar antenna to operate in two linear polarizations. In a dual-linear polarization approach the important measurement variables are horizontal (Z_{HH}) and vertical radar reflectivities (Z_{VV}). These are defined by

DOI 10.1007/978-981-4560-75-7_30-1

Fig. 35 Reflectors antennas in the Australian SKA pathfinder (ASKAP) (Courtesy CSIRO). This array consists of 36 12 m-diameter antennas, which are fed with 188 element phased array feeds, operate over a frequency range from 0.7 to 1.8 GHz (Picture courtesy CSIRO Australia)

$$Z_{pp} = \frac{32}{k^5|\chi|^2}\int \sigma_{pp} N(D) dD \tag{115}$$

where p is either the horizontal (H) or vertical (V), σ_{pp} is the radar cross section for either co-polar state, $N(D)$ is the particle-size distribution, and $\chi = (\varepsilon_r - 1)/(\varepsilon_r + 1) - 1$ where ε_r is the dielectric constant of the particle. Of importance is the differential reflectivity, which is defined by

$$Z_{DR} = 10\log\left(\frac{Z_{HH}}{Z_{VV}}\right) \tag{116}$$

and the linear depolarization ratio

$$LDR = 10\log\left(\frac{Z_{HV}}{Z_{VV}}\right). \tag{117}$$

Therefore, to accurately measure all parameters, Eqs. 115, 116, and 117, it was important for the upgraded antenna to have low cross-polarization relative to the co-polarized in both vertical and horizontal planes. A way of ensuring this is to have a symmetrical feed system with low cross-polarization in both diagonal planes. It is useful also to have similar antenna pattern in V- and H-directions to ensure Z_{DR} is a true measure of the differential reflectivity.

An existing single-channel Doppler radar was upgraded by introducing a dual-polarized feed that would give overall low cross-polarization. The antenna consisted of a paraboloid reflector and rotating mount. In the upgrade it was decided to design a new high performance axially symmetric feed with a ring slot flange (shown in Fig. 18c). As well, the two orthogonally polarized signals were transmitted from the same source with a short time delay, and, therefore, the upgrade included a high-power-microwave switch, and also an orthomode transducer (OMT) with high isolation between transmit and receive ports. The resulting reflector antenna is shown in Fig. 34 (Hayman et al. 1998). The target performance for the

DOI 10.1007/978-981-4560-75-7_30-1

antenna was specified by the end-user and included well matched co-polarized beam patterns, low sidelobes, low cross polarization taken in the diagonal planes and high isolation between channels. It was found for the feed that a worst case cross-polar isolation of −28 dB in the diagonal planes was sufficient to meet the required isolation in the V- and H-directions. The overall performance obtained from the antenna system is summarized in Table 2. The antenna has been used for a number of years in regions north of Australia to monitor tropical thunderstorms to use polarimetric measurements to provide quantitative rainfall data.

Radio Astronomy

As in astronomy at optical frequencies, microwave reflectors find significant use is in radio astronomy. There are a number of major radio astronomy observatories around the world. For observations above 800 MHz, reflectors are invariably involved. They can vary in size from around 20 m to over 300 m for the Arecibo spherical reflectors. Sometimes they can involve basic geometries such as a paraboloid, sphere or Cassegrain. Some antennas such as the Australia telescope array use a shaped dual reflector system. There are advantages and disadvantages of reflector shaping. One is that shaping can increase the gain and hence sensitivity for a point focus and use with a single feed. Reduced sidelobes and interference from nearby directions can be accommodated in the design as well. However, such a design may limit the possibilities for improvement by methods such as an extended array feed in the focal. Such feeds will have very poor gain performance for off-axis feeds than say a paraboloid or a classical Cassegrain antenna. The development of the international square kilometre array (SKA) observatory has stimulated much work on reflectors both in design and in manufacture as many antennas of a particular type need to be fabricated. Techniques being trialed vary from standard approaches with shaped aluminium panels and steel construction to a combination of steel backing structure and a carbon fibre reflector surface to reduce weight. There is competition between standard reflector shapes and some shaping. A range of feed options are available and this may ultimately resolve this as well as adaptability in the future. Figure 35 shows the Australian SKA pathfinder (ASKAP) system, which uses low-cost, low-power reflectors and phased array feeds. The 36 12 m-diameter antennas provide a 30 square degree field-of view of the southern sky over a frequency range from 0.7 to 1.8 GHz. The digital beamformer, which has a significant 300 MHz bandwidth, also has the potential for zooming the beam.

An earlier development was the introduction of array feeds in radio astronomy. In the first instance with an array of N elements, it enabled the data collection time for significant survey of the sky to be reduced by a factor of 1/N. In addition, they enabled through cross-correlation techniques to uncover additional information about the region of the sky. Phased array feeds became possible with the use of beamforming techniques that did not significantly contribute to the noise temperature. Details of both approaches with individual and phased array feeds are given in a chapter of this book entitled "► Applications of Phased Array Feeders for Reflector Antennas".

Conclusion

This chapter is an introduction to the theory and practice of reflector antennas. The material provided could find use in an advanced course in antennas at the advanced undergraduate or postgraduate level as well as for the practising telecommunications engineer. Also, the fundamentals are provided to allow the reader to understand the current literature on reflectors and to consider making initial progress on their own either in research or engineering design. The brief history shows that many of the basic reflector

DOI 10.1007/978-981-4560-75-7_30-1

geometries have been known from ancient times. However, many improvements have been made in the last 80 years. These vary over a considerable range of topics and include new reflector geometries as well as shaping, methods of feeding, computational methods, integrated design, correction of imperfections that result from manufacturing or misalignment and, of course, the materials used.

Advances in reflector system design have been steady, and this has often come with improvements in computational tools, which have been fundamental to many of the developments mentioned in the previous paragraph. With improved speed and accuracy of computation, there is no less reliance now on measured results than in earlier times. However, measured results may indicate unexpected difficulties due to certain assumptions, and, of course, they are important for confirming the accuracy or otherwise of the modelling. There are some specific algorithms for reflector shaping, but most of these use assumptions about the geometry; e.g., the geometrical optics method covered in this chapter neglects edge diffraction. Integrated design comes through the use of rigorous methods together to produce a specific design. For example, this may involve both reflector as well as horn profile shaping, both of which take into account interactions with a support structure. Correction of reflector misalignments, possibly over time, through adjustment of the structure is also possible through techniques such as in situ sensors and measurement processes of an operating reflector system. Finally, and most importantly, the progress being made in other areas with advanced materials, both physical and electromagnetic-based structures, opens up the potential of lightweight, rigid structures that are less prone to manufacturing tolerances than in the past. For example, fibre glass and carbon fibre reflectors have been in use for many years in several applications, but they have proved less reliable than might have been expected. Nevertheless, the fact that normally unwieldy reflectors made lightweight can alter the cost structure of the whole system particularly with the possibility of the use of less expensive towers is an important option. Taking just some of these possibilities into account indicates that the area of reflector antennas will continue into the future to be an exciting field to work in.

Cross-References

► Applications of Phased Array Feeders for Reflector Antennas

References

Ashmead J, Pippard AB (1946) The use of spherical reflectors as microwave scanning aerials. J Inst Elec Eng 93(pt. III-A):627–632

Balanis CA (1982) Antenna theory. Harper & Sons, New York

Bird TS, James GL (1999) Design and practice of reflector antennas and feed systems in the 1990s. In: Stone WR (ed) The review of radio science 1996–1999. URSI, Oxford University Press, New York, Chap. 4

Bird TS, Love AW (2007) Horn antennas. In: Volakis J (ed) Antenna engineering handbook, 4th edn. McGraw-Hill, New York, Chap. 14

Bird TS, Sprey MA, Greene KJ, James GL (1995) A circularly polarized X-band feed system with high transmit/receive port isolation, IEE conference antennas propagation ICAP, 4–7 Apr, Eindhoven, pp 322–326

Galindo V (1964) Design of dual-reflector antennas with arbitrary phase and amplitude distributions. IEEE Trans Antennas Propag AP-12:403–408

DOI 10.1007/978-981-4560-75-7_30-1

Hayman DB, Bird TS, James GC (1998) Fresnel-zone measurement and analysis of a dual-polarized meteorological radar antenna. AMTA'98, Montréal, 26–30 Oct, pp 127–132

Hertz H (1893) Electric waves, English translation by D.E. Jones, Macmillan, London, Chap. XI

Holt FS, Bouche EL (1964) A Gregorian corrector for spherical reflectors. IEEE Trans Antennas Propag AP-12:223–226

Li T (1959) A study of spherical reflectors as wide-angle scanning antennas. IRE Trans Antennas Propag AP-7:44–47

Olver AD, Clarricoats PJB, Kishk AA, Shafai L (1994) Microwave horns and feeds. IEEE Press, New York

Rudge (1975) Multiple-beam antennas: offset reflectors with offset feeds. IEEE Trans Antennas Propag AP-23:224–239

Rudge AW, Adatia NA (1975) A new class of primary-feed antennas for use with offset parabolic-reflector antennas. Electron Lett 11:597–599

Rusch WVT (1984) The current state of the reflector antenna art. IEEE Trans Antennas Propag AP-32:313–329

Ruze J (1966) Antenna tolerance theory – a review. Proc IEEE 54:633–640

Silver S (1946) Microwave antenna theory and design, first published by McGraw-Hill Book, New York. Reprint published by Peter Peregrinus Ltd., London, 1985

Wood PJ (1980) Reflector analysis and design. Peter Peregrinus, London

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_33-2

Spiral, Helical, and Rod Antennas

Hisamatsu Nakano* and Junji Yamauchi
Faculty of Science and Engineering, Hosei University, Koganei, Tokyo, Japan

Abstract

Various antennas radiating a circularly polarized wave are reviewed on the basis of numerical techniques. The method of moments, the finite-difference time-domain method, the beam-propagation method, and the finite element method are employed to clarify the operation principles of spiral, helical, and rod antennas. A novel type of spiral antenna (a metamaterial-based spiral antenna) can provide a dual-band counter-circularly polarized wave. Since the operation principle of an endfire-mode helical antenna is very close to that of a dielectric rod antenna, the so-called discontinuity-radiation concept of a surface-wave antenna is discussed using a dielectric rod. Modification of the dielectric rod is studied to achieve a higher gain. A rod antenna based on an artificial dielectric is also investigated using a periodic structure of circular metal disks.

Keywords

Array antenna; Artificial dielectric; Circularly polarized wave; Dielectric antennas; Discontinuity-radiation concept; Dual circular polarization; Endfire antennas; Low-profile antennas; Metamaterial-based transmission line; Primary feed; Surface-wave antennas; Tapered rod; Thin-wire antennas; Traveling wave antennas; Wideband antennas

Introduction

One of the broadband radiators of a circularly polarized (CP) wave is a spiral antenna, whose geometry is classified into Archimedean and equiangular types. An almost complete frequency independency can be realized by the equiangular spiral (Mushiake 1996; Rumsey 1966), and similar wideband characteristics are also achieved by an Archimedean spiral (Kaiser 1960). In addition to the frequency independency, its advantage of a low-profile structure encourages us to employ the spiral for a space-constraint case, such as mobile and spacecraft applications. Although a slot spiral also operates as an efficient radiator of the CP wave, an emphasis is primarily put on a spiral antenna consisting of conducting wires. In section "Spiral Antennas," the wideband operation of the Archimedean spiral antenna is discussed on the basis of the numerical analysis using the method of moments (Harrington 1968; Mei 1965; Nakano 1987), followed by the introduction of a metamaterial-based spiral antenna of a left-handed property (Nakano et al. 2011, 2013).

A CP wave can also be efficiently radiated using a helical structure. There are several types of helical antennas; Kraus' helix invented in 1947 (Kraus and Marthefka 2002) can radiate the CP wave toward the endfire direction. The basic characteristics of an endfire helical antenna will be summarized in section "Helical Antennas." Note that the endfire helical antenna can also be categorized into a surface-wave

*Email: hymat@hosei.ac.jp
*Email: nakano@hosei.ac.jp

DOI 10.1007/978-981-4560-75-7_33-2

antenna such as a dielectric rod antenna, since the helical wire supports a traveling wave current with almost constant amplitude. Accordingly, the operation principle of both antennas are similar to each other, i.e., the radiation pattern can be viewed as a combination of direct radiation from the feed and surface-wave radiation from the transverse plane passing through the terminal discontinuity (Collin and Zucker 1969). In section "Rod Antennas," various rod antennas, constructed of a dielectric material or an artificial dielectric such as a corrugated-metal surface, will be reviewed. The finite-difference time-domain (FDTD) method (Taflove and Hagness 2005) is used for the numerical investigation. Section "Conclusion" concludes this chapter with the obtained results.

Spiral Antennas

A spiral antenna is known as a frequency-independent antenna which provides essentially constant directivity, gain, beamwidth, and impedance across a broad frequency range. For airborne applications, a compact, flush-mounted antenna is absolutely needed together with a wide CP beam across a broad bandwidth. A flat spiral antenna meets all of these requirements, so that the spiral antenna becomes irreplaceable components of many communication platforms. In the following, the basic characteristics of the Archimedean spiral antenna will be discussed. A finding is also provided on a new type of spiral antenna, so-called metamaterial-based spiral antenna (metaspiral), which is capable of radiating right- and left-handed CPs at different frequencies.

Archimedean Spiral Antenna

The radiation mechanism of a spiral antenna floating in free space was first qualitatively explained using current band theory (Kaiser 1960) and later quantitatively described using the current distribution (Nakano 1987). A round Archimedean spiral arm is defined by $r = a\phi$, where r is the radius from the spiral center, ϕ is the winding angle measured in radians, and a is the constant which controls a spiral pitch. A rectangular Archimedean spiral is also defined, as will be mentioned in the next subsection.

The current band theory is created on the basis of the fact that the two-wire spiral behaves as if it were a two-wire transmission line which gradually transforms itself into a radiating structure. For example, neighboring current elements start antiphase at the feed points and gradually come into phase as they proceed outward along the two-wire spiral line. These currents are precisely in phase when r is $\lambda/2\pi$ (λ: the wavelength at the operating frequency). It follows that radiation from the spiral is centered in an annular ring of turns of one wavelength mean circumference, often called an active region. This radiation is termed the first mode, since this represents the first occasion for which conditions are correct for radiation.

The spiral radiates a CP wave in the two directions normal to the spiral plane (bidirectional radiation), due to the outgoing current decaying from the feed point to the arm ends. The antenna exhibits almost constant input impedance across a wide frequency range. For practical applications, the bidirectional radiation is transformed into unidirectional radiation by placing a conducting reflector (or a cavity) behind the spiral. Generally, as the distance between the spiral and the reflector (antenna height) is decreased, the inherent wideband antenna characteristics degrade. However, this degradation can be mitigated by connecting resistors to the antenna arm ends or by placing an absorbing strip (ABS: see Fig. 1) behind the outermost portion of the arms. A spiral antenna having an antenna height of less than $\lambda/10$ attains a frequency bandwidth of more than 1:3, satisfying both a 3-dB axial ratio criterion and a VSWR = 2 criterion.

One of the interesting and practical features of the spiral antenna lies in generation of higher modes. In contrast to the conventional first mode with maximum radiation toward the axial direction (axial beam),

DOI 10.1007/978-981-4560-75-7_33-2

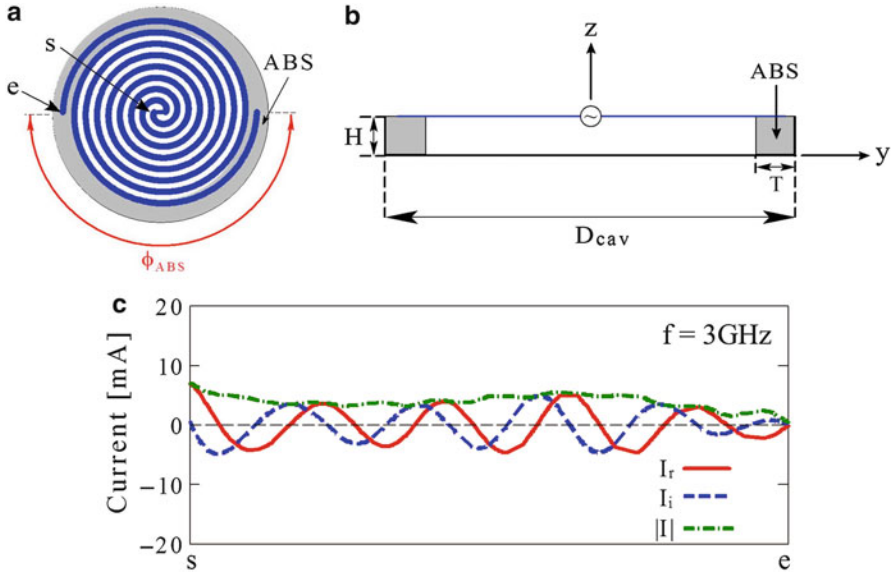


Fig. 1 Spiral antenna backed by a shallow cavity. (**a**) Perspective view. (**b**) Side view. (**c**) Current distribution

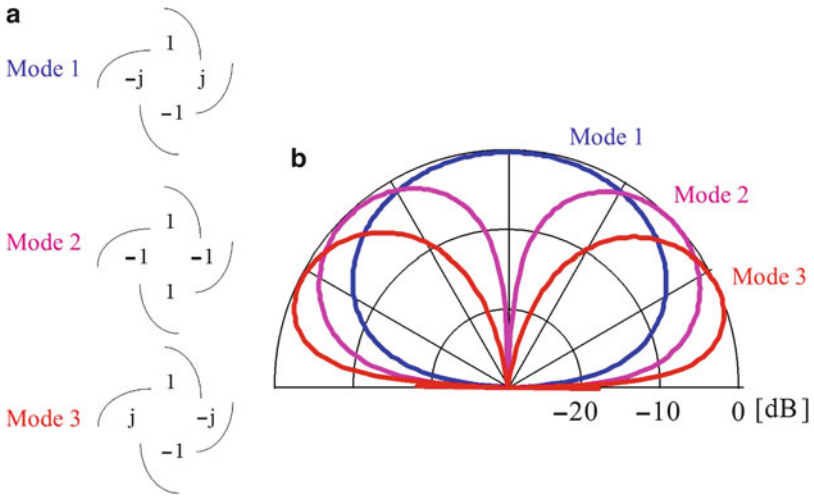


Fig. 2 (**a**) Excitation modes of a four-arm spiral. (**b**) Far-field patterns for respective modes versus elevation

the higher modes, such as the second and third modes, generate a conical beam. For example, the second mode can be generated when the two arms are excited in phase. The current band theory indicates that the axial beam is suppressed when the two arms are excited in phase. Note, however, that all the modes are active when a single-arm spiral is employed. Care, therefore, should be taken in the beam tilt for the single-arm spiral, since the conical beam generated from the second mode has the opposite phase relation with respect to the *z*-axis. Application of a single-arm spiral to a dual-band counter-circularly polarized wave will be described in the next subsection.

Another way of efficiently realizing the higher modes is to use a multi-arm spiral. As an example, the excitation condition of a four-arm spiral is illustrated in Fig. 2a. It should be noted that the first mode spiral generates a complete CP wave toward the axial direction due to the 90° phase difference in each arm. Furthermore, the third mode also generates a conical beam, like the second mode, but the field phase is the same with respect to the *z*-axis (Fig. 2b). Owing to the phase difference, the combination of the first and

DOI 10.1007/978-981-4560-75-7_33-2

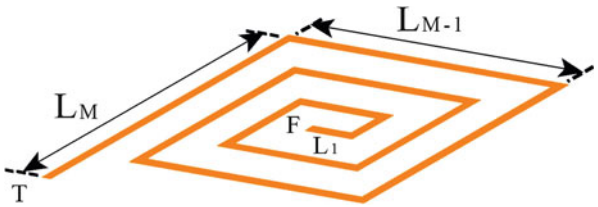


Fig. 3 Spiral antenna composed of a single continuous arm. The number of straight filaments is M, and the filament lengths are $L_1, L_2, \ldots, L_M$. The coordinate origin is denoted as F

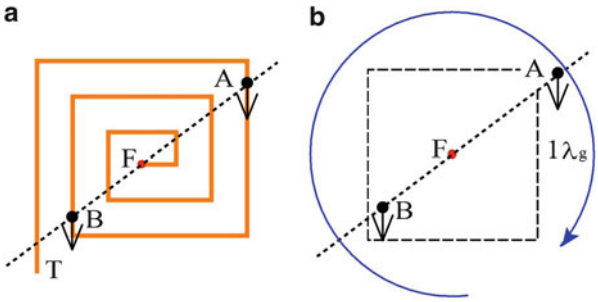


Fig. 4 TF case. (**a**) In-phase current elements. (**b**) LH CP radiation in the forward direction (z-direction). The length of the *dotted ring line* is one guided wavelength at frequency f_{LH}. The coordinate origin is denoted as F

second modes can be applied to radar direction finding of an incoming wave (Volakis 2007; Penno and Pasala 2001).

Metamaterial-Based Spiral Antenna (Metaspiral)

The rotational sense of the CP radiation from the conventional spiral antenna presented in the previous subsection is uniquely determined depending on the winding sense of the wire (single CP radiation). It is not possible for the conventional spiral to radiate a left-handed CP wave at a specific frequency f_{LH} and a right-handed CP wave at a different frequency f_{RH} ($f_{LH} \neq f_{RH}$). This limitation can be overcome when a new concept, a left-handed property (Nakano et al. 2011, 2013), is introduced.

Figure 3 shows a single-arm spiral antenna. The continuous straight strip filaments constituting the spiral arm have lengths L_1, L_2,. . ., L_M starting from the center. Let the current along the arm smoothly travel in only one direction, either from point T to point F (the coordinate origin) or from point F to point T, and that there is no reflected current.

First, a situation is considered where the current flows from point T to point F at frequency f_{LH}. This situation is designated as the TF case. Let points A and B be on the arm, both being on a straight line passing through the coordinate origin F and being quasi-point-symmetric with respect to point F (see Fig. 4a). If the path length along the arm from point A to point B is $\lambda_g/2$ (with λ_g being the guided wavelength), then the current elements at points A and B, illustrated by the arrows, are spatially in the same direction (i.e., the current elements are in phase), because of a 180° phase shift from point A to point B. As a result, the fields generated from these two current elements are added in the forward direction. The in-phase current elements travel along the arm near a square loop region of $1\lambda_g$ on the spiral plane, shown by a dotted line in Fig. 4b, and hence the polarization of the resultant field is LH in the $+z$ space.

Three facts are noted here: (1) current elements away from $1\lambda_g$ square loop region make less of a contribution to the resultant field, because they are not in phase; (2) the in-phase current elements are not located symmetrically with respect to the coordinate origin F due to the asymmetry of the spiral structure,

DOI 10.1007/978-981-4560-75-7_33-2

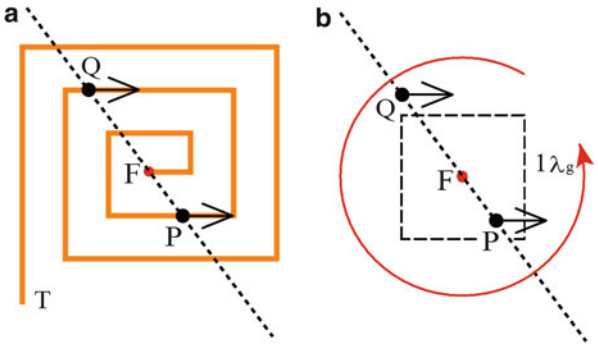


Fig. 5 FT case. (**a**) In-phase current elements. (**b**) RH CP radiation in the forward direction (*z*-direction). The length of the *dotted ring line* is one guided wavelength at frequency f_{RH}. The coordinate origin is denoted as F

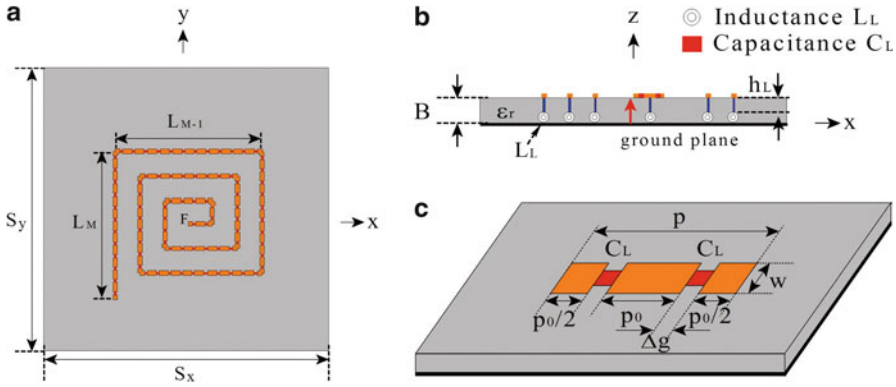


Fig. 6 Metamaterial-based single-arm spiral antenna (1-MTM-SPA). (**a**) Top view. (**b**) Side view. (**c**) Unit cell with inductance L_L and capacitance C_L loading

and hence, the radiation is not exactly symmetric with respect to the *z*-axis; and (3) the current experiences a *regressive* phase shift as it travels along the arm from point T to point F; conversely, the current experiences a *progressive* phase shift from point F to point T, i.e., the phase constant of the current along the arm, β, is negative with respect to the coordinate along the arm starting at point F and ending at point T (let us call this the arm coordinate).

Next, a case is considered where the current travels smoothly from point F to point T, opposite to the TF case. This is designated as the FT case. Figure 5a shows a $1\lambda_g$ loop region at frequency $f_{RH}(\neq f_{LH})$, where points P and Q are quasi-point-symmetric with respect to the coordinate origin F. Like the TF case, the current element arrows at P and Q are spatially in phase. These in-phase current elements rotate, producing CP radiation, whose rotational sense is RH in the forward direction. Note that the phase shift with respect to the arm coordinate (starting at point F and ending at point T along the arm, as already defined) is regressive, i.e., the phase constant β is positive with respect to the arm coordinate.

The aforementioned TF case (Fig. 4) is realized by locating the feed point at point T, while the FT case (Fig. 5) is realized by locating the feed point at point F. It follows that a switching circuit is needed to select the feed point for counter-CP radiation; it is not possible for a spiral having a single *fixed* feed point to realize counter-CP radiation. In order for the spiral with a single *fixed* feed point to realize dual-band counter-CP radiation, the phase constant β with respect to the arm coordinate must have a negative value within a specific frequency band and a positive value within a different frequency band. Based on this requirement for the phase constant, a novel spiral arm is introduced in the following.

DOI 10.1007/978-981-4560-75-7_33-2

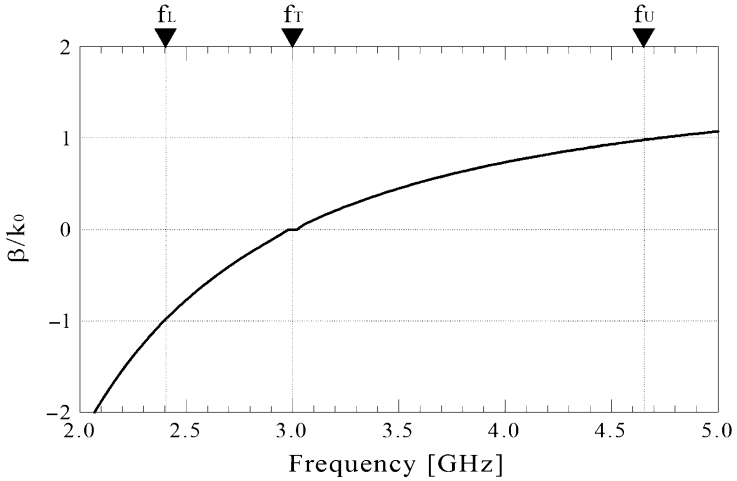


Fig. 7 Dispersion diagram for a unit cell. The notation f_L and f_U are the lower and upper bound frequencies for a fast wave, respectively

Table 1 Parameters

	Symbol	Value
Cell	ϵ_r	2.6
	B	1.6 mm
	w	2 mm
	p	10 mm
	p_0	4 mm
	Δg	1 mm
	ρ	0.5 mm
	h_L	0.6 mm
Ground plane	s_x	110 mm
	s_y	110 mm
LH elements	C_L	1.07 pF
	L_L	3.74 nH
Transition frequency	f_T	3.0 GHz

As shown in Fig. 6, the antenna arm shown in Fig. 3 is modified, fixing the feed point at the innermost point F. Each filament is composed of numerous strip conductors and is printed on a dielectric substrate (of thickness B and relative permittivity ϵ_r) backed by a conducting ground plane having an area of $s_x \times s_y$. The cell, defined by length $p(=2p_0 + 2\Delta g)$ and width w, has a conducting pin (radius ρ) extending toward the ground plane. The p is designated as the cell periodicity. A shunt inductance is inserted between the end of the pin and the ground plane, and a capacitance is inserted between neighboring arm conductors. Thus, the spiral arm has a metamaterial (left-handed) property (Engheta and Ziolkowski 2006) together with its inherent right-handed property. This single-arm metamaterial-based spiral antenna is abbreviated as the 1-MTM-SPA. Note that the outermost point T is terminated with a Bloch impedance (Collin 1966).

Figure 7 shows the dispersion diagram for a unit cell, where $\beta(=2\pi/\lambda_g)$ is the phase constant of the current along the spiral arm and $k_0(=2\pi/\lambda)$ is the phase constant in free space. The lower and upper bound frequencies for a fast wave are marked by f_L and f_U, respectively, where $\beta/k_0 = -1$ at f_L and $\beta/k_0 = +1$ at f_U. The parameters used for this dispersion diagram are shown in Table 1, leading to a transition frequency

DOI 10.1007/978-981-4560-75-7_33-2

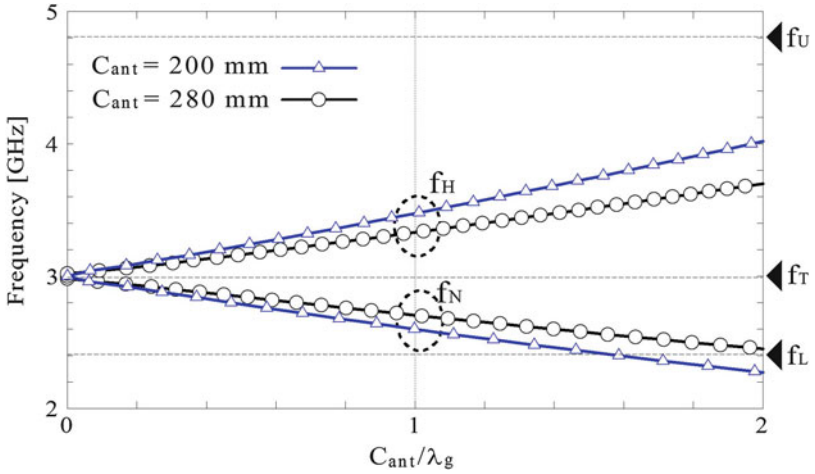


Fig. 8 Normalized antenna size C_{ant}/λ_g, where f_N and f_H denote the N and H frequencies, respectively, at which $C_{ant}/\lambda_g = 1$ is satisfied

of 3 GHz (Eleftheriades and Balmain 2005), which is denoted as f_T. Note that negative and positive phase constants for dual-band counter-CP radiation are prepared below and above f_T, respectively. Also note that the cell periodicity p in Table 1 is chosen to be less than $\lambda_g/4$ across a fast-wave frequency band of f_L to f_U, taking into account the homogeneity condition of a composite right- and left-handed transmission line (CRLH TL) (Caloz and Itoh 2006).

A square loop of $1\lambda_g$ (see the dotted lines in Figs. 4b and 5b) acts as an active region for CP radiation. Therefore, the 1-MTM-SPA must have an area sufficiently large to support this active region. For this, it is required that the last filament length L_M be larger than $\lambda_g/4$ so that the antenna size (defined by peripheral length $C_{ant} = 4L_M$) is larger than $1\lambda_g$.

Figure 8 shows the antenna size normalized to the guided wavelength, C_{ant}/λ_g, as a function of frequency, with the last filament length L_M as a parameter. A frequency that is below the transition frequency and satisfies $C_{ant}/\lambda_g = 1$ is designated as the N frequency (f_N), and a frequency that is above the transition frequency and satisfies $C_{ant}/\lambda_g = 1$ is designated as the H frequency (f_H). It is found that increasing the antenna size C_{ant} leads to a narrowing of the separation between f_N and f_H.

Based on the physical meanings of f_L, f_N, f_H, and f_U, the following four predictions are made: (1) the 1-MTM-SPA will exhibit maximum gain for LH CP radiation at a frequency between the lower bound frequency f_L and the N frequency; (2) the 1-MTM-SPA will exhibit maximum gain for RH CP radiation at a frequency between the H frequency f_H and the upper bound frequency f_U; (3) as the frequency is increased from the N frequency toward the transition frequency ($f_T = 3$ GHz), the gain for the LH CP radiation decreases, due to the disappearance of the active region for CP radiation, resulting from an increase in the guided wavelength; and (4) conversely, as the frequency is decreased from the H frequency toward the transition frequency f_T, the gain for the RH CP radiation decreases, again due to the disappearance of the active region for CP radiation.

For confirming the predictions mentioned above, a 1-MTM-SPA is analyzed, where a small number of filaments ($M = 12$) are used for the spiral arm to reduce the computation time. The filament lengths are denoted as $L_{2n-1} = L_{2n} = nL_1 (n = 1, 2 \ldots)$, where $L_1 = 10$ mm, and result in an $f_N = 2.7$ GHz and $f_H = 3.4$ GHz (note: $f_L = 2.4$ GHz and $f_U = 4.8$ GHz from Fig. 7).

Figure 9 shows the gain in the z-direction at frequencies below $f_T = 3$ GHz, which is obtained using the finite element method (HFSS 2015). Note that G_L and G_R denote the gains for LH CP radiation and RH CP radiation, respectively. These gains do not include the effect of input impedance mismatch. The actual gain is obtained by reducing the amount of input impedance mismatch from the G_L and G_R. For example,

DOI 10.1007/978-981-4560-75-7_33-2

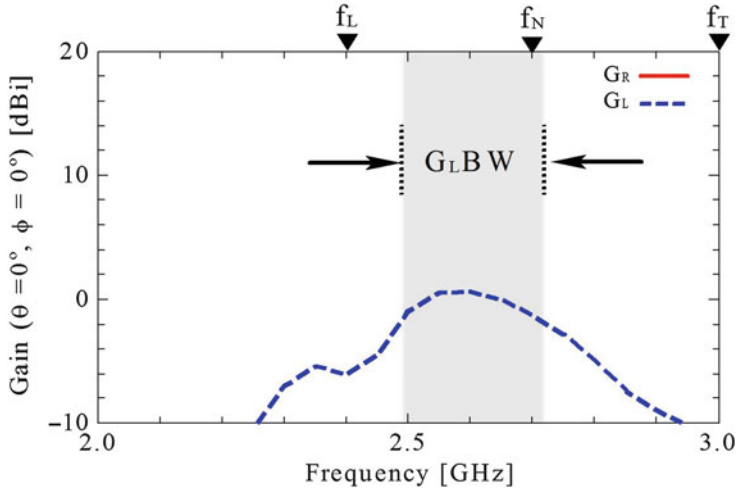


Fig. 9 Gain in the *z*-direction below $f_T = 3$ GHz. The LH CP radiation is dominant. G_LBW denotes a 3-dB gain reduction bandwidth for LH CP radiation. G_R is less than −10 dB and not illustrated

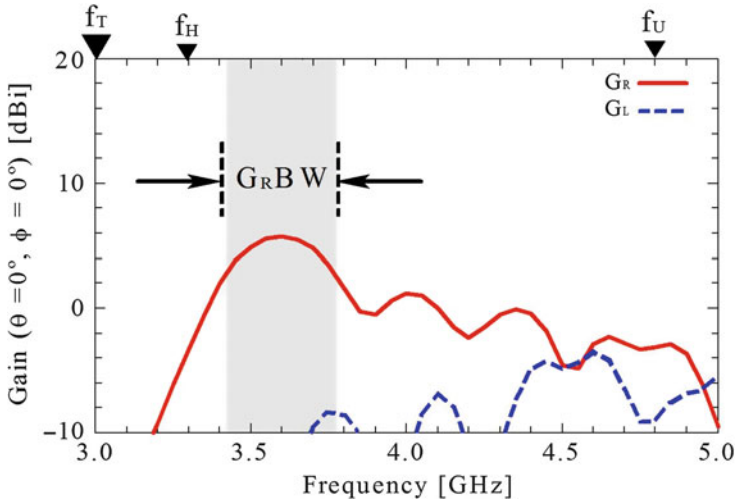


Fig. 10 Gain in the *z*-direction above $f_T = 3$ GHz. The RH CP radiation is dominant. G_RBW denotes a 3-dB gain reduction bandwidth for RH CP radiation

the amount of mismatch for VSWR = 2 is 0.51 dB. As predicted, the radiation at frequencies between the lower bound frequency f_L and the transition frequency f_T is dominated by LH CP radiation (G_R is less than −10 dB and is not illustrated), and the maximum gain of G_L for LH CP radiation, G_{Lmax}, appears between the lower bound frequency f_L and the N frequency f_N.

The prediction that as the frequency is increased from the N frequency f_N toward the transition frequency $f_T = 3$ GHz, the gain G_L will decrease is also confirmed. The 3-dB gain reduction bandwidth for LH CP radiation in the *z*-direction, G_LBW, is calculated to be 9.6 %.

Figure 10 shows the gain in the *z*-direction at frequencies above $f_T = 3$ GHz. Again, as predicted, the maximum gain of G_R for RH CP radiation, G_{Rmax}, appears at a frequency between the upper bound frequency f_U and the H frequency f_H. In addition, the gain G_R decreases with decreasing frequency toward the transition frequency f_T. The 3-dB gain reduction bandwidth for RH CP radiation in the *z*-direction, G_RBW, is calculated to be 9.7 %. Thus, the abovementioned RH CP radiation is confirmed. As a result, the

DOI 10.1007/978-981-4560-75-7_33-2

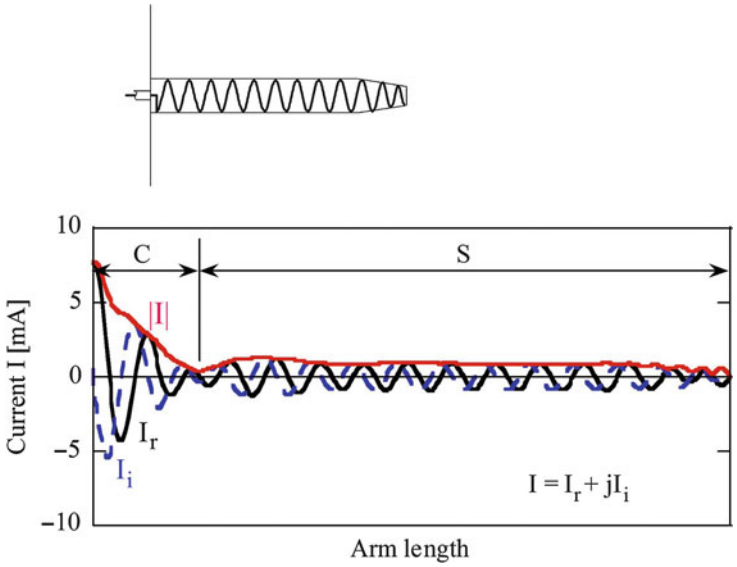


Fig. 11 Current distribution of an axial-mode helical antenna

predicted dual-band counter-CP radiation (LH CP radiation within a specific frequency band and RH CP radiation within a different frequency band) is numerically confirmed.

Helical Antennas

A helical antenna generally consists of a single wire or metallic strip wound into a helical shape. There are several operation modes, e.g., endfire (axial), backfire, sidefire, and normal modes. These modes basically depend on a difference in the size of the helix circumference. Since maximum CP-wave radiation along the helical axis is provided by the endfire mode, first, attention will be focused to this mode, which is closely related to the dielectric rod antenna described in later sections. Next, the CP-wave radiation toward the opposite endfire direction will be discussed, i.e., a backfire mode.

Endfire-Mode Helical Antenna

The current distribution of an axial-beam helical antenna has been analyzed using an integral equation (Nakano 1987). It is found that there are two distinct regions: one region (C-region) from the feed point to a point near the end of the second turn and the remaining region (S-region) just after the C-region, as shown in Fig. 11. The current distributed along the C-region generates backfire radiation toward the conducting flat reflector/plate. This backfire radiation is reflected by the flat reflector and then excites the S-region, inducing a traveling wave current whose amplitude is almost constant, except near the arm end. In other words, the C-region acts as an exciter and the S-region acts as a director (a waveguide element).

There are two findings for this antenna: (1) the helical antenna can work even when these two regions are disconnected, and (2) a helical antenna with a small number of turns constituting only the C-region radiates a CP wave. The latter finding leads to realizing a low-profile helix as a CP element, using the combination of a small number of turns and a low pitch angle.

The low-profile helices can be arrayed to construct a flat-type antenna, as shown in Fig. 12. The wave regarded as a local TEM mode in the radial direction propagates in the radial waveguide provided the spacing of the two parallel plates is sufficiently small compared with the wavelength, i.e., $S_w = 7.5\ \text{mm} = 0.3\lambda_{12}$ (λ_f: the wavelength at frequency f GHz). The feed wire of each helix is inserted

DOI 10.1007/978-981-4560-75-7_33-2

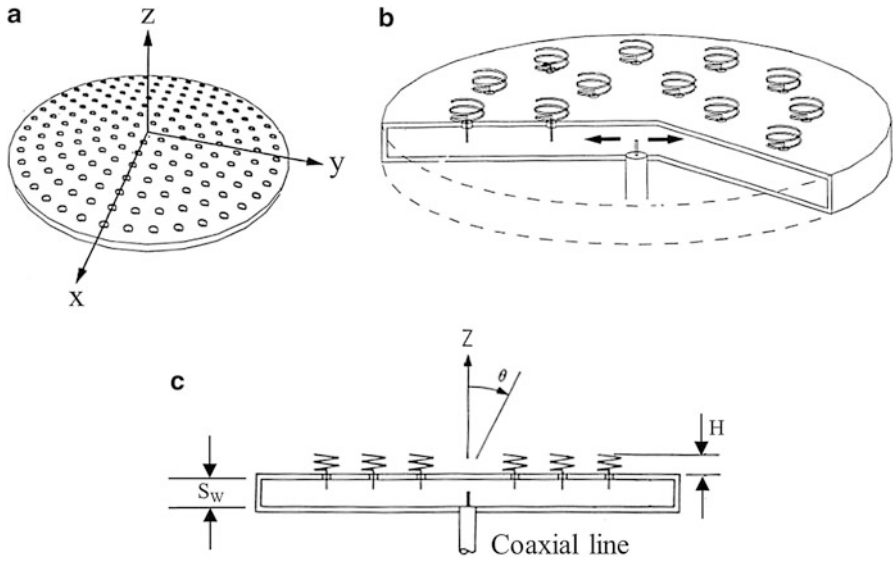


Fig. 12 Flat-type antenna consisting of low-profile helices

into the radial waveguide through a small hole and excited by a traveling wave flowing from the radial waveguide center (coaxial line) toward the waveguide edge. The radius of the feed wire is the same as that of the helix proper. It should be noted that the axial length H of the helical element above the surface of the radial waveguide is extremely low ($H = 4.7$ mm), so that the array antenna has a thickness of 15 mm ($\simeq S_w + H +$ plate thickness).

Using a two-turn helix of 4° pitch angle, which exhibits a 12 % bandwidth for a 3 dB axial ratio criterion, a flat-type antenna is fabricated for direct reception of the broadcasting satellite TV programs (DBS) in Japan (Nakano et al. 1992). The total number of helices is 368. It should be noted that the insertion length of the feed wire is adjusted to control an amplitude distribution. To minimize the backward traveling wave, the feed wires of the outermost helices are inserted at a distance of a quarter-wavelength from the waveguide edge, where the standing wave is maximal. The backward traveling wave can be absorbed by the outermost helices and becomes negligible over the waveguide except near the edge. In addition, it should be noted that the forward traveling wave in the waveguide decays as it progresses toward the edge. Therefore, to maintain a uniform amplitude distribution (or a uniform power distribution) on the array surface, the feed wires are gradually lengthened as the radial distance increases. Impedance matching to the feed coaxial line located in the center of the lower plate of the waveguide is made by appropriate insertion of the inner conductor of the coaxial line into the waveguide. As a result, good impedance matching is obtained, where the measured return loss is less than −18 dB in a frequency band of 11.7–12 GHz.

The radiation beam can be tilted by realizing the required phase condition. The amplitude distribution remains almost unchanged regardless of the helix rotation, so that there is no need to readjust the insertion length of the feed wire. Since the relative phase of each helix can be determined by measurement (Mano and Katagi 1982), the only thing to be done is to rotate each helix around its axis according to the phase requirement determined by the array theory. Figure 13 shows the gain at a frequency of 11.85 GHz, as a function of the beam tilt angle. The gain at a beam tilt angle of 30° is found to be only 1.1 dB lower than that without the beam tilt (normal beam).

DOI 10.1007/978-981-4560-75-7_33-2

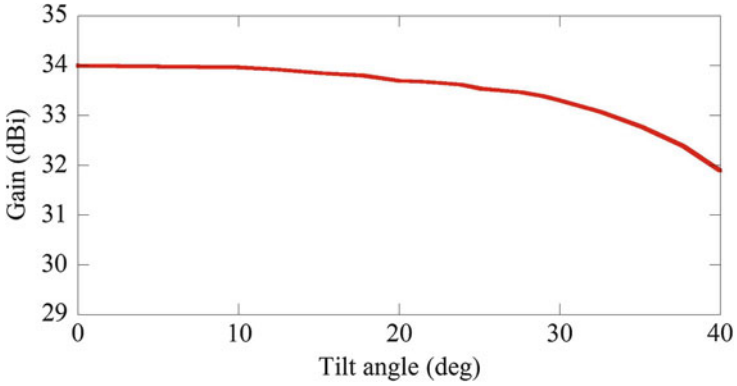


Fig. 13 Gain at 11.85 GHz

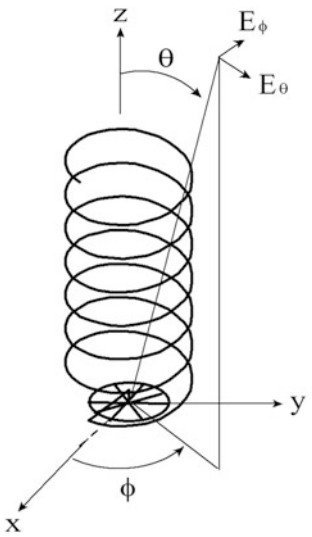


Fig. 14 Configuration and coordinate system of a helical antenna with a small ground plane

Backfire Helical Antenna

The endfire mode of operation in the previous subsection is most widely used. Another interesting mode of operation was reported by Patton in 1962 (Walter 1965). By using a bifilar helical antenna, he found a backfire mode, i.e., a CP wave radiated in the backward direction. It should be noted that the bifilar helix has an advantage over a conventional endfire helix in that a ground plane (reflector) is not needed. Attention is, therefore, directed toward a front-to-back (F/B) ratio. One of the means of realizing a sufficient F/B ratio is to modify the helix proper. A tapered backfire helix realizes a good F/B ratio across a wide frequency range, with smaller occupied space as compared with that of a conventional endfire helix mounted on a ground plane.

A similar effect is also observed in a monofilar helix with the small ground plane shown in Fig. 14 (Nakano et al. 1988). It should be recalled that the current in the C-region mainly radiates in the backfire direction. Since the reflector is small, the surface wave is not excited. This fact leads to another advantage that the phase center of the backfire helix is approximately and consistently fixed (note that the far-field from the endfire helix is composed of two fields radiated from spatially separated locations: the radiation from the feed and that from the open end). Thus, it is expected that the backfire helix be used as a primary feed for a paraboloidal reflector.

DOI 10.1007/978-981-4560-75-7_33-2

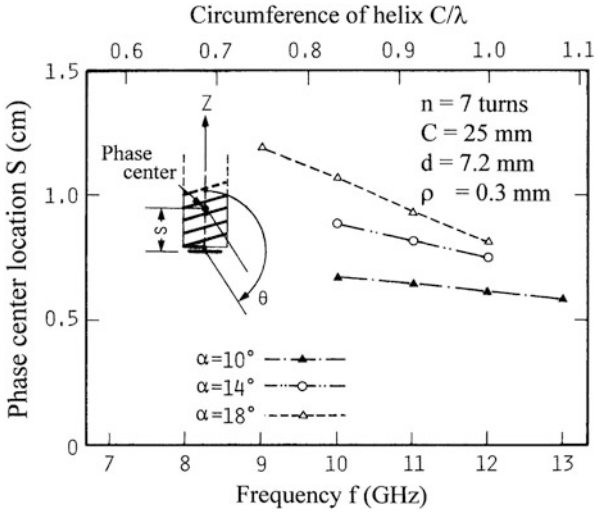


Fig. 15 Phase center location

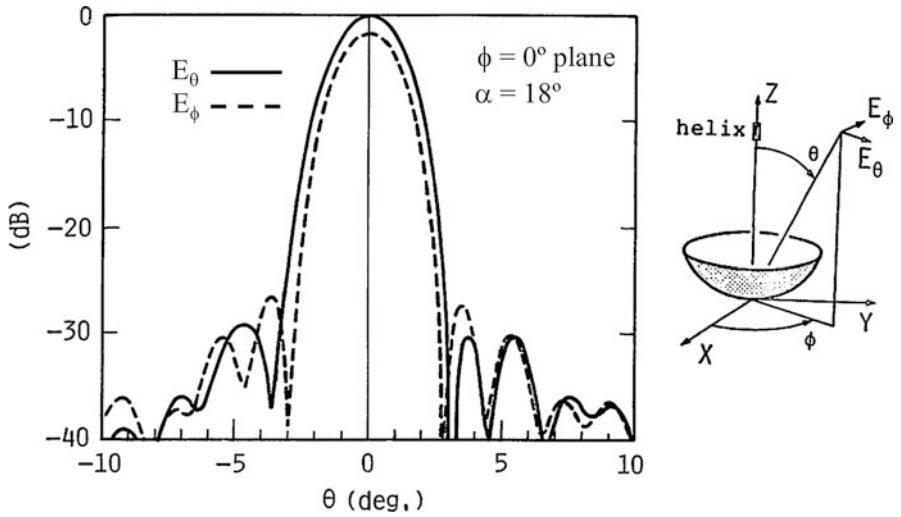


Fig. 16 Radiation pattern of a paraboloidal reflector antenna fed by a monofilar backfire helix

Now the monofilar backfire helix is employed as a primary feed for a symmetrical front-fed paraboloidal reflector. The reflector to be considered here has an aperture diameter of $D = 75$ cm and a focal length of $L = 22$ cm. The backfire helix is suitable for a feed of a front-fed reflector, since the configuration of the backfire helix is very small compared with a horn antenna and a conventional axial-mode helix with a ground plane. In addition, the backfire helix has an advantage that the feed line can be placed on the focal axis, resulting in a small aperture blockage.

To use the helix as an efficient feed for the reflector, it is imperative that the phase center location be known. The phase center is found by plotting the phase change of the radiation field against cos θ. The distance s from the helix origin to the phase center is shown in Fig. 15. Strictly speaking, the distance s varies according to the azimuth cut. Figure 15 presents the average of the two distances obtained in the $\phi = 0^\circ$ and 90° planes. As the frequency is increased, the distance s moves toward the feed point. This is due to the fact that the attenuation rate of the current becomes larger with an increase in frequency. The

DOI 10.1007/978-981-4560-75-7_33-2

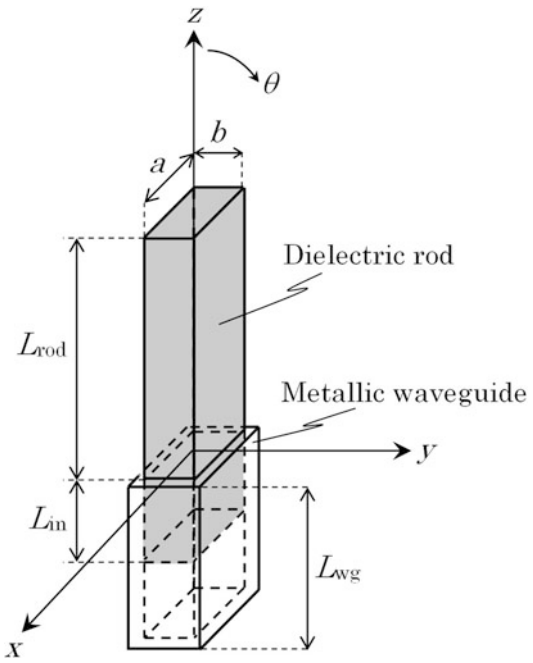


Fig. 17 Configuration of a dielectric rod fed by a metallic waveguide

change in the distance s as a function of frequency is small for a helix with a lower pitch angle, because the wire of the lower pitch helix is wound more tightly.

Figure 16 shows the radiation pattern of the paraboloidal reflector antenna. A seven-turn backfire helix with a pitch angle of $\alpha = 18^\circ$ and a ground plane diameter of 7.2 mm is situated at the focal point on the basis of the results shown in Fig. 15. The pattern calculation is performed at 12 GHz using the physical optics approximation. It is seen that the half-power beamwidth is 2.4° and that the first sidelobe level is less than −26 dB. The radiation field is circularly polarized with an axial ratio of 2.0 dB. The gain is calculated to be 37.2 dBi.

Rod Antennas

In this section, the characteristics of several rod antennas are described. First, the basic radiation properties of a rectangular rod antenna are studied. Emphasis is placed on the mechanism of the gain variation with rod length. The amplitude and phase distributions of the field near the rod are investigated. The gain increase with rod length is explained in terms of the expansion of the equiphase region in the terminal aperture regarded as a secondary Huygens plane. Then, a circular rod antenna is studied, since the circular geometry is preferable to radiate a CP wave. Modifications of dielectric rod antennas are discussed in terms of generation of a high gain, which exceeds the so-called Hansen-Woodyard condition (Hansen and Woodyard 1938; Andersen 1971). Final subsection is devoted to an artificial dielectric rod antenna consisting of a corrugated-metal surface.

Dielectric Rod Antenna

Figure 17 shows the configuration and coordinate system of a dielectric rod (Ando et al. 2002). A rectangular dielectric rod with a relative permittivity of $\epsilon_r = 2.05$ (Teflon) is fed by a metallic waveguide (WR-90) with internal dimensions of $(a, b) = (22.9\text{ mm}, 10.2\text{ mm})$. The length of the waveguide is fixed at $L_{wg} = 20$ mm for the analysis. The dimensions of the cross section of the rod are the same as the internal

DOI 10.1007/978-981-4560-75-7_33-2

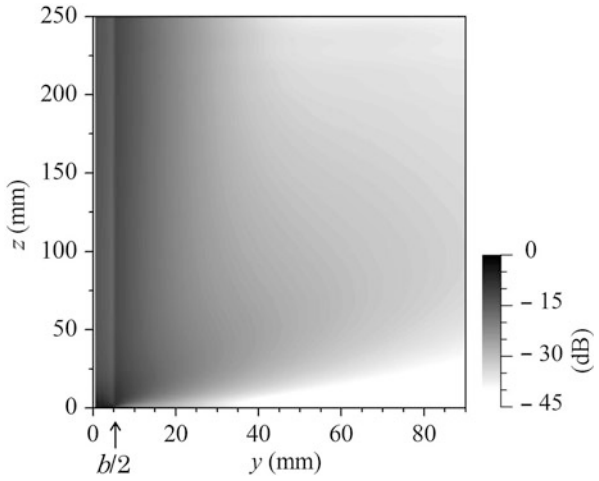


Fig. 18 Total field distribution of E_y in the y–z plane at $x = 0$ mm (half region) (From Ando et al. 2002, ©IET)

dimensions of the waveguide. The waveguide is excited with the TE_{10} mode, whose major electric field component is E_y. For impedance matching between the rod and the waveguide, a portion of the rod with length L_{in} is inserted into the waveguide without any modification. At the outer boundary of the computation region, the absorbing layers are placed to avoid spurious reflection toward the antenna.

The information on the field distribution near the rod is important to evaluate the radiation characteristics. For this, the field distribution is analyzed by the FDTD method. The field is decomposed into the surface and unguided waves, using a technique combining the FDTD method and the Yee-mesh-based beam-propagation method (YM-BPM) (Yamauchi 2003).

Figure 18 shows the total field distribution of E_y in the y–z plane at $x = 0$ mm. The field profile of the TE_{10} mode differs from that of the E_{11}^y mode, so that the power excited at the feed end is transformed into the surface- and unguided-wave power. At a sufficiently long z-distance, where the unguided-wave power is almost dissipated, the field along the rod can be regarded as the surface wave. At $z = 250$ mm, the ratio of power in the surface wave to the total radiation power is calculated to be approximately 52 %.

Figure 19a illustrates the distribution of the surface wave. By subtracting the surface wave in Fig. 19a from the field in Fig. 18, the unguided wave in Fig. 19b is obtained. As is well known, the phase velocity of the surface wave is slower than that of the unguided wave. Therefore, the phase interaction between the two waves occurs near the rod, as shown in Fig. 20. The phase interaction is used to interpret the gain variation with rod length L_{rod} in Fig. 24.

The directivity of a long dielectric rod with L_{rod} can be efficiently evaluated on the basis of the discontinuity-radiation concept (Collin and Zucker 1969). The radiation pattern is formed by superimposing the feed pattern onto the terminal pattern: the feed pattern generated from the unguided wave radiated near the feed end and the terminal pattern from the surface wave diffracted at the free end.

Figures 21 and 22 show the feed $D_{feed}(\theta)$ and terminal $D_{terminal}(\theta)$ patterns, which are calculated using the unguided and surface waves in Fig. 19, respectively. As can be seen, the unguided wave is diffracted in all directions, whereas the surface wave is in the endfire direction. Taking into account the phase relationship between the two patterns, the following superimposed pattern can be obtained:

$$D_{total}(\theta) = D_{feed}(\theta) + D_{terminal}(\theta)\exp[-j(\beta - k_0 \cos\theta)L_{rod}] \quad (1)$$

where k_0 is the wavenumber in free space and β is the propagation constant of the rod waveguide. Although not illustrated, good agreement is found to exist between the superimposed pattern and the

DOI 10.1007/978-981-4560-75-7_33-2

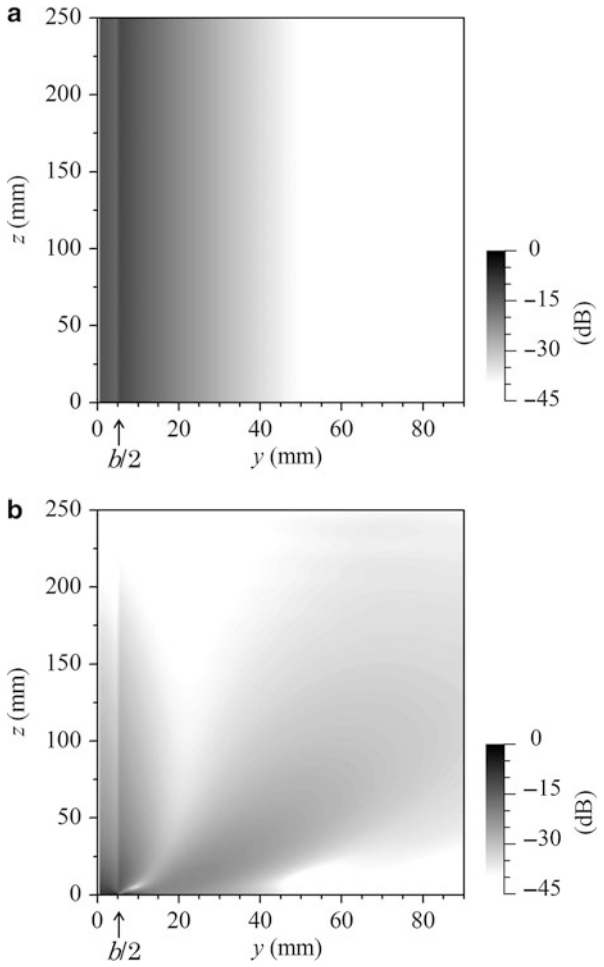


Fig. 19 Field distributions of surface and unguided waves in the y – z plane at x = 0 mm (half region). (**a**) Surface wave. (**b**) Unguided wave. (From Ando et al. 2002, © IET)

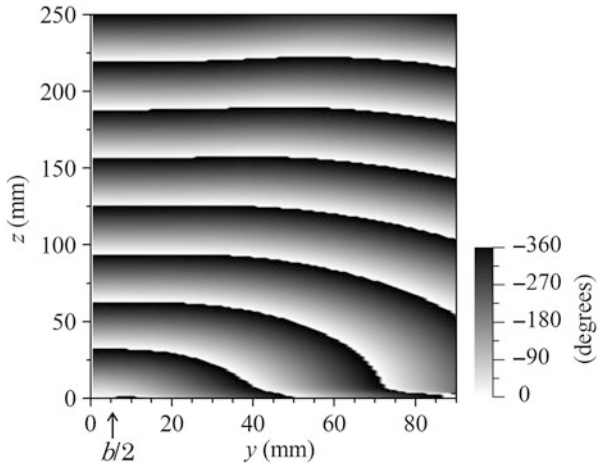


Fig. 20 Phase distribution of E_y near the rod (half region) (From Ando et al. 2002, ©IET)

pattern calculated using the total fields. In other words, superimposing the feed pattern onto the terminal pattern forms the radiation pattern of the dielectric rod antenna.

DOI 10.1007/978-981-4560-75-7_33-2

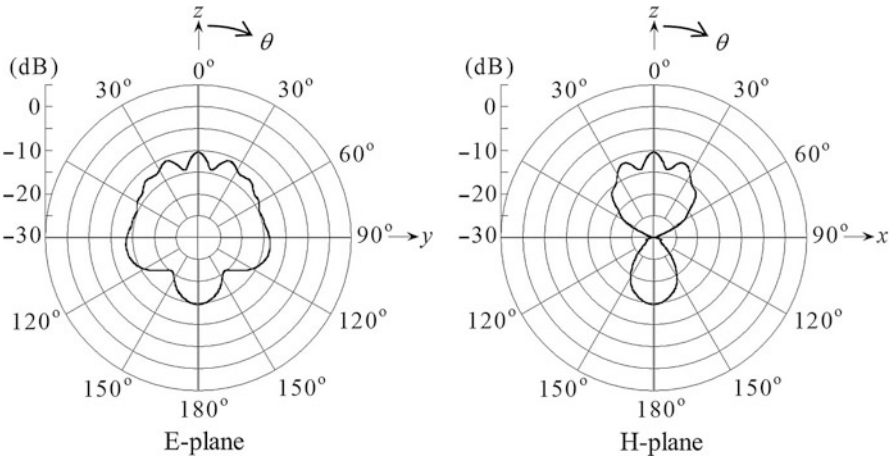


Fig. 21 Feed patterns generated from the unguided wave. Each feed pattern is normalized to a maximum value of corresponding terminal pattern (From Ando et al. 2002, ©IET)

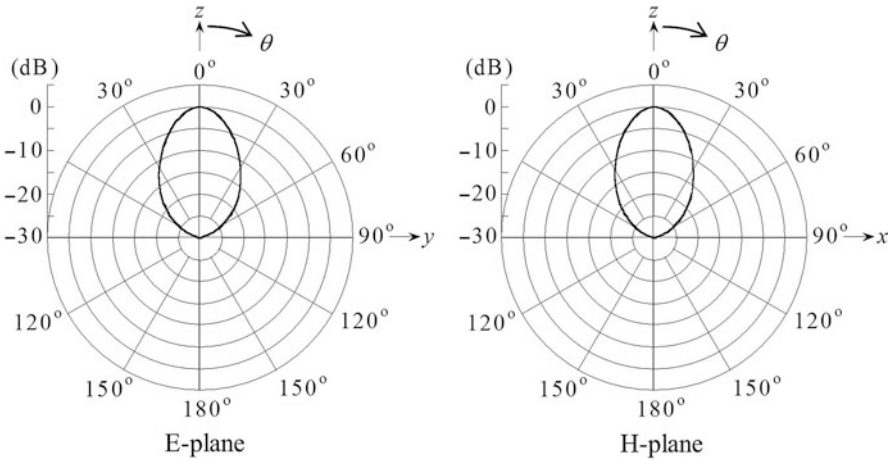


Fig. 22 Terminal patterns generated from the surface wave (From Ando et al. 2002, ©IET)

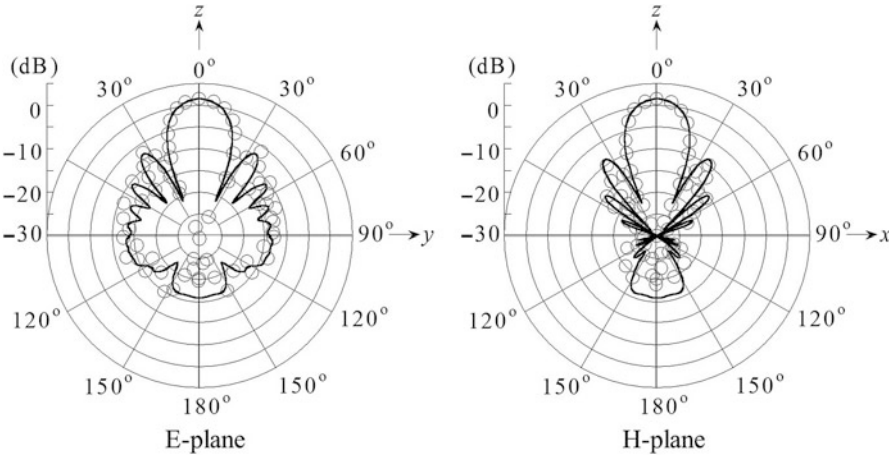


Fig. 23 Radiation patterns for $L_{rod} = 150$ mm; theoretical –, experimental ○ (From Ando et al. 2002, ©IET)

DOI 10.1007/978-981-4560-75-7_33-2

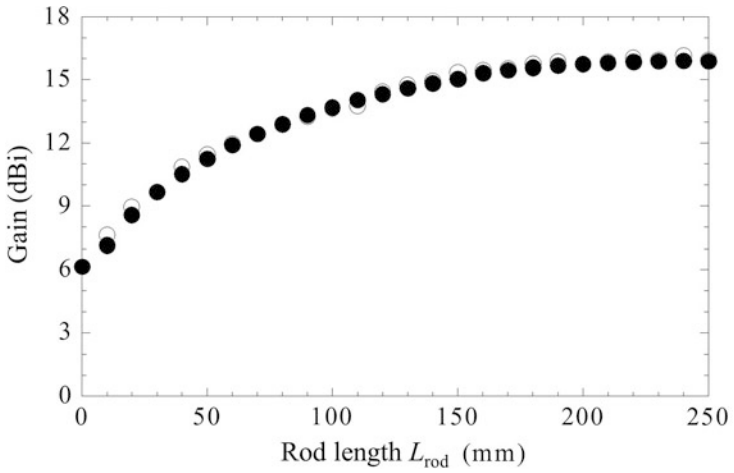


Fig. 24 Gain as a function of rod length L_{rod}; theoretical ●, experimental ○ (From Ando et al. 2002, ©IET)

The typical superimposed pattern for $L_{rod} = 150$ mm is shown as a solid line in Fig. 23. Needless to say, the pattern tends to have a sharper beam as L_{rod} is increased. The half-power beamwidth changes from $\pm 13°$ to $\pm 9°$ in the E- and H-planes, as L_{rod} is increased from 150 to 240 mm. The theoretical results agree well with the results measured in an anechoic chamber. A slight discrepancy in the backward radiation is due to the fact that in the experiment, a feeding system (including a waveguide to coaxial line transition) is located in the $-z$-axis direction.

It seems that the radiation pattern is also evaluated on the basis of the terminal aperture theory (Brown and Spector 1957): the x–y plane at $z = L_{rod}$ is regarded as a secondary Huygens plane where the source field composed of the surface and unguided waves exists. The unguided wave extends into the x–y plane (Fig. 19b), and hence the dimensions of the plane should be taken to be sufficiently large. Auxiliary calculations showed that the major lobes converge to those shown in Fig. 23, with an increase in the dimensions of the x–y plane. However, the minor lobes (in particular, the backward direction) disagree. This is because the source field in the secondary Huygens plane does not include the unguided wave diffracted toward the backward direction from the feed end (Fig. 21).

To take into account the effect of the unguided wave diffracted toward the backward direction on the radiation, the radiation pattern is calculated on the basis of the field on a closed surface containing the x–y plane at $z = L_{rod}$ around the rod. As a result, there is essentially no difference between the radiation pattern generated from the field on the closed surface and that in Fig. 23 obtained by the discontinuity-radiation concept, although the radiation pattern in the backward direction cannot be evaluated on the basis of the terminal aperture theory.

The gain computed from the field on the closed surface also agrees with that evaluated on the basis of the discontinuity-radiation concept. The computed gain in the $+z$-direction against L_{rod} is presented by solid circles in Fig. 24. It is observed that the gain increases from 15.1 to 15.9 dBi, as L_{rod} is lengthened from 150 to 240 mm. The gain variation with L_{rod} is validated by the experimental results presented by open circles.

As mentioned previously, the directivity in the endfire direction is dominated by the field in the secondary Huygens plane at the free end. Hence, the mechanism of the increase in the gain shown in Fig. 24 must be understood by the field distribution in the x–y plane at $z = L_{rod}$. Figure 25 illustrates the phase distributions of E_y ($x = 0, y, z = L_{rod}$) for several values of L_{rod}. The phase is normalized to a value at $y = \Delta y/2$ (Δy is the FDTD cell size in the y direction). The expansion of the equiphase region in the secondary Huygens plane corresponds to the gain increase observed in Fig. 24.

DOI 10.1007/978-981-4560-75-7_33-2

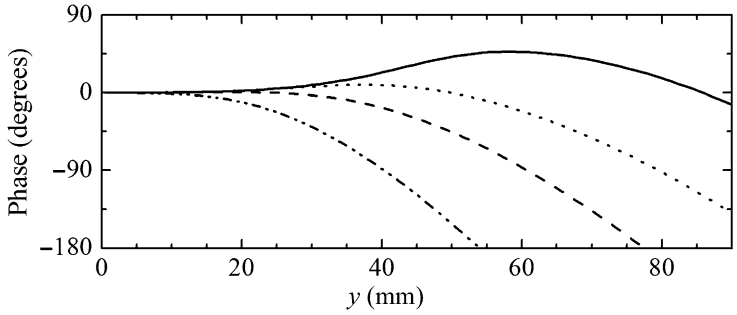


Fig. 25 Phase distribution of E_y at the free end (half region); $L_{rod} = 240$ mm —, $L_{rod} = 150$ mm · · ·, $L_{rod} = 100$ mm - - -, $L_{rod} = 50$ mm – · · – (From Ando et al. 2002, ©IET)

In summary, the terminal aperture theory is effective, provided that the diffracted wave toward the backward direction generated at the feed end is negligible.

Modified Dielectric Rod Antenna

It is known that the gradual taper of a dielectric rod leads to an increase in the gain. So far, the directivity of a tapered rod has been estimated using traditional design guidelines (Chatterjee 1985; Lo and Lee 1988). Zucker presented a design principle for a maximum-gain antenna, in which a tapered rod is regarded as an endfire array of discrete elements (Volakis 2007). James described a semiempirical approach to the selection of an optimum taper profile, in which a tapered rod is regarded as a series of noninteracting planar radiating apertures (James 1972). The attainable gain of a practical rod antenna may be limited to 20 dBi, although the traditional design guidelines predict the gain increase with an increase in rod length.

In this subsection, a tapered cylindrical-dielectric-rod fed by a metallic waveguide with a launching horn is analyzed, and the effect of the guided-mode conversion in the rod on the directivity is discussed (Ando et al. 2005). The waves propagating along linearly and curvilinearly tapered rods are evaluated using the body-of-revolution (BOR) FDTD method (Taflove and Hagness 2005). The use of the BOR technique enables us to efficiently calculate long cylindrical-dielectric rods.

Figure 26a shows the configuration of a cylindrical-dielectric rod fed by a circular metallic waveguide. The rod with a relative permittivity of $\varepsilon_r = 2.05$ (Teflon) is assumed to be a lossless medium and is composed of a uniform section (L_{uni}) and a tapered section (L_{tap}). The inner diameter of the metallic waveguide, which is the same as the diameter of the uniform rod, is $2\rho_{feed} = 17.475$ mm ($=0.64\lambda$) where λ ($=27.3$ mm) is the wavelength at a test frequency of $f = 11$ GHz. The metallic waveguide is excited with the TE_{11} mode, whose cutoff frequency is 10 GHz. The uniform rod operates as a single mode waveguide at the test frequency, since the cutoff frequency of higher-order modes (Balanis 1989) is 13 GHz.

To obtain smooth transition from the TE_{11} mode of the metallic waveguide to the HE_{11} mode of the uniform rod, the feeding system illustrated in Fig. 26b is introduced. A portion of the uniform rod is tapered and inserted into the metallic waveguide. A launching horn is placed at the feed end.

Typical taper profiles to be investigated are shown in Fig. 26c, in which the diameter at the free end is taken to be $2\rho_{free} = \rho_{feed}$ and the overall length of the tapered section is fixed to be $L_{tap} = 20\lambda$. The change in the radius of the tapered section, $\rho(z_{tap})$, is expressed as

$$\rho\left(z_{tap}\right) = \rho_{feed} - \left(\rho_{feed} - \rho_{feed}\right)\left(\frac{z_{tap}}{L_{tap}}\right)^{\frac{1}{n}}, \tag{2}$$

where z_{tap} is the axial distance of the tapered section (see Fig. 26a). The rod of $n = 1$ is linearly tapered and that of $n > 1$ is curvilinearly tapered. For the curvilinear taper, the change in the radius near the feed end

DOI 10.1007/978-981-4560-75-7_33-2

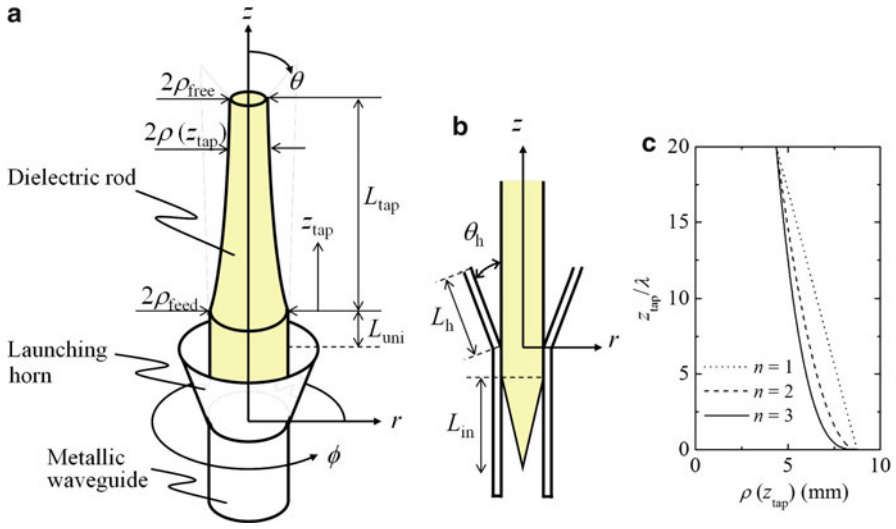


Fig. 26 Configuration of a tapered cylindrical-dielectric-rod antenna ($\varepsilon_r = 2.05$ (Teflon), $2\rho_{feed} = 17.475$ mm (=0.64λ), $2\rho_{free} = \rho_{feed}$, λ = 27.3 mm ($f$ = 11 GHz)). (**a**) Perspective view. (**b**) Feeding system. (**c**) Taper profiles for $L_{tap} = 20\lambda$; linear ($n = 1$) · · ·, curvilinear ($n = 2$) - - -, curvilinear ($n = 3$) —

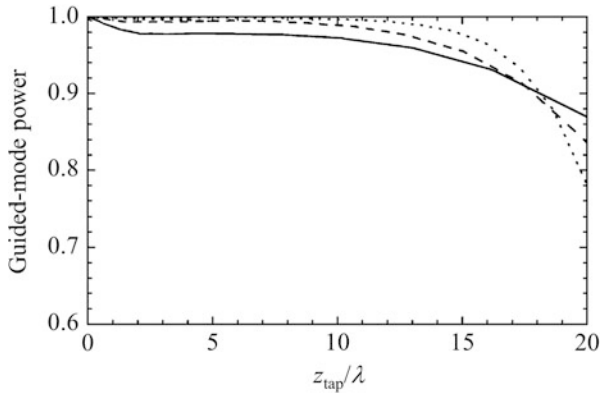


Fig. 27 Guided-mode power for $L_{tap} = 20\lambda$ as a function of z_{tap}; linear (n =1) . . ., curvilinear ($n = 2$) - - -, curvilinear ($n = 3$ —)

becomes large with an increase in n, while that near the free end becomes small. The guided-mode conversion properties and directivities will be evaluated for the linearly and curvilinearly tapered structures with $L_{tap} = 10\lambda$ and 20λ.

The gain of a dielectric rod increases with an expansion of the equiphase field region in the terminal aperture. As is well known, the field profile of the HE_{11} mode is extended into the air region by reducing the diameter of a rod. Hence, the enhanced gain is obtained, if the HE_{11} mode excited at the feed end is smoothly converted into that at the free end where the diameter is reduced. To achieve smooth guided-mode conversion, the diameter in the forward direction should be gradually reduced.

The diameter of the tapered rod connected with the uniform rod ($L_{uni} = \lambda$) is reduced from $2\rho_{feed} = 0.64\lambda$ to $2\rho_{free} = \rho_{feed} = 0.32\lambda$, in which the phase constant (k_z) of the HE_{11} mode wave at the free end is reasonably close to the free-space wave number (k_0), i.e., $k_z/k_0 = 1.005$. The directivity depends on a taper profile (James 1972), so that several profiles of tapered rods are investigated by changing the parameter n in Eq. 2.

DOI 10.1007/978-981-4560-75-7_33-2

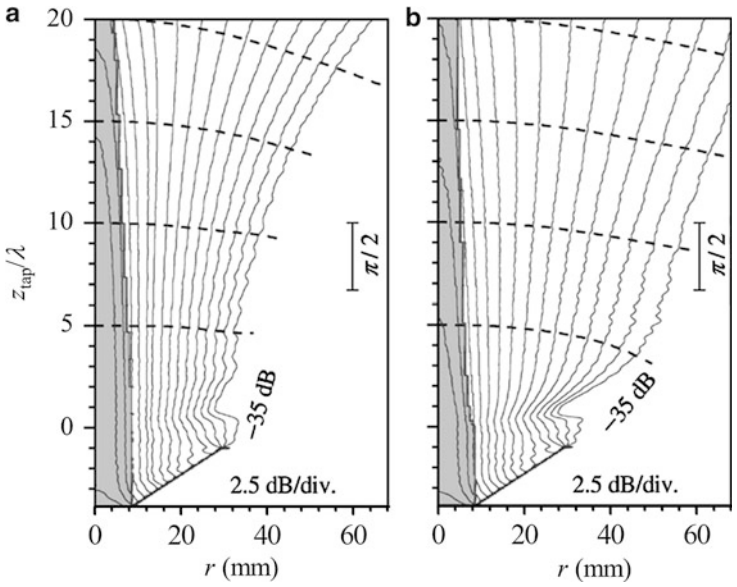


Fig. 28 Field distributions near tapered rods with $L_{tap} = 20\lambda$ (half region); amplitude —, phase - - -. (**a**) linear ($n = 1$). (**b**) curvilinear ($n = 3$)

Note that Ladouceur and Love (1996) intuitively described a low-loss criterion for the taper profile of a dielectric waveguide. Radiation loss will be small if the taper length is large compared with the coupling length between the fundamental guided-mode and the radiation field. The limit of the local taper angle [$\Omega(z_{tap})$] between the z-axis and the tangent to the dielectric interface was expressed as

$$\Omega(z_{tap}) = \tan^{-1}\frac{\rho(z_{tap})\left[k_z(z_{tap}) - k_0\right]}{2\pi} \simeq \frac{\rho(z_{tap})\left[k_z(z_{tap}) - k_0\right]}{2\pi}, \tag{3}$$

where $k_z(z_{tap})$ is the phase constant of the HE_{11} mode at z_{tap}. This equation implies that the change in $\rho(z_{tap})$ should become smaller for larger z_{tap}. Therefore, the guided-mode power is expected to be maintained in a curvilinear taper compared with a linear taper.

Figure 27 shows the guided-mode power along z_{tap} for $L_{tap} = 20\lambda$, in which the overlap integral between the eigenmode field of the tapered rod and the numerically determined field is used. The power is normalized to the HE_{11} mode power excited at the feed end. It is found that the guided-mode power for the linear taper drastically decreases near the free end. On the other hand, the guided mode for the curvilinear taper can smoothly be converted with an increase in L_{tap}, so that the considerable amount of the power is maintained at the free end. This tendency becomes more appreciable for a longer tapered rod. A value of 87 % is obtained in a curvilinearly tapered rod of $n = 3$ with $L_{tap} = 20\lambda$ (the reflected power generated in the tapered section is negligible and is calculated to be less than 0.05 %). As a result, smooth guided-mode conversion can be achieved in a curvilinearly tapered rod, whose radius slightly varies near the free end.

Figure 28a, b compares the field distributions along the tapered rods with $L_{tap} = 20\lambda$. The solid lines present the amplitude of $|E_r|$ and dashed lines present the phase at $z_{tap} = 5\lambda$, 10λ, 15λ, and 20λ. It is clearly observed that the fields along the curvilinearly tapered rods gradually extend into the air region near the free end in comparison to the field along the linearly tapered rod. The gradual extension of the field results in the expansion of the equiphase field region in the terminal aperture.

DOI 10.1007/978-981-4560-75-7_33-2

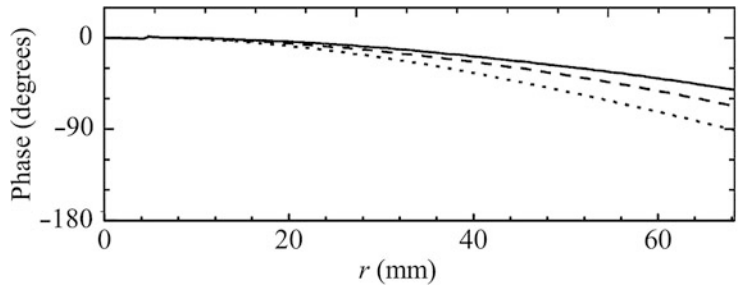


Fig. 29 Phase distributions for $L_{tap} = 20\lambda$ at free end (half region); linear ($n = 1$) · · ·, curvilinear ($n = 2$) - - -, curvilinear ($n = 3$ —)

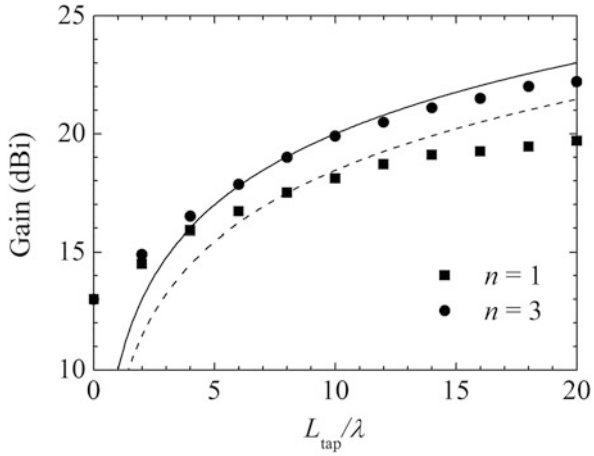


Fig. 30 Gain as a function of L_{tap}. HWG - - -, MG —

The phase distribution in the terminal aperture for tapered rods is plotted for $L_{tap} = 20\lambda$ in Fig. 29. The equiphase region is expanded, as n and L_{tap} are increased. The expansion of the equiphase region corresponds to an increase in the guided-mode power observed in Fig. 27. Owing to the expansion of the equiphase field region, a narrowed radiation pattern is expected with a subsequent higher gain.

Figure 30 plots the gain in the endfire direction ($\theta = 0°$) against L_{tap}. As can be seen, the gain increases with an increase in L_{tap}. The gain for $n = 3$ at $L_{tap} = 20\lambda$ is calculated to be 22.2 dBi, which is a 2.5 dB increase compared with that for $n = 1$. Auxiliary calculations show that a gain of greater than 20 dBi is maintained in a tapered rod of $n = 3$ with $L_{tap} = 20\lambda$ across a frequency range of 10–12.7 GHz (24 % bandwidth).

Finally, the gain obtained from the FDTD analysis is compared with that from the design guideline presented by Zucker (Volakis 2007). The gain variation with L_{tap} is expressed as

$$G \cong m\frac{L_{tap}}{\lambda}, \tag{4}$$

where m is a variable. The data for $m = 7$ corresponds to the so-called Hansen-Woodyard gain (HWG), and that for $m = 10$ corresponds to the maximum gain (MG). The HWG and MG are plotted by dashed and solid lines in Fig. 30, respectively. It is found that a value of higher than the HWG can be obtained in a curvilinearly tapered rod of $n = 3$. On the other hand, the MG indicates that the gain of a rod with $L_{tap} = 20\lambda$ is approximately 23 dBi. For further reference, curvilinearly tapered rods for larger values of n are also evaluated. Although not presented, the gain reaches a value of 22.6 dBi when n is chosen to be

DOI 10.1007/978-981-4560-75-7_33-2

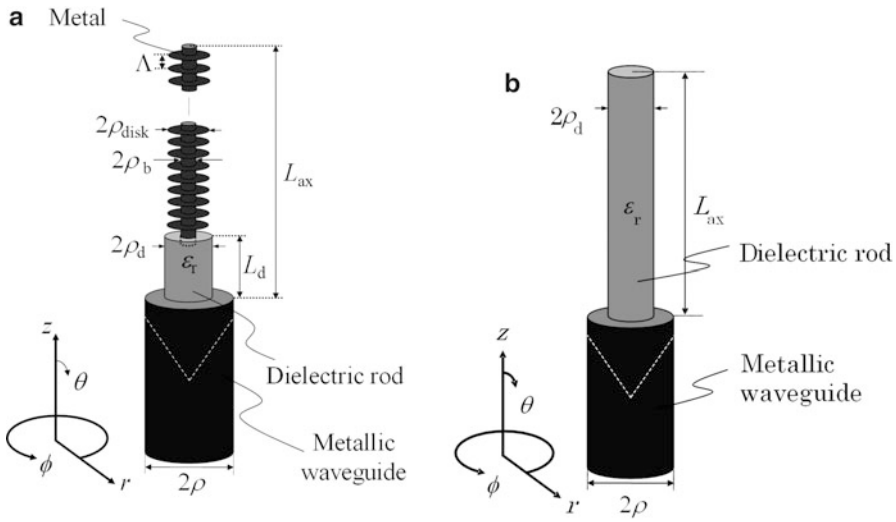


Fig. 31 Configurations of (**a**) a cigar antenna and (**b**) a reference dielectric rod antenna (From Yamauchi et al. 2011, ©KIEES)

5, while a guided-mode power of 82 % is maintained at the free end. Consequently, the design guideline is effective in predicting the gain increase with an increase in L_{tap} and roughly estimating the gain of a long tapered rod.

Artificial Dielectric Rod Antenna

A periodic structure consisting of an array of metal plates supports a surface wave. Therefore, an endfire antenna without using any real dielectric can be constructed. In other words, these metal structures can be viewed as an artificial dielectric that is often lighter than dielectrics such as Teflon and polyethylene. A pioneering work on the endfire antenna using a periodic structure of circular metal disks, called a cigar antenna, has been performed by Simon and Weill (1953). Although several works have been carried out experimentally and/or theoretically, recent numerical techniques such as the FDTD method allow more efficient and practical investigation of this type of antenna. The radiation characteristics regarding the cigar antenna will be summarized, where a circular metallic waveguide is adopted to excite a disk-on-rod structure.

Figure 31a shows the configuration of a cigar antenna fed by a metallic waveguide (WCI-120) whose inner diameter is $2\rho = 17.475$ mm. The waveguide is excited with the TE_{11} mode at a center frequency of 11 GHz ($\lambda_{11} = 27.3$ mm). The tapered dielectric rod whose relative permittivity is $\epsilon_r = 2.05$ is inserted into the metallic waveguide so that the impedance matching may be made between the air-filled and dielectric-filled regions.

The effective relative permittivity of the cigar structure depends on a period length Λ, a disk diameter $2\rho_{disk}$, and a boom diameter $2\rho_b$. These parameters are adjusted so as to realize almost the same effective relative permittivity as that observed in the reference dielectric rod antenna shown in Fig. 31b. Note that the cigar and dielectric rod antennas are compared under the condition of the same axial length L_{ax}. Two types of cigars #1 and #2 with $2\rho_b \simeq 0.11\lambda_{11}$ are treated: #1 is the model in which the periodicity is chosen to be small, while #2 is the model with a relatively large periodicity. To be specific, the parameters for #1 are $\Lambda = 0.09\lambda_{11}$ and $2\rho_{disk} \simeq 0.28\lambda_{11}$, while those for #2 are $\Lambda = 0.27\lambda_{11}$ and $2\rho_{disk} \simeq 0.32\lambda_{11}$. It is assumed that the cigar structure is perfectly conducting. To support the cigar, a short dielectric rod with a diameter of $2\rho_d \simeq 0.38\lambda_{11}$ and a length of $L_d = 0.54\lambda_{11}$ is introduced at the junction between the metallic waveguide and the cigar.

DOI 10.1007/978-981-4560-75-7_33-2

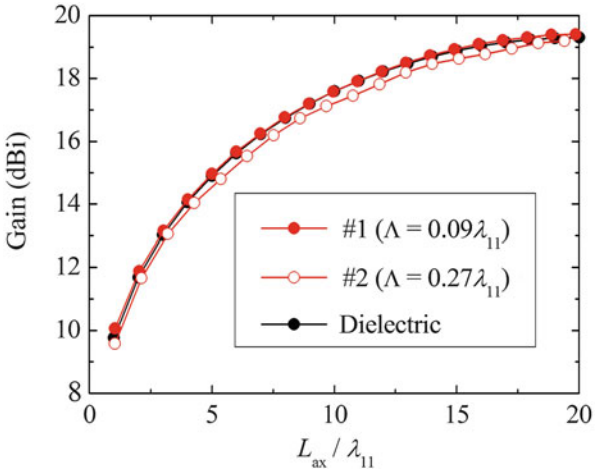


Fig. 32 Gain characteristics as a function of L_{ax} (From Yamauchi et al. 2011, ©KIEES)

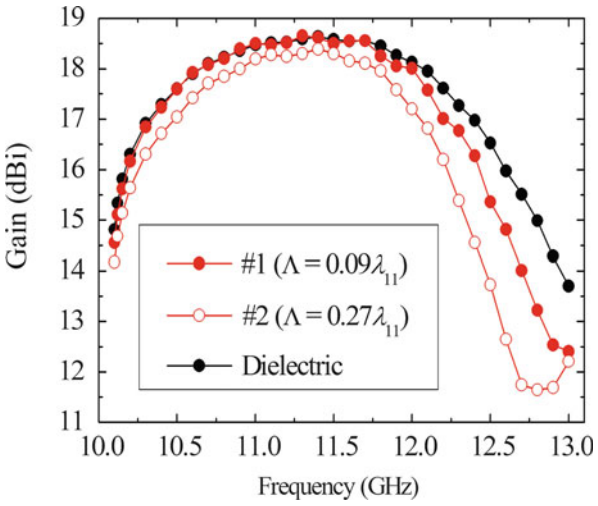


Fig. 33 Frequency response of gain (From Yamauchi et al. 2011, ©KIEES)

Figure 32 shows the gain characteristic at 11 GHz as a function of axial length L_{ax}. The data presented by red solid and open circles show the calculated gains for #1 and #2, respectively, while a black solid circle shows for the reference dielectric rod. In all the results, the gain increases as L_{ax} is increased. As can be seen, the gain property of the cigar is in a good correlation with that of the dielectric rod. It is confirmed that the cigar structure operates as an artificial dielectric. The gain reaches a maximum value of 19.4 dBi, when L_{ax} is taken to be around $20\lambda_{11}$. In the following analysis, the structures with $L_{ax} \simeq 13\lambda_{11}$ are treated.

The frequency response of the gain is shown in Fig. 33. A gain of more than 17 dBi is obtained for #1 across a frequency range of 10.4–12.2 GHz. A maximum gain of 18.6 dBi is obtained at 11 GHz. It is found that the gain bandwidth for #1 is wider than that for #2 and is close to that observed for the reference dielectric rod. In other words, when the periodicity of the circular disk is sufficiently small compared with the wavelength, the bandwidth becomes almost the same as that for the dielectric rod. The return loss for #1 is more than 15 dB across a frequency range of 10.5–13 GHz.

Figure 34a, b illustrates the currents $I(=I_r + jI_i)$ distributed along the cigar structure at 11 GHz and 13 GHz, respectively. For convenience, the current values obtained on the wall surfaces of disks are

DOI 10.1007/978-981-4560-75-7_33-2

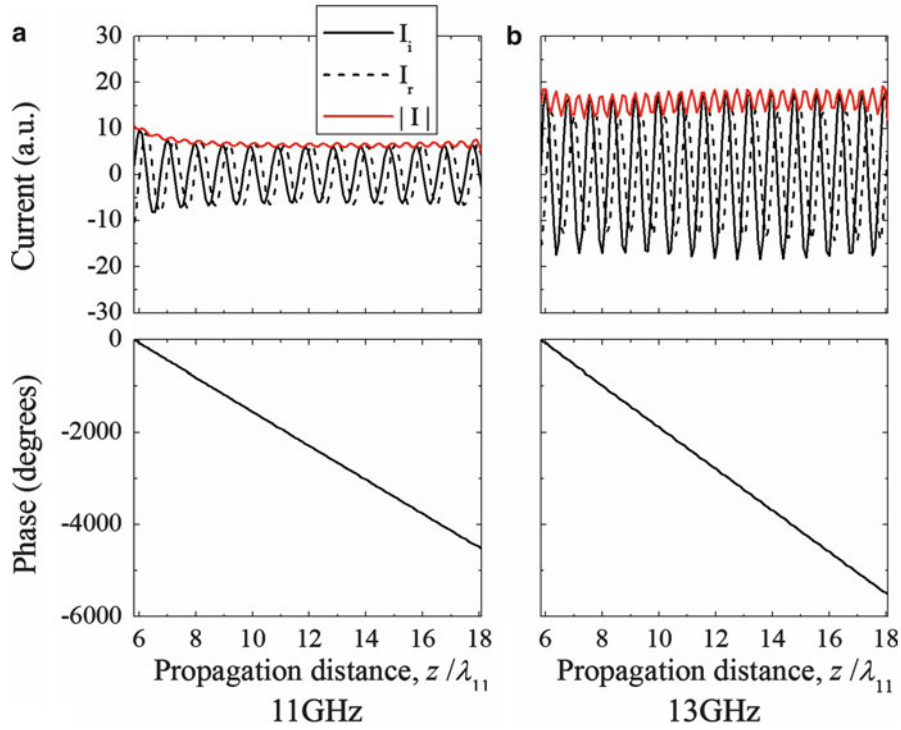


Fig. 34 Current distribution. (**a**) 11 GHz. (**b**) 13 GHz (From Yamauchi et al. 2011, ©KIEES)

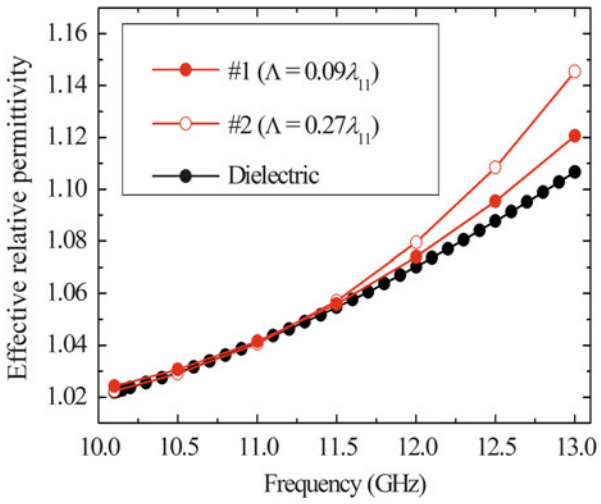


Fig. 35 Effective relative permittivity as a function of frequency (From Yamauchi et al. 2011, ©KIEES)

sampled and continuously plotted. It is found that the currents are of traveling wave type, as seen from the phase progressions. The amplitude tends to increase with subsequent generation of a standing wave, as the frequency is increased.

The effective relative permittivity as a function of frequency is shown in Fig. 35. The result for the cigar is estimated by the current distribution illustrated in Fig. 34. On the other hand, that for the dielectric rod is calculated from the eigenmode solver using the YM-BPM (Yamauchi 2003). Good agreement is found to exist between the results for #1 and for the dielectric rod. It should be noted that the values for #2 rapidly deviate from those of the dielectric rod as the frequency is increased. This results in the deterioration of the

radiation pattern with an increase in frequency. It can be said that the choice of a sufficiently short period length is extremely important for the wideband properties.

Conclusion

Antennas radiating a circularly polarized wave have been discussed. After reviewing the radiation mechanism of a flat Archimedean spiral antenna, a novel spiral structure is discussed, where a composite left- and right-handed transmission line is employed. The antenna radiates a dual circularly polarized wave. Then, discussion is made for a helical antenna radiating toward the endfire direction. Some applications of the helix to a flat-type antenna for direct reception of the broadcasting satellite TV programs have been introduced. An application of a backfire helix to the primary feed for a paraboloidal reflector has also been presented. Finally, the radiation properties of a dielectric rod antenna have been studied. The discontinuity-radiation concept has been confirmed by the numerical approach. A technique for obtaining a higher gain has been explained by introducing a tapered rod. As an artificial dielectric, a periodic structure of circular metal disks has been used, and their characteristics have been compared with those for the conventional dielectric rod antenna.

References

Andersen JB (1971) Metallic and dielectric antennas. Polyteknisk Forlag, Lyngby
Ando T, Yamauchi J, Nakano H (2002) Rectangular dielectric-rod fed by metallic waveguide. IEE Proc Microw Antennas Propag 149(2):92–97
Ando T, Ohba I, Numata S, Yamauchi J, Nakano H (2005) Linearly and curvilinearly tapered cylindrical-dielectric-rod antenna. IEEE Trans Antennas Propag 53(9):2827–2833
Balanis CA (1989) Advanced engineering electromagnetics. Wiley, New York
Brown J, Spector JQ (1957) The radiating properties of end-fire aerials. Proc IEE 104B:27–34
Caloz C, Itoh T (2006) Electromagnetic metamaterials. Wiley, New York
Canonsburg PA (2015) HFSS ANSYS. [Online] Available: http://www.ansys.com/Products/Simulation+Technology/Electronics/Signal+Integrity/ANSYS+HFSS
Chatterjee R (1985) Dielectric and dielectric-loaded antennas. Research Studies Press, Letchworth
Collin R (1966) Foundation for microwave engineering. McGraw-Hill, New York
Collin R, Zucker FJ (1969) Antenna theory pt 2. McGraw-Hill, New York
Eleftheriades G, Balmain K (2005) Negative-refraction metamaterials: fundamental principles and applications. Wiley, New York
Engheta N, Ziolkowski RW (eds) (2006) Metamaterials. Wiley, New York
Hansen WW, Woodyard JR (1938) A new principle in directional antenna design. Proc IRE 19:1184–1215
Harrington RF (1968) Field computation by moment methods. Macmillan, New York
James JR (1972) Engineering approach to the design of tapered dielectric-rod and horn antennas. Radio Electron Eng 42(6):251–259
Kaiser JA (1960) The Archimedean two-wire spiral antenna. IRE Trans AP-8(3):312–323
Kraus JD, Marthefka RJ (2002) Antennas, 3rd edn. McGraw-Hill, Boston
Ladouceur F, Love JD (1996) Silica-based buried channel waveguides and devices. Chapman & Hall, London
Lo YT, Lee SW (1988) Antenna handbook. Van Nostrand Reinhold, New York

Mano S, Katagi T (1982) A method of measuring amplitude and phase of each radiating element of a phase array antenna. Trans IECE Jpn J-65-B:555–560
Mei KK (1965) On the integral equations of thin wires antennas. IEEE Trans AP AP-13:374–378
Mushiake A (1996) Self-complementary antennas. Springer, New York
Nakano H (1987) Helical and spiral antennas – a numerical approach. Research Studies Press, Letchworth
Nakano H, Yamauchi J, Mimaki H (1988) Backfire radiation from a monofilar helix with a small ground plane. IEEE Trans Antennas Propag 36(10):1359–1364
Nakano H, Takeda H, Kitamura Y, Mimaki H, Yamauchi J (1992) Low-profile helical array antenna fed from a radial waveguide. IEEE Trans Antennas Propag 40(3):279–284
Nakano H, Miyake J, Oyama M, Yamauchi J (2011) Metamaterial spiral antenna. IEEE Antennas Wirel Propag Lett 10:1555–1558
Nakano H, Miyake J, Sakurada T, Yamauchi J (2013) Dual-band counter circularly polarized radiation from a single-arm metamaterial-based spiral antenna. IEEE Trans Antennas Propag 61(6):2938–2947
Penno RP, Pasala KM (2001) Theory of angle estimation using a multiarm spiral antenna. IEEE Trans Aerosp Electron Syst 37(1):123–133
Rumsey VH (1966) Frequency independent antennas. Academic, New York
Simon JC, Weill G (1953) A new type of antenna for endfire radiation. Ann Radioelecricite 8:183–193
Taflove A, Hagness SC (2005) Computational electrodynamics the finite-difference time-domain method, 3rd edn. Artech House, Boston
Volakis JL (ed) (2007) Antenna engineering handbook, 4th edn. McGraw-Hill, New York
Walter CH (1965) Traveling wave antennas. Dover Pub, New York
Yamauchi J (2003) Propagating beam analysis of optical waveguides. Research Studies Press, Letchworth
Yamauchi J, Harada H, Nakano H (2011) Numerical analysis of a cigar antenna. Int Symp Antennas Propag, Jeju FrD2-4

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_35-1

Dielectric Resonator Antennas

Eng Hock Lim[a], Yong-Mei Pan[b] and Kwok Wa Leung[c]*
[a]Electrical and Electronic Engineering, Universiti Tunku Abdul Rahman, Kajang, Selangor, Malaysia
[b]Electronic and Information Engineering, South China University of Technology, Guangdong, China
[c]Electronic Engineering, City University of Hong Kong, Hongkong, China

Abstract

Over the past 30 years, many interesting developments have been seen in the field of dielectric resonator antenna (DRA). Analytical and numerical models for different shapes of DRAs were established in the 1980s and 1990s to understand their radiation properties. A couple of excitation schemes have also been proposed so that the DRAs can be excited efficiently. With the rapid advancement of dielectrics and microfabrication technologies in recent years, the DRA can now be made very compact for applications in portable wireless communication and millimeter-wave systems. In the first part of this chapter, miniaturization techniques of the DRA are discussed. It is found that the DRA can be integrated with power dividers for designing the circularly polarized and differential DRAs. Use of ground miniaturization technique has enabled realization of the omnidirectional CP DRA and the quasi-isotropic DRA for the first time. A review of the recent achievements in millimeter-wave DRA is then given. In this chapter, different dielectrics such as polymer and glass are also explored for designing the DRA. Transmission lines such as microstrip and substrate-integrated waveguide are deployed for exciting the DRA in the millimeter-wave spectrum. Elucidation is made on the design procedures and other considerations of the miniature and millimeter-wave DRAs.

Keywords

Dielectric resonator antenna; Miniature DRA; Omnidirectional CP DRA; Quasi-isotropic DRA; Millimeter-wave DRA

Introduction

It was first shown in the late 1930s (Richtmyer 1939) that an unmetalized dielectric object can also behave like a metallic cavity resonator, and it was subsequently given the name of dielectric resonator (DR). But DR was only explored for designing various microwave circuits beginning in the 1960s when technologies for making low-loss and temperature-stable dielectrics were becoming more mature. Since then, the DR has been broadly used for designing filters and oscillators because it can provide high unloaded Q-factor (Abe et al. 1978; Cohn 1968; Fiedziuszko 1986; Kajfez and Guillon 1998; Plourde and Ren 1981), which is usually in the range of 20–10,000, making it an excellent candidate for designing energy storage components and for achieving high frequency selectivity. As the wavelength of electromagnetic waves in the DR is smaller by a factor of $1/\sqrt{\varepsilon_r}$ compared to that in free space, the component size can usually be scaled down when a DR is deployed. A higher degree of compactness is also achievable as the DR can be easily integrated with different microwave integrated circuits (MICs) using simple coupling

*Email: eekleung@cityu.edu.hk

DOI 10.1007/978-981-4560-75-7_35-1

schemes (Trans-Tech 2013). Although it was first found in the 1960s (Gastine et al. 1967; Sager and Tisi 1968) that the Q factors of certain resonances in a spherical dielectric sphere can be very low, meaning that there is a possibility of coupling energy into an external medium, the pioneering work of exploring the DR as an electromagnetic radiating element was only reported in 1983 (Long et al. 1983), where Professor Long systematically showed that the TM and TE modes of a cylindrical DR could be easily excited to make it an efficient microwave radiator. Since then, the dielectric resonator antenna (DRA) has been explored for a myriad of applications at high frequencies due to the absence of conductive loss. The development of the DRA is always underpinned by the advancement of dielectrics technology. Among all dielectrics, ceramic is one of the most commonly used as it is able to provide a high dielectric constant (>20) and low-loss tangent. Ceramics, however, are very hard in nature, and machining them can be very tough. Glass is another good low-loss dielectric, but its permittivity range is only 4–8. Recently, different types of soft materials such as polymers and compounds (Rashidian and Klymyshyn 2010) have been developed for the DRA, even though their dielectric constants are usually less than 5. In the past decade, the advancement of microfabrication and nanoscale technologies (Madou 2011) such as ceramic stereography, optical lithography, and laser milling has also enabled significant miniaturization of the DRA, making it viable for millimeter waves or even higher.

In the first part of this chapter, the DRA is explored for making various miniature antennas. Here, the DRA is integrated with different microwave power dividers for designing the circularly polarized (CP) and differential antennas. Ground miniaturization techniques will subsequently be used for the design of the omnidirectional CP DRA and the quasi-isotropic DRA. In the second part, a study is performed to excite the DRA using different transmission lines at millimeter-wave frequencies. It is found that the higher-order modes of the DRA can be used to relax fabrication tolerances.

Developments and Issues

Since a couple of good reviews are already available on the theoretical and experimental developments of the DRA (Huitema and Monediere 2012; Kishk 2007; Luk and Leung 2003; Petosa 2007), this section will only briefly describe the important developments and milestones throughout the years. A good introduction of some of the early works conducted in the 1980s and 1990s can also be found in (Mongia and Bhartia 1994; Petosa et al. 1998). In the 1990s, research was conducted to analyze the characteristics of the cylindrical, hemispherical (Leung et al. 1993, 1995; McAllister and Long 1984), rectangular (Ke and Cheng 2001; Mongia 1992; Mongia and Ittipiboon 1997), and triangular DRAs (Lo et al. 1999). Studies were conducted to understand their resonance modes, radiation patterns, and excitation schemes. The hemispherical DRA was the earliest to receive attention because of the existence of exact solutions. With the use of method of moments (MoM) and Green's function, the input impedance of the hemispherical DRA can be determined in a rigorous way (Leung et al. 1993, 1995). Later, analytical methods were also applied to establish the theoretical foundations for the cylindrical (Junker et al. 1994, 1996) and rectangular (Liu et al. 2002; Takashi et al. 2004) DRAs, which are more complicated to be analyzed because they have more boundaries. With the availability of powerful computers, some of the memory-intensive numerical methods such as the finite-difference time domain (FDTD), finite volume time domain (FVTD), finite element method (FEM), etc., were widely used for characterizing various DRAs (Fumeaux et al. 2004; Li and Leung 2005; Sangiovanni et al. 2004). Commercial electromagnetic softwares such as Ansys HFSS and CST Microwave Studio, which are designed from numerical methods, have also become popular tools for designing and analyzing the radiation performances of various DRAs in the recent decades.

DOI 10.1007/978-981-4560-75-7_35-1

Many excitation schemes have been developed for the DRA, and they are quite straightforward. Most of the excitation methods for the microstrip antennas can also be used to excite a DRA, including the use of coaxial probe, microstrip-fed aperture, microstrip line, and coplanar waveguide. Their design considerations will now be discussed. Probe excitation method may require drilling a hole in the DR for efficient coupling. This is a tough job as the DR can be very hard in nature. Also, a metallic probe has a larger ohmic loss and self-reactance at millimeter-wave frequencies. Using the microstrip and aperture to feed a DRA is attractive as the DR can be directly placed on the top of the excitation source. Microstrip-based feeding methods can be easily integrated with the monolithic microwave integrated circuits (MMIC). Coplanar waveguide (Al-Salameh et al. 2002), which is made on a single layer of metal, is useful for exciting the DRA when the operating wavelength goes beyond millimeter. For all of the aforementioned excitation methods, however, the formation of air gap in between the DR and ground plane is sometimes unavoidable in practice, and it can seriously affect the coupling efficiency (Junker et al. 1995). To overcome this problem, the conformal-strip feeding method was proposed in (Leung 2000), where a metallic strip was closely attached on the DR surface to excite the DRA. This excitation scheme inherits most of the advantages of the coaxial-probe feeding method. Substrate-integrated waveguide is another new form of transmission line that can be used to excite the DRA at the millimeter-wave frequencies (Hou et al. 2014).

The DRA has been explored for designing a variety of linearly polarized (LP) and circularly polarized (CP) antennas (Luk and Leung 2003; Petosa 2007) to achieve wide bandwidth and to provide multiple frequency passbands. Bandwidth widening is no doubt one of the most popular research topics in the past decades. A number of techniques such as stacked- (Kishk et al. 1989; Chair et al. 2004), polygon- (Hamsakutty et al. 2007; Kishk 2003), embedded- (Ong et al. 2004; Sangiovanni et al. 1997), and hybrid-DR (Esselle and Bird 2005; Guha et al. 2006) were explored for achieving wide bandwidth in the linearly and circularly polarized DRAs. Perforated DRA was found to be able to produce higher antenna bandwidth (Chair and Kishk 2006) because of its low Q-factor. Also, excitation of multiple higher-order modes is also a good way that can be used to extend the impedance bandwidth of a DRA as much as 40 % (Li and Leung 2005). In the past, one of the challenges that deters the DRA from practical usage is inflexibility of frequency tuning. This is because it is almost not possible to change the shape of the DR once it is made. To solve this problem, it was shown in (Ng and Leung 2005, 2006) that the resonant frequency of a DRA can be easily tuned by adding in loading strips or caps.

In recent years, the applications of the DRA have extended to portable wireless communication systems, where compactness is one of the most important criteria. To miniaturize the RF frontends, multifunction DRAs (Lim and Leung 2012) that can be used as packaging cover, oscillator load, filter element, and light cover have been proposed. Having multiple functions in a single piece of DRA will not only reduce the circuit footprint, it will also reduce the hardware costs significantly.

Miniaturized and Integrated DRAs

It is always very desirable to make an antenna very compact. The circuit size of a DRA can be miniaturized by integrating it with other components such as circulators, filters, and couplers. Having a three-dimensional (3D) structure, the DRA can accommodate its feeding circuit beneath its DR or inside the hollow region of the DR. Such a design strategy is very desirable for the circularly polarized (CP) or differentially fed DRAs since a hybrid 90° or an 180° coupler is usually required to provide two forms of the signal, which have equal magnitude with quadrature or differential in phase. Embedding the coupler beneath the DRA does not increase the footprint of the antenna, thus making the system very compact. Design examples will now be presented for a CP DRA and a differentially fed DRA.

DOI 10.1007/978-981-4560-75-7_35-1

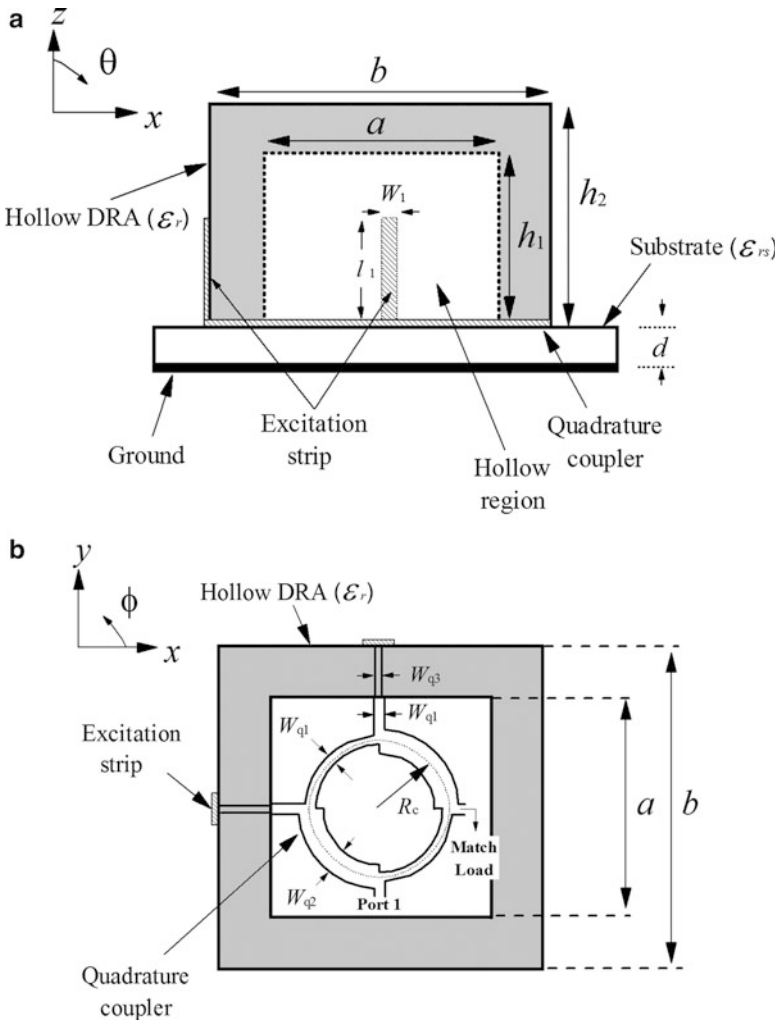


Fig. 1 CP DRA with an underlaid quadrature coupler. An external 50-Ω load is used for the matching port of the coupler. (**a**) *Front view.* (**b**) *Top view* (From Lim et al. (2011), copyright © 2011 IEEE, with permission)

CP DRAs with Underlaid Quadrature Coupler

CP system allows a more flexible orientation between the transmitting and receiving antennas, and they are very popular for certain applications such as satellite communication and global positioning system (GPS). Therefore, various singly fed or dual-fed CP DRAs have been proposed and investigated in the past two decades (Haneishi and Takazawa 1985; Oliver et al. 1995). Generally, the singly fed CP DRAs including the chamfered-corner DRA (Haneishi and Takazawa 1985), cross-slot-fed DRA (Oliver et al. 1995; Huang et al. 1999), and strip-loaded DRA (Leung and Ng 2003; Leung et al. 2000) have simpler feed networks, but their axial-ratio (AR) bandwidths are relatively narrow, with a typical value given by ~4 % only. The dual-fed DRAs can provide much wider AR bandwidths, but they usually need external quadrature couplers and therefore have larger system sizes (Mongia et al. 1994).

A compact CP rectangular DRA that employs an underlaid printed quadrature coupler was proposed in (Lim et al. 2011) for the first time. Figure 1 shows the antenna configuration. The rectangular DRA has a square cross section of side length $b = 31.8$ mm, a height of $h_2 = 20.5$ mm, and a dielectric constant of $\varepsilon_r = 10$. A hollow block with side length $a = 22$ mm and height $h_1 = 14$ mm is introduced at its bottom for housing a quadrature coupler, which is printed on a grounded substrate with a dielectric constant of $\varepsilon_{rs} = 6.15$ and a thickness of $d = 0.63$ mm. Since the coupler is located inside the hollow region of the

DOI 10.1007/978-981-4560-75-7_35-1

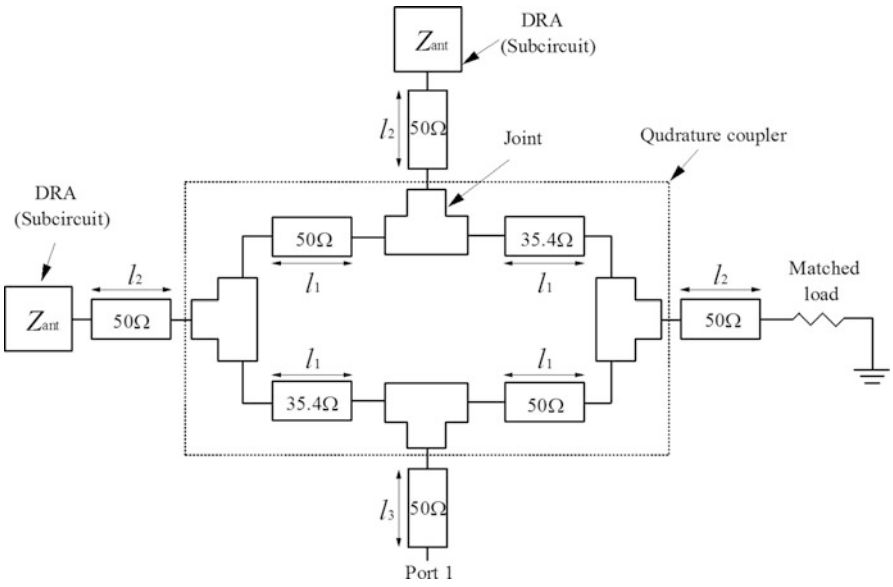


Fig. 2 Network model for the integrated CP DRA (From Lim et al. (2011), copyright © 2011 IEEE, with permission)

DRA, it can be designed without affecting the structure of the DRA. Also, because virtually no extra space is needed for accommodating the coupler, the CP antenna is very compact. A pair of adhesive conducting strips which connect to the 0°- and 90°-output ports of the coupler are stuck on two adjacent side walls of the DRA to excite the degenerate TE^x_{111} and TE^y_{111} modes. Each strip has a width of $W_1 = 0.93$ mm and a length of $l_1 = 12.5$ mm. An external load of 50 Ω is used for the matching port of the coupler.

A network model has been obtained to simplify the design of the integrated CP antenna. Figure 2 shows the network model, in which the quadrature coupler is represented by its equivalent transmission lines and the functional block of the isolated DRA is extracted with the coupler removed. Using Ansys HFSS, the input impedance (Z_{ant}) is simulated by looking into one of the two excitation strips with the remaining excitation strip connected to a 50-Ω matched load. The whole model was then simulated in Microwave Office.

Figure 3 shows the measured and simulated reflection coefficients as a function of frequency. With reference to the figure, the reflection coefficient obtained by using the network model agrees reasonably well with the HFSS simulation and measurement. The discrepancy is expected because the network model does not take into account the mutual couplings. Since a hollow DRA is used, the measured 10 dB impedance bandwidth is as wide as 24.95 %. The measured and simulated ARs in the boresight direction ($\theta = 0^0$) are shown in Fig. 4. Also, a very wide 3-dB AR bandwidth of 33.8 % is obtained, ranging from 2.02 to 2.8 GHz. Figure 5 shows the antenna gains. The measured and simulated antenna gains are 6.18 dBi and 6.45 dBi at resonance frequencies, respectively. As can be seen from the figure, the measured gain is generally lower than the simulated one due to imperfections of the DRA and coupler. Figure 6 shows the broadside radiation patterns of the CP DRA at 2.40 GHz. The measured left-handed CP (LHCP) fields are stronger than the right-handed CP (RHCP) fields by 25 dB in the boresight direction.

Since a strip stuck on a DRA can inherently provide an impedance, the external 50-Ω resistor can also be replaced by a loading strip at the matching port (Lim et al. 2011). Figure 7 shows the proposed antenna configuration. In this case, a third strip that is attached on the corner of DRA is connected to the match port through an extended microstrip line. The strip has a width of $W_2 = 0.58$ mm and a length of $l_2 = 4$ mm. It was found that the impedance level can be easily adjusted by varying l_2, and a good match can be observed when a suitable l_2 is used. Figure 8 shows the simulated and measured ARs of the antenna. The measured

DOI 10.1007/978-981-4560-75-7_35-1

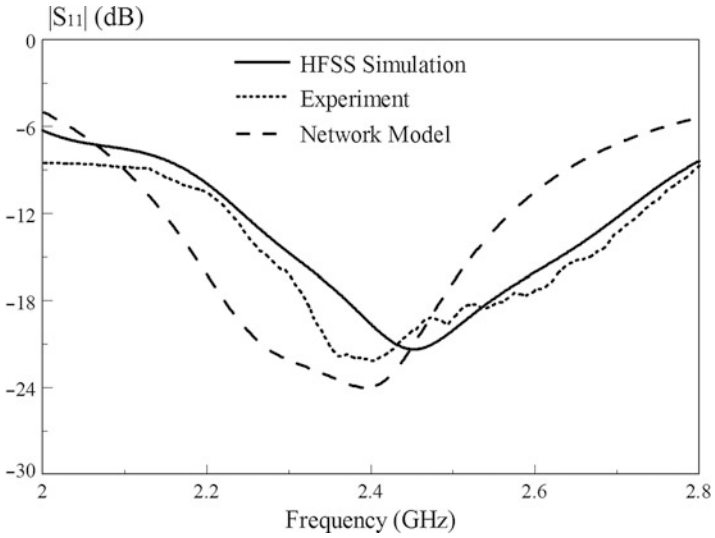


Fig. 3 Measured and simulated reflection coefficients of the CP DRA with external 50-Ω load (From Lim et al. (2011), copyright © 2011 IEEE, with permission)

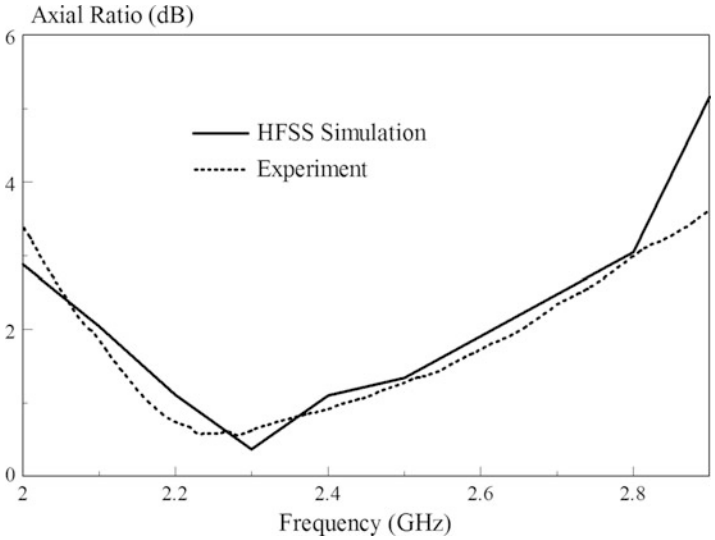


Fig. 4 Measured and simulated ARs of the CP DRA with external 50-Ω load (From Lim et al. (2011), copyright © 2011 IEEE, with permission)

AR bandwidth is found to be ∼10 %, narrower than that for the former configuration. This is as expected since the DRA-loaded strip is unable to provide a constant 50-Ω load.

Differential DRA with Underlaid 180° Hybrid Coupler

The design idea was also used to realize a differential DRA (Fang et al. 2010), which is differentially fed by a 180° hybrid coupler to suppress the common-mode interference, and consequently to increase the signal-to-noise ratio. Figure 9 shows the antenna configuration. The hollow DRA with parameters of $a = 22$ mm, $b = 31.8$ mm, $d_1 = 14$ mm, $d_2 = 20.5$ mm, and $\varepsilon_r = 10$ is excited in its TE^y_{111} mode. In the hollow region of the DRA, there is a rat-race coupler printed on a Duroid substrate with a dielectric constant of $\varepsilon_{rs} = 6.15$ and a thickness of $h = 0.63$ mm. A pair of conducting strips of width $w = 0.58$ mm and length $l = 10.5$ mm are connected to the two output ports of the rat-race coupler.

DOI 10.1007/978-981-4560-75-7_35-1

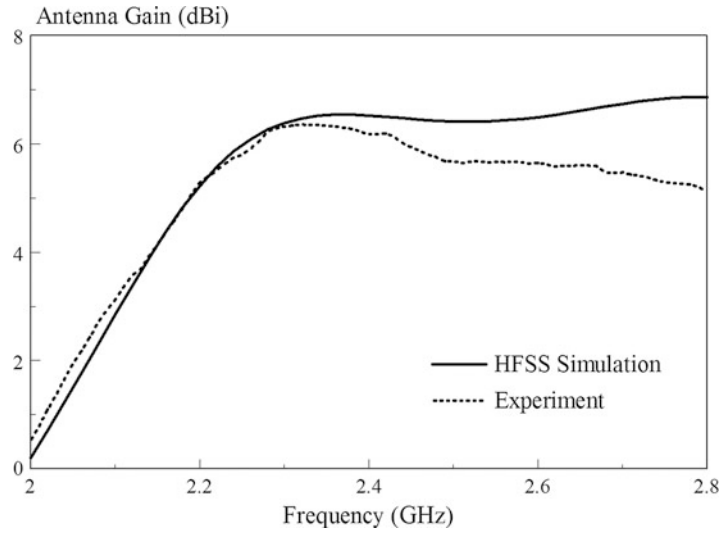


Fig. 5 Measured and simulated antenna gain of the CP DRA with external 50-Ω load (From Lim et al. (2011), copyright © 2011 IEEE, with permission)

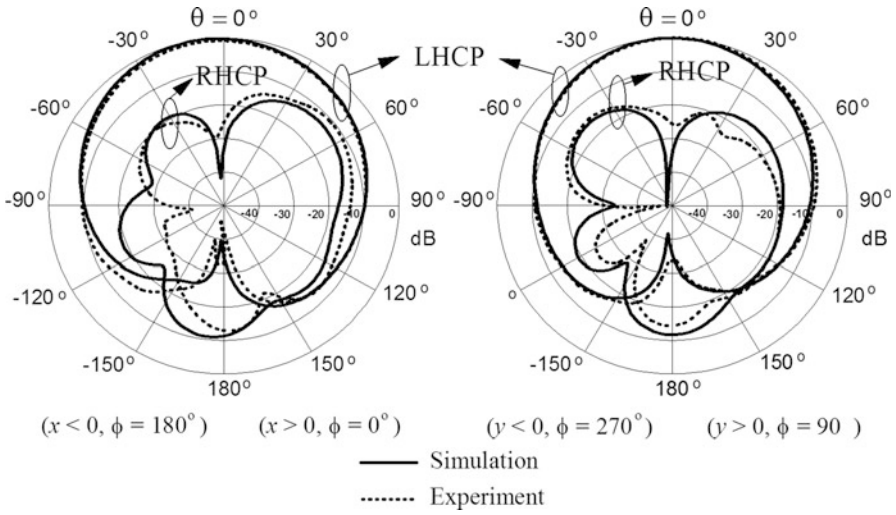


Fig. 6 Measured and simulated normalized radiation patterns of the CP DRA with external 50-Ω load (From Lim et al. (2011), copyright © 2011 IEEE, with permission)

They are stuck on the two opposite walls of the DRA to feed the antenna differentially. The isolation port of the rat-race coupler is terminated by an external matched resistor of 50 Ω.

Figure 10 shows the measured and simulated reflection coefficients of the differential DRA as a function of frequency. As can be seen from the figure, the measured and simulated impedance bandwidths are given by 11.8 % (2.24–2.52 GHz) and 10.6 % (2.24–2.49 GHz), respectively. Figure 11 shows the measured and simulated radiation patterns at 2.4 GHz. Broadside radiation patterns are observed for both of the E- and H-planes as expected, since the fundamental TE^{y}_{111} mode is excited. In each plane, the cross-polarized field is weaker than its co-polarized counterpart by more than 26 dB in the boresight direction. For comparison, a single feeding strip was also used to feed the same hollow DRA. It was found that the cross-polarized fields of the differential DRA are generally lower than those of the single-ended design, especially for field pattern in the H-plane.

DOI 10.1007/978-981-4560-75-7_35-1

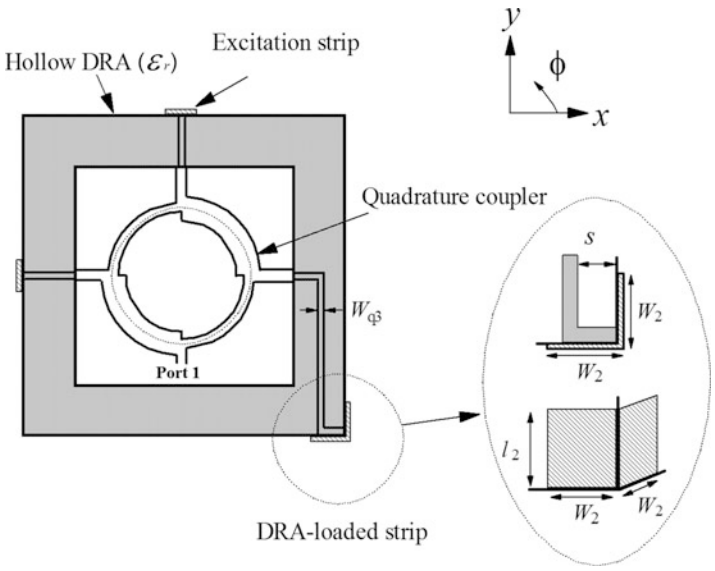


Fig. 7 *Top view* of the proposed CP DRA. A loading strip is added at the DRA corner to provide a 50-Ω load for the matching port of the coupler (From Lim et al. (2011), copyright © 2011 IEEE, with permission)

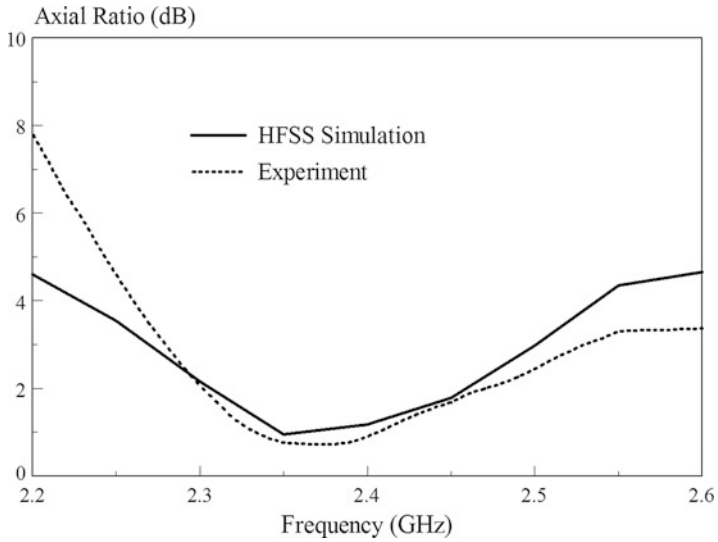


Fig. 8 Measured and simulated ARs of the CP DRA with DRA-loaded strip (From Lim et al. (2011), copyright © 2011 IEEE, with permission)

It should be mentioned that the underlaid technique can also be extended to solid DRAs. One factor to be noted is that the design of the coupler should take into account the loading effect of the upper DRA.

Compact-Ground DRAs

In the past three decades, various antenna geometries, excitation schemes, and bandwidth enhancement techniques for the design of DRA have been developed. Thus far, studies of the DRA have mainly focused on the main radiating element, and relatively much less attention has been paid to the other important constituent part – the ground plane. However, since the time-varying current formed on the ground plane

DOI 10.1007/978-981-4560-75-7_35-1

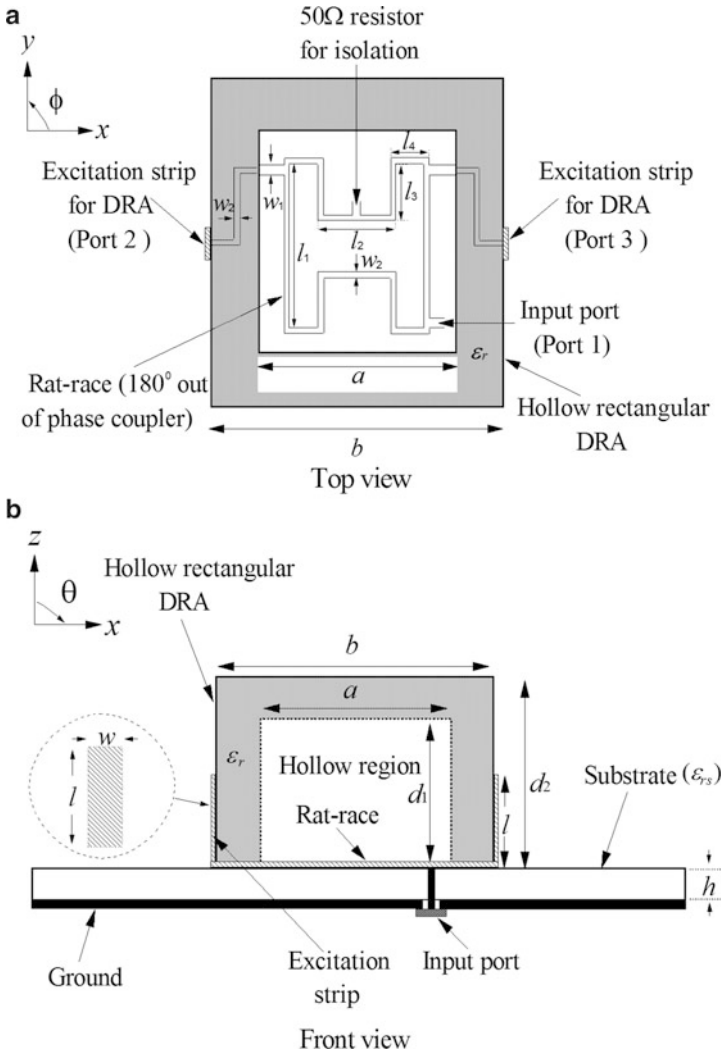


Fig. 9 Configuration of the differential rectangular hollow DRA with an underlaid rat-race coupler. (**a**) *Top view*. (**b**) *Front view* (From Fang et al. (2010), copyright © 2010 IEEE, with permission)

can also generate radiation, the dimension and shape of the ground plane may also play a crucial role on the antenna performance . Actually, it was found that very small ground planes are essential for some specific antennas such as the following omnidirectional CP DRA and isotropic DRA to provide a satisfactory performance. Having a smaller ground will also definitely make the DRA more compact.

Omnidirectional CP DRA

CP antennas are widely used in modern wireless communication systems because they can mitigate multipath and fading problems. On the other hand, omnidirectional radiation patterns are generally desirable since they can provide larger signal coverage and stabilize the signal transmission. As a result, the investigation of omnidirectional CP antennas has attracted increasing attention in recent years (Kawakami et al. 1997; Nakano et al. 2000; Park and Lee 2011; Quan et al. 2013; Row and Chan 2010).

Figure 12 shows the configuration of the first omnidirectional CP DRA (Pan et al. 2012), which is a rectangular DRA with four inclined slots fabricated on its sidewall. The DRA has a length of a, a width of b, a height of h, and a dielectric constant of ε_r, and each slot has a width of w and a depth of d. A coaxial probe of length l and radius r_1 is used to centrally feed the antenna. The probe is extended from the inner

DOI 10.1007/978-981-4560-75-7_35-1

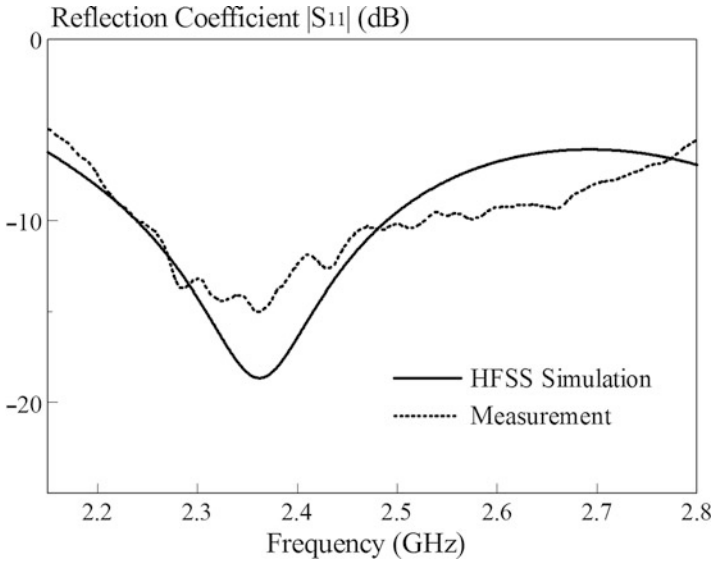


Fig. 10 Measured and simulated reflection coefficients of the differential rectangular hollow DRA versus frequency (From Fang et al. (2010), copyright © 2010 IEEE, with permission)

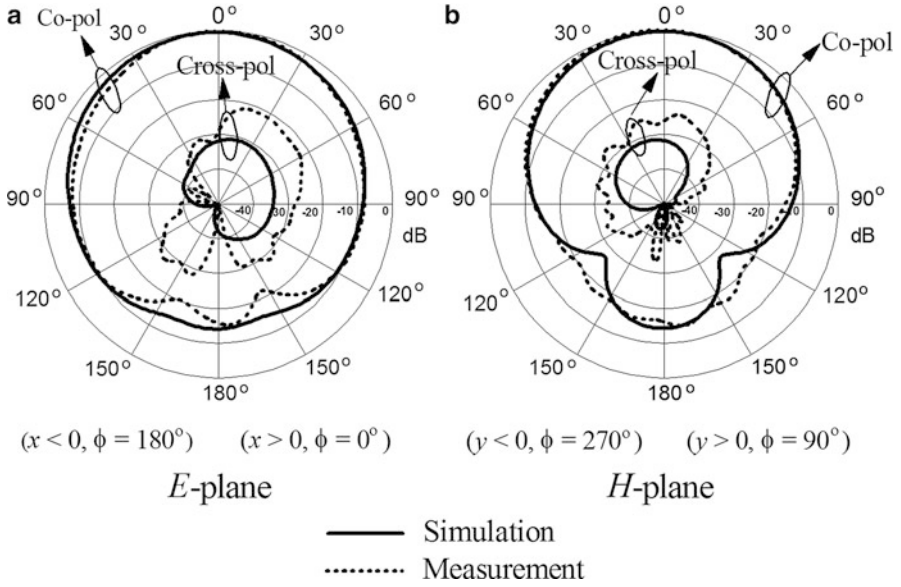


Fig. 11 Measured and simulated radiation patterns of the proposed differential rectangular hollow DRA at 2.4 GHz (From Fang et al. (2010), copyright © 2010 IEEE, with permission)

conductor of an SMA connector. To enable maximum radiation around $\theta = 90^{o}$ and minimize the radiated fields from tilting, the flange of the SMA is used as the (small) ground plane. The working principle of the antenna is hereby explained. When a rectangular DRA is centrally fed by a coaxial probe, it will generate monopolar linearly polarized (LP) fields. By introducing a slot on each DRA sidewall, the vertical E_z field from the vertical monopole is perturbed and can be resolved into orthogonal $E_{\perp}$ and $E_{//}$ components which are perpendicular and parallel to the inclined slot. Since the slotted DR behaves effectively as an anisotropic dielectric (Kirschbaum and Chen 1957), the two components will travel at

DOI 10.1007/978-981-4560-75-7_35-1

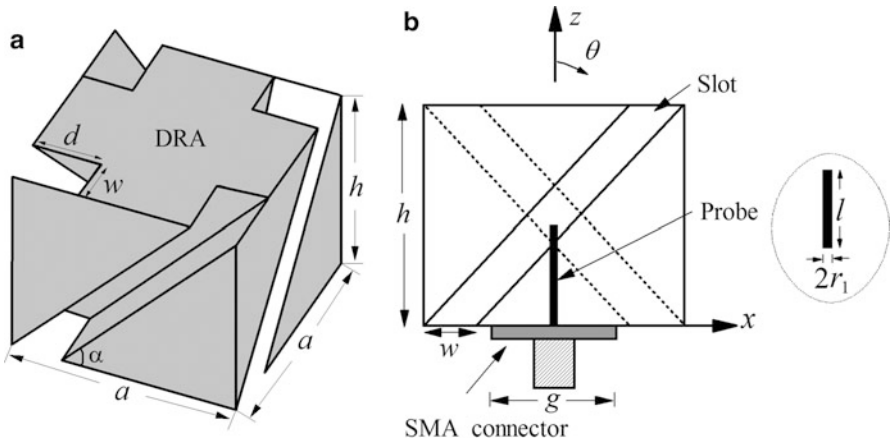


Fig. 12 Configuration of the omnidirectional CP DRA. (**a**) *Perspective view*. (**b**) *Front view* (From Pan et al. (2012), copyright © 2012 IEEE, with permission)

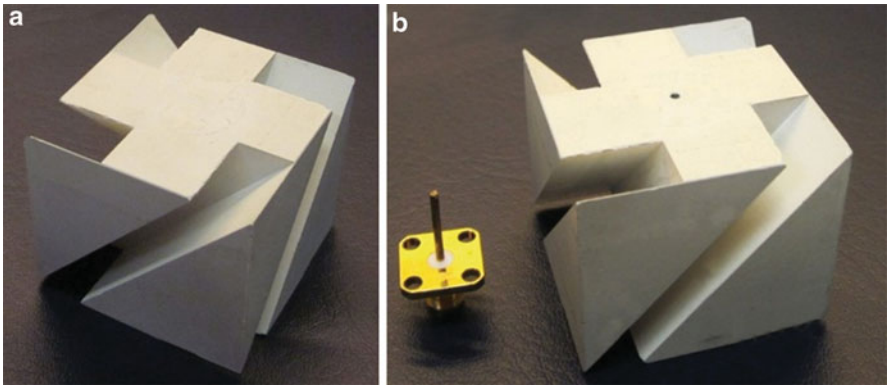


Fig. 13 Prototype of the proposed omnidirectional CP DRA. (**a**) Photograph showing the top face and sidewalls. (**b**) Photograph showing the bottom face of the DRA (From Pan et al. (2012), copyright © 2012 IEEE, with permission)

different velocities, causing a phase difference between them. When the phase difference reaches 90° and $|E_{\perp}| = |E_{/\!/}|$, a CP wave emerges from the DRA.

Figure 13 shows two photographs of a prototype operating at 2.4-GHz WLAN frequency range, with parameters given by $\varepsilon_r = 15$, $a = 39.4$ mm, $h = 33.4$ mm, $w = 9.4$ mm, $d = 14.4$ mm, $l = 12.4$ mm, $r_1 = 0.63$ mm, and $g = 12.7$ mm. The reflection coefficients and axial ratios of the proposed CP DRA are shown in Fig. 14a, b, respectively. With reference to the figures, the measured 10-dB impedance bandwidth and 3-dB AR bandwidth are 24.4 % (2.30−2.94 GHz) and 7.3 % (2.39−2.57 GHz), respectively. The measured AR passband falls within the impedance passband, and therefore the entire AR passband is usable. Figure 15 shows the simulated and measured radiation patterns in the elevation (*xz*) and azimuth (*xy*) planes. As can be observed from the figure, the elevation pattern has a null in the boresight direction and the maximum radiation occurs at $\theta = 90°$, whereas the azimuthal pattern is omnidirectional. In the whole azimuth plane, the co-polarized LHCP field is at least ~20 dB stronger than the corresponding cross-polarized RHCP counterpart, showing good CP performance.

For an omnidirectional CP antenna, the ground-plane size usually affects the antenna performance considerably. Figure 16 shows the results for different side lengths of ground plane. As can be observed from the figure, the impedance and AR bandwidths are reduced from 20.3 % to 10.6 % and from 8.2 % to 4.2 %, respectively, with an increase of g from 12.7 to 32.7 mm. But when g exceeds 42.7 mm, the AR

DOI 10.1007/978-981-4560-75-7_35-1

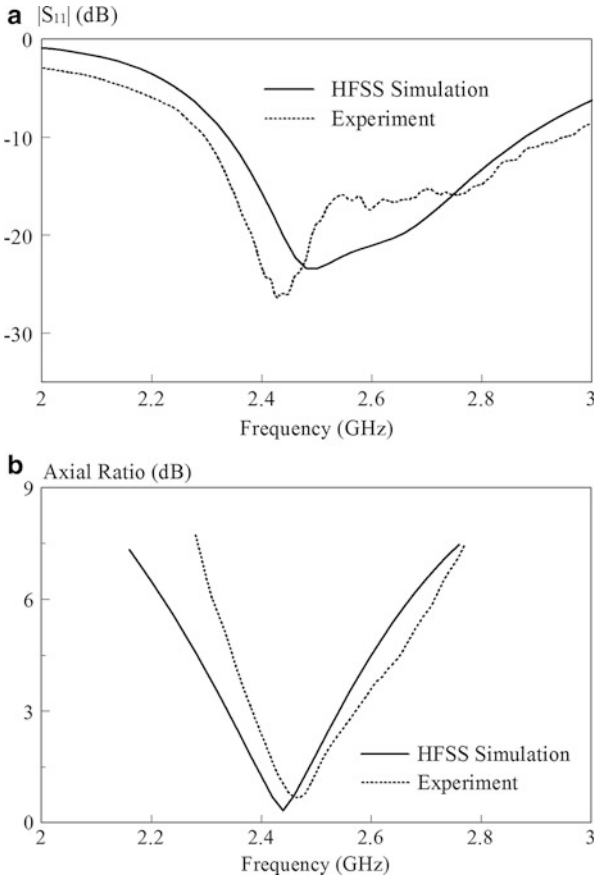


Fig. 14 Measured and simulated reflection coefficients and ARs of the omnidirectional CP DRA. (**a**) Reflection coefficients. (**b**) ARs (From Pan et al. (2012), copyright © 2012 IEEE, with permission)

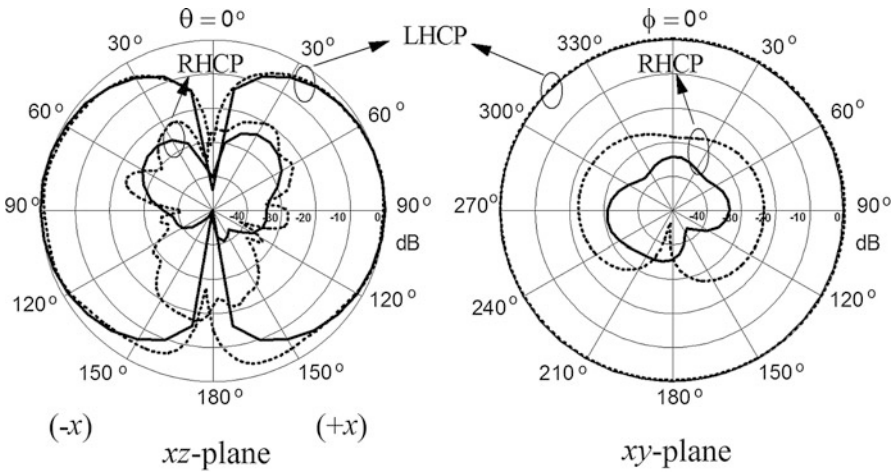


Fig. 15 Measured and simulated radiation patterns of the omnidirectional CP DRA (From Pan et al. (2012), copyright © 2012 IEEE, with permission)

DOI 10.1007/978-981-4560-75-7_35-1

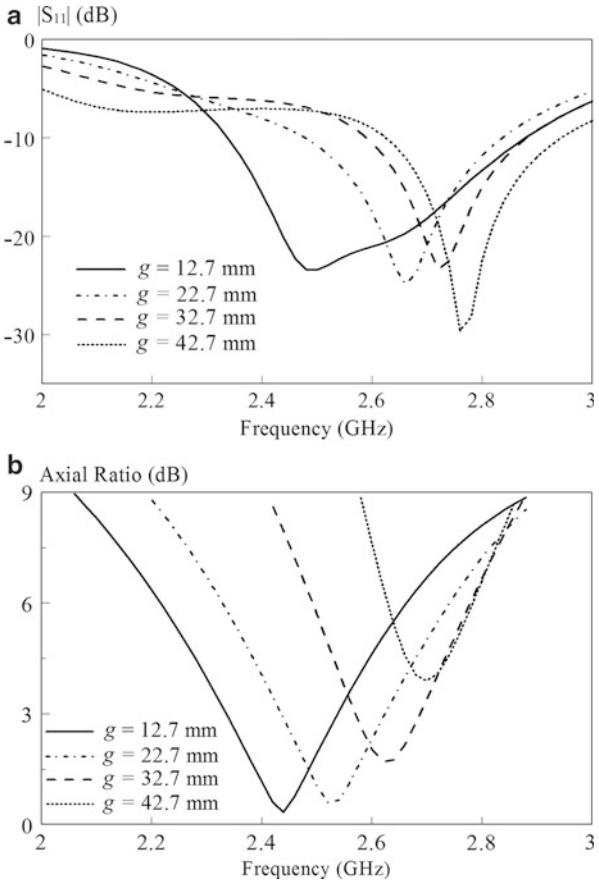


Fig. 16 Simulated reflection coefficient and AR of the omnidirectional CP DRA as a function of frequency for different side lengths of the ground plane. (**a**) Reflection coefficient. (**b**) AR (From Pan et al. (2012), copyright © 2012 IEEE, with permission)

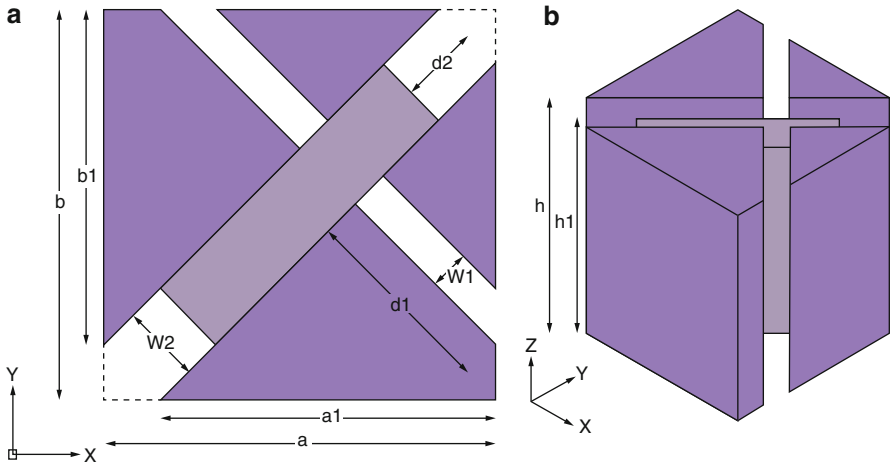


Fig. 17 Configuration of the omnidirectional CP DRA. (**a**) *Top view*. (**b**) *Perspective view* (From Khalily et al. (2014), copyright © 2014 IEEE, with permission)

DOI 10.1007/978-981-4560-75-7_35-1

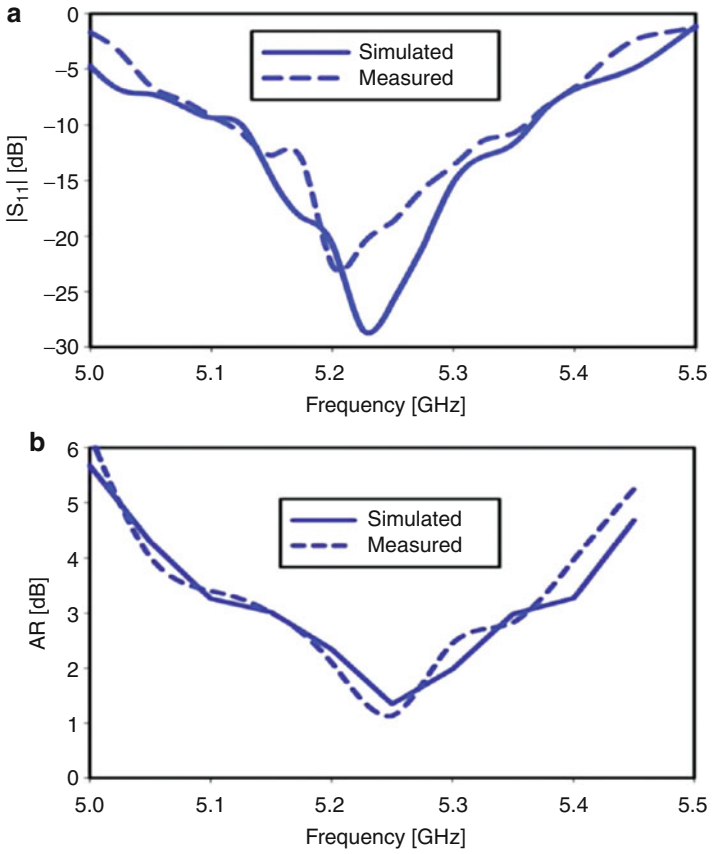


Fig. 18 Measured and simulated reflection coefficients and ARs of the omnidirectional CP DRA: $\varepsilon_r = 10$, $a = 31$ mm, $h = 24$ mm, $h_1 = 19$ mm, $w_1 = 5$ mm, $d_1 = 9.2$ mm, $w_2 = 10$ mm, $d_2 = 5$ mm, $l = 13.25$ mm, $r_1 = 0.63$ mm, and $g = 12.7$ mm. (**a**) Reflection coefficients. (**b**) ARs (From Khalily et al. (2014), copyright © 2014 IEEE, with permission)

deteriorates greatly, and the entire AR curve becomes higher than the 3-dB level. Poor AR is obtained because two orthogonal components are usually required to generate CP fields, but the boundary condition of a large ground plane requires that the tangential *E*-field component be zero on its surface and only the perpendicular *E*-field component remains. Therefore, a small ground plane is preferable in designing an omnidirectional CP antenna.

A similar slotted omnidirectional CP DRA has also been presented in (Khalily et al. 2014). Figure 17 shows the corresponding antenna configuration. In this case, vertical slots rather than inclined ones are introduced onto the sidewall of the DRA to generate CP fields. Again, the antenna is centrally fed by a coaxial probe, and the flange of the SMA acts as the small ground plane. Figure 18 shows the reflection coefficient and AR of the CP DRA. With reference to the figure, the measured 10-dB impedance bandwidth is given by 4.57 % (5.13–5.37 GHz), and the 3-dB AR bandwidth is 4.18 % (5.15–5.37 GHz). The omnidirectional radiation patterns of the DRA are similar to those shown in Fig. 15.

For the above two antennas, slots need to be fabricated on dielectric surfaces which can be difficult especially when a cylindrical or hemispherical DRA is used. Therefore, a novel omnidirectional CP DRA loaded with a modified Alford loop was proposed in (Li and Leung 2013). Figure 19 shows the configuration of the cylindrical DRA with a diameter of D, a height of H, and a dielectric constant of ε_r. The DRA is axially fed by a coaxial probe at the center of its bottom, exciting its fundamental $TM_{01\delta}$ mode. A modified Alford loop which consists of a center patch and four conducting curved branches is placed on top of the DRA for generating CP fields. The center piece of the patch has a diameter of d_1, and

DOI 10.1007/978-981-4560-75-7_35-1

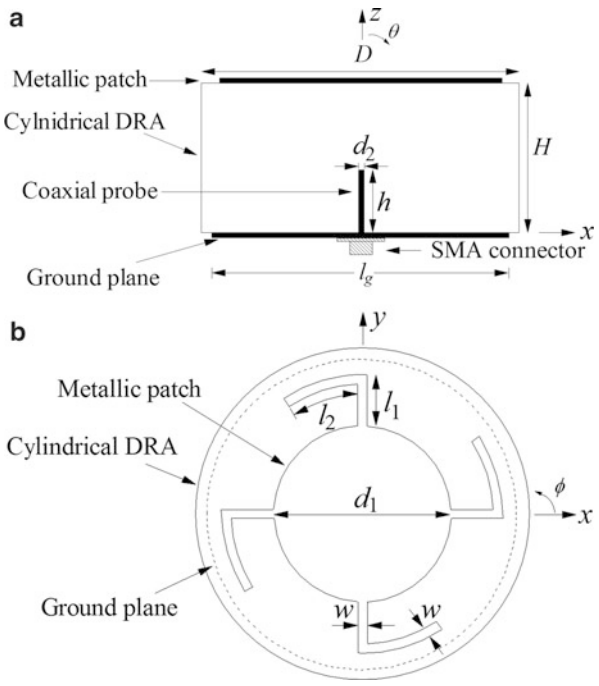


Fig. 19 Configuration of the omnidirectional CP cylindrical DRA. (**a**) *Side view* and (**b**) *top view* (From Li and Leung (2013), copyright © 2013 IEEE, with permission)

each of the extended branches has a width, radial length, and arc length of w, l_1, and l_2, respectively. Figure 20 shows the simulated and measured reflection coefficients and ARs of the CP antenna. Two modes, one is caused by the DRA while another is caused by the feeding probe, have merged together, giving an impedance bandwidth of 13.2 % (2.34–2.67 GHz). The measured 3-dB AR bandwidth is 10.5 % (2.26–2.51 GHz). Therefore, the usable overlapping bandwidth of the AR and impedance passbands is 7.0 % (2.34–2.51 GHz), which is almost the same as the previous slotted DRA (Pan et al. 2012).

Another omnidirectional CP DRA having parasitic patches on its sidewalls has been proposed in (Leung et al. 2013). Figure 21 shows the configuration of the proposed design. In this case, four patches are inclined along the diagonals of the sidewalls to effectively perturb the DRA fields, thus exciting nearly degenerate modes that lead to the generation of CP fields. As shown in Fig. 22, the measured impedance bandwidth and AR bandwidth of the CP DRA are comparable, given by 6.6 % and 6.9 %, respectively.

As shown in the above designs, both of the inclined slots and parasitic patches (or strips) can generate CP fields. Therefore, it is possible to obtain two CP modes by placing a parasitic strip inside each slot of the original slotted DRA (Pan et al. 2012). Figure 23 shows the configuration of the wideband omnidirectional CP antenna (Pan and Leung 2012). Four identical conducting strips of length l_s and width w_s are introduced at a depth of x_0 from the DR surface. It was found that by properly tuning the length and location of the strip, a new CP mode induced by the strips can be obtained and combined with the DRA CP mode to broaden the AR bandwidth. Moreover, a circular cylinder of radius r can be removed from the center of DR to accommodate the feeding probe and also to enhance the impedance bandwidth. The simulated and measured reflection coefficients and ARs are shown in Fig. 24. As can be seen from the figure, both of the impedance and AR bandwidths of the antenna read ∼24 %. The usable overlapping bandwidth of the antenna is given by 22.0 % (3.16−3.94 GHz), which is more than sufficient for many wireless communication applications. Furthermore, the omnidirectional radiation patterns are found to be very good and stable across the entire passband.

DOI 10.1007/978-981-4560-75-7_35-1

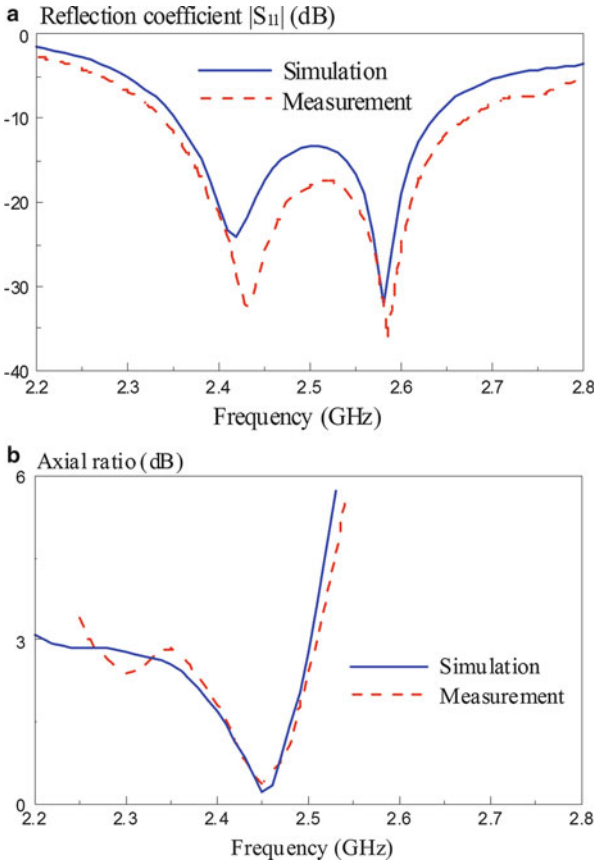


Fig. 20 Measured and simulated reflection coefficients and ARs of the omnidirectional CP DRA. $H = 22$ mm, $D = 49$ mm, $\varepsilon_r = 10$, $h = 10$ mm, $d_2 = 1.27$ mm, $d_1 = 26$ mm, $l_1 = 8$ mm, $l_2 = 11.5$ mm, $w = 1.4$ mm, and $l_g = 47$ mm. (**a**) Reflection coefficients. (**b**) ARs (From Li and Leung (2013), copyright © 2013 IEEE, with permission)

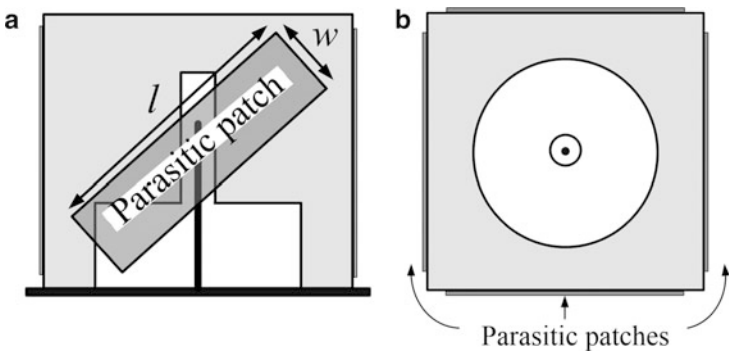


Fig. 21 Configuration of the proposed omnidirectional CP glass DRA with parasitic patches. (**a**) *Front view*. (**b**) *Top view* (From Leung et al. (2013), copyright © 2013 IEEE, with permission)

Compact Quasi-Isotropic DRA

For certain applications such as radio frequency identification (RFID) and wireless access points, it is very desirable to have an isotropic antenna which can provide full coverage to maintain good communication links at all angles. An isotropic antenna is usually obtained by arraying a number of discrete elements in a circle (Chen et al. 2012; Zhang et al. 2011) or by using two perpendicular half-wave linear dipoles with a

DOI 10.1007/978-981-4560-75-7_35-1

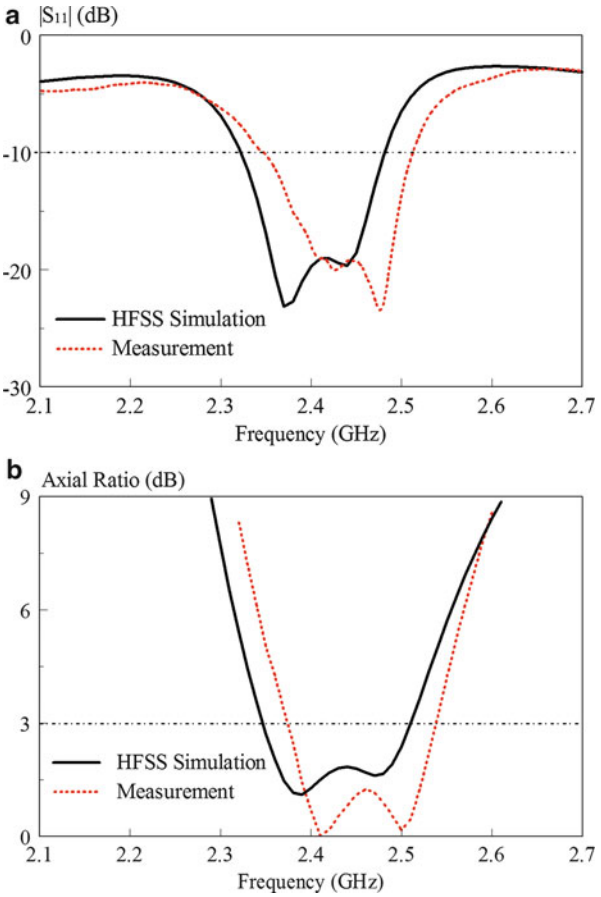


Fig. 22 Measured and simulated reflection coefficients and ARs of the omnidirectional CP DRA. (**a**) Reflection coefficients. (**b**) ARs (From Leung et al. (2013), copyright © 2013 IEEE, with permission)

phase difference of 90° (Radnovic et al. 2010; Kraus and Marhefka 2003), but the two methods involve either complex structures or complicated feeding networks.

A very simple quasi-isotropic DRA was investigated in (Pan et al. 2014) by using the idea of complementary antennas. The scheme of complementary antenna concept is shown in Fig. 25. From the figure, the E- and H-plane radiation patterns of an electric dipole are found to be of the figure-8 and figure-O shapes, respectively, and the two patterns are interchanged for a magnetic dipole. Therefore, when an electric dipole and a magnetic dipole are combined perpendicularly, the null-field direction of one of them is along the maximum-field direction of the other. Consequently, there is no blind spot, and it is possible to provide a quasi-isotropic antenna.

Figure 26 shows the configuration of the quasi-isotropic DRA. It is a rectangular DRA with a square cross section, which has a side length of a, a height of d, and a dielectric constant of ε_r. The DRA is fed by a coaxial probe of length l and radius r, located at a displacement of y_0 from the center. A small square ground plane with a side length of g is used for the antenna. The required equivalent magnetic and electric dipoles for obtaining a quasi-isotropic antenna are provided by the DR and ground plane, respectively. The working principle of the antenna is illustrated in Fig. 27, which shows the surface current on the ground plane and the fields in the DRA. With reference to Fig. 27a, the ground plane has a nearly uniform current distribution, and it can be approximated as an electric dipole along y axis. With reference to Fig. 27b, c, the fields of the DRA are very similar to those produced by a short x-directed magnetic dipole.

DOI 10.1007/978-981-4560-75-7_35-1

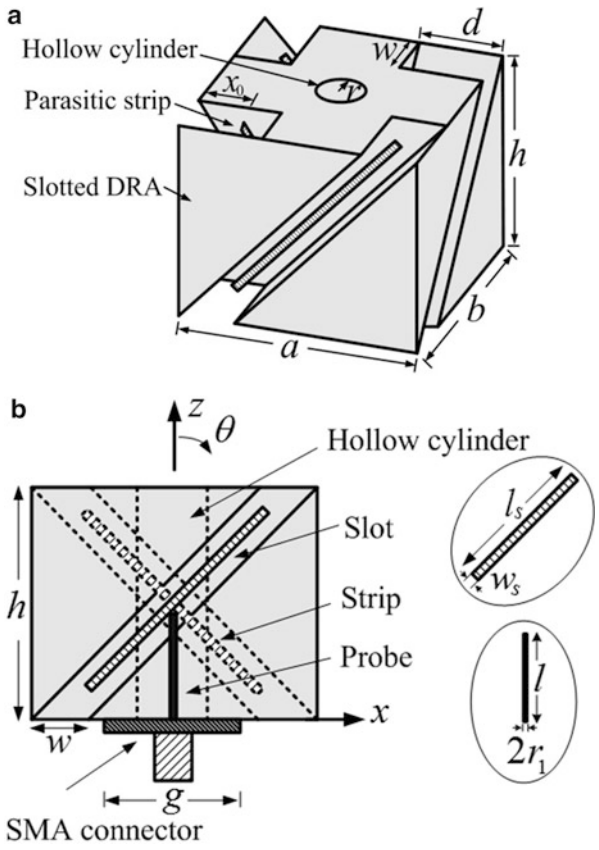


Fig. 23 Configuration of the proposed wideband omnidirectional CP DRA. (**a**) *Perspective view.* (**b**) *Front view* (From Pan and Leung (2012), copyright © 2012 IEEE, with permission)

As a result, the DRA with small ground plane is equivalent to a pair of perpendicular magnetic and electric dipoles, providing quasi-isotropic fields.

Figure 28 shows the measured and simulated reflection coefficients of the quasi-isotropic DRA. Since the coaxial feed line is an unbalanced structure, a choke (balun) is used to obtain a balanced current in the measurement. The measured 10-dB impedance bandwidth of the antenna is 6.9 % (2.38–2.55 GHz), agreeing well with the simulated result. Figure 29 shows the three-dimensional (3D) pattern of the total field E_T. It can be seen from Fig. 29a that the simulated E_T is independent of ϕ and the same field pattern is obtained for all elevation planes. The difference between the maximum and minimum radiation power densities is about 5.6 dB over the entire pseudo-spherical radiating surface, which is 2.6 dB higher than the theoretical value of 3 dB. This is because the two dipole sources are not ideal. Similar pattern has been obtained in the measurement, as shown in Fig. 29b. A much more isotropic pattern can be realized by using a narrower rectangular ground plane. Figure 30 shows the simulated surface current on the ground plane (24×6 mm^2) of the DRA with $\varepsilon_r = 10$, $a = 24$ mm, $d = 17$ mm, $l = 4.5$ mm, and $y_0 = 9$ mm. It can be clearly seen that the current distribution on the narrower ground plane is quite similar to that of a dipole antenna and thus the gain difference of the antenna can be reduced to 4.5 dB. An even smaller difference of 3.4 dB can be obtained using a DRA with a higher dielectric constant of ε_r=15 .

DOI 10.1007/978-981-4560-75-7_35-1

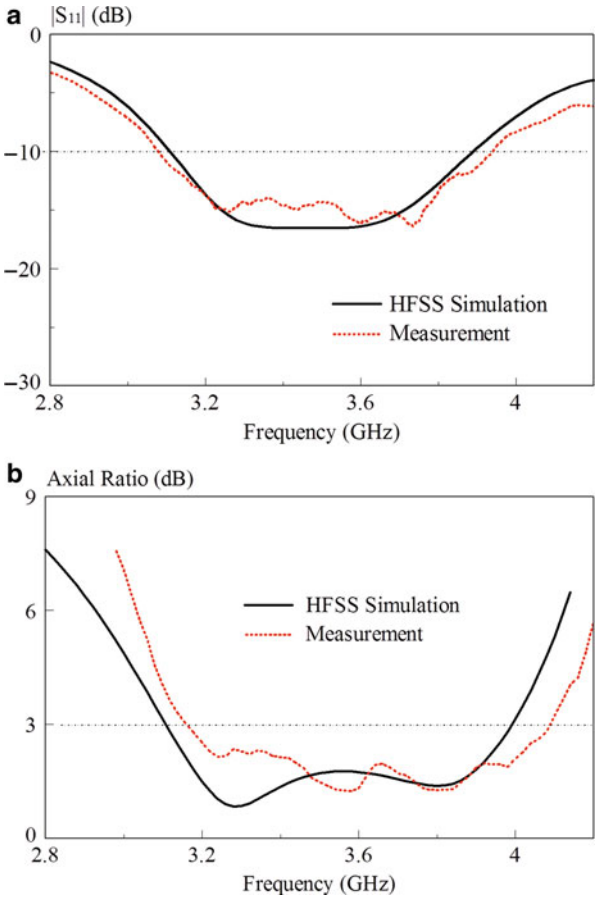


Fig. 24 Measured and simulated reflection coefficients and ARs of the wideband omnidirectional CP DRA. $\varepsilon_r = 15$, $a = b = 30$ mm, $h = 25$ mm, $r = 3$ mm, $w = 7$ mm, $d = 10.5$ mm, $l_s = 30.5$ mm, $w_s = 1$ mm, $x_0 = 6.4$ mm, $r_1 = 0.63$ mm, and $l = 19$ mm. (**a**) Reflection coefficients. (**b**) ARs (From Pan and Leung (2012), copyright © 2012 IEEE, with permission)

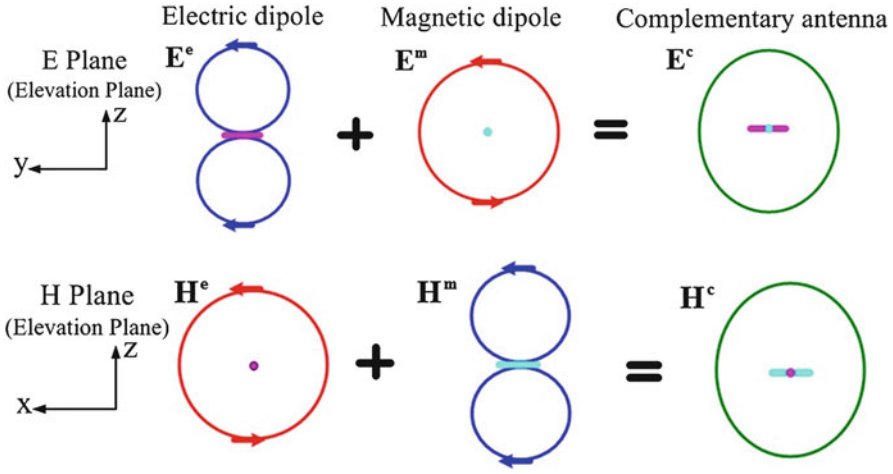


Fig. 25 Scheme of complementary antenna

DOI 10.1007/978-981-4560-75-7_35-1

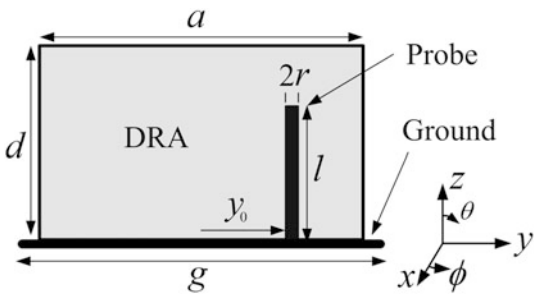


Fig. 26 Probe-fed rectangular DRA with small ground plane (From Pan et al. (2014), copyright © 2014 IEEE, with permission)

Millimeter-Wave Miniature DRAs

Millimeter-wave frequency ranges such as Ka (26.5–40 GHz), Q (40.5–43.5 GHz), V (40–75 GHz), and W (75–110 GHz) bands are traditionally used for military, astronomy, and remote sensing. Particularly, the International Telecommunication Union (ITU) has designated 30–110 GHz as the extremely high frequency (EHF) band. The key advantage of using millimeter-wave communication link is the possible availability of large amount of spectral bandwidth which may enable the transmission of high volume of data (Kay 1966; Rappaport et al. 2011). Nevertheless, millimeter waves are easily absorbed by rain, gases, and water vapor in the atmosphere (Weibel and Dressel 1967). Signals in the frequency range of 57–64 GHz can also be seriously attenuated by the resonances of oxygen molecules. As a result, in practice, millimeter-wave communication links require the use of high-gain antennas (Li and Luk 2014). This frequency range has already been successfully deployed for automotive radar sensing systems (Menzel and Moebius 2012) which do not require far radiation coverage. Parabolic reflectors and horn-fed lens are among the earliest nonplanar structures that are able to provide high antenna gain in the millimeter-wave ranges (Kay 1966). Later, planar antennas such as slots, microstrip patches, and printed dipoles were also explored (Pozar 1983; Vilar et al. 2014). Due to its small wavelength, the main challenges of making a millimeter-wave antenna are fabrication accuracy and feed line loss (Schwering 1992). Different from the microstrip-based resonators, the DR itself does not possess conductive loss. This feature is very desirable for enhancing radiation efficiency. Although it is stretching into a three dimensional, the DRA is still much more compact compared to reflectors and lenses.

The development of the DRA in the millimeter-wave ranges is boosted by the availability of microfabrication technology and low-loss dielectrics. Advancement in micromachining techniques such as photolithography and laser milling has enabled the fabrication of very small antenna structures. Ceramic stereography was deployed for fabricating periodic structures such as DRA arrays in (Buerkle et al. 2006; Brakora et al. 2007). Effort has also been made to explore low-loss materials for millimeter-wave applications. Ceramic is conventionally used for the DRA because it is able to provide high dielectric constant ($\varepsilon_r > 20$). However, ceramic material is hard in nature and difficult to be machined, making the fabrication and tuning processes extremely difficult. Soft materials such as polymer and plastic compounds have been developed for ease of machining, but such materials can only achieve a permittivity range of 3–5 (Zou et al. 2002; Wasylyshyn 2005; Koulouridis et al. 2006; Rashidian and Klymyshyn 2010). Having low dielectric constant may not be a problem as the antenna size is usually small at millimeter-wave frequency. It allows for the flexibility of making the antenna size larger to overcome fabrication imperfections. But the key issue is that a DRA with low dielectric constant may not be able to be excited efficiently.

Polymer materials such as liquid crystal polymer (LCP), polydimethylsiloxane (PDMS), and polyoxymethylene (POM) are among the low-loss dielectrics that can be used for making the

DOI 10.1007/978-981-4560-75-7_35-1

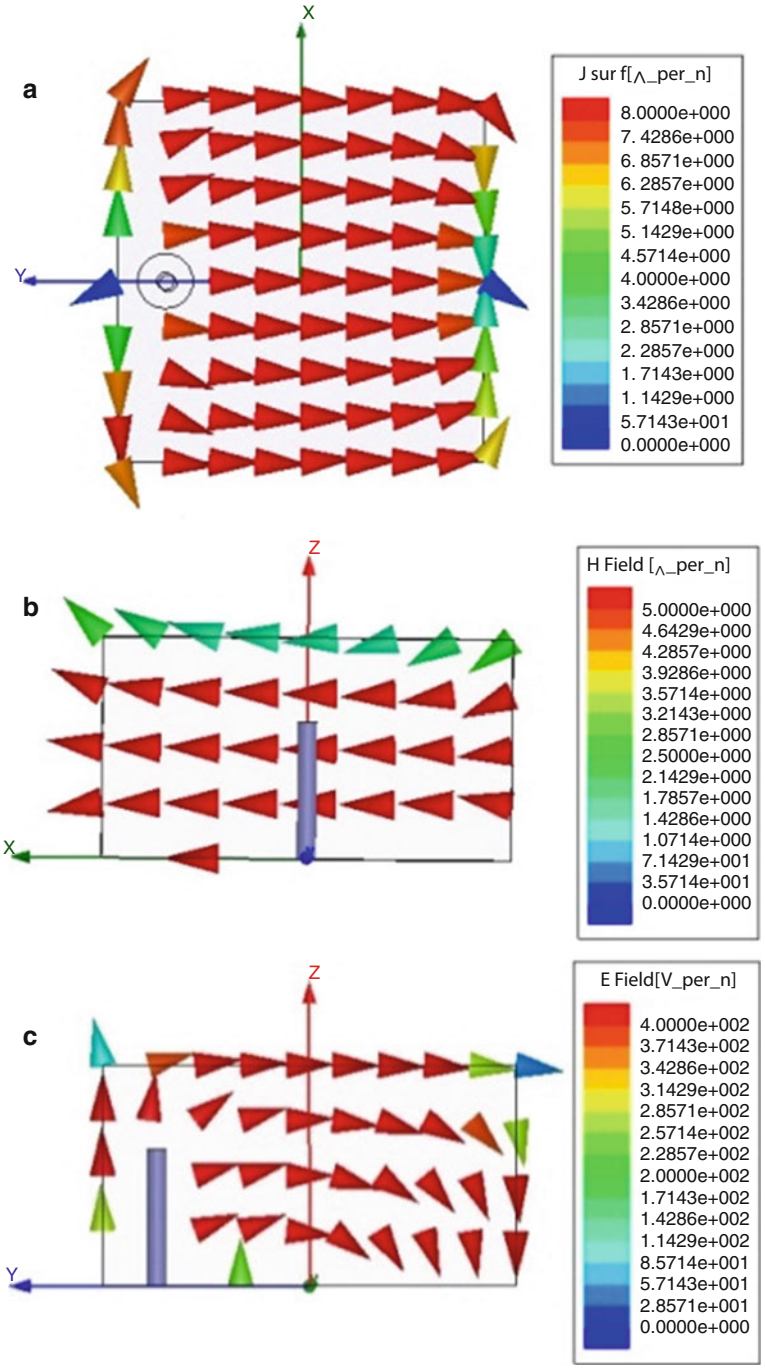


Fig. 27 Surface current on the ground plane and the fields in the quasi-isotropic DRA. (**a**) Surface current on the ground plane. (**b**) Magnetic field in the DRA. (**c**) Electric field in the DRA

DRA. Recently, one type of photosensitive polymer that can be processed using the X-ray lithography has been tested for designing a millimeter-wave DRA (Rashidian and Klymyshyn 2010). This section explores the possibility of applying low-permittivity dielectrics in the millimeter-wave spectrum. Figure 31 shows the configuration of the aperture-fed DRA, which has a dielectric constant of $\varepsilon_r = 4.2$, a footprint of $a = b = 5.7$ mm, and a height of $d = 2$ mm. Microwave signal is coupled from the 50-Ω microstrip feed line to the DRA through an aperture located on the reverse side of a piece of Duroid substrate with a dielectric constant of 2.2. The line stub L_s can be adjusted to improve the

DOI 10.1007/978-981-4560-75-7_35-1

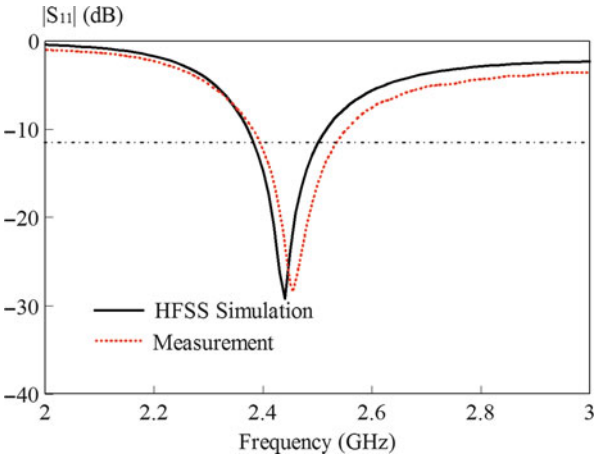


Fig. 28 Simulated and measured reflection coefficients of the quasi-isotropic DRA: ε_r=10, a = 27 mm, d = 14.5 mm, g = 27 mm, l = 9 mm, and r = 0.63 mm (From Pan et al. (2014), copyright © 2014 IEEE, with permission)

impedance matching. The antenna configuration was simulated using the commercial software HFSS simulator, and experiment was conducted on an Agilent 8722ES vector network analyzer. An Inter-Continental Microwave WK-3001-G test fixture is used for the measurement. As can be seen from the measured and simulated reflection coefficients in Fig. 32, the aperture-coupled DRA is able to provide a frequency passband from 21.25 to 29.25 GHz, covering a bandwidth of 32 %. In the same figure, for comparison, a ceramic DRA with ε_r = 10, $a = b$ = 5.7 mm, and d = 2 mm is also simulated and measured. It has a bandwidth of 9 % (covering 19–20.7 GHz) only, which is much less compared to that for the polymer. Figure 33 shows the measured and simulated radiation patterns in the *xz*- and *yz*-planes at the resonant frequency of 26.1 GHz. Broadside radiation has been observed for the TE_{111} resonance, where the co-polarized fields are at least 20 dB larger than their cross-polarized counterparts in both of the major cutplanes. The polymer DRA is found to have an antenna gain of 4.9 dBi in the boresight direction. This work shows that polymers can be used for designing the DRAs at millimeter-wave frequencies.

To work in the millimeter-wave ranges, the DRA has been conventionally incorporated with different transmission feed lines such as microstrip, image line, and waveguide. Recently, the DRA is incorporated with a new type of waveguide-like transmission line, called substrate-integrated waveguide (SIW), which works at the TE modes (Wu et al. 2012). The SIW structure was first proposed in 1998 (Uchimura et al. 1998), and it has a lower attenuation constant than microstrip since its propagating electric fields are confined in between the two backing metal plates and trenches of vias (Lai et al. 2009). But it has higher loss than the conventional metallic waveguide due to field leakage through the vias. To have compact size, as shown in Fig. 34, a half-mode substrate-integrated waveguide (HMSIW) is used to design a slot-coupled linearly polarized DRA (Lai et al. 2010), working at 60 GHz. The HMSIW is made on a Rogers 5880 substrate with a dielectric constant of ε_r = 2.2, a loss tangent of tanδ = 0.001, and a thickness of t = 0.127 mm. In this case, the DRA is machined from the TM10i dielectric material with $\varepsilon_r = 10.2 \pm 0.2$ and loss tangent of tanδ = 0.002. The DRA is placed on top of the slot using the Araldite resin which has a dielectric constant of 3.6 and a loss tangent of 0.06. The resin layer has a thickness of h_g = 15 μm. With reference to Fig. 34a, b, it can be seen that the HMSIW is fed by a tapered microstrip at one end on the bottom surface. Another end of the waveguide is shorted with a row of metallic vias. With a transverse slot etched on the top surface of the HSIW, as can be seen in Fig. 34a, a linearly polarized DRA can be easily designed. The optimized dimension of the linearly polarized DRA is w_{ms} = 0.35 mm, l_{ms} = 10 mm, w_{tr} = 0.9 mm, l_{tr} = 1.4 mm, w = 2.2, l = 8.9 mm, d = 0.4 mm, s = 0.5 mm,

DOI 10.1007/978-981-4560-75-7_35-1

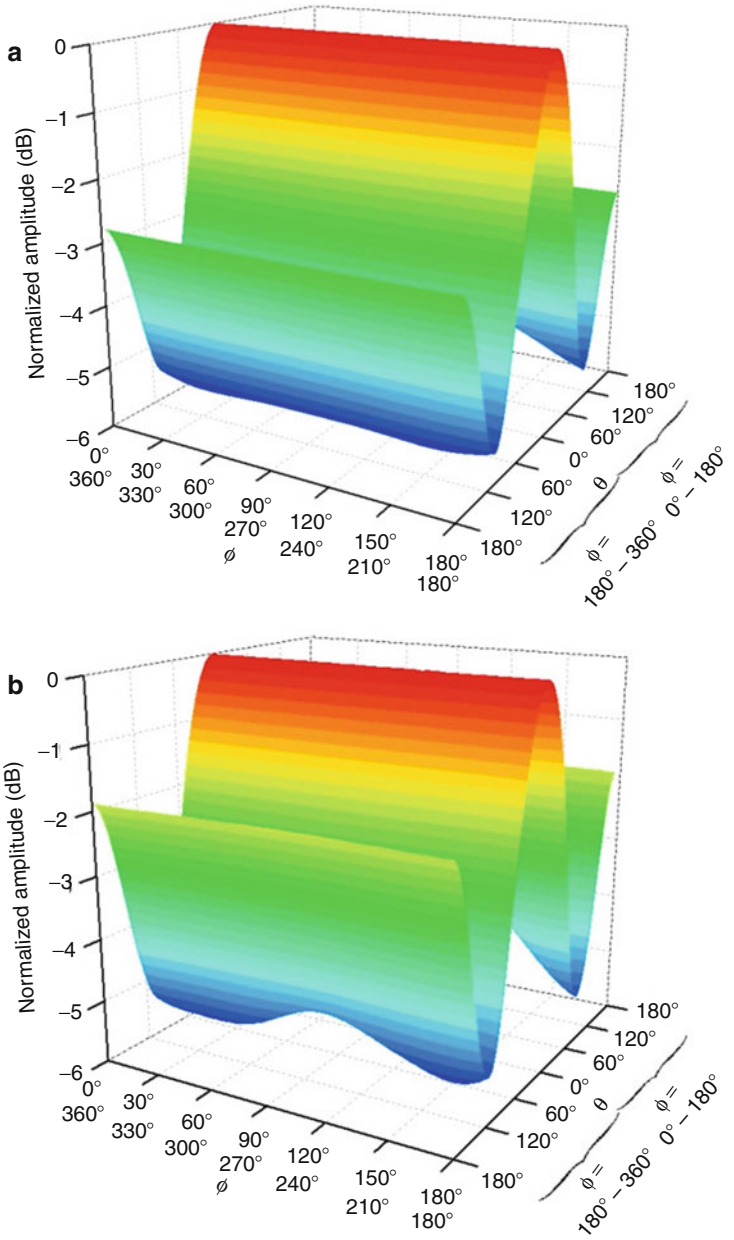


Fig. 29 Simulated (2.44 GHz) and measured (2.46 GHz) 3D radiation patterns of the quasi-isotropic DRA. (**a**) Simulation. (**b**) Measurement (From Pan et al. (2014), copyright © 2014 IEEE, with permission)

$D = 3$ mm, $h = 0.5$ mm, $x_{slot} = 1.0$ mm, $z_{slot} = 2.3$ mm, $w_{slot} = 0.2$ mm, $l_{slot} = 2.0$ mm, $h_g = 15 \pm 5$ μm, and $t = 5$mil. When a cross slot is used, as shown in Fig. 35, the same DRA can be made to generate circularly polarized waves. It has an optimized dimension of $w_{ms} = 0.35$ mm, $l_{ms} = 10$ mm, $w_{tr} = 0.$ m3m, $l_{tr} = 2.4$ mm, $w = 2.35$, $l = 8.55$ mm, $d = 0.4$ mm, $s = 0.5$ mm, $D = 2$ mm, $h = 0.7$ mm, $x_{slot} = 1.0 \pm 0.02$ mm, $z_{slot} = 2.75$ mm, $w_{slot} = 0.2$ mm, $l_{slot} = 2.0$ mm, $h_g = 15 \pm 5$ μm, and $t = 5$mil. Figure 36 shows the fabricated prototypes of the DRA samples.

The simulated and measured reflection coefficients of the linearly polarized DRA are shown in Fig. 37. Three resonances ($HEM_{11\delta}$ and slot modes) are clearly seen in the figure. It has a measured bandwidth of 24 %, which is slightly narrower than its simulated counterpart of 25 %. Figure 38 shows the simulated and measured antenna gains of the linearly polarized DRA, which are greater than 5.5 dBi in the frequency

DOI 10.1007/978-981-4560-75-7_35-1

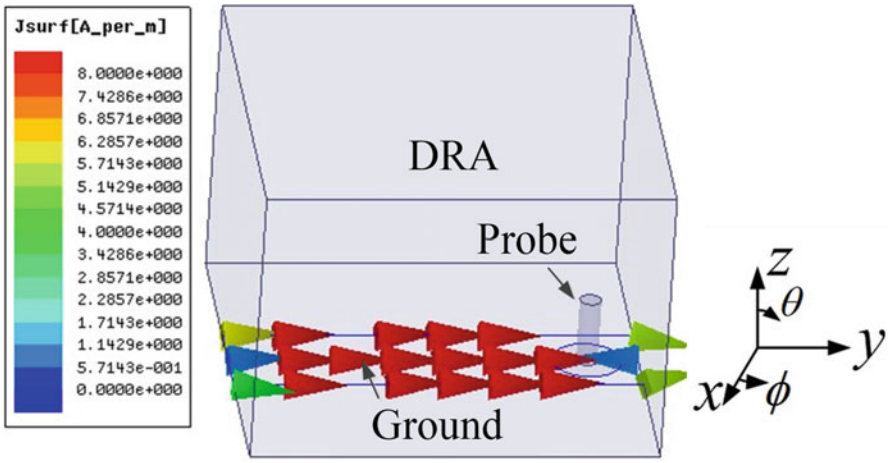


Fig. 30 Simulated surface current on the ground plane (24 × 6 mm^2) of the DRA

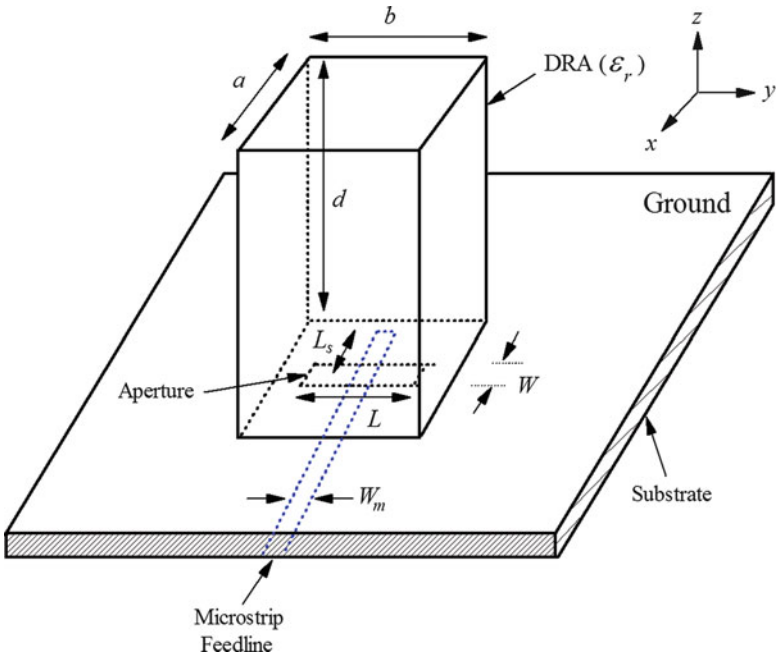


Fig. 31 Configuration of the aperture-coupled DRA

range of 49–62 GHz. Broadside radiation pattern is observed at 55 GHz, as can be seen in Fig. 39, which is expected for the HEM and slot modes. An antenna gain of ∼5 dBi is obtained in the boresight direction ($\theta = 0°$). Figure 40 is the simulated and measured reflection coefficients of the circularly polarized DRA. The impedance matching has achieved its optimum point around ∼ 60 GHz, achieving a wide 10-dB impedance bandwidth of 2.7 GHz (4.5 %). The axial ratio in Fig. 41 shows that the CP antenna has a 3-dB AX bandwidth of 4.0 %, covering 58.6–61.0 GHz. Figure 42 shows the radiation patterns in the *xy*- and *yz*-planes at 59.4 GHz. The 3-dB beamwidth is 99° in the *xy*-plane and 81° in the *yz*-plane.

Most studies on the millimeter-wave DRAs mainly focus on the fundamental modes (Keller et al. 1998; Lai et al. 2008; Svedin et al. 2007; Wahab et al. 2009). For example, a rectangular DRA with a dielectric constant of $\varepsilon_r = 9.8$ has a dimension of 1.91 mm, 0.635 mm, and 1.91 mm at 40 GHz. It is not easy to fabricate the DRA precisely when the antenna size becomes very small, making such antenna structure

DOI 10.1007/978-981-4560-75-7_35-1

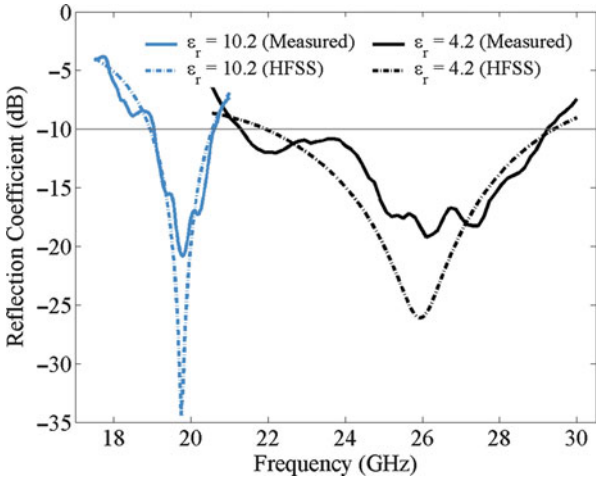


Fig. 32 Reflection coefficient for the polymer DRA with $\varepsilon_r = 4.2$, with dimensions of $a = b = 5.7$ mm, and $d = 2$ mm. Also shown is the reflection coefficient for a ceramic DRA with $\varepsilon_r = 10$, which has the same dimension (From Rashidian and Klymyshyn (2010))

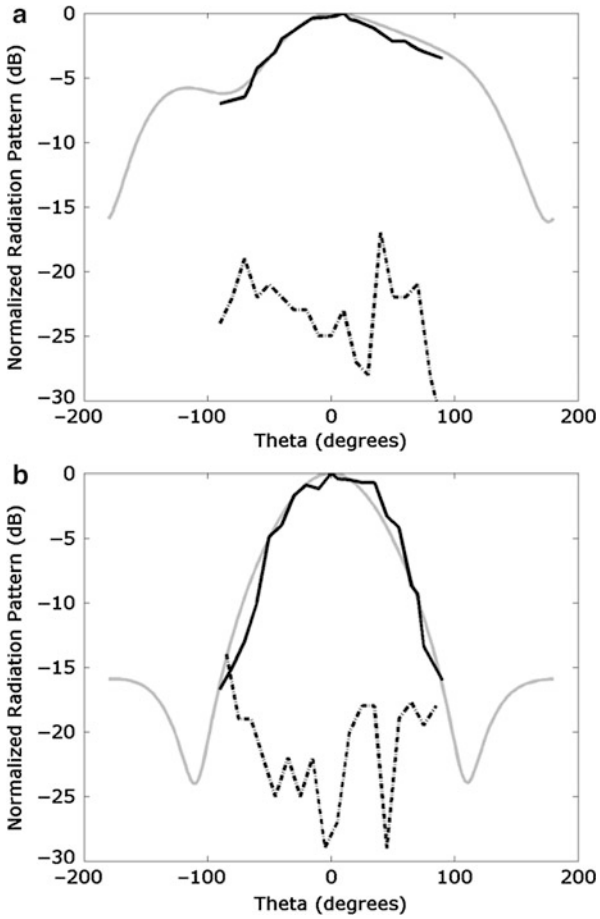


Fig. 33 Measured and simulated radiation patterns of the polymer DRA in the (**a**) *yz*-plane and (**b**) *xz*-plane at the resonant frequency of 26.1 GHz (From Rashidian and Klymyshyn (2010))

DOI 10.1007/978-981-4560-75-7_35-1

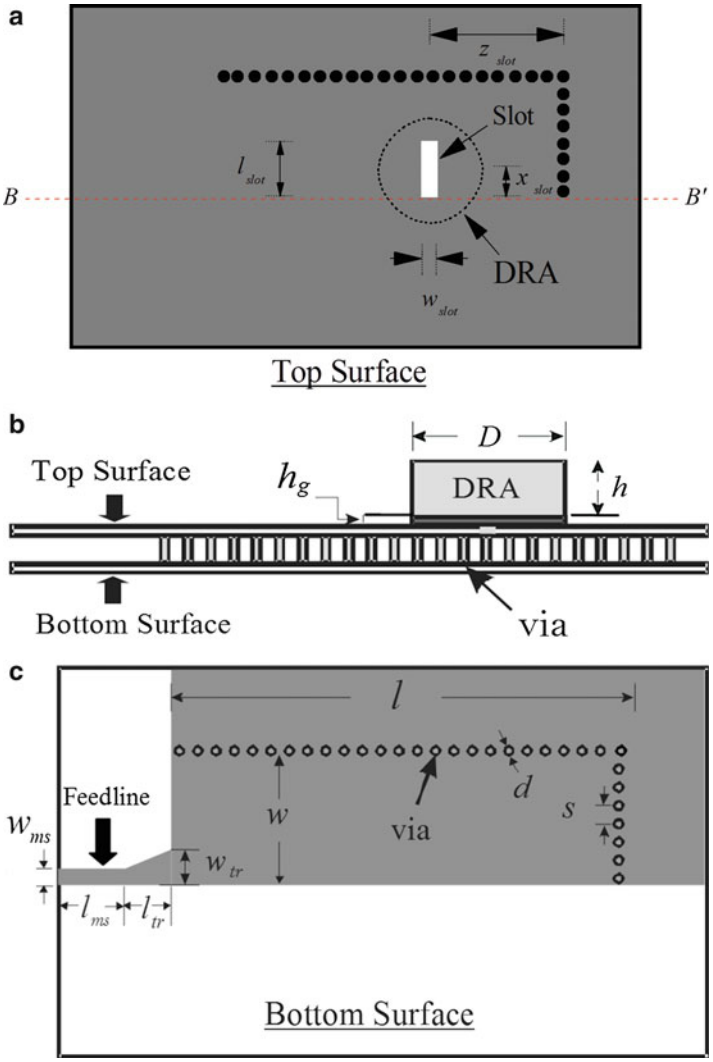


Fig. 34 Configuration of the linearly polarized HMSIW-fed DRA. (**a**) *Top* surface of the linearly polarized DRA. (**b**) *Side view* showing the cross section of the HMSIW and the DRA. (**c**) *Bottom* surface showing the microstrip-fed HMSIW cavity. Metal layer is shaded in *gray* color

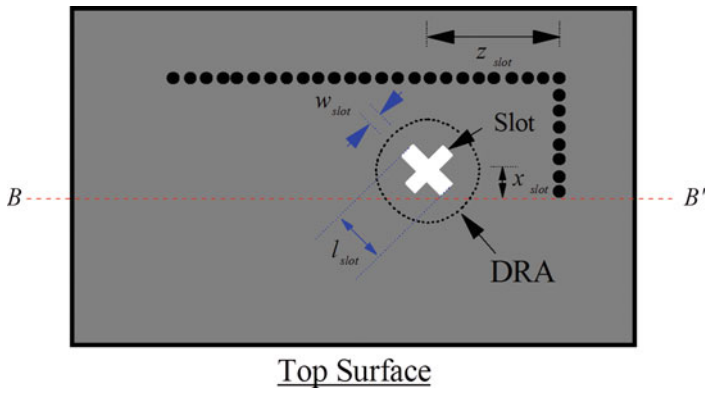


Fig. 35 *Top* surface of the circularly polarized DRA

DOI 10.1007/978-981-4560-75-7_35-1

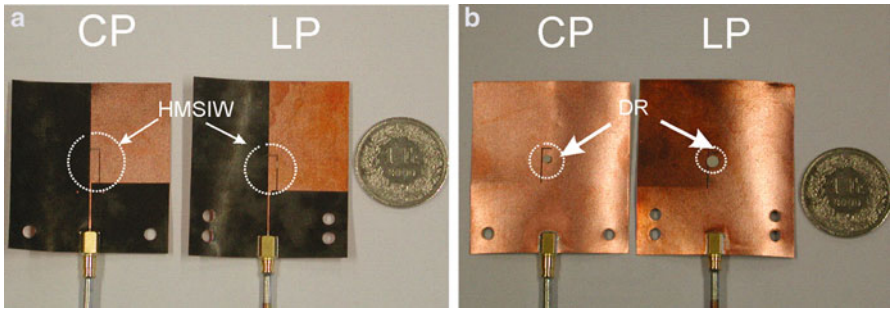


Fig. 36 Fabricated prototypes of the linearly and circularly polarized DRAs. (**a**) *Bottom* surface. (**b**) *Top* surface

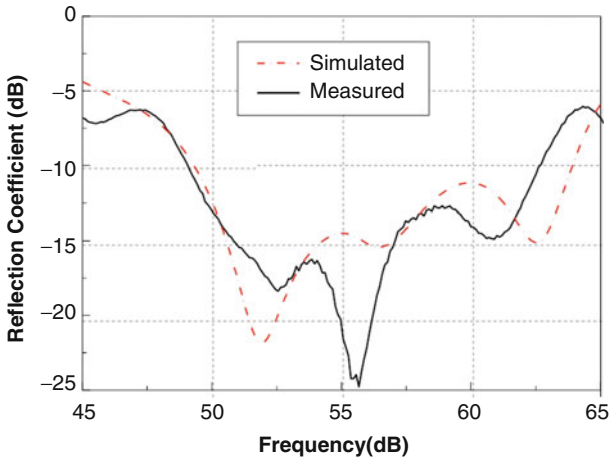


Fig. 37 Simulated and measured reflection coefficients of the linearly polarized HMSIW-fed DRA

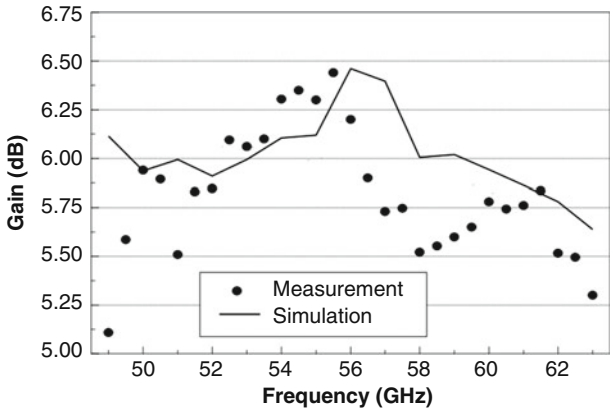


Fig. 38 Simulated and measured antenna gains of the linearly polarized HMSIW-fed DRA

difficult to be realized at millimeter-wave frequencies. The use of the higher-order modes of the DRA has been proven to be useful for relaxing the precision problem in the fabrication of the DRA samples at high frequencies (Petosa et al. 2009, 2011; Pan et al. 2011; Hou et al. 2014).

Three aperture-coupled rectangular DRAs (referring to the antenna structure in Fig. 31: DRA1, $a = b = 7$ mm and $d = 10$ mm; DRA2, $a = b = 6$ mm and $d = 15$ mm; DRA3, $a = b = 5$ mm and

DOI 10.1007/978-981-4560-75-7_35-1

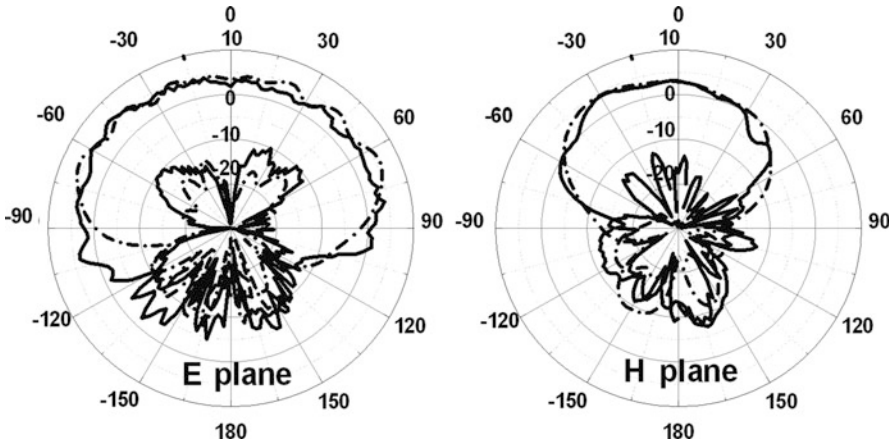


Fig. 39 Measured (*solid line*) and simulated (*dashed line*) co- and cross-polarized radiation patterns of the linearly polarized HMSIW-fed DRA at 55 GHz (From Lai et al. (2010), copyright © 2010 IEEE, with permission)

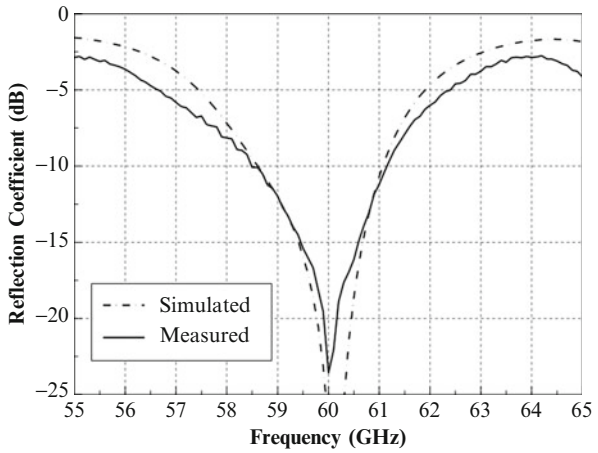


Fig. 40 Reflection coefficient of the circularly polarized HMSIW-fed DRA

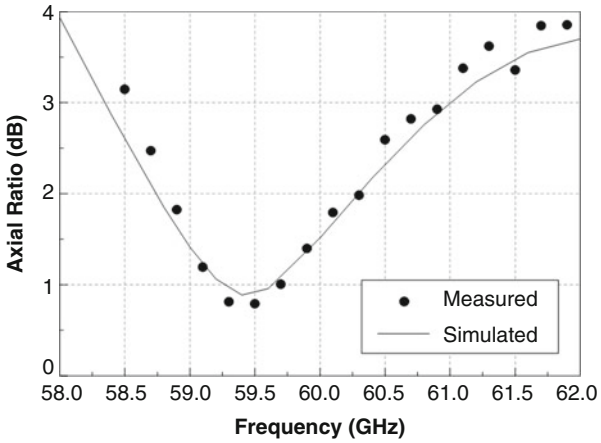


Fig. 41 Axial ratio of the circularly polarized HMSIW-fed DRA in the boresight direction (From Lai et al. (2010), copyright © 2010 IEEE, with permission)

DOI 10.1007/978-981-4560-75-7_35-1

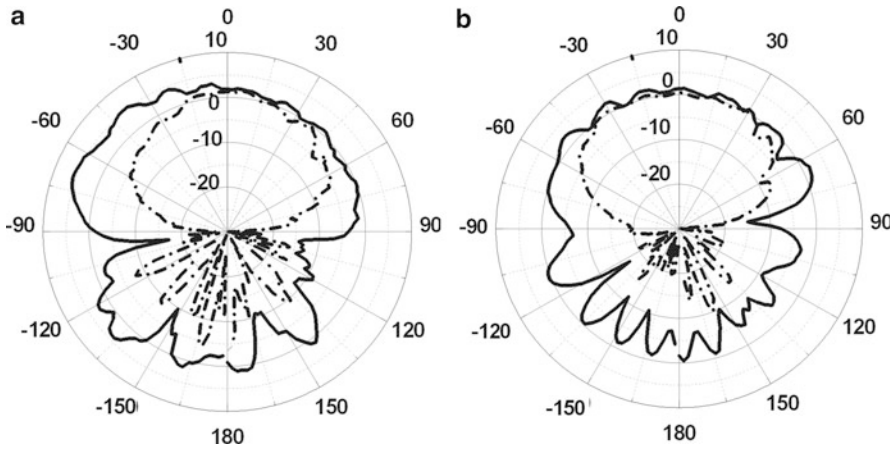


Fig. 42 Measured electric fields in the major and minor axes of the circularly polarized HMSIW-fed DRA at 59.4 GHz. (**a**) *yz*-plane and (**b**) *xy*-plane (From Lai et al. (2010), copyright © 2010 IEEE, with permission)

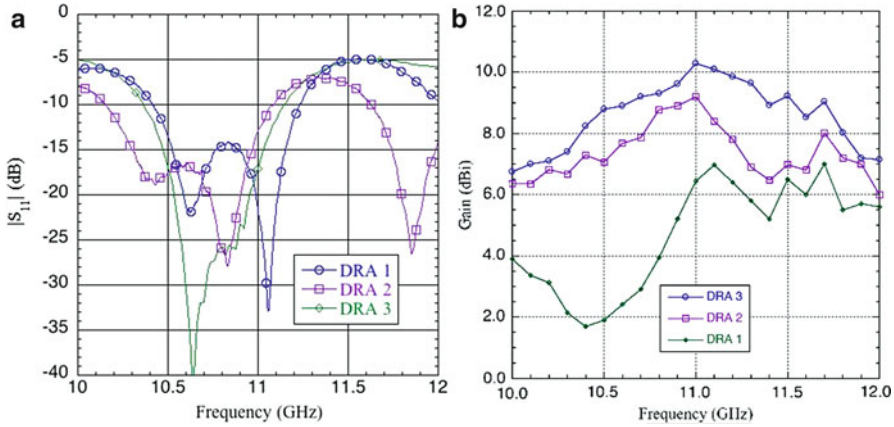


Fig. 43 (**a**) Reflection coefficients and (**b**) antenna gains of DRA1, DRA2, and DRA3 (From Petosa and Thirakoune (2011), copyright @ 2011 IEEE, with permission)

$d = 30$ mm) are studied in (Petosa et al. 2009; Petosa and Thirakoune 2011), where the higher-order $TE_{\delta 11}$, $TE_{\delta 13}$, and $TE_{\delta 15}$ modes are excited in the rectangular DRA1, DRA2, and DRA3, respectively. The measured reflection coefficients and antenna gains are depicted in Fig. 43. At 11 GHz, the respective antenna gains are found to be 5.5, 8.2, and 10.2 dBi. This work has shown that the rectangular DRA is able to provide higher antenna gain with the excitation of the higher-order modes. Figure 44a shows the magnetic field distribution of the $TE_{\delta 15}$ mode, which can be represented by a couple of equivalent magnetic dipoles, as shown in Fig. 44b. In this case, image theory is applied to remove the ground plane. Due to the existence of the ground plane, only odd modes are excitable. Even modes are not possible as they are short-circuited by the ground. The separation between the magnetic dipoles is determined by the dimension of the DRA. It can affect the antenna gain, beamwidth, and sidelobe directly. All of the DRA1, DRA2, and DRA3 are radiating in the broadside. Figure 45 shows the measured and simulated radiation pattern of the $TE_{\delta 13}$ mode, whose beamwidth is narrower than that for the $TE_{\delta 11}$ mode but broader than for the $TE_{\delta 15}$.

The properties of the higher-order TE^{y}_{pqr} modes of the DRA are further explored in (Pan et al. 2011) at the resonant frequency of 24 GHz. Again, the antenna structure shown in Fig. 31 is used for the design.

DOI 10.1007/978-981-4560-75-7_35-1

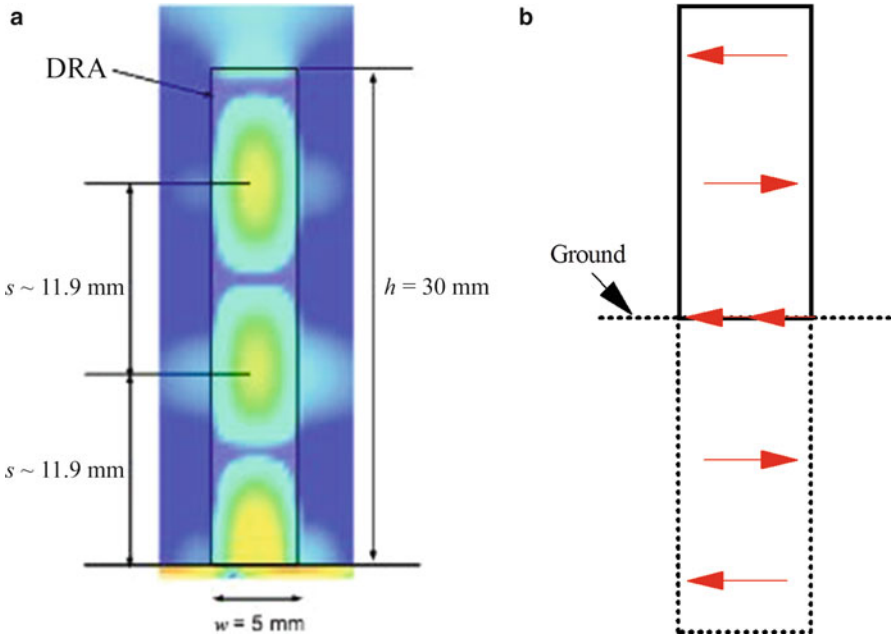


Fig. 44 (**a**) Magnetic field distribution for DRA3 working in the $TE_{\delta 15}$ mode. (**b**) Radiation model for the $TE_{\delta 15}$ mode (From Petosa and Thirakoune (2011), copyright @ 2011 IEEE, with permission)

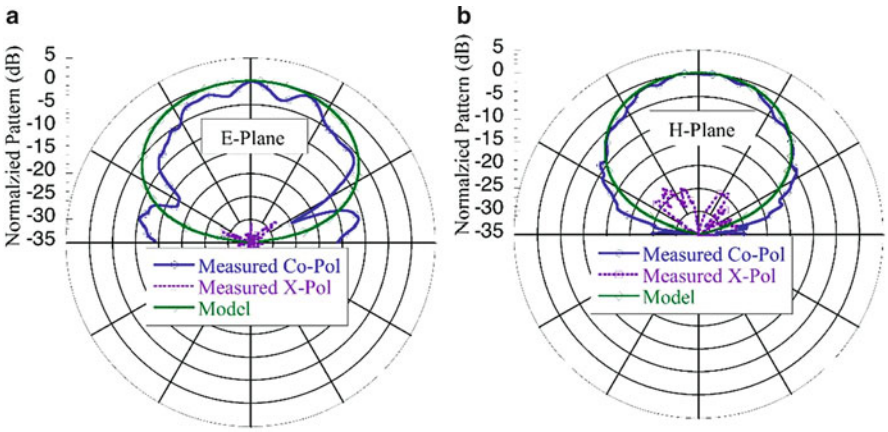


Fig. 45 Radiation patterns for the $TE_{\delta 13}$ mode. (**a**) *E*-plane. (**b**) *H*-plane

A *y*-directed slot is etched underneath the DRA for exciting it. For an aperture-fed rectangular DRA that is placed on a ground plane, it has been proven that the indices *p*, *q*, *r* have to be odd numbers. Figure 46 shows the reflection coefficient of an aperture-fed rectangular DRA with the design parameters of $a = 4.8$ mm, $b = 6.4$ mm, $d = 3$ mm, $\varepsilon_r = 10$, $W = 0.5$ mm, $L = 2.8$ mm, $W_m = 0.78$ mm, and $L_s = 2.6$ mm. Obviously, only odd modes are excitable. Resonant modes with even indices (TE^y_{211}, TE^y_{121}, TE^y_{112}) are not seen in the figure. Microstrip feed line has been used because it can be designed and fabricated easily. Coplanar waveguide and dielectric image guide can be used if the operating frequency goes higher.

The characteristics of the higher-order TE^y_{115} and TE^y_{119} modes are now studied (Pan et al. 2011). The optimized dimension for the DRA with the TE^y_{115} mode is $a = b = 4.0$ mm, $d = 6.1$ mm, $\varepsilon_r = 10$, $W = 0.5$ mm, $L = 2.2$ mm, $W_m = 0.78$ mm, and $L_s = 1.4$ mm, while the optimized dimension for the DRA with the TE^y_{119} mode is $a = b = 4.2$ mm, $d = 10.7$ mm, $\varepsilon_r = 10$, $W = 0.5$ mm, $L = 1.8$ mm,

DOI 10.1007/978-981-4560-75-7_35-1

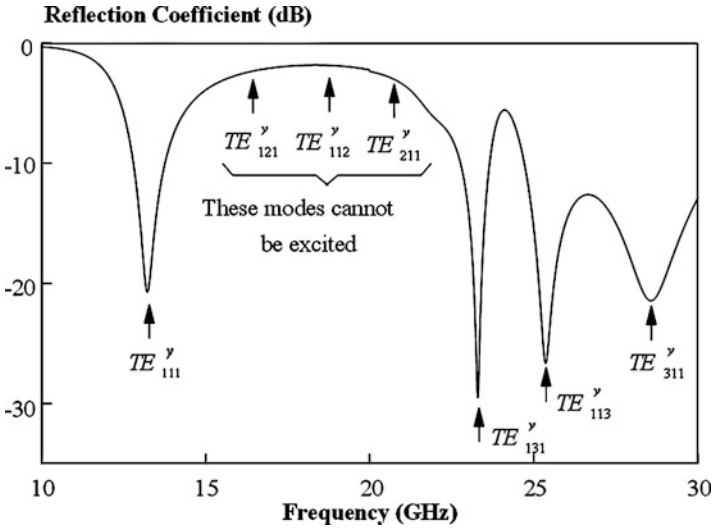


Fig. 46 Simulated reflection coefficient of the aperture-fed rectangular DRA as a function of frequency

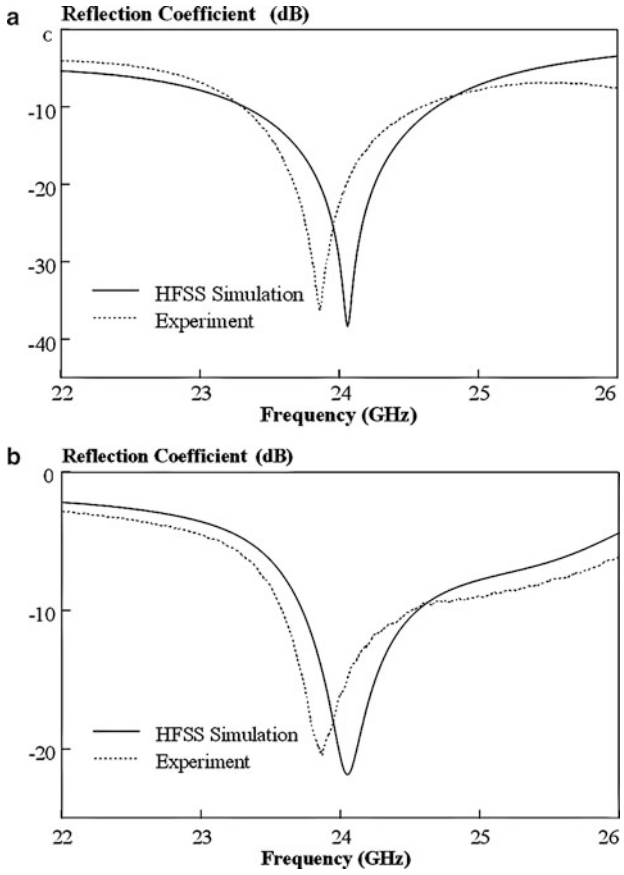


Fig. 47 Simulated and measured reflection coefficients of the DRAs excited in the (**a**) TE^{y}_{115} mode and (**b**) TE^{y}_{119} mode

$W_m = 0.78$ mm, and $L_s = 0.9$ mm. It can be observed from the reflection coefficients in Fig. 47 that the resonant frequencies are measured to be 23.86 and 23.87 GHz, respectively, for the TE^{y}_{115} and TE^{y}_{119} modes. The corresponding 10-dB bandwidths are 5.39 % (covering 23.30–24.59 GHz) and 3.87 % (23.58–24.51 GHz). Figure 48 shows the simulated and measured antenna gains of the DRAs,

DOI 10.1007/978-981-4560-75-7_35-1

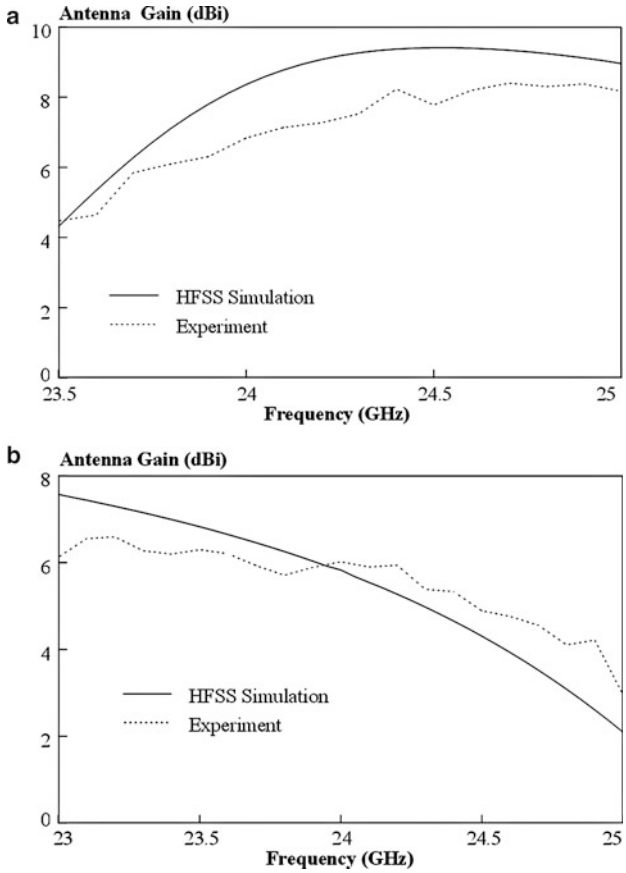


Fig. 48 Simulated and measured antenna gains of the DRAs excited in the (**a**) TE^{y}_{115} mode and (**b**) TE^{y}_{119} mode

which are ~5.8 dBi (TE^{y}_{115} mode) and ~6.3 dBi (TE^{y}_{119} mode) at around 24 GHz. Higher gain is observed in the latter. It should be mentioned that the beamwidth, sidelobe, and antenna gain of a higher-order mode can change with the DRA dimension. It can be seen from the measured and simulated radiation patterns in Fig. 49 that the higher-order modes are radiating into the boresight direction, as expected. For both of the E (xz-plane) and H (yz-plane) planes, the co-polarized fields are stronger than their cross-polarized counterparts by at least 20 dB.

Miniature DRAs can be combined with various integrated circuits (IC) for achieving compact footprint. Integrating an antenna directly on the RF circuit eliminates the use of interconnects, which may improve signal loss at high frequencies, especially in the millimeter-wave ranges. For a silicon-integrated antenna, the substrate conductivity is the most crucial parameter that decides its radiation efficiency. In (Hou et al. 2014), an on-chip DRA is integrated with a half-mode SIW cavity for the excitation of the higher-order $TE^{x}_{\delta 13}$ and $TE^{x}_{\delta 15}$ modes at the frequency of 135 GHz. Figure 50 shows the configuration of the on-chip DRA, and Fig. 51a shows the feeding half-mode SIW cavity. For comparison, the configuration of a standard full-mode SIW cavity is also depicted in Fig. 51b. It can be seen from the figures that the half-mode SIW cavity has the same resonance mode as the full-mode one. The entire structure is fabricated using the standard 0.18 μm CMOS technology. A DRA with a dimension of $a \times b \times h$ is directly placed on an aperture that is made on top of the feeding structure. Other design parameters are given here: $w = 10$ μm, $t = 300$ μm, $d = 100$ μm, $d_1 = 210$ μm, $d_2 = 160$ μm, $s = 30$ μm, $l = 750$ μm, $l_1 = 800$ μm, $th = 10$ μm, and $h_1 = 6.6$ μm. For the DRA that is working at the $TE^{x}_{\delta 13}$

DOI 10.1007/978-981-4560-75-7_35-1

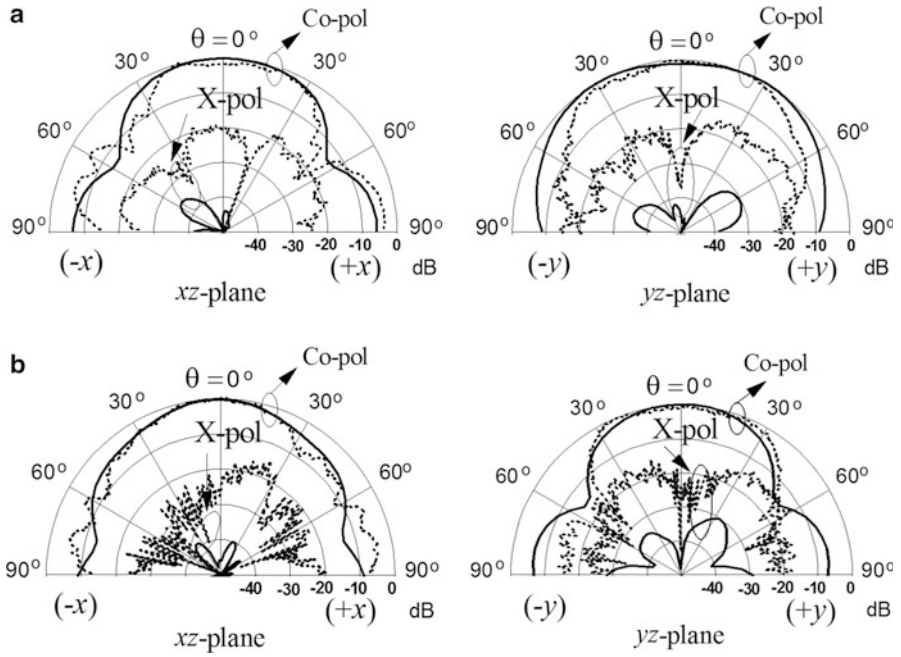


Fig. 49 Simulated and measured radiation patterns for the DRAs excited in the (**a**) TE^{y}_{115} mode and (**b**) TE^{y}_{119} at 24 GHz (From Pan et al. (2011), copyright @ 2011 IEEE, with permission)

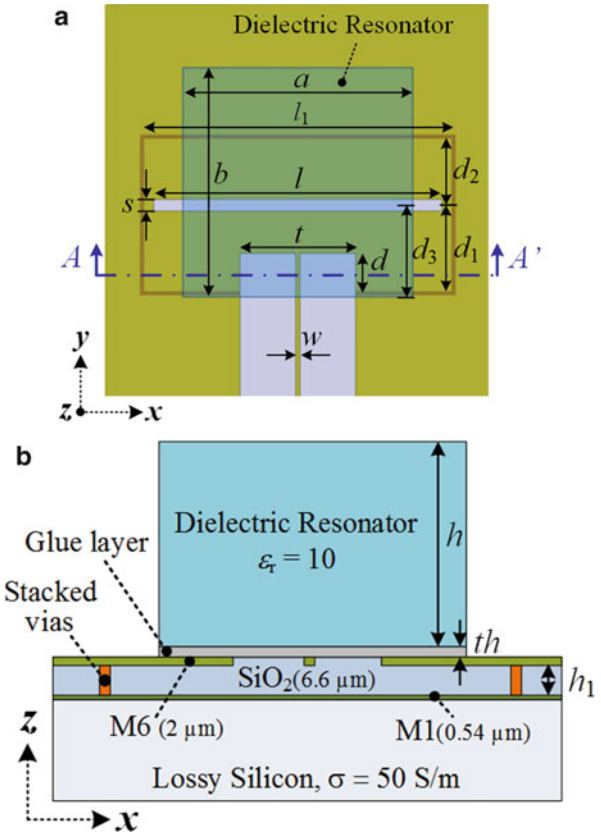


Fig. 50 The on-chip dielectric resonator antenna: (**a**) *top view*. (**b**) *Side view* (A− A' plane) (From Hou et al. (2014), copyright © 2014 IEEE, with permission)

DOI 10.1007/978-981-4560-75-7_35-1

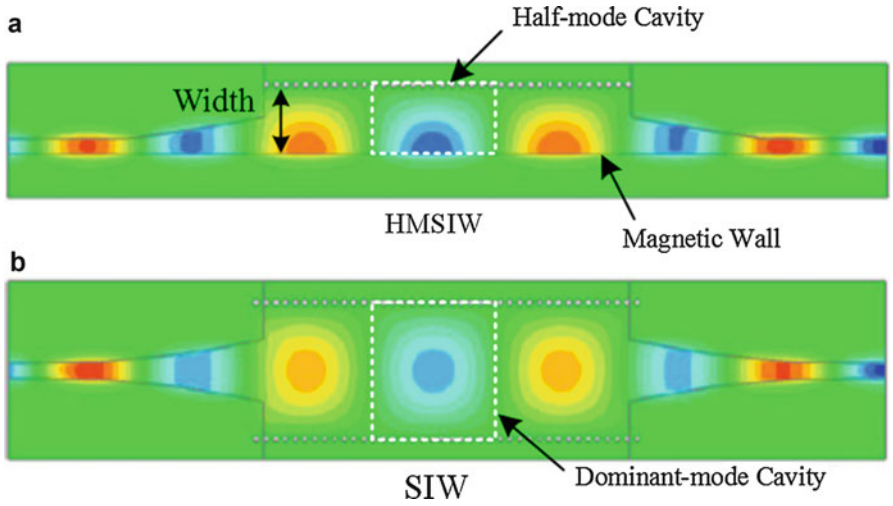


Fig. 51 (**a**) The half-mode SIW cavity. (**b**) The full-mode SIW cavity (From Hou et al. (2014), copyright © 2014 IEEE, with permission)

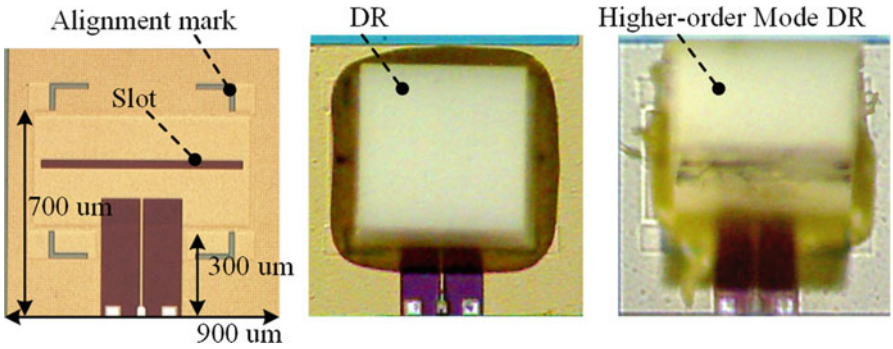


Fig. 52 Photographs of the on-chip DRA and its feeding half-mode SIW cavity (From Hou et al. (2014), copyright © 2014 IEEE, with permission)

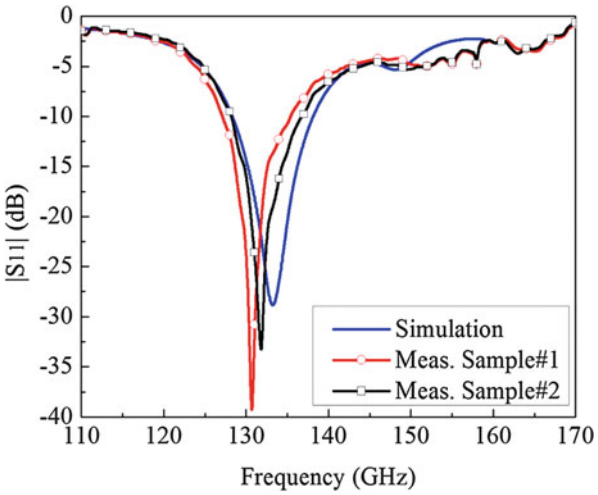


Fig. 53 The simulated and measured reflection coefficients of the on-chip DRA working at the $TE^{x}_{\delta 15}$ mode (From Hou et al. (2014), copyright © 2014 IEEE, with permission)

DOI 10.1007/978-981-4560-75-7_35-1

mode, it has an optimized dimension of $a = 650$ μm, $b = 380$ μm, $h = 1300$ μm, and $d_3 = 140$ μm. The DRA that operates at the $TE^x_{\delta 15}$ mode has a dimension of of $a = 650$ μm, $b = 380$ μm, $h = 2200$ μm, and $d_3 = 140$ μm. The photographs of the DRA and its feeding structure are illustrated in Fig. 52. Two samples have been fabricated. The simulated and measured reflection coefficients for the $TE^x_{\delta 15}$ case are shown in Fig. 53. It has an antenna bandwidth of 8 %. With reference to (Hou et al. 2014), the simulated radiation efficiency and antenna gain are 42 % and 7.6 dBi, respectively. It is clear that the main challenge of the silicon antenna is always its low radiation efficiency.

Conclusion

In the first part of this chapter, the recent developments of the miniature DRA are discussed. Miniature DRAs that are designed by integrating the DRA with different power dividers are studied. It is found that the DRA can be integrated with a hybrid coupler for making a wideband CP antenna. When integrated with the rat-race coupler, a differential DRA can be easily obtained. In the second part, ground miniaturization technique has been explored for reducing the footprint of the DRA. Case studies are presented to show the design processes of the omnidirectional CP DRA and the quasi-isotropic DRA, with minimum ground area. Finally, the application of the DRA is explored for millimeter-wave ranges using different feeding methods. Higher-order modes of the DRA are found to be useful for relaxing the fabrication tolerances at millimeter-wave frequencies.

References

Abe H, Takayama Y, Higashisaka A, Takamizawa H (1978) A highly stabilized low-noise GaAs FET integrated oscillator with a dielectric resonator in the C band. IEEE Trans Microw Theory Tech 26(3):156–162

Al-Salameh MS, Antar YMM, Seguin G (2002) Coplanar waveguide fed slot-coupled rectangular dielectric resonator antenna. IEEE Trans Antennas Propag 50(10):1415–1419

Brakora KF, Halloran J, Sarabandi K (2007) Design of 3-D monolithic MMW antennas using ceramic stereolithography. IEEE Trans Antennas Propag 44(3):790–797

Buerkle A, Brakora KF, Sarabandi K (2006) Fabrication of a DRA array using ceramic stereolithography. IEEE Antennas Wirel Propag Lett 5:479–482

Chair R, Kishk AA (2006) Experimental investigation for wideband perforated dielectric resonator antenna. Electron Lett 42(3):15–16

Chair R, Kishk AA, Lee KF, Smith CE (2004) Wideband flipped staired pyramid dielectric resonator antennas. Electron Lett 40(10):581–582

Chen ZN, Qing XM, See TSP, Toh WK (2012) Antennas for WiFi connectivity. Proc IEEE 100(7):2322–2329

Cohn SB (1968) Microwave bandpass filters containing high-Q dielectric resonators. IEEE Trans Microw Theory Tech 16(4):218–227

Esselle KP, Bird TS (2005) A hybrid-resonator antenna: experimental results. IEEE Trans Antennas Propag 53(2):870–871

Fang XS, Leung KW, Lim EH (2010) Compact differential rectangular dielectric resonator antenna. IEEE Antennas Wirel Propag Lett 9:662–665

Fiedziuszko SJ (1986) Microwave dielectric resonators. Microw J 29:189–200

Fumeaux C, Baumann D, Leuchtmann P, Vahldieck R (2004) A generalized local time-step scheme for efficient FVTD simulations in strongly inhomogeneous meshes. IEEE Trans Microw Theory Tech 52(3):1067–1076

Gastine M, Courtois L, Dormann JL (1967) Electromagnetic resonances of free dielectric spheres. IEEE Trans Microw Theory Tech 15(12):694–700

Guha D, Antar YMM, Ittipiboon A, Petosa A, Lee D (2006) Improved design guidelines for the ultra wideband monopole-dielectric resonator antenna. IEEE Antennas Wirel Propag Lett 5(1):373–376

Hamsakutty V, Kumar A, Yohannan J, Mathew KT (2007) Hexagonal dielectric resonator antenna for 2.4GHz WLAN applications. Microw Opt Technol Lett 49:162–164

Haneishi M, Takazawa H (1985) Broadband circularly polarized planar array composed of a pair of dielectric resonator antennas. Electron Lett 21(10):437–438

Hou D, Hong W, Goh WL, Chen J, Xiong YZ, Hu S, Madihian M (2014) D-band on-chip higher-order-mode dielectric-resonator antennas fed by half-mode cavity in CMOS technology. IEEE Antennas Propag Mag 56(3):80–89

Huang CY, Wu JY, Wong KL (1999) Cross-slot-coupled microstrip antenna and dielectric resonator antenna for circular polarization. IEEE Trans Antennas Propag 47(4):605–609

Huitema L, Monediere T (2012) Dielectric materials for compact dielectric resonator antenna applications. In: Dielectric Material. Intech (ISBN 978-953-51-0764-4)

Junker GP, Kishk AA, Glisson AW (1994) Input impedance of dielectric resonator antennas excited by coaxial probe. IEEE Trans Antennas Propag 42(7):960–966

Junker GP, Kishk AA, Glisson AW, Kajfez D (1995) Effect of fabrication imperfections for ground-plane-backed dielectric-resonator antennas. IEEE Antennas Propagation Magazine 37(1):40–47

Junker GP, Kishk AA, Glisson AW (1996) Input impedance of aperture-coupled dielectric resonator antennas. IEEE Trans Antennas Propag 44(5):600–607

Kajfez D, Guillon P (1998) Dielectric resonators. Noble, Atlanta

Kawakami H, Sato G, Wakabayashi R (1997) Research on circularly polarized conical-beam antennas. IEEE Antennas Propag Mag 39(6):27–39

Kay AF (1966) Millimeter wave antennas. Proc IEEE 54(4):641–647

Ke SY, Cheng YT (2001) Integration equation analysis on resonant frequencies and quality factors of rectangular dielectric resonators. IEEE Trans Microw Theory Tech 49(3):571–574

Keller MG, Oliver MB, Roscoe DJ, Mongia RK, Antar YMM, Ittipiboon A (1998) EHF dielectric resonator antenna array. Microw Opt Technol Lett 17(6):345–349

Khalily M, Kamarudin MR, Mokayef M, Jamaluddin MH (2014) Omni-directional circularly polarized dielectric resonator antenna for 5.2-GHz WLAN applications. IEEE Antennas Wirel Propag Lett 13:443–446

Kirschbaum HS, Chen L (1957) A method of producing broadband circular polarization employing an anisotropic dielectric. IRE Trans Microw Theory Tech 5(3):199–203

Kishk AA (2003) Wide-band truncated tetrahedron dielectric resonator antenna excited by a coaxial probe. IEEE Trans Antennas Propag 51(10):2913–2917

Kishk A (2007) Chapter 17, Dielectric resonator antenna. In: Antenna engineering handbook. McGraw-Hill Education, New York

Kishk AA, Ahn B, Kajfez D (1989) Broadband stacked dielectric resonator antennas. Electron Lett 25(18):1232–1233

Koulouridis S, Kiziltas G, Zhou Y, Hansford DJ, Volakis JL (2006) Polymer-ceramic composites for microwave applications: fabrication and performance assessment. IEEE Trans Microw Theory Tech 54(12):4202–4208

Kraus JD, Marhefka RJ (2003) Antennas for all applications, 3rd edn. McGraw-Hill, New York

Lai QH, Almpanis G, Fumeaux C, Benedickter H, Vahldieck R (2008) Comparison of the radiation efficiency for the dielectric resonator antenna and the microstrip antenna at Ka band. IEEE Trans Antennas Propag 56(11):3589–3592

Lai QH, Fumeaux C, Hong W, Vahldieck R (2009) Characterization of the propagation properties of the half-mode substrate integrated waveguide. IEEE Trans Microw Theory Tech 57(8):1996–2004

Lai QH, Fumeaux C, Hong W, Vahldieck R (2010) 60 GHz aperture-coupled dielectric resonator antennas fed by a half-mode substrate integrated waveguide. IEEE Trans Antennas Propag 58(6):1856–1864

Leung KW (2000) Conformal strip excitation of dielectric resonator antenna. IEEE Trans Antennas Propag 48(6):961–967

Leung KW, Ng HK (2003) Theory and experiment of circularly polarized dielectric resonator antenna with a parasitic patch. IEEE Trans Antennas Propag 51(3):405–412

Leung KW, Luk KM, Lai KYA, Lin D (1993) Theory and experiment of a coaxial probe fed hemispherical dielectric resonator antenna. IEEE Trans Antennas Propag 41(10):1390–1398

Leung KW, Luk KM, Lai KYA, Lin D (1995) Theory and experiment of an aperture-coupled hemispherical dielectric resonator antenna. IEEE Trans Antennas Propag 43(11):1192–1198

Leung KW, Wong WC, Luk KM, Yung EKN (2000) Circular-polarised dielectric resonator antenna excited by dual conformal strips. Electron Lett 36(6):84–486

Leung KW, Pan YM, Fang XS, Lim EH, Luk KM, Chan HP (2013) Dual-function radiating glass for antennas and light covers – part I: omnidirectional glass dielectric resonator antennas. IEEE Trans Antennas Propag 61(2):578–586

Li B, Leung KW (2005) Strip-fed rectangular dielectric resonator antennas with/without a parasitic patch. IEEE Trans Antennas Propag 53(7):2200–2207

Li WW, Leung KW (2013) Omnidirectional circularly polarized dielectric resonator antenna with top-loaded alford loop for pattern diversity design. IEEE Trans Antennas Propag 61(2):563–570

Li MJ, Luk KM (2014) A low-profile unidirectional printed antenna for millimeter-wave applications. IEEE Trans Antennas Propag 62(3):1232–1237

Lim EH, Leung KW (2012) Compact multi-functional antennas for wireless systems. Wiley, Hoboken

Lim EH, Leung KW, Fang XS (2011) The compact circularly polarized hollow rectangular dielectric resonator antenna with an underlaid quadrature coupler. IEEE Trans Antennas Propag 59(1):288–293

Liu Z, Chew WC, Michielssen E (2002) Numerical modeling of dielectric-resonator antennas in a complex environment using the method of moments. IEEE Trans Antennas Propag 50(1):79–82

Lo HY, Leung KW, Luk KM, Yung EKN (1999) Low profile equilateral- triangular dielectric resonator antenna of very high permittivity. Electron Lett 35(25):2164–2166

Long SA, McAllister MW, Shen LC (1983) The resonant cylindrical dielectric cavity antenna. IEEE Trans Antennas Propag 31(3):156–162

Luk KM, Leung KW (2003) Dielectric resonator antennas. Research Studies Press, London

Madou MJ (2011) Fundamentals of microfabrication and nanotechnology, 3rd edn. CRC Press, Boca Raton

McAllister MW, Long SA (1984) Resonant hemispherical dielectric antenna. Electron Lett 20(16):657–659

Menzel W, Moebius A (2012) Antenna concepts for millimeter-wave automotive radar sensors. Proc IEEE 100(7):2372–2379

Mongia RK (1992) Theoretical and experimental resonant frequencies of rectangular dielectric resonators. IEE Proc H Microw Antennas Propag 1:98–104

Mongia RK, Bhartia P (1994) Dielectric resonator antenna – a review and general design relations to resonant frequency and bandwidth. Int J Microw Millim Wave Comput Aided Eng 4:230–247

Mongia RK, Ittipiboon A (1997) Theoretical and experimental investigations on rectangular dielectric resonator antennas. IEEE Trans Antennas Propag 45(9):1348–1355
Mongia RK, Ittipiboon A, Cuhaci M, Roscoe D (1994) Circularly polarized dielectric resonator antenna. Electron Lett 30(17):1361–1362
Nakano H, Fujimori K, Yamauchi J (2000) A low-profile conical beam loop antenna with an electromagnetically coupled feed system. IEEE Trans Antennas Propag 48(12):1864–1866
Ng HK, Leung KW (2005) Frequency tuning of the dielectric resonator antenna using a loading cap. IEEE Trans Antennas Propag 53(3):1229–1232
Ng HK, Leung KW (2006) Frequency tuning of the linearly and circularly polarized dielectric resonator antennas using multiple parasitic strips. IEEE Trans Antennas Propag 54(1):225–230
Oliver MB, Antar YMM, Mongia RK, Ittipiboon A (1995) Circularly polarized rectangular dielectric resonator antenna. Electron Lett 31(3):418–419
Ong SH, Kishk AA, Glisson AW (2004) Rod-ring dielectric resonator antenna. Int J RF Microw Comput Aided Eng 14(5):441–446
Pan YM, Leung KW (2012) Wideband omnidirectional circularly polarized dielectric resonator antenna with parasitic strips. IEEE Trans Antennas Propag 60(6):2992–2997
Pan YM, Leung KW, Luk KM (2011) Design of the millimeter-wave rectangular dielectric resonator antenna using a higher-order mode. IEEE Trans Antennas Propag 59(8):2780–2788
Pan YM, Leung KW, Lu K (2012) Omni-directional linearly and circularly polarized rectangular dielectric resonator antennas. IEEE Trans Antennas Propag 60(2):751–759
Pan YM, Leung KW, Lu K (2014) Compact quasi-isotropic dielectric resonator antenna with small ground plane. IEEE Trans Antennas Propag 62(2):577–585
Park BC, Lee JH (2011) Omnidirectional circularly polarized antenna utilizing zeroth-order resonance of epsilon negative transmission line. IEEE Trans Antennas Propag 59(7):2717–2721
Petosa A (2007) Dielectric resonator antenna handbook. Artech House, Norwood
Petosa A, Thirakoune S (2011) Rectangular dielectric resonator antennas with enhanced gain. IEEE Trans Antennas Propag 59(4):1385–1389
Petosa A, Ittipiboon A, Antar YMM, Roscoe D, Cuhaci M (1998) Recent advances in dielectric-resonator antenna technology. IEEE Antennas Propag Mag 40(3):35–48
Petosa A, Thirakoune S, Ittipiboon A (2009) Higher-order modes in rectangular DRAs for gain enhancement. In: 13th international symposium on antenna technology and applied electromagnetics and the Canadian radio sciences meeting, Toronto, ON
Plourde JK, Ren CL (1981) Application of dielectric resonators in microwave components. IEEE Trans Microw Theory Tech 29(8):754–770
Pozar DM (1983) Considerations for millimeter wave printed antennas. IEEE Trans Antennas Propag 31(5):740–747
Quan XL, Li RL, Tentzeris MM (2013) A broadband omnidirectional circularly polarized antenna. IEEE Trans Antennas Propag 61(5):2363–2370
Radnovic I, Nesic A, Milovanovic B (2010) A new type of turnstile antenna. IEEE Antennas Propag Mag 52(5):168–171
Rappaport TS, Murdock JN, Gutierrez F (2011) State of the art in 60-GHz integrated circuits and systems for wireless communications. Proc IEEE 99(8):1390–1436
Rashidian A, Klymyshyn DM (2010) Development of polymer-based dielectric resonator antennas for millimeter-wave applications. Prog Electromagn Res C 13:203–216
Richtmyer RD (1939) Dielectric resonator. J Appl Phys 10:391–398
Row JS, Chan MC (2010) Reconfigurable circularly-polarized patch antenna with conical beam. IEEE Trans Antennas Propag 58(8):2753–2757

Sager O, Tisi F (1968) On eigenmodes and forced resonance-modes of dielectric spheres. Proc IEEE 56(9):1593–1594

Sangiovanni A, Dauvignac JY, Pichot C (1997) Embedded dielectric resonator antenna for bandwidth enhancement. Electron Lett 33(25):2090–2091

Sangiovanni A, Garel PY, Dauvignac JY, Pichot C (2004) Numerical analysis of dielectric resonator antennas. Int J Numer Model 13(2–3):199–215

Schwering FK (1992) Millimeter wave antennas. Proc IEEE 80(1):92–102

Svedin J, Huss LG, Karlen D, Enoksson P, Rusu C (2007) A micromachined 94 GHz dielectric resonator antenna for focal plane array applications. In: Proceedings of the IEEE MTT-S international microwave symposium, Honolulu, HI, pp 1375–1378

Takashi I, Naoki I, Nobuyoshi K (2004) Application of modal polarization current model method to dielectric resonator antennas. Electron Commun Jpn (Part I Commun) 87(5):42–51

Trans-Tech (2013) Application note 202805B: introduction to dielectrics, pp 1–2

Uchimura H, Takenoshita T, Fujii M (1998) Development of a laminated waveguide. IEEE Trans Microw Theory Tech 46(12):2438–2443

Vilar R, Czarny R, Lee ML, Loiseaux B, Sypek M, Makowski M, Martel C, Crepin T, Boust F, Joseph R, Herbertz K, Bertuch T, Marti J (2014) Q-band millimeter-wave antennas. IEEE Microw Mag 15(4):122–130

Wahab WMA, Safavi-Naeini S, Busuioc D (2009) Low cost low profile dielectric resonator antenna (DRA) fed by planar waveguide technology for millimeter-wave frequency applications. In: Proceedings for the radio and wireless symposium, San Diego, CA, pp 27–30

Wasylyshyn DA (2005) Effects of moisture on the dielectric properties of polyoxymethylene (POM). IEEE Trans Dielectr Electr Insul 12(1):183–193

Weibel GE, Dressel HO (1967) Propagation studies in millimeter-wave link systems. Proc IEEE 55(4):497–513

Wu K, Cheng YJ, Djerafi T, Hong W (2012) Substrate-integrated millimeter-wave and terahertz antenna technology. Proc IEEE 100(7):2220–2232

Zhang X, Gao X, Chen W, Feng Z, Iskander MF (2011) Study of conformal switchable antenna system on cylindrical surface for isotropic coverage. IEEE Trans Antennas Propag 59(3):776–783

Zou G, Groenqvist H, Starski JP, Liu J (2002) Characterization of liquid crystal polymer for high frequency system-in-a-package applications. IEEE Trans Adv Packag 25(4):503–508

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_36-1

Circularly Polarized Antennas

Lot Shafai[a]*, Z. A. Pour[a], S. Latif[b] and A. Rashidian[a]
[a]Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada
[b]University of South Alabama, Mobile, AL, USA

Abstract

This chapter is focused on circularly polarized antennas. Key definitions and governing equations of circular polarization are given. Infinitesimal dipole sources are considered to establish circularly polarized radiation. First, radiation patterns of cross dipoles are mathematically reviewed, from which the condition of circularly polarized waves is concluded. Later, the idea is extended to four displaced sequentially rotated dipole antennas, resulting in circularly polarized waves within a wide angular range in space. The extension of the concept to the magnetic source counterparts and Huygens sources is briefly discussed. Other than point sources, also known as one-dimensional current sources, sources of circularly polarized radiation are further investigated for two-dimensional cases, such as microstrip patch antennas, and three-dimensional structures, such as volumetric current sources existing in dielectric resonator antennas. For these cases, the creation of circularly polarized radiation using single-feed and dual-feed, perturbed structures and sequentially rotated method is described. As a design example, numerical and measurement results of circularly polarized square patch ring antennas are extensively discussed and presented in this chapter. The square-ring microstrip antenna is selected as it closely approximates the sequentially rotated currents, and also it has not been widely studied in the literature.

Keywords

Circular polarization; Antennas; Electric and magnetic sources; Sequential rotation technique; Axial ratio; Two- and three-dimensional current sources; Microstrip patches; Dielectric resonator antennas; Square-ring patch antennas

Introduction

This chapter presents circularly polarized antennas. Initially, circular polarization and its related field vectors, in both rectangular and spherical coordinate systems, are defined and presented in section "Basic Formulas and Definitions." Traditionally, circular polarization is defined in rectangular coordinates. However, antenna far-field radiation components are best described in spherical coordinates and are more natural for antenna applications. Thus, spherical components are used in this chapter for subsequent presentations. Following this section, the source concept is introduced, and elementary electric dipoles are used to develop the fundamental expressions for radiated field of electric-type sources. The results are used for determining the conditions for circular polarization. By way of these elementary sources, the quality of circular polarization is studied and compared, using the axial ratio of the polarization vector. Hybrid electric and magnetic sources are also presented and discussed. Next, the results are used to present practical circularly polarized antenna types with two-dimensional and three-dimensional current

*Email: lot.shafai@umanitoba.ca

DOI 10.1007/978-981-4560-75-7_36-1

distributions, using microstrip and dielectric resonator-type antennas, respectively. Finally, the case of a microstrip square-ring antenna is selected, as the most general radiation source, and studied in detail. A parametric study is conducted to relate the quality of circular polarization to the antenna dimensions and the substrate permittivity. Methods for improving the antenna impedance and axial ratio beamwidths are also studied and presented.

Basic Formulas and Definitions

Polarization of electromagnetic (EM) waves is simply specified by the behavior of the instantaneous electric fields in the time-space domain (Kraus 1988; Stutzman 1993). The tip of the electric field vector, as traveling along the direction of the propagation in time, leaves a spatial trace or footprint, which is called polarization. If the resultant trace is a line, the polarization is linear, which is the simplest case to visualize the polarization concept. The most general form of the aforementioned spatial trace is an ellipse, which defines the elliptical polarization. Under a special condition, where the axial ratio of the ellipse, i.e., the ratio of the major to minor axes, becomes unity, the polarization is circular, which is the main focus of this chapter. Mathematically, linear polarization is also a special case of elliptic polarization, when the axial ratio becomes infinity.

In antenna engineering, the polarization is determined by that of the radiated EM waves, and it plays an important role, as the polarization of transmit and receive antennas in any wireless communication must be matched in order to maximize the reception. For example, the reception of two orthogonally polarized antennas, such as vertical and horizontal dipoles, is zero. In order to attain a maximum signal in the receiver, the orientation of the dipoles must be parallel in either of the wireless-link terminals. Other than the loss due to the antenna misalignments, the Faraday rotation loss (Brookner et al. 1985) and the multipath interference loss (Counselman 1999) are inevitable, when a linear polarization is used for the wireless channel. The latter raises the noise floor of the system due to the detrimental summation of the unwanted received signals from multiple paths. The former is associated with the EM wave propagation in the Earth's ionosphere. Interestingly enough, the circular polarization (CP) can overcome all of these deleterious effects. This explains why CP antennas are in great demand in satellite communications, radars, and Global Positioning Systems.

Antennas radiate EM waves into the far-field zone, where they can locally be considered as plane waves containing only transverse fields, which are normal to the propagation direction. Generally, these waves can be decomposed into two orthogonal components. Assuming the plane waves propagate along the $+z$-axis and the time-harmonic term is $e^{j\omega t}$; the corresponding instantaneous electric field can be expressed in terms of x- and y-components as follows (Stutzman 1993):

$$\vec{E}\ (t,z) = E_x(t,z)\hat{x} + E_y(t,z)\hat{y} \tag{1}$$

where E_x and E_y are given by

$$E_x(t,z)\ = e_x \cos(\omega t - kz) \tag{2}$$

$$E_y(t,z)\ = e_y \cos(\omega t - kz + \delta) \tag{3}$$

where ω, k, and δ are the angular frequency, wave number, and the phase shift of the E_y component with respect to the E_x, respectively. Also, e_x and e_y are real numbers, representing the amplitudes of the x- and

DOI 10.1007/978-981-4560-75-7_36-1

y-components of the instantaneous electric fields. After substituting Eqs. 2 and 3 into Eq. 1 and enforcing z to zero for simplicity, one can write

$$\vec{E}(t) = e_x \cos \omega t \hat{x} + e_y \cos(\omega t + \delta)\hat{y} \tag{4}$$

In general, the tip of the above electric field, expressed by Eq. 4, forms an elliptical trace over a full period of 1/f, where f is the frequency. This is the most generalized case of EM polarization. Now, let us consider the special cases. If the phase shift between the decomposed electric field components is nπ, i.e., $\delta = n\pi$, where n = 0, 1, 2, . . ., the total electric field vector is expressed as

$$\vec{E}(t) = \cos \omega t \left(e_x \hat{x} \pm e_y \hat{y}\right) \tag{5}$$

which represents linear polarization.

The polarization is circular, when there is a quadrature phase shift between the y- and x-components of electric fields, i.e., $\delta = \pm(n + 0.5)\pi$, where n = 0, 1, 2, . . . and the magnitudes of them are equal, i.e., $e_x = e_y$. Thus, the corresponding total electric field vector traces a circle over a full oscillation period, and it is expressed as

$$\vec{E}(t) = e_x (\hat{x} \pm j\hat{y}) \tag{6}$$

Therefore, two orthogonal waves, which are 90° apart both spatially and electrically, with equal amplitudes, are required to create CP waves.

Since the focus of this chapter is on CP antennas, it is instructive to express the antenna polarization characterizations in terms of the spherical components of electric fields at the far-field region. First, let us define the senses of CP polarization. Based on the IEEE standard (IEEE Standard 1979), the sense of the polarization is determined by that of the rotation of the electric field tip forming the circle or the ellipse. It is right-handed (RH) if the electric field rotates clockwise, when one looks in the direction of the propagation. Similarly, it is left-handed (LH) if the rotation is counterclockwise (IEEE Standard 1979). Mathematically, based on Eq. 4, the RHCP and LHCP waves are realized when δ is equal to −90° and +90°, respectively. The far-field electric field can be expressed in terms of the spherical components as

$$\vec{E} = E_\theta \hat{\theta} + E_\phi \hat{\phi} \tag{7}$$

Thus, it can also be decomposed in terms of the RHCP and LHCP waves as (Milligan 2005)

$$\vec{E} = E_R \hat{a}_R + E_L \hat{a}_L \tag{8}$$

where $\hat{a}_R$ and $\hat{a}_L$ are the unit vectors of the RHCP and LHCL vectors. They are orthogonal to each other and expressed by

$$\begin{cases} \hat{a}_R = \dfrac{1}{\sqrt{2}}\left(\hat{a}_\theta - j\hat{a}_\phi\right) \\ \hat{a}_\mathrm{L} = \dfrac{1}{\sqrt{2}}\left(\hat{a}_\theta + j\hat{a}_\phi\right) \end{cases} \tag{9}$$

DOI 10.1007/978-981-4560-75-7_36-1

Then, E_R and E_L are simply obtained by projecting the electric field of Eq. 7 into the above unit vectors. That is, the inner product of Eq. 7 with the complex conjugate of either unit vectors given by Eq. 9. They are therefore expressed by

$$\begin{cases} E_R = \dfrac{1}{\sqrt{2}}\left(E_\theta + jE_\phi\right) \\ E_{\mathbf{L}} = \dfrac{1}{\sqrt{2}}\left(E_\theta - jE_\phi\right) \end{cases} \tag{10}$$

The axial ratio, denoted by AR, specifies the purity of the CP waves and is the ratio of the major axis to the minor axis of the polarization ellipse. The major and minor axes are maximum and minimum of the electric field, respectively, expressed by (Hollis et al. 1969)

$$E_{\max} = \frac{1}{\sqrt{2}}\left(|E_R| + |E_L|\right) \tag{11}$$

$$E_{\min} = \begin{cases} \dfrac{1}{\sqrt{2}}\left(|E_R| - |E_L|\right), & \text{RHCP} \\ \dfrac{1}{\sqrt{2}}\left(|E_L| - |E_R|\right), & \text{LHCP} \end{cases} \tag{12}$$

In antenna applications, the axial ratio is given in dB by the following expression:

$$AR(dB) = 20\log\left(\frac{E_{\max}}{E_{\min}}\right) \tag{13}$$

Therefore, a 0 dB axial ratio demonstrates pure CP field.

So far, a brief introduction on the circular polarization is provided. More in-depth material on the subject may be found in (Stutzman 1993; Hollis et al. 1969) for interested readers. The rest of the chapter deals with circularly polarized antennas. In particular, different techniques will be discussed to realize CP fields mainly using dipole, microstrip patch, and dielectric resonator antennas.

Sources of Circularly Polarized Radiation

The previous section provided the key definitions and fundamental field vectors that govern the circular polarization. In this section, the basic electric current sources that can radiate circularly polarized fields are presented. The magnetic current sources can similarly be defined and studied. However, their corresponding radiated field expressions can be easily inferred from the duality principle and using the results of the electric current sources. A unique case is the combination of both electric and magnetic sources, and for this reason, it is also presented and discussed. Apertures and slots are typical sources of equivalent magnetic currents. However, they radiate in the presence of the object they are located on. Other nonconducting sources, such as dielectric resonators, resemble the electric fields more closely and are thus more realistic magnetic sources.

Considering the electric current sources, different arrangements of the sources are considered and studied. The simplest case of two infinitesimal dipoles is first investigated, and then, it is generalized to two finite-sized electric currents. Subsequently, the case of four displaced electric currents is discussed,

DOI 10.1007/978-981-4560-75-7_36-1

and the concept of sequential rotation for improving the quality of circular polarization is introduced, by investigating the axial ratio of the radiated field. In each case, the magnetic vector potentials are derived and used to determine the field vectors. Since the circular polarization is normally defined and used in the far-field regions, only far-field vectors are determined and discussed.

Case I: Two Collocated Infinitesimal Electric Dipoles

Figure 1 shows the geometry of two orthogonal electric dipoles, along *x*- and *y*-axes. Their current vectors are

$$\boldsymbol{I}_1 = I_1 \hat{\boldsymbol{a}}_x \tag{14}$$

$$\boldsymbol{I}_2 = I_2 \hat{\boldsymbol{a}}_y \tag{15}$$

These sources are infinitesimal dipoles, and their magnetic vector potentials are given by

$$\boldsymbol{A}_1 = \frac{\mu I_1 \ell_1}{4\pi} \frac{e^{-jkr}}{r} \hat{\boldsymbol{a}}_x \quad \ell_1 << \lambda \tag{16}$$

$$\boldsymbol{A}_2 = \frac{\mu I_2 \ell_2}{4\pi} \frac{e^{-jkr}}{r} \hat{\boldsymbol{a}}_y \quad \ell_2 << \lambda \tag{17}$$

And at far-field zone, their radiated electric fields are

$$\boldsymbol{E}_1 = -j\omega \boldsymbol{A}_1 = \frac{-jk\eta I_1 \ell_1}{4\pi} \frac{e^{-jkr}}{r} \hat{\boldsymbol{a}}_x \quad r >> \lambda \tag{18}$$

$$\boldsymbol{E}_2 = -j\omega \boldsymbol{A}_2 = \frac{-jk\eta I_2 \ell_2}{4\pi} \frac{e^{-jkr}}{r} \hat{\boldsymbol{a}}_y \quad r >> \lambda \tag{19}$$

The total electric field then is

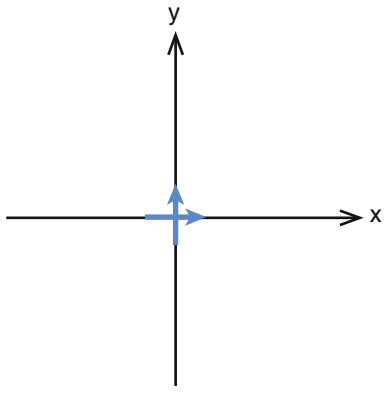


Fig. 1 Geometry of two orthogonal collocated electric dipoles, along *x*- and *y*-axes

DOI 10.1007/978-981-4560-75-7_36-1

$$\boldsymbol{E} = \boldsymbol{E}_1 + \boldsymbol{E}_2 = -\frac{jk\eta}{4\pi}\frac{e^{-jkr}}{r}\left[I_1\ell_1\hat{\boldsymbol{a}}_x + I_2\ell_2\hat{\boldsymbol{a}}_y\right] \tag{20}$$

Thus, from Eq. 6, the condition for circular polarization is

$$\begin{cases} |I_1\ell_1| = |I_2\ell_2| \ = I_o\ell \\ \angle I_1\ell_1 - \angle I_2\ell_2 = \pm\dfrac{\pi}{2} \end{cases} \tag{21}$$

Giving

$$E = -\frac{jk\eta I_o\ell}{4\pi}\frac{e^{-jkr}}{r}\left(\hat{a}_x \pm j\hat{a}_y\right) \tag{22}$$

One can also show that for two collocated magnetic dipoles, the condition for circular polarization is the same. That is,

> Two collocated orthogonal electric, or magnetic, dipoles will radiate a circularly polarized wave, if their magnitudes are equal, and their phase difference is ±90°. The phase difference of the field vectors remains the same as the source currents.

Since the radiation vectors are tangent to the radiation sphere, it is important to determine the circularly polarized vectors in spherical coordinates. They can be found from

$$\begin{aligned} \hat{\boldsymbol{a}}_x &= \hat{\boldsymbol{a}}_r \sin\theta\cos\phi + \hat{\boldsymbol{a}}_\theta \cos\theta\cos\phi - \hat{\boldsymbol{a}}_\phi \sin\phi \\ \hat{\boldsymbol{a}}_y &= \hat{\boldsymbol{a}}_r \sin\theta\sin\phi + \hat{\boldsymbol{a}}_\theta \cos\theta\sin\phi + \hat{\boldsymbol{a}}_\phi \cos\phi \end{aligned} \tag{23}$$

On the radiation sphere the radial component is zero, and thus the total electric field vector of Eq. 22 can be expressed as

$$E = -jk\eta\frac{I_o\ell}{4\pi}\frac{e^{-jkr}}{r}e^{\pm j\phi}\left(\cos\theta\,\hat{a}_\theta \pm j\hat{a}_\phi\right) \tag{24}$$

The RHCP and LHCP components can be found from Eq. 10 as

$$\begin{cases} E_R = \dfrac{1}{\sqrt{2}}\left(-jk\eta\dfrac{I_o\ell}{4\pi}\dfrac{e^{-jkr}}{r}\right)e^{\pm j\phi}(\cos\theta \mp 1) \\ E_L = \dfrac{1}{\sqrt{2}}\left(-jk\eta\dfrac{I_o\ell}{4\pi}\dfrac{e^{-jkr}}{r}\right)e^{\pm j\phi}(\cos\theta \pm 1) \end{cases} \tag{25}$$

Assume $l_1 = l_2 = l$; the resulting radiation will be RHCP if $I_2 = -jI_1$ and LHCP when $I_2 = +jI_1$. That is, when $I_2 = -jI_1$

$$\begin{cases} E_R = \frac{1}{\sqrt{2}}\left(-jk\eta\dfrac{I_o\ell}{4\pi}\dfrac{e^{-jkr}}{r}\right)e^{-j\phi}(\cos\theta\ + 1)\Big|_{\theta=0} \to \max \ \text{RHCP} \\ E_L = \frac{1}{\sqrt{2}}\left(-jk\eta\dfrac{I_o\ell}{4\pi}\dfrac{e^{-jkr}}{r}\right)e^{-j\phi}(\cos\theta\ - 1)\Big|_{\theta=0} = 0 \end{cases} \tag{26}$$

DOI 10.1007/978-981-4560-75-7_36-1

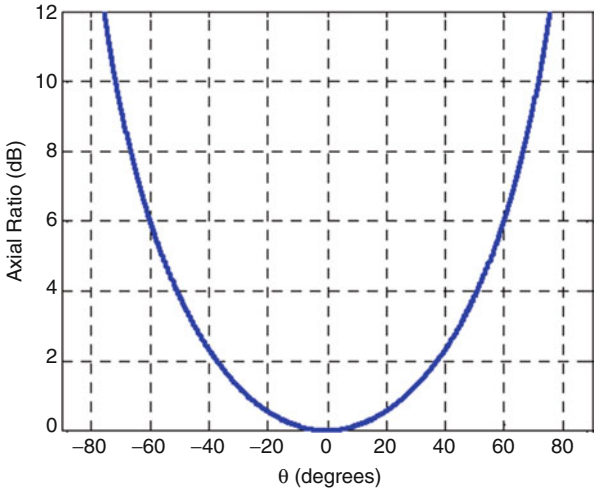


Fig. 2 Variation of AR as a function of θ, for two orthogonal crossed electric dipoles, with a phase difference of 90°

And for $I_2 = +jI_1$

$$\begin{cases} E_R = \frac{1}{\sqrt{2}}\left(-jk\eta\frac{I_o\ell}{4\pi}\frac{e^{-jkr}}{r}\right)e^{+j\phi}(\cos\theta\ -1)\Big|_{\theta=0} = 0 \\ E_L = \frac{1}{\sqrt{2}}\left(-jk\eta\frac{I_o\ell}{4\pi}\frac{e^{-jkr}}{r}\right)e^{+j\phi}(\cos\theta\ +1)\Big|_{\theta=0} \rightarrow \max\ \ \text{LHCP} \end{cases} \tag{27}$$

Thus, using Eq. 13, it becomes evident that a perfect circular polarization occurs only on the z-axis, where $\theta = 0$ with $AR = 1$ (0 dB). Away from the z-axis the axial ratio becomes

$$AR = \frac{|E_R| + |E_L|}{|E_R| - |E_L|} = \frac{|\cos\theta + 1| + |\cos\theta - 1|}{|\cos\theta + 1| - |\cos\theta - 1|} \tag{28}$$

Figure 2 shows a plot of the AR as a function of θ, the angle off the z-axis.

Case II: Two Finite Orthogonal Electric Currents

Figure 3 shows two orthogonal electric currents, along the x- and y-axes and centered at the origin. Their current distributions are assumed to be sinusoidal. For the x-oriented dipole, it can be written in the following form:

$$\begin{aligned} \boldsymbol{I}_x &= \hat{\boldsymbol{a}}_x I_0 \sin\left[k\left(\frac{\ell}{2} - x\right)\right] \quad 0 \le x \le \frac{\ell}{2} \\ &= \hat{\boldsymbol{a}}_x I_0 \sin\left[k\left(\frac{\ell}{2} + x\right)\right] \quad -\frac{\ell}{2} \le x \le 0 \end{aligned} \tag{29}$$

For the x-directed dipole, the magnetic vector potential is given by

DOI 10.1007/978-981-4560-75-7_36-1

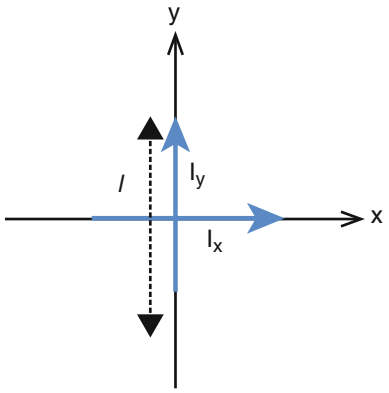


Fig. 3 Geometry of two orthogonal sinusoidal electric currents of finite lengths, centered at origin

$$A_x = \frac{\mu I_0}{4\pi}\frac{e^{-jkr}}{r}\left\{\int_{-\frac{\ell}{2}}^{0}\sin\left[k\left(\frac{\ell}{2}+x'\right)\right]e^{jkx'\,\sin\theta\,\cos\phi}\;dx' + \int_{0}^{-\frac{\ell}{2}}\sin\left[k\left(\frac{\ell}{2}-x'\right)\right]e^{jkx'\,\sin\theta\,\cos\phi}\;dx'\right. \tag{30}$$

After some manipulations one finds

$$A_x = \frac{\mu I_{o_x}}{2\pi k}\frac{e^{-jkr}}{r}\left[\frac{\cos\left(k\frac{\ell}{2}\,\sin\theta\,\cos\phi\right)-\cos\left(k\frac{\ell}{2}\right)}{1-\sin^2\theta\,\cos^2\phi}\right] = \frac{\mu I_{o_x}}{2\pi k}\frac{e^{-jkr}}{r}F_x \tag{31}$$

where F_x represents the angular-dependent function in Eq. 31 as

$$F_x = \frac{\cos\left(k\frac{\ell}{2}\,\sin\theta\,\cos\phi\right)-\cos\left(k\frac{\ell}{2}\right)}{1-\sin^2\theta\,\cos^2\phi} \tag{32}$$

Similarly for a finite-length dipole along the y-axis, the magnetic vector potential is expressed as

$$A_y = \frac{\mu I_{o_y}}{2\pi k}\frac{e^{-jkr}}{r}\left[\frac{\cos\left(k\frac{\ell}{2}\,\sin\theta\,\sin\phi\right)-\cos\left(k\frac{\ell}{2}\right)}{1-\sin^2\theta\,\sin^2\phi}\right] = \frac{\mu I_{o_y}}{2\pi k}\frac{e^{-jkr}}{r}F_y \tag{33}$$

where F_y is defined as

$$F_y = \frac{\cos\left(k\frac{\ell}{2}\,\sin\theta\,\sin\phi\right)-\cos\left(k\frac{\ell}{2}\right)}{1-\sin^2\theta\,\sin^2\phi} \tag{34}$$

and the far-field electric field vectors are

$$\begin{cases} E_x = -j\omega A_x \\ E_y = -j\omega A_y \end{cases} \tag{35}$$

Again the condition for circular polarization is

$$\begin{cases} |I_{o_x}| = |I_{o_y}| = I_o \\ \angle I_{o_x} - \angle I_{o_x} = \pm\dfrac{\pi}{2} \end{cases} \tag{36}$$

In terms of the spherical coordinate systems, the far-field electric field can be shown to be

$$\begin{cases} E_\theta = E_x \cos\theta \cos\phi + E_y \cos\theta \sin\phi \\ E_\phi = -E_x \sin\phi + E_y \cos\phi \end{cases} \tag{37}$$

After substituting Eqs. 31, 32, 33, 34, 35, and 36 into Eq. 37, the total electric field can be expressed as

$$E = -j\frac{\eta I_o}{2\pi}\frac{e^{-jkr}}{r}\left[\cos\theta\left(F_x \cos\phi \pm jF_y \sin\phi\right)\hat{a}_\theta \pm j\left(F_y \cos\phi \pm jF_x \sin\phi\right)\hat{a}_\phi\right] \tag{38}$$

which on the z-axis, i.e., $\theta = 0$, it is simplified to

$$E = -j\frac{\eta I_o}{2\pi}\frac{e^{-jkr}}{r}e^{\pm j\phi}\left[1 - \cos\left(k\frac{\ell}{2}\right)\right]\left[\hat{a}_\theta \pm j\hat{a}_\phi\right] \tag{39}$$

From Eq. 10, the corresponding circularly polarized vectors will have the following forms

$$\begin{cases} E_R = \dfrac{1}{\sqrt{2}}\left(-j\eta\dfrac{I_o}{2\pi}\dfrac{e^{-jkr}}{r}\right)e^{\pm j\phi}\left[\cos\theta\left(F_x \cos\phi \pm jF_y \sin\phi\right) \mp \left(F_y \cos\phi \pm jF_x \sin\phi\right)\right] \\ E_L = \dfrac{1}{\sqrt{2}}\left(-j\eta\dfrac{I_o}{2\pi}\dfrac{e^{-jkr}}{r}\right)e^{\pm j\phi}\left[\cos\theta\left(F_x \cos\phi \pm jF_y \sin\phi\right) \pm \left(F_y \cos\phi \pm jF_x \sin\phi\right)\right] \end{cases} \tag{40}$$

and the axial ratio can be calculated from Eq. 13. Figure 4 shows the effect of electric current length l on the axial ratio. The 3 dB axial ratio beamwidths are plotted in Fig. 5 against the dipole length. The beamwidth becomes narrower, as the dipole enlarges in length.

Case III: Four Displaced Sequentially Rotated Electric Dipole Sources

In this section, electric current sources displaced from the origin are considered. There are two possibilities in this case: Case IIIa when dipole currents are normal to the axes at their locations and Case IIIb when dipole currents are along the axes. Figure 6 shows Case IIIa, where four dipoles are located normal to x- and y-coordinate axes at their locations, and their parameters are given as

$$\begin{aligned} \boldsymbol{I}_1 &= \boldsymbol{I}_1\hat{\boldsymbol{a}}_y \\ \boldsymbol{I}_2 &= \boldsymbol{I}_2\left(-\hat{\boldsymbol{a}}_x\right) \\ \boldsymbol{I}_3 &= \boldsymbol{I}_3\left(-\hat{\boldsymbol{a}}_y\right) \\ \boldsymbol{I}_4 &= \boldsymbol{I}_4\hat{\boldsymbol{a}}_x \end{aligned} \tag{41}$$

DOI 10.1007/978-981-4560-75-7_36-1

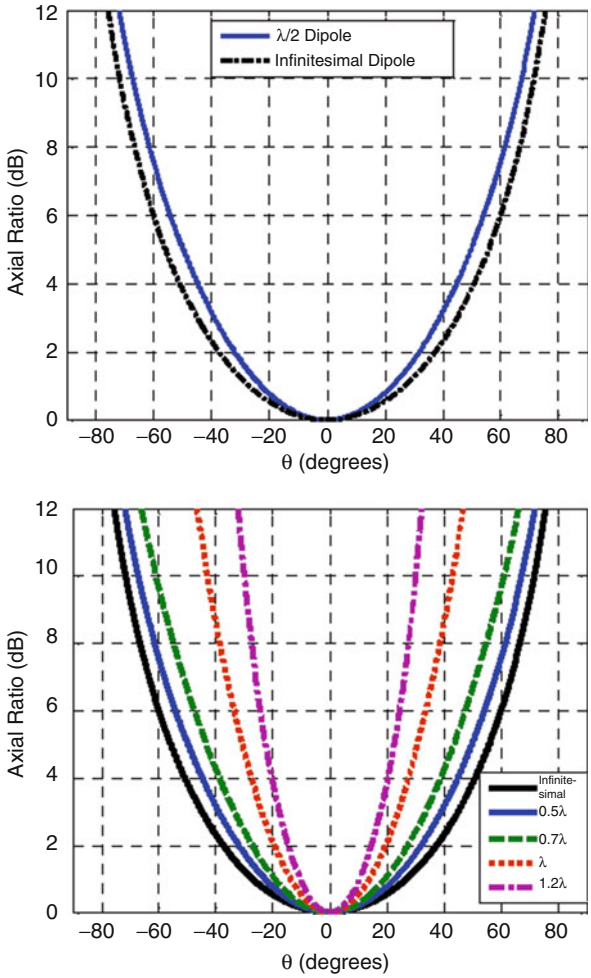


Fig. 4 Effect of electric current length on the axial ratio of two crossed dipoles of finite lengths

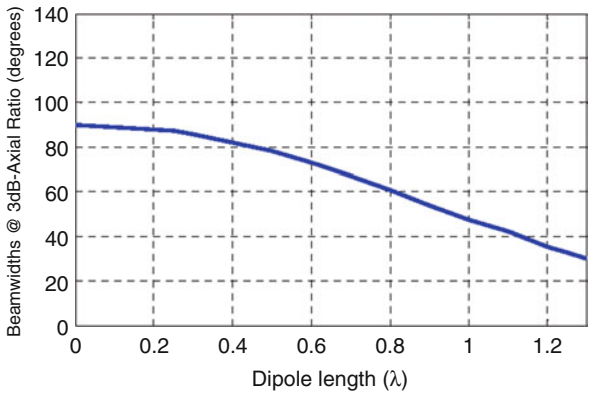


Fig. 5 Beamwidths of 3 dB *AR* of two crossed dipoles of finite lengths with sinusoidal current distributions

DOI 10.1007/978-981-4560-75-7_36-1

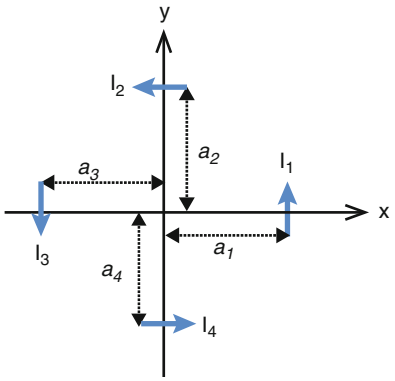


Fig. 6 Geometry of Case IIIa, four electric dipoles perpendicular to *x*- and *y*-axes at their locations

Their corresponding far zone electric fields are

$$\begin{cases} E_1 = -jk\eta \dfrac{I_1\ell_1}{4\pi}\dfrac{e^{-jkr}}{r} e^{jka_1 \sin\theta \cos\phi}\hat{a}_y \\ E_2 = jk\eta \dfrac{I_2\ell_2}{4\pi}\dfrac{e^{-jkr}}{r} e^{jka_2 \sin\theta \sin\phi}\hat{a}_x \\ E_3 = jk\eta \dfrac{I_3\ell_3}{4\pi}\dfrac{e^{-jkr}}{r} e^{-jka_3 \sin\theta \cos\phi}\hat{a}_y \\ E_4 = -jk\eta \dfrac{I_4\ell_4}{4\pi}\dfrac{e^{-jkr}}{r} e^{-jka_4 \sin\theta \sin\phi}\hat{a}_x \end{cases} \tag{42}$$

Thus, the total electric field is given by

$$E = -jk\eta \frac{e^{-jkr}}{4\pi r}\left[\left(I_1\ell_1 e^{jka_1 \sin\theta \cos\phi} - I_3\ell_3 e^{-jka_3 \sin\theta \cos\phi}\right)\hat{a}_y - \left(I_2\ell_2 e^{jka_2 \sin\theta \sin\phi} - I_4\ell_4 e^{-jka_4 \sin\theta \sin\phi}\right)\hat{a}_x\right] \tag{43}$$

It is clear that for finite radiated field on the *z*-axis, one must have

$$\begin{cases} |I_1\ \ell_1| = |I_3\ \ell_3| = I_1\ell \\ |I_2\ \ell_2| = |I_4\ \ell_4| = I_2\ell \\ \angle I_3\ \ell_3 - \angle I_1\ \ell_1 = \pi \\ \angle I_4\ \ell_4 - \angle I_2\ \ell_2 = \pi \end{cases} \tag{44}$$

With these conditions and $a_1 = a_3$ and $a_2 = a_4$, the electric field becomes

$$E = -jk\eta\ell \frac{e^{-jkr}}{2\pi r}\left[-I_2 \cos\left(ka_2 \sin\theta \sin\phi\right)\hat{a}_x + I_1 \cos\left(ka_1 \sin\theta \cos\phi\right)\hat{a}_y\right] \tag{45}$$

For equally displaced dipole sources, i.e., $a_1 = a_2 = a_3 = a_4 = a$, the condition for circular polarization is

$$\begin{cases} |I_1| = |I_2| = I_o \\ \angle I_2 - \angle I_1 = \pm\dfrac{\pi}{2} \end{cases} \tag{46}$$

DOI 10.1007/978-981-4560-75-7_36-1

Combining the conditions of Eqs. 44 and 46, one can arrive at the condition

$$\angle I_1 = 0, \angle I_2 = \pm\frac{\pi}{2}, \angle I_3 = \pm\pi, \angle I_4 = \pm\frac{3\pi}{2} \tag{47}$$

The condition for circular polarization is:

For four equally displaced electric currents around the z-axis the phase progression for the currents must be 2π.

It can be shown that this condition holds for any number of current elements around the z-axis. That is, if there are n numbers of equally displaced currents around the z-axis, as long as the total phase progression is 2π, the radiated field will be circularly polarized. In the limit, when n approaches infinity, the current elements converge to an electric current loop, i.e., a loop antenna.

In terms of the spherical coordinates, one can find the electric field

$$E = -jk\eta I_o \ell \frac{e^{-jkr}}{2\pi r} \left\{ \begin{array}{l} \cos\theta[\cos(ka\sin\theta\cos\phi)\ \sin\phi \mp j\cos(ka\sin\theta\sin\phi)\cos\phi]\,\hat{a}_\theta \\ +[\cos(ka\sin\theta\cos\phi)\ \cos\phi \pm j\cos(ka\sin\theta\sin\phi)\ \sin\phi]\,\hat{a}_\phi \end{array} \right\} \tag{48}$$

On the z-axis it becomes

$$E = k\eta I_o \ell \frac{e^{-jkr}}{2\pi r} e^{\pm j\phi}\left[\,\hat{a}_\theta \pm j\,\hat{a}_\phi\right] \tag{49}$$

In general for off z-axis points, Eq. 48 can be expressed in terms of the array factor terms, due to the displaced radiating elements along the x- and y-axes. They are defined as

$$\left\{ \begin{array}{l} AF_x = \cos(ka\sin\theta\cos\phi) \\ AF_y = \cos(ka\sin\theta\sin\phi) \end{array} \right. \tag{50}$$

Thus, one can write

$$E = -jk\eta I_o \ell \frac{e^{-jkr}}{2\pi r}\left\{\cos\theta\left(AF_x\sin\phi \mp jAF_y\cos\phi\right)\hat{a}_\theta + \left(AF_x\cos\phi \pm jAF_y\sin\phi\right)\hat{a}_\phi\right\} \tag{51}$$

Axial ratios for different element spacing d, i.e., $d = 2a$, between the dipoles are shown below in Figs. 7 and 8. As the dipole spacing increases, initially, the axial ratio beamwidth increases. The best spacing appears to be around $d = 0.443\lambda$, where the 3 dB axial ratio beamwidth becomes about 170° and covers nearly the entire half space, except for an angular range near the plane of the dipoles. This sequentially rotated dipole arrangement, therefore, appears to be the best source arrangement for circular polarization. It is also easy to implement using microstrip patch or other planar antennas. However, the finite size of the patch antenna will influence the axial ratio results, but the configuration is still the optimum source for a broad axial ratio beamwidth. This sequentially rotated configuration is also known to give much wider input impedance bandwidths, for the array elements, than other patch arrangements (Teshirogi et al. 1985; Huang 1986).

For Case IIIb, the geometry of the sequentially rotated elements is illustrated in Fig. 9, where the tip of the current vector of each element points toward the coordinate origin, i.e., each element is parallel to its respective axis. Under the condition given by Eq. 44, the corresponding electric field can be expressed as

DOI 10.1007/978-981-4560-75-7_36-1

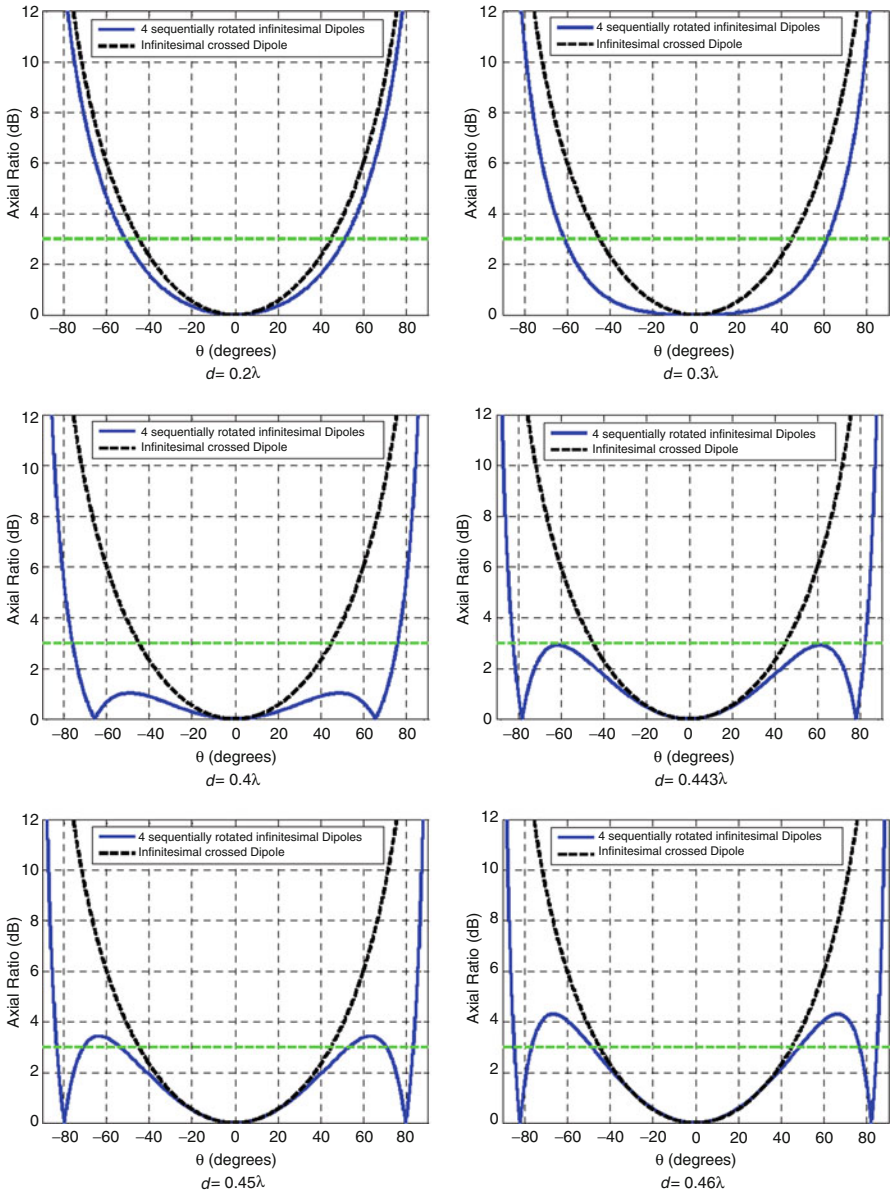


Fig. 7 *AR* of four sequentially rotated infinitesimal dipoles with different spacing, compared with crossed infinitesimal dipoles

$$E = jk\eta\ell \frac{e^{-jkr}}{2\pi r}\left[I_1 \cos\left(ka_1 \sin\theta \cos\phi\right)\hat{a}_x + I_2 \cos\left(ka_2 \sin\theta \cos\phi\right)\hat{a}_y\right] \tag{52}$$

Comparing the above equation with its counterpart electric field in Eq. 45, the magnitudes of the *x*- and *y*-components are swapped. The rest of the analysis can be carried out following Eqs. 46, 47, 48, 49, 50, and 51.

Case IV: Hybrid Electric and Magnetic Sources

An interesting antenna configuration is the combination of parallel electric and magnetic dipoles, collocated or adjacent. The far field of the electric dipole was provided earlier in Eqs. 16 and 24. The same can be generated similarly for a magnetic dipole, and in the spherical coordinate it is given by

DOI 10.1007/978-981-4560-75-7_36-1

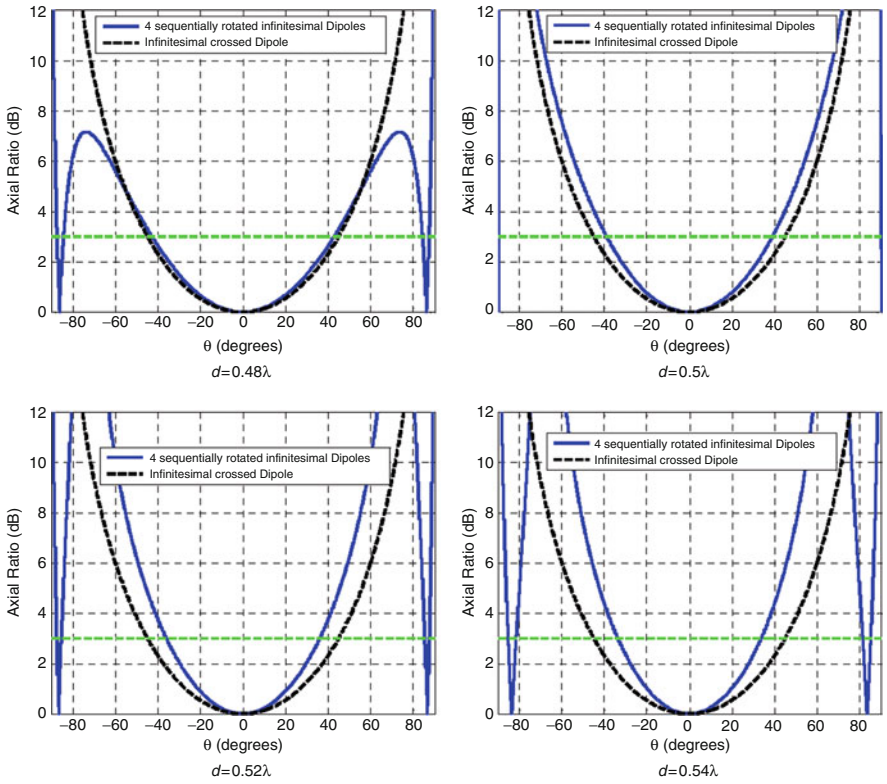


Fig. 8 *AR* of four sequentially rotated infinitesimal dipoles with different spacing, compared with crossed infinitesimal dipoles

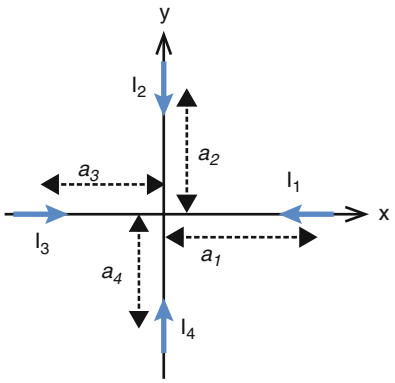


Fig. 9 Geometry of Case IIIb, four electric dipoles parallel to *x*- and *y*-axes, each element vector pointing toward the coordinate origin

$$\begin{aligned} \boldsymbol{I}^m &= \boldsymbol{I}^m\ \hat{\boldsymbol{a}}_x \\ \boldsymbol{E}^m &= \left(-\frac{jkI^m\ell}{4\pi}\frac{e^{-jkr}}{r}\right)\left[-\sin\phi\,\hat{\boldsymbol{a}}_\theta - \cos\theta\,\cos\phi\,\hat{\boldsymbol{a}}_\phi\right] \end{aligned} \tag{53}$$

If the magnetic source is related to the electric one by the characteristic impedance of space and a phase of 90°, the total field can be found from

$$\begin{aligned} E &= E^e + E^m \\ I^m &= j\eta I^e \end{aligned} \tag{54}$$

DOI 10.1007/978-981-4560-75-7_36-1

in the following form

$$\boldsymbol{E} = \left(-\frac{jk\eta I\ell}{4\pi}\frac{e^{-jkr}}{r}\right)(\cos\theta\cos\phi - j\sin\phi)\left(\hat{\boldsymbol{a}}_\theta - j\hat{\boldsymbol{a}}_\phi\right) \tag{55}$$

which is a circularly polarized wave in whole space. Thus, the conditions for circular polarization by a combination of electric and magnetic sources are

$$\begin{aligned} |I^m| &= \eta\,|I^e| \\ \angle I^m &= \angle I^e \pm \frac{\pi}{2} \end{aligned} \tag{56}$$

In practice the realization of a far field given by Eq. 55 is not realistic, since there are no perfect magnetic dipoles. It may be approximated by an aperture or more exactly by a narrow slot. However, slots are present on a conducting surface, and its presence alters the slot's radiation patterns. Thus, while the combination of an electric and magnetic dipole is an ideal source for circular polarization, its practical implementation is approximate and must be investigated for its axial ratio and impedance bandwidths, similar to other antennas.

Practical Implementations

Having established the mathematical foundations for circular polarization, the practical antenna configurations are now reviewed for generating circular polarization. From an electrical point of view, they can be divided into two different types:

- Traveling wave antennas
- Resonant antennas

From a current distribution point of view, they can be divided into three categories:

- One-dimensional currents
- Two-dimensional currents
- Three-dimensional currents

The traveling wave antennas are mostly made of continuous conducting structures, such as spirals, helices, and microstrip transmission lines over ground planes. Their current can be distributed in two dimensions such as planar spiral antennas, or in three dimensions similar to conical spiral antennas, and radiates as it travels along the structure. Thus, the shape of the antenna geometry is important for circular polarization. These antennas are covered in a separate chapter and will not be repeated here. For this reason, in this chapter, the second subdivision approach, i.e., the dimensional characteristic of the antenna currents, will be used for further discussion. In this manner, traveling wave antennas, based on microstrip transmission line, will be discussed as part of the two-dimensional magnetic current antennas.

Antennas with One-Dimensional Current Distributions

The case of one-dimensional currents is referred to one-dimensional antennas, which in practice are thin wire antennas. This case was already discussed in Case II above, for a general form of "two finite

DOI 10.1007/978-981-4560-75-7_36-1

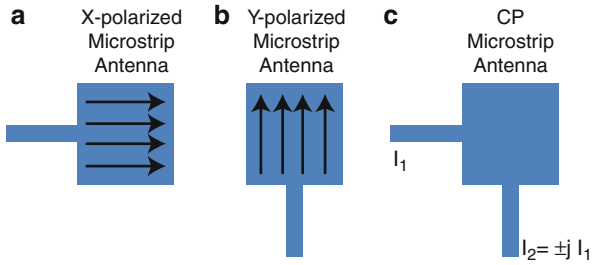


Fig. 10 (**a**) *x*-polarized and (**b**) *y*-polarized square microstrip patch antennas fed by transmission lines and (**c**) CP microstrip patch excited by two orthogonal transmission lines

orthogonal electric currents." A practical example of this case is "two orthogonal half-wave dipole antennas."

Antennas with Two-Dimensional Current Distributions

The two-dimensional current distribution can be generated from the one-dimensional case by using planar surfaces, in lieu of thin wires. The mathematical extension of two orthogonal wires to two orthogonal planar surfaces is obvious and will not be discussed here. Instead, the case of microstrip patch antennas will be considered as the example of two-dimensional currents, because of their popularity and ease of fabrication. The microstrip patch antenna is a resonant structure, similar to a wire antenna. Thus, its most common size is half-wave length, where the current distribution is sinusoidal, but distributed over its surface, as shown in Fig. 10a. For simplicity, the square patch of Fig. 10a is fed by a microstrip transmission line. However, as is well known, it can be also fed by other means such as a probe or a slot. The advantage of a microstrip patch is that it can also be fed in the orthogonal direction, for the orthogonal currents, as shown in Fig. 10b. Thus, for circular polarization a single square patch can be excited with two orthogonal feeds of Fig. 10c. It should also be noted that because of symmetry, one can also use a circular patch, in place of a square patch in Fig. 10.

The purity of CP radiation is measured by the axial ratio value, and the operating bandwidth is defined by the axial ratio bandwidth. For microstrip antennas, the axial ratio bandwidth is calculated using the following equation (Langston and Jackson, 2004):

$$BW_{CP}^{AR} = \frac{AR_{\max} - 1}{\sqrt{AR_{\max}}Q} \tag{57}$$

where $AR_{\max}$ is the maximum axial ratio value and Q is the quality factor.

Figure 11 shows different geometries for circular polarization. There are two fundamental approaches, the two-feed approach, which is discussed above, and a single-feed approach. The single-feed approach generates the orthogonal currents with a geometric deformation, which for symmetry reason is located at 45°, with respect to the feed. It is more advantageous in practical applications, because of its feed system simplicity. They are discussed separately below.

Dual-Orthogonal-Fed Circularly Polarized Patch

Dual-orthogonal feeds are used to generate two orthogonal modes with equal amplitude but in-phase quadrature. An external power divider is used in this case. Several power dividers have been successfully used for CP generation such as the quadrature hybrid, the Wilkinson power divider, the 180° hybrid, and the T-junction power divider (Garg et al. 2001). The configuration of dual-fed CP patches with a quadrature hybrid is shown in Fig. 12. Depending on the sense of CP rotation, one port is used as the

DOI 10.1007/978-981-4560-75-7_36-1

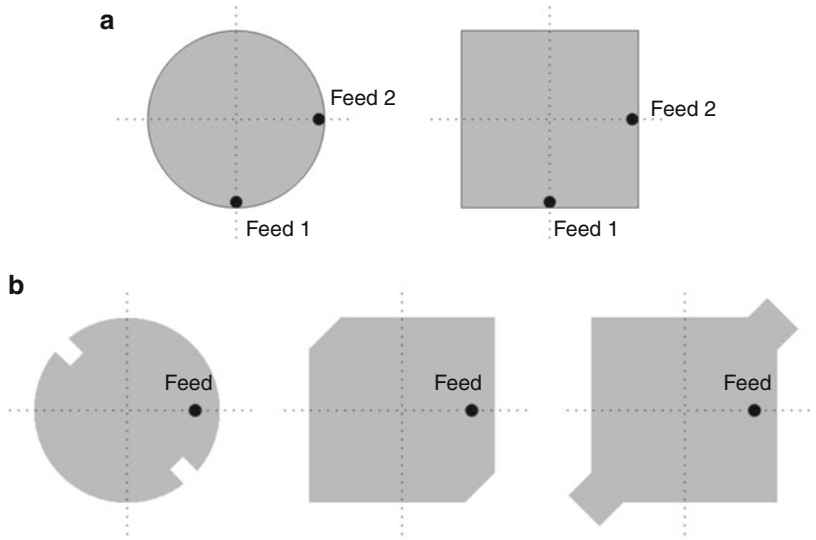


Fig. 11 (**a**) Dual-fed CP patch antenna requiring an external polarizer and (**b**) singly-fed CP patch antenna with perturbations

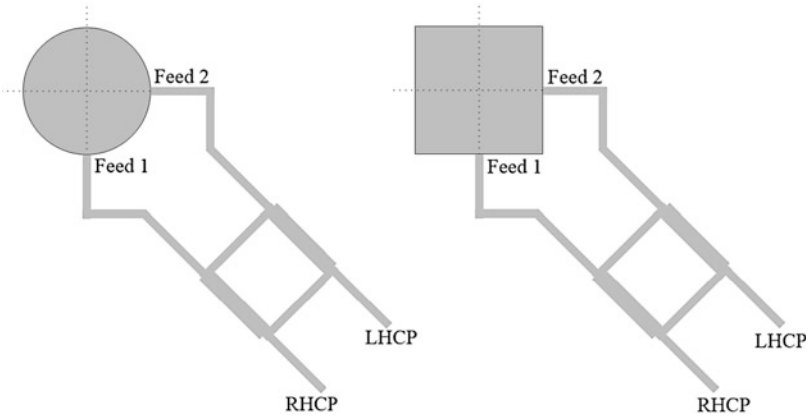


Fig. 12 Dual-fed CP patch antenna configuration with a quadrature hybrid power divider

input, and the other port is terminated with a matched load. The output ports are then connected to two feed points on the patch. The input signal is thus divided into two equal amplitudes, but in-phase quadrature, with good isolation, generally exceeding 20 dB, between two output ports. In dual-fed configurations, the 3 dB axial ratio bandwidth of the patch antenna is broad since a quadrature hybrid is broadband (Langston and Jackson 2004). Its operating bandwidth is limited by the impedance bandwidth of the patch antenna itself.

Singly-Fed Circularly Polarized Patch

Dual feed for CP requires external power dividers, thus needs extra space, and, in many cases, is difficult to accommodate. In these scenarios, a single-point feed patch capable of generating CP radiation is very attractive. A microstrip patch with a single feed generates linear polarization at the dominant mode. To induce two orthogonal patch modes with equal amplitude and in-phase quadrature, slight perturbation of the patch with respect to the feed is necessary. Various perturbation schemes for generating CP from patch antennas have been reported in the literature (Garg et al. 2001). However, the main principle in these schemes is the detuning of the degenerate mode in a symmetrical patch by introducing small perturbation segments at appropriate locations. The generated mode of a patch antenna can be separated into two

DOI 10.1007/978-981-4560-75-7_36-1

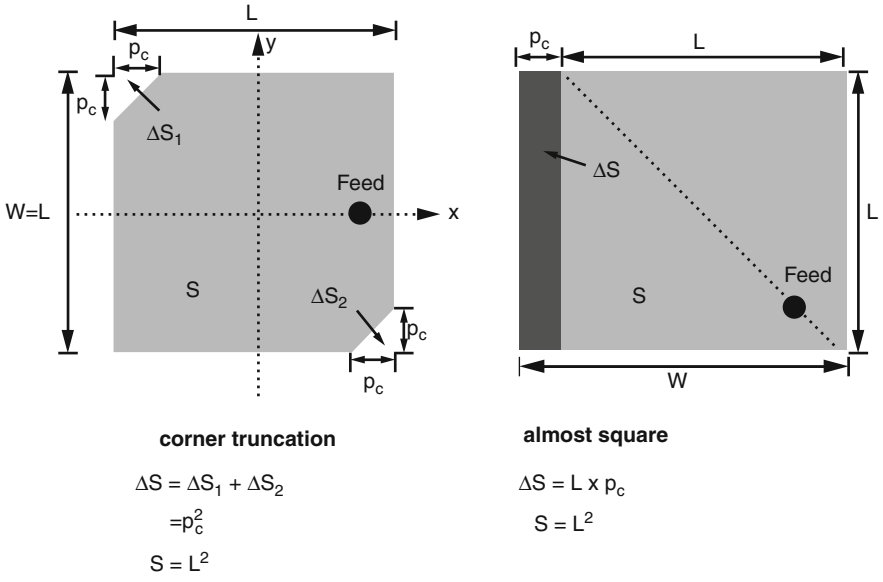


Fig. 13 Single feed CP patch antenna configuration with perturbations. Depending on the feed location, two types are possible: "corner truncation" and "almost square"

orthogonal modes by a perturbation segment in the form of a slit, a slot, a truncated segment, or an added stub. The radiation fields excited by these modes are typically perpendicular to each other. When the amount of perturbation is selected to the optimum value, these two modes are excited in equal amplitude and phase quadrature at the center frequency. This enables the CP radiation from the patch antenna even with a single feed.

The analysis for determining the amount of segmentation in the case of rectangular patch is described in detail in (Hall and James 1989). Two types of configuration based on the feed location, as shown in Fig. 13, are used in the analysis. In this figure, "corner truncation" is the configuration when the feed is either on the *x*- or *y*-axis, and it is "almost square" when the feed is located on the diagonal axis of the patch. ΔS represents the total sum of the perturbation segments, which may consist of a single (e.g., "almost square") or multiple segments (e.g., "corner truncation"). By introducing symmetric perturbations on opposite diagonal corners in a corner truncation patch, CP radiation can be obtained. In this type, the amount of perturbation can be determined using the simplified design equation

$$\left|\frac{\Delta S}{S}\right| = \frac{1}{2Q_o} \tag{58}$$

where S represents the total area of the patch and Q_o is the unloaded quality factor of the cavity resonator produced by the patch and the ground plane. Q_o depends on the dimension of the patch, the substrate thickness, and the substrate dielectric constant. For better accuracy, Q_o should be selected such that the patch radiation efficiency is greater than 90 % (Garg et al. 2001).

In the case of the "almost square" patch antenna, the perturbation is introduced by adding a strip to the patch, which, in fact, provides a nearly square patch antenna, and when it is fed diagonally, CP is generated. For this type, the amount of perturbation can be obtained using the following equation:

DOI 10.1007/978-981-4560-75-7_36-1

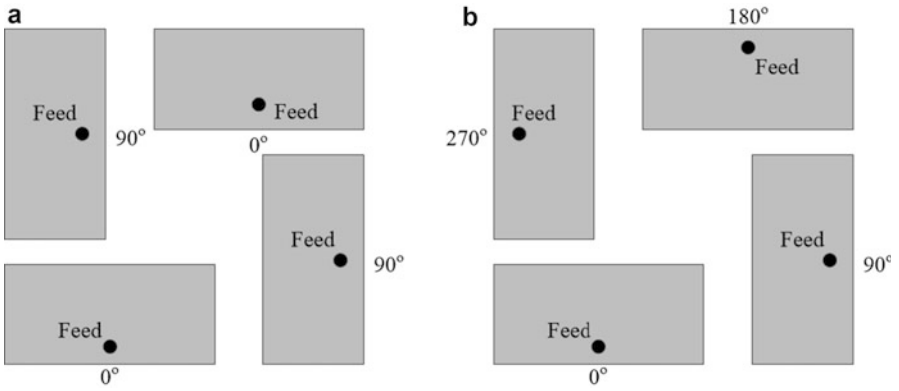


Fig. 14 Microstrip array with linearly polarized elements sequentially rotated for CP radiation: (**a**) narrow band arrangement and (**b**) wide band arrangement

$$\left|\frac{\Delta S}{S}\right| = \frac{1}{Q_o} \tag{59}$$

Sequentially Rotated Microstrip Arrays for CP Radiation

In section "Sources of Circularly Polarized Radiation," it was shown that four sequentially rotated dipoles can generate circular polarization with very broad axial ratio beamwidth, which is an ideal method for generating circular polarization. This configuration can easily be implemented with four microstrip patches, by replacing the dipoles with microstrip patches. It has been also shown that microstrip patches in this configuration have wide impedance bandwidths (Teshirogi et al. 1985; Huang 1986). Two arrangements are presented in Fig. 14, where an external power divider with appropriate phase shift is necessary to generate CP radiation from these linearly polarized patches. The configuration in Fig. 14a does not meet the phase requirement of Eq. 47 and has a narrow axial ratio beamwidth. The configuration of Fig. 14b, on the other hand, meets the phase requirement of Eq. 47 and has a substantially wider axial ratio beamwidth, with significantly low cross-polarization (Garg et al. 2001).

Sequentially Rotated Magnetic Dipoles

A simple way to produce a magnetic dipole is to bend a microstrip line, as shown in Fig. 15a. The electric field discontinuity at the bend, measured from the two sides of the line, generates an equivalent magnetic dipole, at 45° with each side of the angle. Thus, it radiates an electric field that bisects the corner. Since an undulated line has four corners, it is equivalent to four sequentially rotated magnetic dipoles. If four corners generate 90° and 180° phase shifts, the opposite corners are out of phase, and adjacent corners are at phase quadrature, and the four corners radiate circular polarization. This condition occurs when the lengths $h = \lambda/2$, $w = \lambda/4$, and $p = \lambda$. With these conditions the period of the corner unit cells becomes λ, and the corners of the undulated traveling line radiate in-phase circular polarization.

Hybrid Electric and Magnetic Dipole Antennas

A microstrip line can also be used to design hybrid electric and magnetic dipole antennas for circular polarization. Figure 15b shows one such case, where a straight microstrip line is etched on a substrate, over a ground plane. On the upper surface an electric dipole is placed orthogonal to the microstrip line. On the opposite side, a resonant slot is etched on the ground plane, which represents a magnetic dipole. If the spacing between the slot and dipole is $\lambda/4$, then the phase shift between the two will satisfy the condition for circular polarization. For better performance, the slot should be over a cavity, or a reflecting plate, to cause unidirectional radiation, similar to the dipole. In this design, the amplitude of the dipole and slot

DOI 10.1007/978-981-4560-75-7_36-1

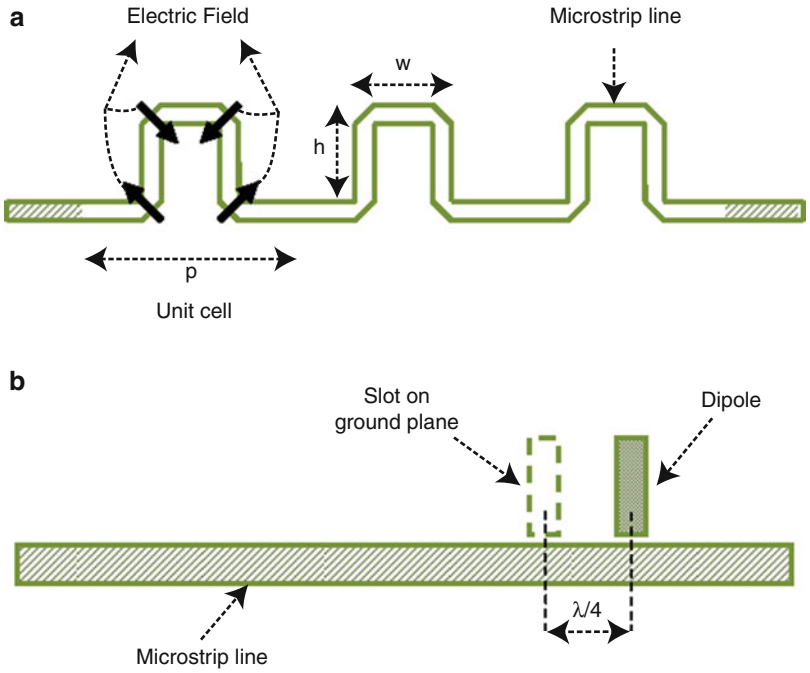


Fig. 15 (**a**) Geometry of a meandered microstrip line. *Solid arrows* at the corners of the line represent electric field. (**b**) Geometry of hybrid electric and magnetic dipoles

excitations can be controlled by adjusting their coupling with the microstrip line, i.e., adjusting their distance from the microstrip line. In this design, the slot and dipole are not collocated, and thus their circular polarization is not perfect, as the case of two hybrid electric and magnetic dipoles, but the antenna configuration is simple and can be designed satisfactorily.

Antennas with Three-Dimensional Current Distributions, Dielectric Resonator Antennas

Three-dimensional current distributions can be generated by volumetric antennas such as cones, spheres, and other shapes. However, in conducting shapes the currents are on the antenna surface and can be represented by the characteristic modes of the antenna. Circular polarization can therefore be generated similar to orthogonal wire antennas. An antenna that is unique in its electromagnetic characteristics is the *dielectric resonator* antenna, in which the characteristic modes resonate within the volume of the antenna. These antennas, depending on the resonating mode, can be equivalent to an electric or magnetic dipole antenna. In addition, since the entire antenna volume contributes to the resonance, more geometric possibilities exist for circular polarization. They are summarized in this section.

Dielectric resonator antennas (DRAs) are resonator-type antennas made of low-loss microwave materials and usually designed for single-mode operation with linear polarization. Different modes having specific internal and associated external field distribution can be excited in DRAs using different feeding methods such as microstrip, probe, and aperture. These modes are equivalent to electric and magnetic dipoles with different orientations. For instance, the field configurations inside the rectangular and cylindrical DRAs for the lowest-order modes are plotted in Fig. 16, where the subscript δ ($\delta = \sim 1$) (Luk and Leung, 2003) denotes the variation of fields along the axial direction inside the cylindrical resonator. These plots are very useful to predict the far-field radiation patterns and to design coupling schemes for various modes. To produce circular polarization, these equivalent dipoles should be considered in similar arrangements as described in section "Sources of Circularly Polarized Radiation."

DOI 10.1007/978-981-4560-75-7_36-1

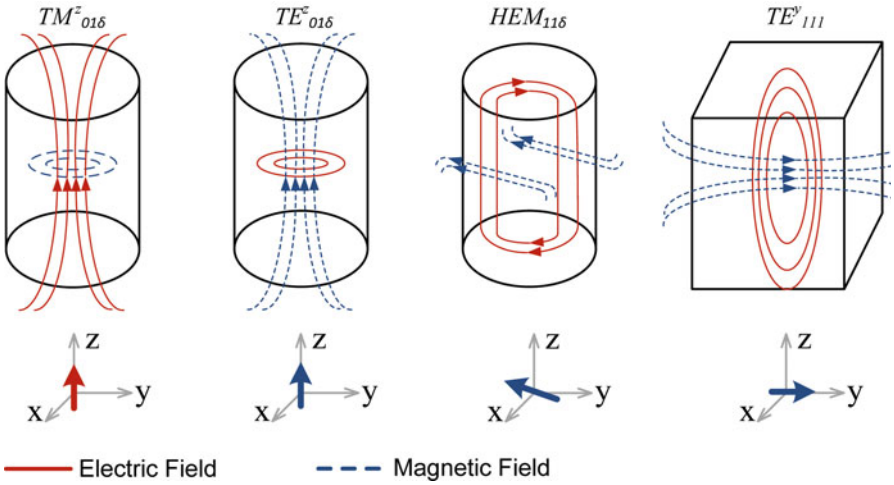


Fig. 16 Electromagnetic field distributions and equivalent dipoles of different modes in cylindrical and rectangular DRAs; δ ($\delta = \sim 1$) denotes the variation of fields along the axial direction inside the cylindrical resonator

An exact analytical closed-form solution does not exist for dielectric resonator geometries except hemispherical resonators, for which separation of variables is possible. For a dielectric resonator of arbitrary shape, general conclusions can be achieved to find approximate solutions. Most of these conclusions are based on the fundamental work of Van Bladel (Van Bladel 1975a, b). Numerical techniques such as finite element and finite difference time domain methods can produce more accurate electromagnetic solutions for these dielectric structures and are mostly used to achieve CP DRAs.

It is worth mentioning that achieving circular polarization usually requires special complicated geometries (e.g., cross shape) or minor changes in the resonator body which are difficult to realize due to the abrasive nature of ceramic, the number one material used for DRA applications. However, it has been recently investigated that DRAs can be fabricated by lithographic processes using ceramic-photoresist microcomposites (Rashidian et al. 2010, 2012). This fabrication technology can be very useful for CP DRAs in which special and complicated geometries may be required.

Similar to the microstrip case, circular polarization can be generated by three different ways, dual-feed system, single-feed system, and sequential rotation feed. For completeness, they are discussed again separately below, for the cases where a ground plane is used to support the antenna. Also, since the principle is the same for any volumetric resonator antenna, in the following the *dielectric resonator antenna* is simply referred to as *resonator antenna.*

Dual-Feed Approach

Dual-feed approach is the most straightforward method to realize CP radiations from inherently linearly polarized resonator antennas such as DRA. Two equal amplitude signals, 90° out of phase, are delivered in two different feed points on the resonator body to excite two orthogonal modes. The power divider splits the input signal into two outputs with equal amplitudes, and an extra quarter-wavelength transmission line is considered in only one of the outputs to generate a 90° phase difference. Alternatively, a directional coupler may be utilized to perform all these tasks.

Figure 17a shows a dual-feed system used to excite the two orthogonal $HEM_{11\delta}$ modes of a cylindrical ring dielectric resonator placed on a circular metallic ground plane (Mongia et al. 1994). A 3 dB quadrature coupler assembled in the backside was utilized to deliver equal amplitude and phase quadrature signals to two vertical probes. The probes are positioned in the proper places to excite the dominant hybrid mode ($HEM_{11\delta}$) in perpendicular directions. As shown in Fig. 16, the near-field distribution of this

DOI 10.1007/978-981-4560-75-7_36-1

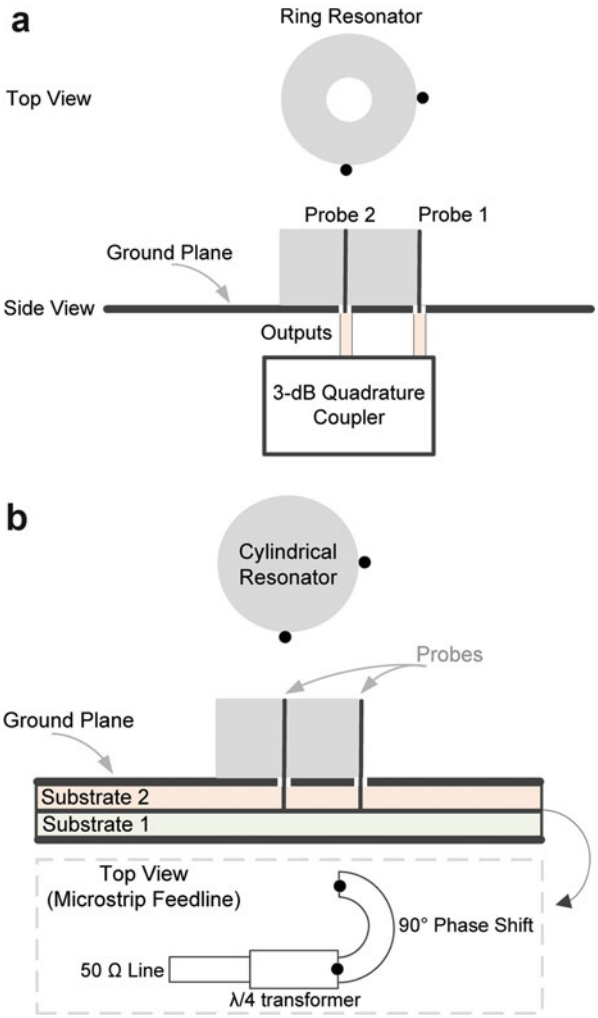


Fig. 17 (**a**) Configuration of a dual-feed ring CP DRA with an external coupler; (**b**) configuration of a dual-feed cylindrical CP DRA with microstrip feed line

type of mode is equivalent to a horizontal magnetic dipole, and since two of them are orthogonally excited with equal amplitudes and in-phase quadrature, CP radiations are resulted in the broadside direction. It should be noted that the 3 dB ellipticity bandwidth can be a large value and even more than the impedance bandwidth when a wideband directional coupler is used. In other words, the ellipticity bandwidth is proportional to the bandwidth of the external directional coupler, which provides in-phase quadrature signals of nearly equal amplitudes in a wide frequency band.

In Fig. 17b, the combination of microstrip line design and probe excitation is employed to achieve circular polarization from a cylindrical dielectric resonator antenna (Drossos et al. 1996). The 3 dB quadrature coupler is realized using a microstrip design. The two orthogonal $HEM_{11\delta}$ modes of the cylindrical resonator are excited by two vertical probes, which are connected to the microstrip design in the backside. A semicircle $\lambda/4$ microstrip line was designed to provide the 90° phase shift between the two excitation probes. Another $\lambda/4$ microstrip line was considered to transform the 50Ω input impedance to the 25Ω output impedance, which is because of the two 50Ω loads in parallel. Not a large ellipticity bandwidth is expected from this configuration due to the bandwidth limitation of the microstrip line design. To achieve a larger axial ratio bandwidth in the broadside direction, a microstrip 90° hybrid coupler can be designed to deliver the quadrature signals to the dual feeds.

DOI 10.1007/978-981-4560-75-7_36-1

Single-Feed Methods

Single-feed DRAs generally radiate linear polarization. To generate circular polarization in a single-feed configuration, three methods can be followed:

- Specific geometries such as cross shape or circular sector are employed, while the feed point position is adjusted so that two orthogonal modes with equal amplitudes and in-phase quadrature are excited.
- Regular shapes such as rectangles and circles that usually produce linear polarization in single-feed arrangements are slightly perturbed at proper locations with respect to the feed point. The perturbation configuration generates two orthogonal degenerate modes, one of which is detuned such that it is of the same magnitude but 90° out of phase with respect to the other mode.
- Feed mechanism is modified to excite two orthogonal modes and generate circular polarization. For instance, a cross-shaped aperture with unequal arms is used to initiate in-phase quadrature orthogonal modes.

The frequency distance between the resonant frequencies of two orthogonal modes must be adjusted so that the requirement of 90° phase difference between these modes can be properly fulfilled. To realize this condition, one can assume that at a particular frequency point between the resonant frequencies of these modes, the lower resonance is driven at its upper −3 dB point, and the upper resonance is driven at its lower −3 dB point. These provide +45°/−45° phase shift and therefore satisfy the 90° phase shift between modes in order to produce circular polarization. Therefore, if f_1 is the resonant frequency of the lower mode, f_2 is the resonant frequency of the upper mode, and Δf is the 3 dB bandwidth of the reflection coefficient of each resonant mode, the quadrature phase condition is fulfilled when

$$f_1 + \frac{\Delta f_1}{2} = f_2 - \frac{\Delta f_2}{2} \tag{60}$$

By choosing appropriate dimensions/geometry for the antenna, the above equation is satisfied, and a circular polarized antenna can be achieved in a single-feed configuration.

In the past several decades, numerous circularly polarized designs with a compact single feed have been reported. Microstrip line, probe, and slot are mostly used to feed these antennas. These designs are organized under different categories and briefly described in the following subsections.

Truncated (Perturbation) Shape Basic geometries such as rectangular and circular shapes with slight or no modifications in their structures can be utilized to generate circular polarization. Although narrow bandwidths for the axial ratio are achieved, the geometry is simple and compact due to single-feed excitation. Figure 18 shows the top view of several examples of this type of CP antennas fed by coaxial probe.

In Fig. 18a no modification is performed, and the coaxial feed is positioned at the corner of the rectangular resonator. In this case, if the ratio of the width to the length is slightly larger/smaller than one, two orthogonal modes of TE^{x}_{111} and TE^{y}_{111} with resonant frequencies in close vicinity are excited. As

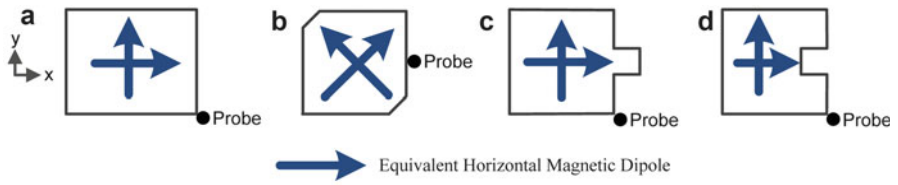


Fig. 18 Top view of probe-fed rectangular shape resonator antennas perturbed at certain locations to generate CP radiations

DOI 10.1007/978-981-4560-75-7_36-1

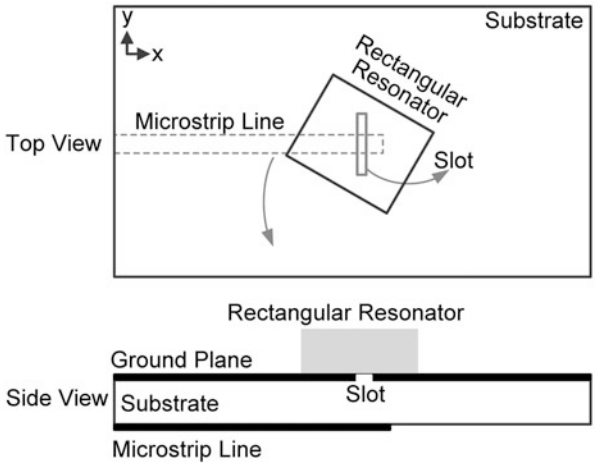


Fig. 19 Slot-fed CP rectangular dielectric resonator antenna

illustrated in Fig. 16, these modes are equivalent to magnetic dipoles in *x*- and *y*-directions, respectively. The aspect ratio bandwidth of the antenna is usually less than a few percent for the optimum design parameters (Malekabadi et al. 2008).

The rectangular-shaped antenna can be also excited using a single slot feed to produce circular polarization as shown in Fig. 19. A narrow nonresonant aperture in the ground plane side of a microstrip line is used to couple the electromagnetic energy to the rectangular-shaped resonator, which is inclined by around 45° with respect to the slot. The two orthogonal modes of TE^{x}_{111} and TE^{y}_{111} are simultaneously excited, and the length and the width of the dielectric resonator can be adjusted to optimize the resonant frequencies to achieve the required 90° phase difference between these modes for circular polarization. Generally, a near-square-shaped resonator provides the best result (Oliver et al. 1995).

The next structure, Fig. 18b, is a square resonator with two chamfered edges. In fact, by cutting two opposite corners of the dielectric resonator and locating the coaxial probe feed in the middle of one of the sidewalls, two near-degenerate orthogonal modes with similar magnitude and 90° phase difference can be excited to provide CP radiations. In this case, the equivalent horizontal magnetic dipoles of these modes are along the diagonal of the dielectric resonator. As the depth of the cut increases, the resonant frequency of the mode created along the chamfered edges increases, while the resonant frequency of the mode created along the unaltered corners stays constant. Therefore, the depth of the cut needs to be adjusted in order to achieve the maximum aspect ratio bandwidth.

Figure 18c and d shows the geometries of modified square resonators in which a rectangular perturbation segment and a rectangular notch are, respectively, introduced in the sidewall of the dielectric resonator. In these structures, the probe is located in the corner of the resonator. Two orthogonal resonant modes are excited along the *x*- and *y*-axes. In Fig. 18c, the first resonant belongs to TE^{x}_{111} as the rectangular perturbation in the *x*-direction creates a longer path for the mode in this direction. In Fig. 18d, the second resonant belongs to TE^{x}_{111} as the rectangular notch in the *x*-direction creates a shorter path for the mode in this direction. To achieve the maximum *AR* bandwidth, the rectangular perturbation/notch has to be optimized using numerical methods or simulation tools. Similar to the conventional rectangular-shaped antenna, all of these modified antennas present broadside radiation patterns. Similar perturbation ideas can be used for the circular-shaped antennas (Malekabadi et al. 2008).

Circular Sector Geometries Using a single feed with proper position, circular sector DRAs produce circular polarized radiations with axial ratio less than 3 dB over a wide frequency bandwidth. Figure 20 illustrates a plane view of the geometry of a circular sector antenna that is formed by removing a part (e.g.,

DOI 10.1007/978-981-4560-75-7_36-1

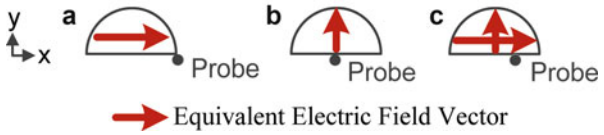


Fig. 20 Top view of circular sector dielectric resonator antennas excited by coaxial probes

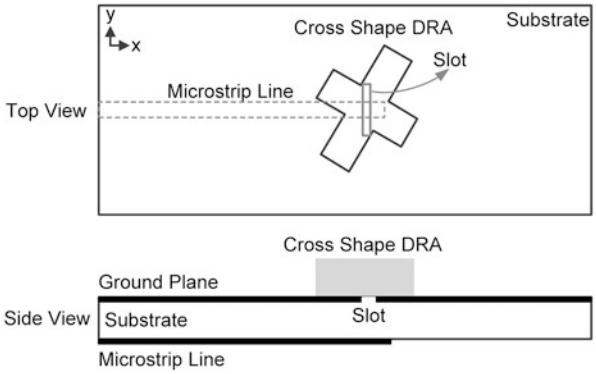


Fig. 21 Geometry of a cross-shaped dielectric resonator antenna excited by microstrip line slot

half) of a circular cylinder. A coaxial probe feed is used to excite the antenna. Two orthogonal modes with similar radiation patterns and 90° phase difference are required to achieve circular polarization from the circular sector antenna. As shown in Fig. 20a, when the feed is located in the corner of the geometry, the fundamental resonant mode of $TM_{11\delta}$ is excited. This mode is equivalent to an electric dipole with radiation polarization along the *x*-axis. Changing the feed position to the center of the geometry results in excitation of the next lowest-order mode of $TM_{21\delta}$ which is equivalent to an electric dipole with radiation polarization along the *y*-axis, as shown in Fig. 20b. These orthogonal dipoles are suitable to generate circular polarized radiation if they are excited simultaneously and are in-phase quadrature. To this end, a proper position for the feed point must be selected, and the antenna dimensions must be adjusted so that the resonant frequencies of these modes satisfy Eq. 60.

To achieve a wide axial ratio bandwidth, deviation from the 90° phase difference must be small over a wide range of frequency. Decreasing the permittivity of the antenna increases the individual bandwidths of the modes and minimizes the deviation over a wider frequency range. For example, an antenna with the radius of 18 mm, height of 15 mm, permittivity of 12, and excited using a probe located between the center and the corner of the geometry shows 3 dB axial ratio bandwidth of 10 % from 2.54 to 2.81 GHz in broadside direction (Tam and Murch 2000).

Cross Shape A cross-shaped dielectric resonator, which is composed of two crossing rectangles with different lengths, can generate CP radiations. Figure 21 shows the top and side view of this structure along with the feeding mechanisms. The cross-shaped dielectric resonator is excited using a single slot at the ground plane of a microstrip line. Due to a small change in lengths, the two crossing rectangular resonators support two orthogonal modes, which are equivalent to two magnetic dipoles operating in close vicinities. To produce maximum axial ratio bandwidth, the length of the crossing rectangles must be adjusted. Using a numerical optimization, over 5 % 3 dB axial ratio bandwidth was achieved when the ratio of the shorter length to the larger length was ∼0.6 (Esselle 1995).

Cross-Aperture Coupled Cross-aperture-coupled DRAs are capable of producing relatively wideband CP radiations. This method facilitates the fabrication procedure as a wide variety of geometries including

DOI 10.1007/978-981-4560-75-7_36-1

simple dielectric structures (e.g., rectangular or circular shapes) without any modification can be used in such an arrangement. Basically, it is relatively difficult to construct three-dimensional dielectric resonators with specific shapes or make small changes in the geometry due to the hardness of the dielectric material. Figure 22 shows different structures that can be examined for circular polarization. The microstrip feed line is printed on the substrate and the coupling cross-aperture of unequal arm lengths is centered on the ground plane of the substrate underneath the dielectric resonator. The cross-aperture is inclined by around 45° with respect to the microstrip feed line. Instead of changing or modifying the dielectric resonator structure, the arms of the cross-aperture are changed to excite two nearly degenerate orthogonal resonant modes with nearly equal amplitudes and 90° phase difference. These two modes are equivalent to orthogonal magnetic dipoles. Depending on which aperture arm is longer, left-hand or right-hand CP operation can be achieved. For instance, the fundamental mode of $HEM_{11\delta}$ in the circular-shaped

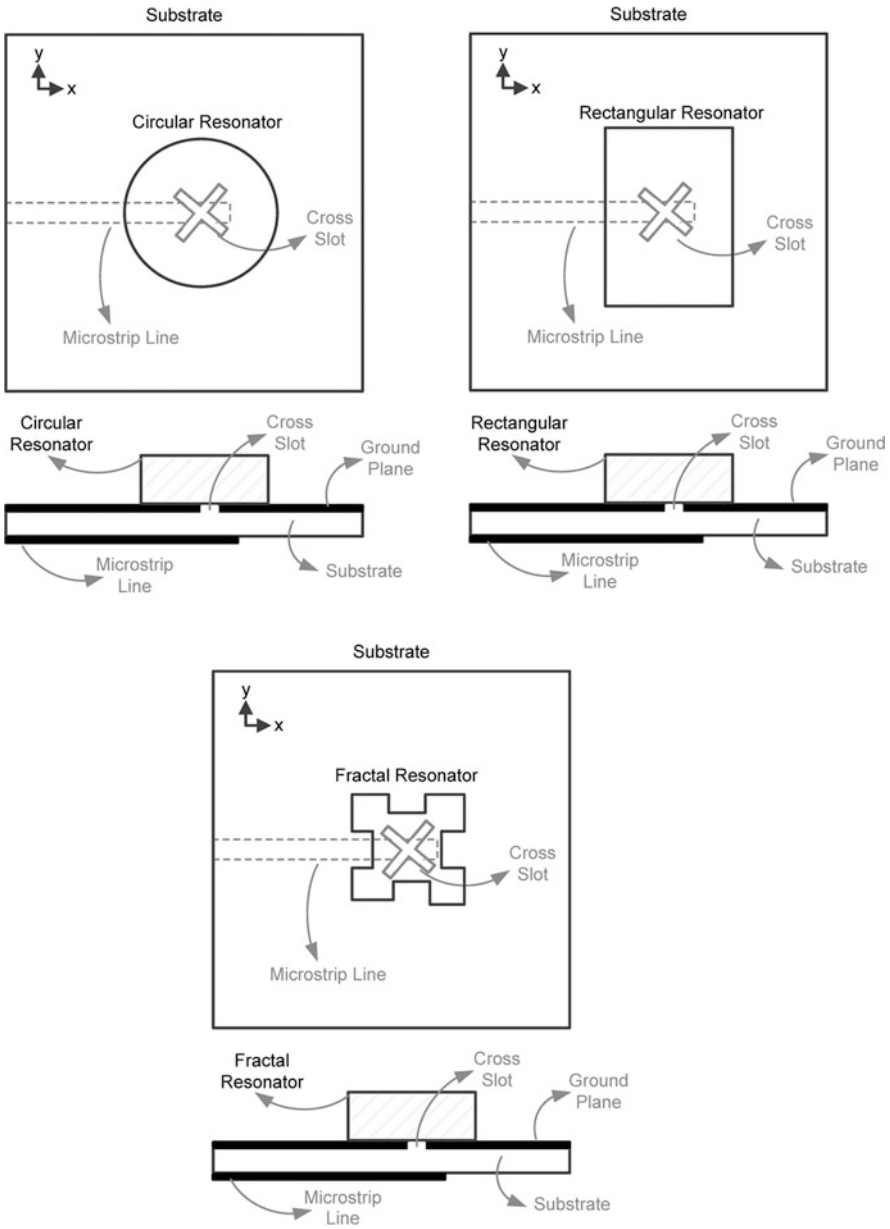


Fig. 22 Cross-aperture-coupled dielectric resonator antennas for circular polarization

DOI 10.1007/978-981-4560-75-7_36-1

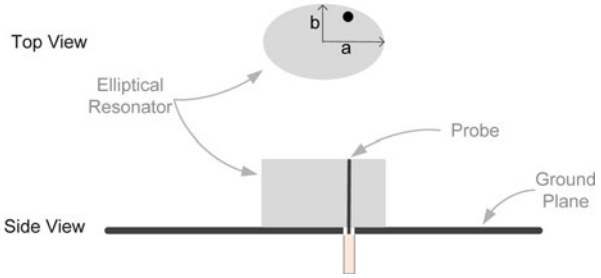


Fig. 23 Probe-fed elliptical dielectric resonator antenna

antenna, shown in Fig. 22, is split into two orthogonal modes and consequently two equivalent orthogonal magnetic dipoles in the slot directions. The resonant frequency of the mode in the direction of the shorter aperture is slightly lower than that of the other mode, and therefore, the right-hand circular polarization is obtained in this example (Huang and Yang 1999). More investigations can be found in (Dhar et al. 2013; Zhang et al. 2014).

Elliptical Shape Elliptical DRA with single feed can produce CP radiations. Figure 23 shows the antenna geometry. The dielectric resonator with elliptic cross section is placed above a finite ground plane and excited by a single coaxial probe feed. The ratio of the radius of the major axis a, to that of the minor axis b, is an important parameter to adjust the resonant frequencies of two orthogonal modes excited inside the resonator. The coaxial probe is located inside the resonator but near the surface and with angular positions around 40°–50° from the major axis. The length of the coaxial probe feed is adjusted to achieve wideband matching between the resonator and the 50Ω coaxial line. Depending on the resonator parameters such as height and permittivity, different optimized values can be achieved for a/b to maximize AR bandwidth. In most cases a/b should be around 1.5 to produce the optimum circular polarization (Kishk 2003; Malhat et al. 2012).

Sequential Rotation Technique

As described in section "Sources of Circularly Polarized Radiation," CP radiations can be achieved from displaced sequentially rotated electric dipoles. This technique can be easily applied for DRAs as they are equivalent to dipoles in their operational modes. Symmetrical geometries such as circular- or square-shaped resonators can also be used in this arrangement without any performance degradation. Since several antenna elements are used in this technique, higher antenna gain is achieved in comparison to the single-element CP antennas. One of the specific advantages of the orthogonal positioning of the adjacent elements in the sequential feeding technique is that the mutual coupling effects that disturb the array performance are considerably reduced. An example that uses aperture-fed cylindrical-shaped DRAs is shown in Fig. 24a. Rectangular slots are used to excite the dominant mode, which radiates like a linearly polarized horizontal magnetic dipole. The elements are employed in a 2 × 2 subarray design in which they are sequentially rotated with relative phases of 0°, 90°, 180°, and 270°.

Instead of linearly polarized, CP antenna elements can be also used in this feeding configuration to increase both the axial ratio bandwidth and the antenna gain (Huang 1986). For instance, the cross-aperture-coupled antenna elements, which generate CP radiations, are employed in a sequential feeding configuration (Pang et al. 2000). The feed network is similar to that shown in Fig. 24a. The only difference is that a cross-aperture of unequal length (e.g., $L_{long}/L_{short} = 1.22$), instead of a rectangular aperture, is used in excitation of the dielectric resonator elements to generate circular polarization. While the 3 dB axial ratio bandwidth is 5.6 % for the single element, it is over 16 % for the sequentially rotated 2 × 2

DOI 10.1007/978-981-4560-75-7_36-1

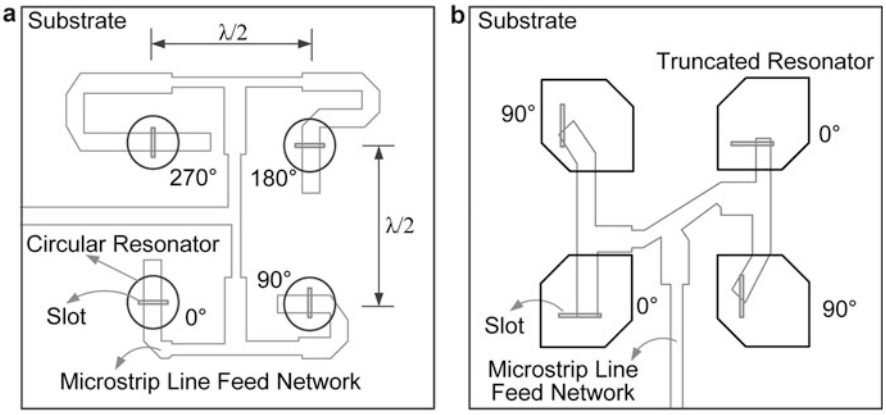


Fig. 24 Generation of circular polarization using sequential rotation technique from (**a**) linearly polarized aperture-coupled cylindrical DRA elements. (**b**) Truncated rectangular-shaped DRA elements

subarray system. Besides, the peak realized gain of the subarray antenna is around 12 dBi, which is 6 dB higher than that of the single CP antenna element.

Truncated rectangular-shaped DRAs result in narrow ellipticity bandwidths, usually not more than a few percent. By applying the sequential rotation technique, one can increase the 3 dB axial ratio bandwidth to the high value of 15 % and beyond (Haneishi and Takazawa 1985). Figure 24b shows the array configuration with microstrip slot-fed coupling. Each element is fed uniformly in amplitude and oriented orthogonally with respect to the neighboring element. The microstrip feed network on the rear plane of the substrate is designed to deliver the uniform power with 90° phase difference through the small apertures in the ground plane on the top of the substrate.

Design Example, Dual-Layer Microstrip Ring Antenna

This section provides a design example to show how the performance of a circularly polarized antenna can be improved. In section "Case III: Four Displaced Sequentially Rotated Electric Dipole Sources," Case IIIa, it was shown that four sequentially rotated currents provided a wide axial ratio beamwidth. Thus, as a design example, a square-ring microstrip antenna is selected, where currents flowing in the arms resemble Case IIIa. For simplicity, the single-feed case is selected, and corner truncation is used for circular polarization. Single layer patch antennas suffer from narrow impedance bandwidth. Although dual-fed patch antennas with external power divider provide wider axial ratio bandwidth, their bandwidth is limited by the impedance bandwidth of the patch. Stacked patches are commonly used to enhance the impedance bandwidth of patch antennas. By introducing appropriate perturbations to both the parasitic and driven patches, two orthogonal modes from each patch are generated. These four modes are combined, by adjusting the patch sizes and their separation, which eventually increase the impedance as well as the axial ratio bandwidths. A 3 dB axial ratio bandwidth of 8 % has been demonstrated in (Chung and Mohan 2003) with an impedance bandwidth equal to 43 % in the C-band.

Thus, a stacked square-ring antenna is selected to improve the bandwidth. What has not been shown so far is that with stacked configuration, one additional sequential rotation can be done, by rotating the upper-stacked patch 90° with respect to the lower patch. The performance of the antenna is provided for each case and discussed to show the reason and degree of improvement in the performance.

The square-ring patch is a miniaturized version of the patch antenna. In the ring patches the currents flow on the ring arms, which are similar to four displaced dipoles of section "Sources of Circularly Polarized Radiation." As it was shown in that section, the axial ratio beamwidth can be improved significantly by selecting the dipole separations of about 0.443λ. However, that requires the use of four

DOI 10.1007/978-981-4560-75-7_36-1

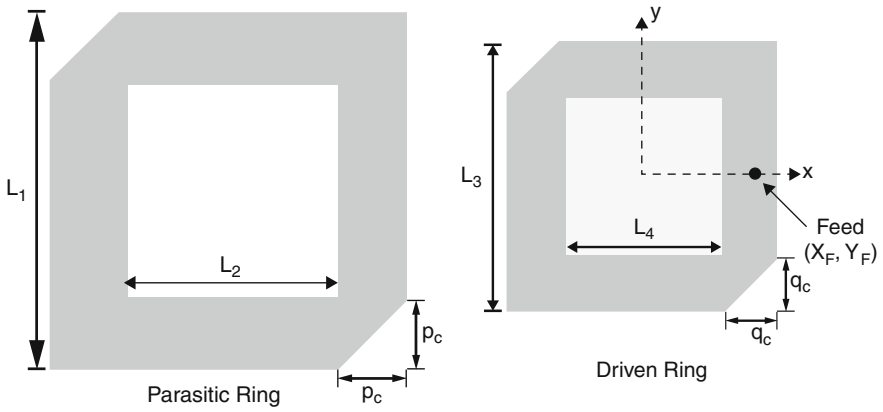


Fig. 25 Singly fed stacked ring patches. Both the driven and parasitic ring patches have perturbations in the form of corner truncation. The antenna parameters are tabulated in Table 1

feeds, one for each arm of the ring, to accomplish the ultrawide axial ratio beamwidth. With a single feed the axial ratio beamwidth is lower, but as it will be shown later, the problem can be remedied with stacked configuration. The axial ratio bandwidths can also be increased which are shown in (Latif and Shafai 2012, 2005), for various perturbation schemes to obtain large axial ratio bandwidths in stacked square-ring patch antennas. They are discussed briefly below. Because of the similarity of this square-ring configuration to that of four dipoles of section "Sources of Circularly Polarized Radiation," it is chosen here for more detailed study.

Stacked Ring Patches for CP by Negative Perturbation

The common form of perturbation, corner truncation, is applied to the diagonal corners of both the driven and parasitic ring patches, as shown in Fig. 25. Corner truncations in the driven ring are denoted by q_c and those in the parasitic ring are by p_c. Here, foam is used to separate these patches, and a grounded dielectric substrate with permittivity, ε_r of 3.2, and thickness (h_1) of 1.6 mm is used for the driven ring patch. The feed is located on the x-axis following the "corner truncation" configuration to obtain right-hand CP. The judicial selection of the ring sizes, the separation between parasitic and driven rings, and the perturbation amount give wide axial ratio and impedance bandwidths. The optimized antenna parameters are tabulated in Table 1. Because of the diagonal corner perturbations on both ring patches, two orthogonal modes are excited on each ring, and the two central adjacent modes are merged together due to the optimized parameters. Thus, three resonances can be noticed on the simulated reflection coefficient plot of Fig. 26a. The simulation has been carried out in Ansoft Ensemble 8.0. The impedance bandwidth is 160 MHz (10.1 %). Due to the perturbation in both rings, the axial ratio plot, shown in Fig. 26b, has two dips close to each other, providing significantly wider 3 dB axial ratio bandwidth to 72 MHz (4.6 %), compared to the singly-fed CP patch antenna. The antenna has right-hand CP (RHCP) gain over 7 dB within this band. The gain patterns at two principal planes of this antenna at 1.6 GHz are shown in Fig. 26c. It shows a boresight RHCP gain of 8.9 dB with cross-polarization level 15 dB below the co-polarization. However, the RHCP gain drops rapidly away from the boresight.

Stacked Ring Patches for CP by Positive Perturbation (Outward)

Positive perturbations, in the form of added metal strips in the diagonal corner of both rings to the outward direction, also provide CP radiation. The diagonal line is at 45° with respect to the feed point in this case, as shown in Fig. 27, for the RH sense of CP. The antenna parameters, optimized for a wide *AR* bandwidth, are indicated in Table 1, and the simulated reflection coefficient and axial ratio with frequency are shown

DOI 10.1007/978-981-4560-75-7_36-1

Table 1 CP stacked ring patch antenna, optimized antenna parameters, and simulated results

Perturbation type	Negative perturbation	Positive perturbation (outward)	Positive perturbation (inward)	Combination
Square ring Patch size (mm)	$L_1 = 59$ $L_3 = 47.3$	$L_1 = 59$ $L_3 = 44.7$	$L_1 = 59$ $L_3 = 47.3$	$L_1 = 59$ $L_3 = 44.7$
Slot size (mm)	$L_2 = 40$ $L_4 = 15.3$	$L_2 = 37.3$ $L_4 = 14.7$	$L_2 = 39$ $L_4 = 15.3$	$L_2 = 39.2$ $L_4 = 14.7$
Perturbation (mm)	$p_c = 6.5$ $q_c = 8.5$	$p_a = 4.5$ $q_a = 6$	$p_i = 5$ $q_i = 6$	$p_c = 5$ $q_a = 6$
$\|S_{11}\| < -10$ dB BW	1,510–1,670 MHz (160 MHz, 10.1 %)	1,510–1,683 MHz (173 MHz, 10.8 %)	1,509–1,686 MHz (177 MHz, 11.1 %)	1,517–1,669 MHz (152 MHz, 9.5 %)
3 dB AR BW	1,541–1,613 MHz (72 MHz, 4.6 %)	1,545–1,624 MHz (79 MHz, 5 %)	1,549–1,627 MHz (78 MHz, 4.9 %)	1,547–1,624 MHz (77 MHz, 4.9 %)
7 dBic RHCP gain	1,505–1,677 MHz (172 MHz)	1,505–1,692 MHz (187 MHz)	1,513–1,699 MHz (186 MHz)	1,515–1,678 MHz (163 MHz)

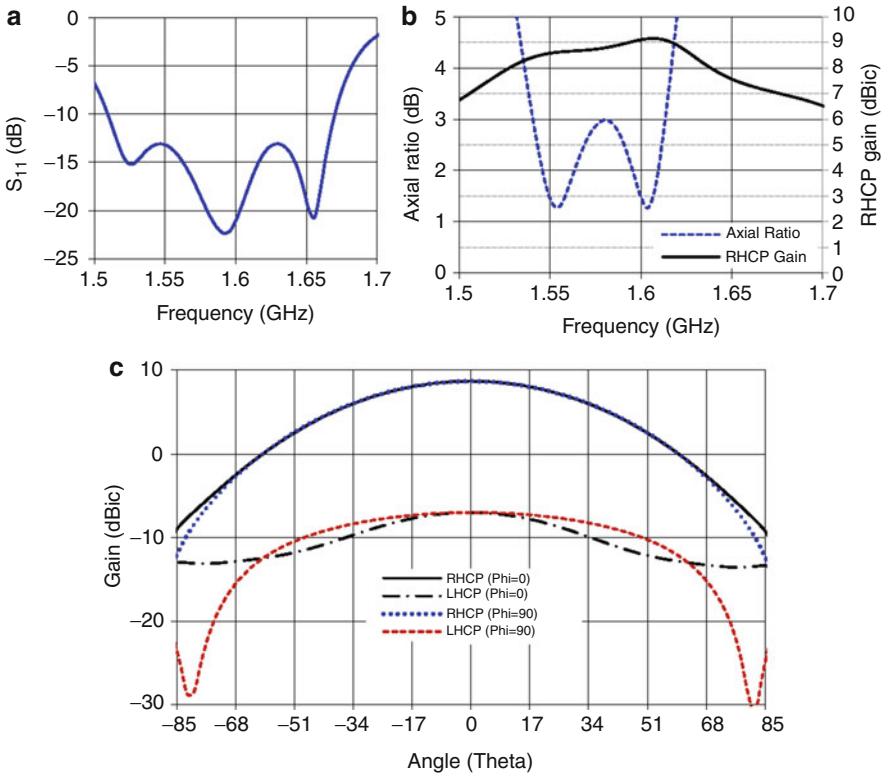


Fig. 26 Simulated (**a**) reflection coefficient and (**b**) axial ratio and RHCP versus frequency plots and (**c**) gain patterns at 1.6 GHz of the stacked ring patch antenna in Fig. 25

in Fig. 28a. The $|S_{11}| < -10$ dB bandwidth is 173 MHz (10.8 %), and 3 dB *AR* bandwidth is 79 MHz (5 %), as shown in Fig. 28b, which are slightly larger than those of the previous case. Because of the added stubs, the overall size of the antenna increases by 25.4 %, with reference to the negatively perturbed case, which is the reason for wider bandwidths. It also slightly increases the RHCP gain with peak value of 9.2 dB at 1.62 GHz.

DOI 10.1007/978-981-4560-75-7_36-1

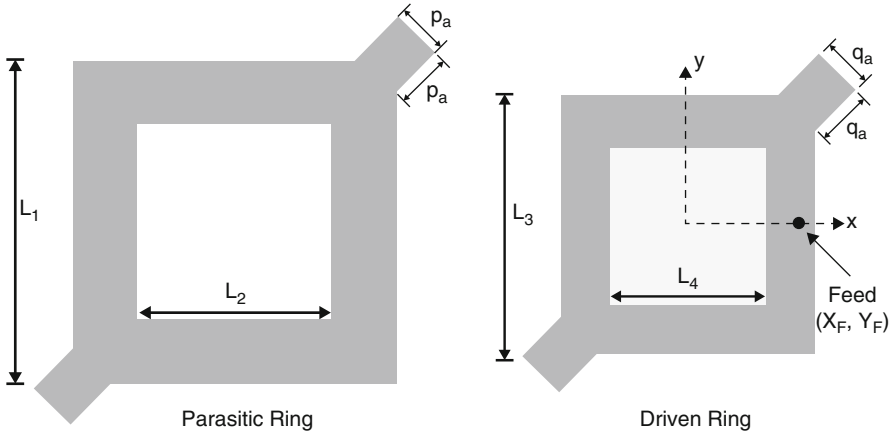


Fig. 27 Singly-fed stacked ring patches. Both the driven and parasitic ring patches have perturbations in the form of added stubs along the diagonal corners. The antenna parameters are tabulated in Table 1

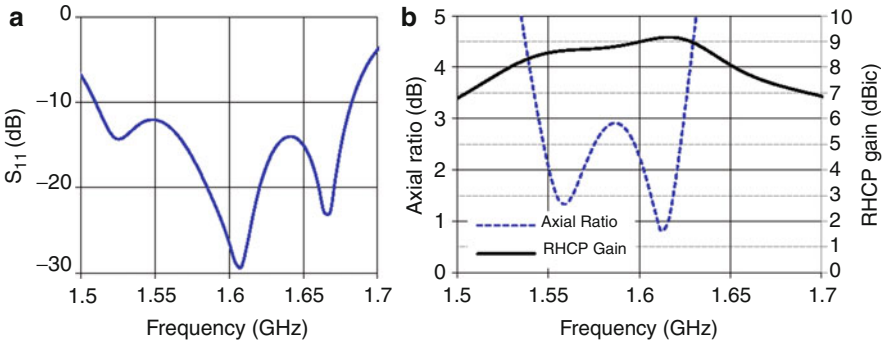


Fig. 28 Simulated (**a**) reflection coefficient and (**b**) axial ratio and RHCP versus frequency plots of the stacked ring patch antenna in Fig. 27

Stacked Ring Patches for CP by Positive Perturbation (Inward)

Ring patches have an added advantage of introducing positive perturbation by adding metal stubs to the inward direction of the diagonal corners of the square rings in the slot region, as shown in Fig. 29. Table 1 summarizes the optimized antenna parameters in this case, along with the obtained impedance and *AR* bandwidths. The simulated reflection coefficient, axial ratio, and RHCP gain plots are shown in Fig. 30. Both types of positive perturbations give similar results; however, in the inward positive perturbation case, the overall size of the antenna is not increasing.

Stacked Ring Patch for CP using Hybrid Perturbations

A similar perturbation type (either positive or negative) is commonly used to achieve CP from microstrip patch antennas. A combination of positive and negative perturbations or hybrid perturbations can also be used to generate CP radiation from the stacked patch antenna (Latif and Shafai 2007). Here, positive perturbation is introduced to the diagonal corners of the driven ring, and negative perturbation is applied to the parasitic ring. This configuration is essentially equivalent to the concept of sequential rotation used in arrays, where elements are physically rotated for correcting the axial ratio (Teshirogi et al. 1985). In other words, the stacked configuration introduces an additional sequential rotation, by rotating the upper-stacked patch 90° with respect to the lower patch. The stacked square-ring antennas can be considered as a two-element array. Instead of physically rotating the elements for phase compensation, the phase

DOI 10.1007/978-981-4560-75-7_36-1

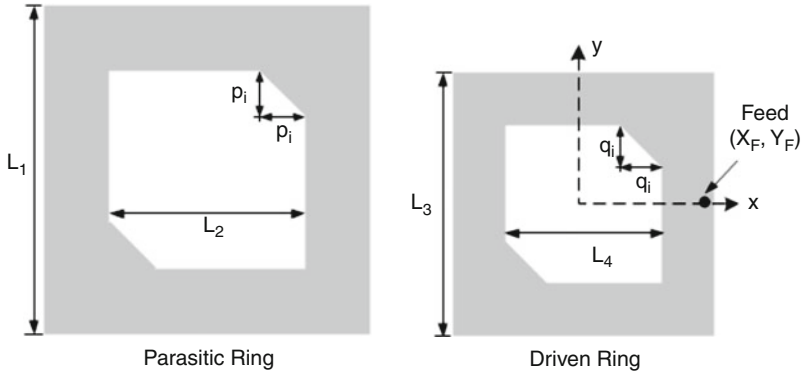


Fig. 29 Singly-fed stacked ring patches. Both the driven and parasitic ring patches have perturbations in the form of added stubs along the diagonal corners in the slot. The antenna parameters are tabulated in Table 1

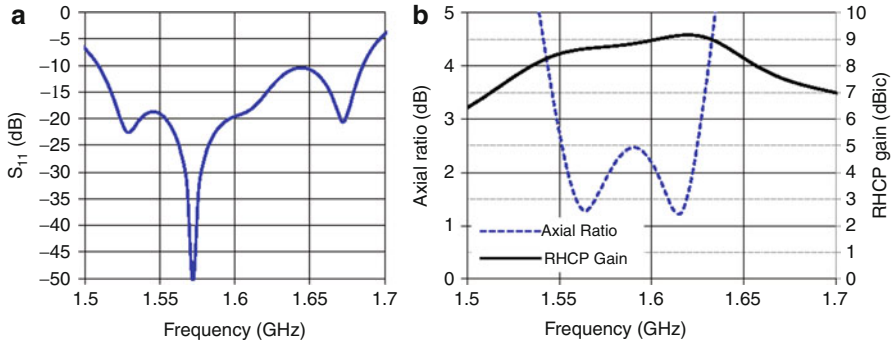


Fig. 30 Simulated (**a**) reflection coefficient and (**b**) axial ratio and RHCP versus frequency plots of the stacked ring patch antenna in Fig. 29

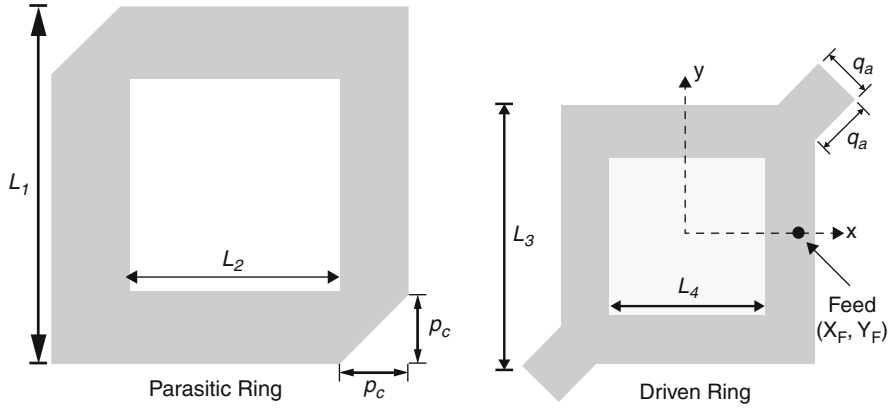


Fig. 31 Singly-fed stacked ring patches. The driven and parasitic patches have positive and negative perturbations, respectively, along the diagonal corners in the slot. The antenna parameters are tabulated in Table 1

correction is obtained here geometrically. The driven ring is smaller in size than the parasitic ring due to the higher permittivity of the related substrate, and metal stubs are added to diagonal corners of the driven ring for generating RHCP, without increasing the antenna size. The perturbations are located with respect to the feed point for "corner truncation" configuration, as shown in Fig. 31. If negative perturbations are introduced along the same diagonal corners of the parasitic ring, this will generate left-hand CP radiation,

DOI 10.1007/978-981-4560-75-7_36-1

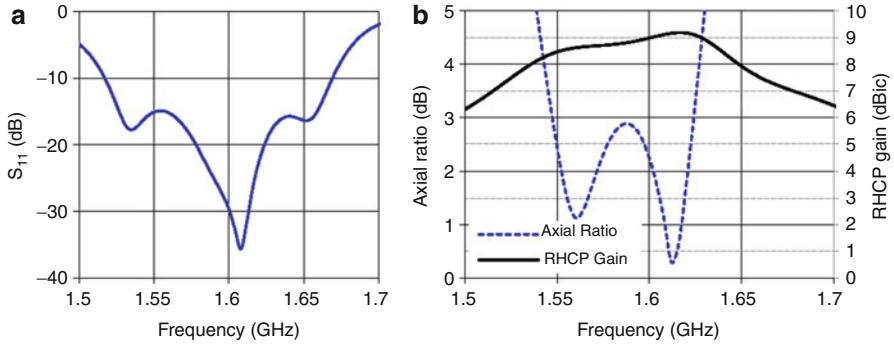


Fig. 32 Simulated (**a**) reflection coefficient and (**b**) axial ratio and RHCP versus frequency plots of the stacked ring patch antenna in Fig. 31

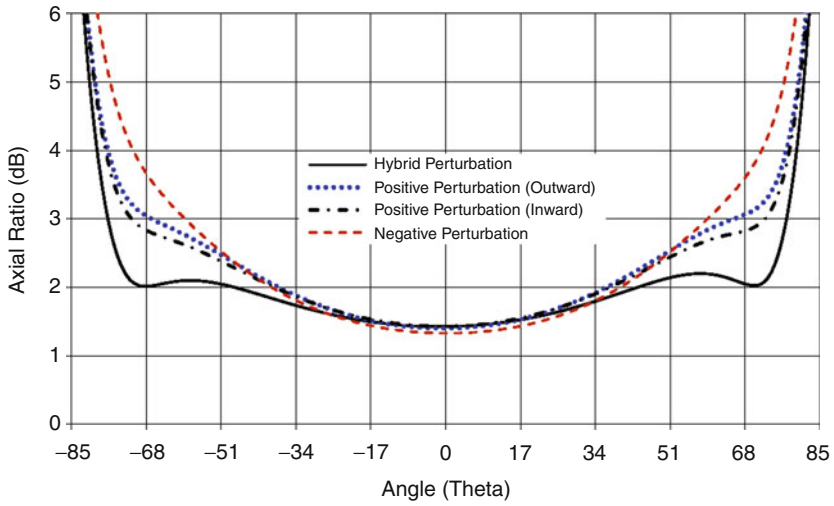


Fig. 33 Axial ratio versus elevation angle plots, in the $\phi = 0°$ plane, for the stacked square-ring antenna with dual-negative corner perturbations (Fig. 25), dual-positive corner perturbations [outward] (Fig. 27), dual-positive corner perturbations [inward] (Fig. 29), and hybrid perturbations (Fig. 31), at 1.56 GHz

and eventually the polarization of the antenna will be linear. However, if the sequential rotation can be enforced, not by physically rotating the parasitic ring, rather by introducing the cuts along the other diagonal corners, as shown in Fig. 31, this will correct the phase relationship between the two waves generated from the two rings and ensure the RHCP. Moreover, cutting corners from the bigger parasitic ring reduces the overall antenna size. The antenna shows wide impedance bandwidth of 152 MHz (9.5 %), as well as broad 3 dB *AR* bandwidth of 77 MHz (4.9 %), as observed in Fig. 32. The antenna also demonstrates RHCP gain over 7 dB in the entire bandwidth. The optimized dimensions of these antennas and a summary of these results are tabulated in Table 1.

Microstrip antennas typically have a narrow CP beamwidth that is restricted around the antenna boresight direction, limiting its use in applications where wide angular coverage is required, such as the Global Positioning System. This hybrid perturbation technique provides a wider angular coverage in the upper hemisphere, similar to a sequentially rotated array (Teshirogi et al. 1985). When compared to different perturbation schemes, the hybrid perturbation scheme provides the largest angular coverage, as illustrated in Fig. 33. The antenna with negative corner perturbations has a 3 dB axial ratio beamwidth of only 118°, in the $\phi = 0°$ plane. However, the antenna with the hybrid perturbations has a wider angular

DOI 10.1007/978-981-4560-75-7_36-1

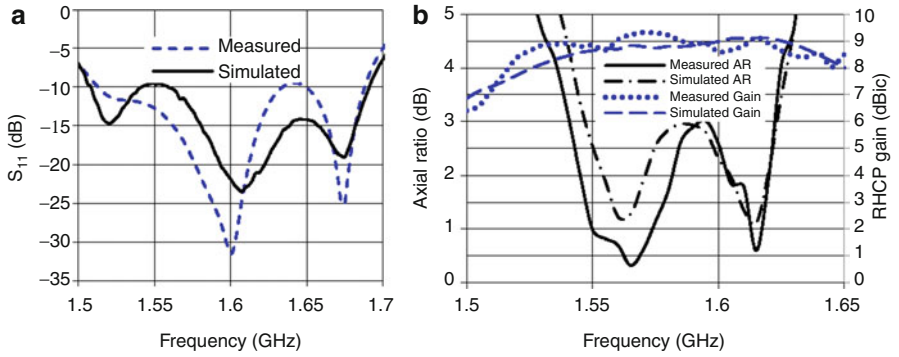


Fig. 34 Comparison between simulated and measured (**a**) reflection coefficient and (**b**) axial ratio and RHCP gain of the stacked ring antenna, with positive perturbation (shown in Fig. 27). The antenna parameters are $\varepsilon_r = 3.2$, $h_1 = 1.6$ mm, $h_2 = 11$ mm, $L_1 = 59$ mm, $L_2 = 37$ mm, $L_3 = 44.5$ mm, $L_4 = 14.5$ mm, $p_a = 5$ mm, $q_a = 6$ mm

coverage with 3 dB axial ratio beamwidth of 154°, from −77° to +77° off the boresight. In the other principal plane, all perturbation types have similar performances.

These results are confirmed by experimental verifications. First, the stacked square-ring antenna with outward positive perturbations, shown in Fig. 27, was fabricated where the driven ring was excited at $X_F = 21$ mm, $Y_F = 0$ mm, using a 50Ω SMA probe. The sizes of the main dielectric substrate ($\varepsilon_r = 3.2$; thickness = 1.6 mm) and the foam, which is the substrate for the parasitic ring, were chosen to be the same, and the rings were placed at the center of the respective substrates for easy alignment. A finite ground plane (160×160 mm^2) was used for the antenna. The fabricated antenna was tested in the Antenna Laboratory at the University of Manitoba. The reflection coefficient was measured by ANRITSU ME7808A Network Analyzer and compared in Fig. 34a with the simulation results. The measured S_{11} is slightly different from simulation, which is due to the difference in dimensions in the handmade antenna. However, the variations are below −10 dB, making the corresponding bandwidths about the same.

The measured and simulated axial ratio and RHCP gains are compared in Fig. 34b. The measured 3 dB axial ratio bandwidth is larger (81 MHz) than the simulated one (76 MHz), and it also has lower minima, as low as 0.32 dB at 1.565 GHz. In other words, the fabrication tolerances have improved the axial ratio performance. The measured RHCP gains are somewhat higher at low frequencies, as seen in the same figure. There are small oscillations in the measured gain, which are due to interactions between the antenna and its support. A sample of the measured radiation patterns, using a rotating linear source, is shown in Fig. 35. The boresight CP performance is quite good, but deteriorates gradually with increasing angle, which is a common characteristic of CP antennas.

Secondly, the stacked square-ring antenna with hybrid perturbation, as shown in Fig. 31, was fabricated and tested, and the comparison with the simulated ones is presented in Fig. 36, which shows good agreement as the first example. Measurement of radiation patterns also confirms the wide angular coverage from this antenna in the $\phi = 0°$ plane, as shown in Fig. 37.

Effects of Finite Ground Plane on CP Performance of Stacked Square-Ring Antenna

Ground Plane Size The effect of the finite ground plane size on the CP performance of the stacked square-ring antenna is also studied. Figure 38 shows the variation of the axial ratio of the antenna with the outward positive perturbation, shown in Fig. 27, versus frequency for different ground plane sizes. Here, the antenna is symmetrically placed at the center of the ground plane. The axial ratio deteriorates as the ground plane size decreases. Between the two dips in the axial ratio plot, caused by the perturbations on both rings, the second dip suffers the most, which moves to higher frequencies with a poor axial ratio. This

DOI 10.1007/978-981-4560-75-7_36-1

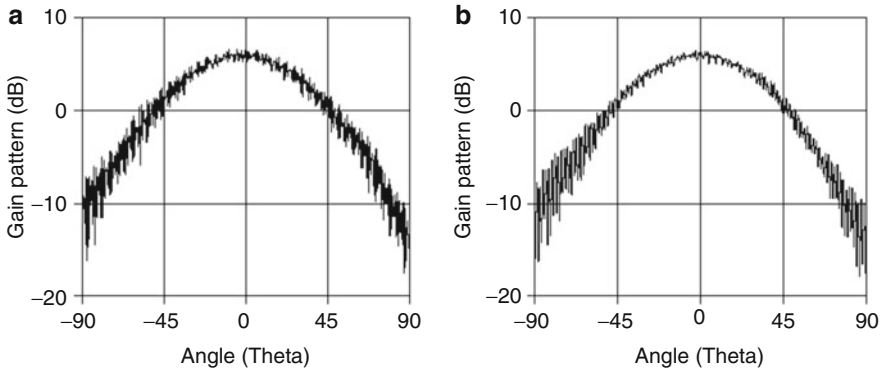


Fig. 35 The measured gain patterns in (**a**) $\phi = 0°$ and (**b**) $\phi = 90°$ planes at 1.56 GHz of the DLSRA antenna, with positive perturbation on both driven and parasitic rings (shown in Fig. 27)

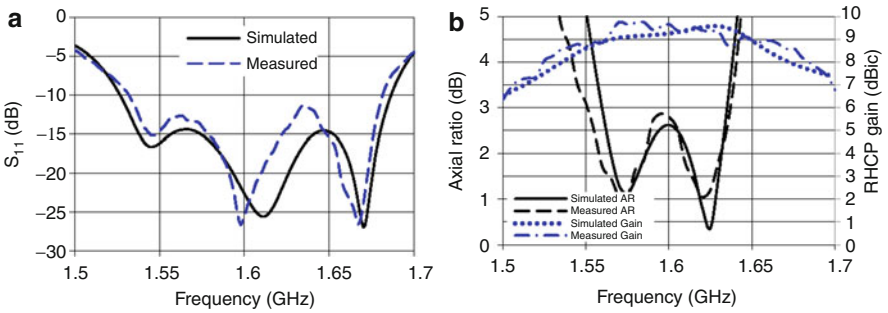


Fig. 36 Comparison between (**a**) simulated and measured reflection coefficient and (**b**) axial ratio and RHCP gain of the stacked square-ring antenna, with hybrid perturbations (shown in Fig. 31). The antenna parameters are $\varepsilon_r = 3.2$, $h_1 = 1.6$ mm, $h_2 = 11$ mm, $L_1 = 59$ mm, $L_2 = 37$ mm, $L_3 = 44.5$ mm, $L_4 = 14.5$ mm, $p_a = 5$ mm, $q_a = 6$ mm

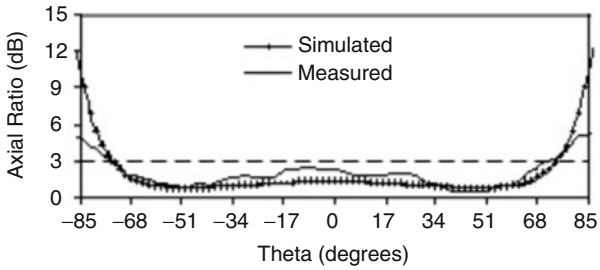


Fig. 37 Measured and simulated axial ratios with elevation angle, in the $\phi = 0°$ plane, of the stacked square-ring antenna with hybrid perturbations in Fig. 31

axial ratio can be improved by adjusting the size of the driven ring that causes this dip and the foam thickness as well. It is also increased to achieve lower values of the axial ratio. The parasitic ring size and perturbations are kept fixed in this case. The simulation results are tabulated in Table 2 for the optimized cases for different ground plane sizes. Figure 39a shows the variation of the axial ratio, and Fig. 39b shows the variations in RHCP gain for different ground plane sizes. As the ground plane size decreases, the RHCP gain decreases. For a ground plane size of 80 × 80 mm^2, the peak RHCP gain is 7.89 dBic, whereas it is 9.2 dBic when the ground plane size is 160 × 160 mm^2.

DOI 10.1007/978-981-4560-75-7_36-1

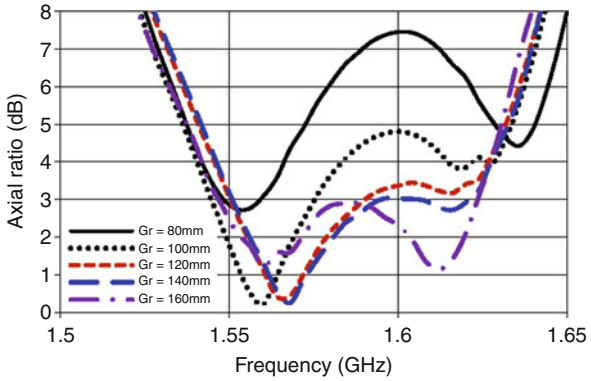


Fig. 38 The variation of axial ratio with frequency, due to different ground plane sizes for the staked square-ring antenna in Fig. 27. Antenna parameters (in mm) are $L_1 = 59$, $L_2 = 37$, $L_3 = 44.5$, $L_4 = 14.5$, $P_a = 5$, $q_a = 6$, $\varepsilon_r = 3.2$, $h_1 = 1.6$, $h_2 = 11$

Table 2 Positively perturbed stacked square-ring antenna with finite ground plane, optimized antenna parameters, and simulated results. Other antenna parameters: $L_1 = 59$ mm, $L_2 = 37$ mm, $p_a = 5$ mm, $q_a = 6$ mm

Driven ring size (mm)	Ground plane (mm^2)	Foam height (mm)	|S11|<−10 dB BW (MHz)	3 dB AR BW (MHz)	7 dBic RHCP gain (MHz)
$L_3 = 44.5$ $L_4 = 14.5$	160 × 160 (0.84 λ_o)	11	1,516–1,688 (172 MHz)	1,547–1,624 (77 MHz)	1,502–1,690 (188 MHz)
$L_3 = 44.7$ $L_4 = 14.7$	140 × 140 (0.74 λ_o)	11.2	1,511–1,682 (171 MHz)	1,545–1,624 (79 MHz)	1,506–1,674 (168 MHz)
$L_3 = 44.8$ $L_4 = 14.8$	120 × 120 (0.63 λ_o)	11.5	1,508–1,679 (171 MHz)	1,543–1,623 (80 MHz)	1,510–1,664 (154 MHz)
$L_3 = 44.9$ $L_4 = 14.9$	100 × 100 (0.53 λ_o)	12.8	1,509–1,673 (164 MHz)	1,538–1,620 (82 MHz)	1,512–1,653 (141 MHz)
$L_3 = 45$ $L_4 = 15$	80 × 80 (0.42 λ_o)	16.7	1,502–1,667 (165 MHz)	1,537–1,618 (81 MHz)	1,520–1,642 (122 MHz)

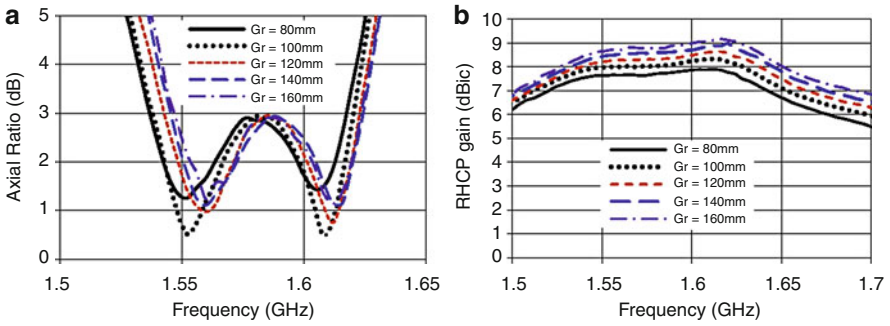


Fig. 39 The effects of finite ground plane sizes on the (**a**) axial ratio and (**b**) RHCP gain of the stacked square-ring antenna in Fig. 27. Antenna parameters are mentioned in Table 2

Asymmetric Antenna Position on the Ground Plane In the study of finite ground plane size discussed above, a symmetric antenna position on the ground plane was considered. In order to see the effects of asymmetric locations, the circularly polarized stacked square-ring antenna with positive perturbations is positioned asymmetrically on the ground plane as shown in Fig. 40a. Considering O as the global origin which is the center of the ground plane, and C(X_g, Y_g) as the center of stacked rings, when the antenna on

DOI 10.1007/978-981-4560-75-7_36-1

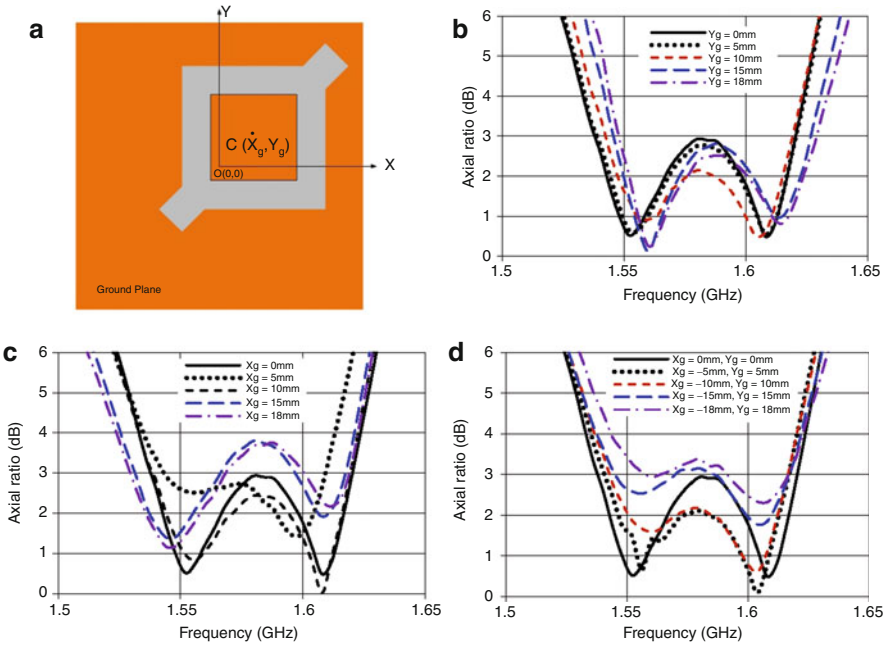


Fig. 40 The effects of asymmetric antenna position on the ground plane on the axial ratio of the stacked square-ring antenna with positive perturbations in Fig. 27. Antenna parameters (in mm) are $L_1 = 59$, $L_2 = 37$, $L_3 = 44.9$, $L_4 = 14.9$, $P_a = 5$, $q_a = 6$, $\varepsilon_r = 3.2$, $h_1 = 1.6$, $h_2 = 12.8$. Ground plane size $= 100 \times 100$ mm^2. (**a**) Geometry, (**b**) when the antenna is asymmetric about *y*-axis, with $X_g = 0$, (**c**) when asymmetric about *x*-axis, and with $Y_g = 0$, (**d**) when asymmetric about both *x*- and *y*-axes

the ground plane is asymmetric about the *y*-axis (with $X_g = 0$ mm), the axial ratio performance does not deteriorate much, as can be seen in Fig. 40b. Here, $X_g = 0$ mm, $Y_g = 0$ mm indicates the symmetric case, and $X_g = 0$ mm, $Y_g = 18$ mm means the stacked ring antenna is at one side of the ground plane. When the antenna is asymmetric about the *x*-axis (keeping $Y_g = 0$ mm), the axial ratio performance degrades significantly, as shown in Fig. 40c. Similarly, when the antenna is asymmetrically positioned about both *x*- and *y*-axes, the axial ratio suffers significantly, as shown in Fig. 40d.

Conclusion

Circularly polarized antennas, categorized into their current sources of radiation, were well elucidated in this chapter. Mathematical expressions and important definitions were first reviewed. The condition of circularly polarized radiation was analytically stated for one-dimensional current sources in the form of crossed electric dipole antennas. It was later extended to four sequentially rotated dipoles, mathematically proving the formation of circularly polarized radiation with beamwidths as wide as 170°. Antennas with two- and three-dimensional current sources were presented. In particular, microstrip patch antennas and dielectric resonator antennas were thoroughly discussed, corresponding to the aforementioned current sources. Different methods for generating circularly polarized waves were given for each of the antennas. More specifically, circularly polarized radiation properties of square-ring microstrip patch antenna were extensively investigated, and the corresponding results were presented for different antenna parameters. The square-ring microstrip antenna was selected as it closely approximates the sequentially rotated currents, and also it has not been widely studied in the literature. The idea presented in this chapter can be applied to other antenna structures for different applications. In summary, circularly polarized antennas

DOI 10.1007/978-981-4560-75-7_36-1

with their attractive features are vital radiating components in most modern satellite communications, radars, and Global Positioning Systems.

References

Brookner E, Hall WM, Westlake RH (1985) Faraday loss for L-band radar and communications systems. IEEE Trans Aerosp Electron Syst 21:459–469

Chung KL, Mohan AS (2003) A systematic design method to obtain broadband characteristics for singly-fed electromagnetically coupled patch antennas for circular polarization. IEEE Trans Antennas Propag 51:3239–3248

Counselman CC (1999) Multipath-rejecting GPS antennas. Proc IEEE 87:86–91

Dhar S, Ghatak R, Gupta B, Poddar DR (2013) Circularly polarized minkowski fractal dielectric resonator antenna. In: URSI international symposium on electromagnetic theory, Japan

Drossos G, Wu Z, Davis LE (1996) Circular polarised cylindrical dielectric resonator antenna. Electron Lett 32:281–283

Esselle KP (1995) Circularly polarized dielectric resonator antenna: analysis of near and far fields using FD-TD. In: URSI symposium digest, Newport Beach

Garg R, Bhartia P, Bahl IJ, Ittipibon A (2001) Microstrip antenna design handbook. Artech House, Boston

Hall PS, James JR (1989) Handbook of microstrip antennas. Peregrinus, London

Haneishi M, Takazawa H (1985) Broadband circularly polarized planar array composed of a pair of dielectric resonator antennas. Electron Lett 21:437–438

Hollis JS, Lyons TJ, Clayton L (1969) Microwave antenna measurements. Scientific Atlanta, Atlanta

Huang CY, Yang CF (1999) Cross-aperture coupled circularly polarized dielectric resonator antenna. In: Proceeding of IEEE international symposium on antennas and propagation, Orlando

Huang J (1986) A technique for an array to generate circular polarization with linearly polarized elements. IEEE Trans Antennas Propag 34:1113–1124

IEEE Standard Test Procedures for Antennas (1979) IEEE standard 149

Kishk A (2003) An elliptical dielectric resonator antenna designed for circular polarization with single feed. Microwave Opt Technol Lett 37:454–456

Kraus JD (1988) Antennas. McGraw-Hill, New York

Langston WL, Jackson DR (2004) Impedance, axial-ratio, and receive-power bandwidths of microstrip antennas. IEEE Trans Antennas Propag 52:2769–2773

Latif S, Shafai L (2005) Dual-layer square-ring antenna (DLSRA) for circular polarization. In: Proceeding of IEEE international symposium on antennas propagation society and USNC/URSI national radio science meeting (APS/URSI), Washington, DC, pp 525–528

Latif S, Shafai L (2007) Hybrid perturbation scheme for wide angle circular polarisation of stacked square-ring microstrip antennas. Electron Lett 43:1065–1066

Latif S, Shafai L (2012) Circular polarization from dual-layer square-ring antennas. IET J Microwaves Antennas Propag 6:1–9

Luk KM, Leung KW (2003) Dielectric resonator antennas. Research Studies Press, Baldock

Malekabadi SA, Neshati MH, Rashed-Mohassel J (2008) Circular polarized dielectric resonator antennas using a single probe feed. Prog Electromagn Res C 3:81–94

Malhat HA, Zainud-Deen SH, Awadalla KH (2012) Circular polarized dielectric resonator antenna for portable RFID reader using a single feed. Int J Eng Bus Manag 4

Milligan TA (2005) Modern antenna design, 2nd edn. Wiley, Hoboken

Mongia RK, Ittipiboon A, Cuhaci M, Roscoe D (1994) Circularly polarised dielectric resonator antenna. Electron Lett 30:1361–1362

Oliver MB, Antar YMM, Mongia RK, Ittipiboon A (1995) Circularly polarized rectangular dielectric resonator antenna. Electron Lett 31:418–419

Pang KK, Lo HY, Leung KW, Luk KM, Yung EKN (2000) Circularly polarized dielectric resonator antenna subarrays. Microwave Opt Technol Lett 27:377–379

Rashidian A, Klymyshyn DM, Boerner M, Mohr J (2010) Deep x-ray lithography processing for batch fabrication of thick polymer-based antenna structures. J Micromech Microeng 20:25026–25036

Rashidian A, Klymyshyn DM, Tayfeh Aligodarz M, Boerner M, Mohr J (2012) Microwave performance of photoresist-alumina microcomposites for batch fabrication of thick polymer-based dielectric structures. J Micromech Microeng 22:1–9

Stutzman WL (1993) Polarization in electromagnetic systems. Artech House, Norwood

Tam MTK, Murch RD (2000) Circularly polarized circular sector dielectric resonator antenna. IEEE Trans Antennas Propag 48:126–128

Teshirogi T, Tanaka M, Chujo W (1985) Wideband circularly polarized array antenna with sequential rotations and phase shift of elements. In: Proceeding of ISAP, pp 117–120

Van Bladel J (1975a) On the resonances of a dielectric resonator of very high permittivity. IEEE Trans Microwave Theory Technol 23:199–208

Van Bladel J (1975b) The excitation of dielectric resonators of very high permittivity. IEEE Trans Microwave Theory Technol 23:208–215

Zhang M, Li B, Lv X (2014) Cross-slot-coupled wide dual-band circularly polarized rectangular dielectric resonator antenna. IEEE Antennas Wirel Propag Lett 13:532–535

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_37-1

Phased Arrays

Takashi Maruyama*, Kazunari Kihira and Hiroaki Miyashita
Antennas Technology Department, Information Technology R&D Center, Mitsubishi Electric Corporation, Kamakura, Kanagawa, Japan

Abstract

This chapter explains the fundamental issues in designing phased array antennas (PAAs). At first, the structure and functions of the PAAs are introduced. Pattern synthesis, which is a feature of array antennas, and array calibration methods are explained. It includes some methods, minimax algorithm, element thinning algorithm, and genetic algorithm. For array calibration, rotating element electric field vector method is also explained. Digital beam forming (DBF) including signal processing and MIMO (multiple input multiple output) techniques are shown in a later part of this chapter.

Keywords

Phased array antennas; Antenna pattern synthesis; Array antenna calibration; Rotating element electric field vector method; Digital beam forming

Introduction

Antennas are defined as devices that convert energy between electromagnetic waves and electric currents in circuits. To increase an antenna's conversion efficiency, it is necessary to ensure the antenna has directivity and align its maximum directivity toward the direction of arrival (or direction of radiation) of the desired radio waves. Such pattern control is performed in aperture antennas and linear antennas by mechanically changing the antenna direction. Phased array antennas (PAAs), which arrange multiple radiating elements on either flat or curved surfaces, have greatly changed this antenna pattern control method. These antennas change the beam direction by controlling the excitation phase and amplitude of the individual antennas. Control of the excitation phase is achieved by adjusting the phase of the signal received at each element (or the signal radiated) using a device called a phase shifter. This opens up the possibility of electronic pattern control in antennas. It also makes high-speed pattern control possible regardless of antenna size or weight.

PAAs were originally developed for radar application. Compared to conventional mechanically driven antennas, they offer dramatically faster beam scanning. As a result, PAAs have become the main type of mobile and fixed radar antennas. Recently, they have been applied to radio communications due the benefits offered by beam scanning, miniaturization of antenna elements and feeding circuits, and cost reduction. Examples reported include an on-satellite beam reforming antenna that switches between beams based on fluctuations in demand, for example, in international satellite communications, and a mobile station antenna for mobile satellite communications. An additional of PAA advantage is their ability to conform to the structure of the items on which they are mounted. This means that the antenna can

*Email: t-maru@ieee.org

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_37-1

be mounted on a satellite or moving vehicle without mechanically altering its form. An even more advanced form of PAA is the active phased array antenna (APAA), which not only has phase shifters for each element but also active elements such as low-noise amplifiers and high-power amplifiers.

The APAA has triggered new changes in antenna control methods.

When the APAA is designed so as to receive the microwave signals of each element and to digitalize them by an analog-to-digital converter, it enables beam forming by using a digital signal processor. A phased array antenna with this kind of functionality is referred to as a digital beam-forming (DBF) antenna. A further extension of the adaptive array technology that has appeared in recent years is MIMO (multiple input multiple output), which uses array antennas to both transmit and receive signals.

Thus, the arrival and subsequent development of PAAs have helped antennas evolve from simple energy converters to comprehensive systems that include such features as control technology, signal processing, and modulation and demodulation technology, the culmination of which are DBF antennas (adaptive array and MIMO).

This chapter explains the fundamental issues in designing PAAs. At first, the structure and functions of the PAAs are introduced. Pattern synthesis, which is a feature of array antennas, and array calibration methods are explained. DBF including signal processing and MIMO techniques are explained in a later part of this chapter.

Function of Array Antenna

The basic structure of an array antenna is depicted in Fig. 1. Each radiating element of the array is called an "element antenna." An array antenna consists of element antennas and a feed network. Various antennas have been used as element antennas, e.g., wire antennas or dipole antennas, slot antennas, microstrip antennas, and even horn antennas. The feed network includes signal distribution/combining circuits, phase shifters, high-power amplifier, low-noise amplifier, etc. Unlike a single antenna, an array antenna can offer a variety of functions by changing antenna type, grid type, and excitation provided by the feed network.

A flagship feature of the array antenna is pattern synthesis; the radiation pattern is optimized so as to meet some requirement. Pattern synthesis offers the following attributes:

(1) Desired radiation pattern
(2) Arbitrary width of main beam
(3) Reduced and controlled sidelobe levels
(4) Designated null point
(5) Desired gain

Using the above attributes allows us to improve the signal-to-noise (S/N) ratio or desired-to-undesired (D/U) ratio.

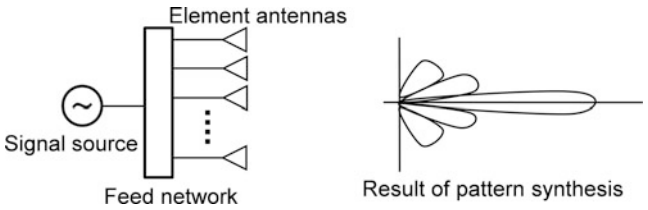


Fig. 1 Fundamental configuration of the array antenna

DOI 10.1007/978-981-4560-75-7_37-1

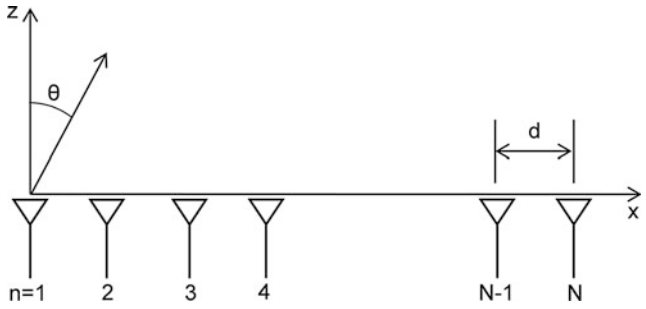


Fig. 2 Linear array antenna with N elements

Another feature of the array antenna is beam scanning which allows the main beam to be arbitrarily directed. For beam scanning, the phases of each element antenna are changed but not their positions. Several methods are available to change the excitation phase:

(1) Inserting phase shifters into the feed networks
(2) Changing frequency by altering length differences in the feed network
(3) Selecting the beam-forming network that corresponds to the desired direction

(1) yields the "phased array antenna (PAA)," which is the de facto choice for electronically scanning antennas.

A phase shifter, either analog or digital, is added to each element feed line. The latter is the main choice recently because of its controllability. Another approach to control the phases of each element antenna is digital beam forming (DBF) (Steyskal 1988; Chiba et al. 1997). In DBF, an intermediate frequency (IF) signal or a baseband signal is converted into a digital signal, and the phase is controlled in the digital stage. The key advantage of DBF is that arbitrary phase and amplitude can be set and multiple beams can be simultaneously created because of the digital operation. The optically controlled array antenna, whose phase and amplitude are controlled in the optical domain, has been also researched (Marpaung et al. 2011; Sumiyoshi et al. 2010). It enables to create multiple beams with small size and light weight.

Characteristics of Array Antenna

Linear Array Antenna

Figure 2 shows a linear array antenna consisting of N element antennas placed in a straight line with equal spacing of d (Hansen 1966; Mailloux 2005). The linear array antenna is the simplest array antenna, and many examples of practical use have been described.

The radiation patterns or input impedance or current distribution (in the case of dipole antennas) of each element antenna is not identical even if all element antennas are excited with the same amplitude and phase. The reason is mutual coupling between element antennas. However, the basic characteristics of array antennas can be explained without addressing these differences. If all element antennas are assumed to receive the same current distributions, the radiation pattern of each element antenna can be expressed by the same function. Given this assumption, the radiation pattern of an array antenna is determined by the radiation pattern function and the phase difference of each element antenna. The phase difference is determined by the position of each element antenna. The radiation pattern of array antenna $F(\theta, \phi)$ is expressed by

DOI 10.1007/978-981-4560-75-7_37-1

$$F(\theta,\phi) = f(\theta,\phi)\sum_{n=1}^{N} a_n \exp(j\varphi_n) \times \exp(jk\mathbf{r}_n \cdot \mathbf{R}) \tag{1}$$

where $\mathbf{r}_n = (x_n,0,0)$ is the position vector of the nth element antenna, a_n is excitation amplitude, p_n is excitation phase, $f(\theta, \phi)$ is the radiation pattern of a single element antenna, vector of observation point is $\mathbf{R} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$, $k = 2\pi/\lambda$ is wave number in free space, and λ is wavelength in free space. For simplicity, $F(\theta, \phi)$ of the linear array antenna in Fig. 2 is expressed as

$$F(\theta,\phi) = f(\theta,\phi)\sum_{n=1}^{N} a_n \exp(j\varphi_n) \times \exp[jk(n-1)d\sin\theta] \tag{2}$$

If the element antenna type is changed, e.g., from dipole antenna to microstrip antenna, radiation pattern $f(\theta, \phi)$ in Eq. 2 is changed. On the other hand, the summation term is unchanged. Therefore, features of the array antenna are determined by the summation term. In order to analyze the basic performance of the array antenna, the pattern that eliminates $f(\theta, \phi)$ is used. It is the same as $f(\theta, \phi) = 1$, the ideal isotropic antenna. This radiation pattern is called the "array factor."

In the array antenna of Fig. 2, assuming that the array antenna is far from the observation point, if all element antennas are excited so that phases from all element antennas are identical at direction $\theta = \theta_0$, excitation phase of the nth element antenna, φ_n, is

$$\varphi_n = -n \cdot u_0 \tag{3}$$

$$u_0 = kd\sin\theta_0 \tag{4}$$

These phases are called "co-phase" and are used when the peak of the radiation pattern is directed to a desired angle. In this case, the array factor is expressed by

$$F(u) = \sum_{n=1}^{N} a_n \exp(jnu) \tag{5}$$

u is used instead of θ,

$$u = kd(\sin\theta - \sin\theta_0) \tag{6}$$

Dropping the radiation pattern of element antenna, $f(\theta, \phi)$, from Eq. 2 yields Eq. 5. θ, d, and θ_0 are determined by just one variable u. This is convenient when analyzing the characteristics of an array antenna.

When all element antennas of the linear array antenna are excited with identical amplitude and co-phase, from Eq. 5, the radiation pattern of the array antenna is expressed as

$$F(u) = \frac{\sin\left(\frac{Nu}{2}\right)}{N\sin\frac{u}{2}} \tag{7}$$

DOI 10.1007/978-981-4560-75-7_37-1

From Eq. 7, F takes maximum value 1 at $u = 2\,m\pi(m = 0, \pm 1, \pm 2, \cdots)$. Local maxima in F, called sidelobes, are found at $Nu/2 = (2\,M + 1)\,\pi/2$ $(M = 0, \pm 1, \pm 2, \cdots)$.

In case of $u \ll 1$, Eq. 7 is approximated by

$$F(u) = \frac{\sin\left(\frac{Nu}{2}\right)}{\frac{Nu}{2}} \tag{8}$$

This approximation is valid when element antenna spacing d is smaller than wavelength or the range of θ is limited. Equation 8 yields the same pattern as that created by an aperture emitting a continuous wave with uniform amplitude distribution. Therefore, the array antenna of this case can be replaced by an aperture antenna with uniform amplitude distribution. Reversing this assumption allows an array antenna to be designed: First, aperture size and amplitude distribution of a continuous wave source are determined so as to satisfy service requirements like gain, beamwidth, and sidelobe level. Second, the number of elements and element antenna spacing are determined by placing the element antennas at the aperture determined in the first step.

Figure 3 is an example of the array factor of a linear array antenna with nine elements. This is a periodic function and yields maximum value 1 at $u = 2\,m\pi(m = 0, \pm 1, \pm 2, \cdots)$. Local maxima are sidelobes.

If the main beam is directed to the z-axis in Fig. 2, the antenna becomes a broadside array, i.e., the main beam and the axis of elements are perpendicular. In this case, θ_0 becomes 0. Because $|\sin(\theta)| \leq 1$, the range of u becomes

$$-kd \leq u \leq kd \tag{9}$$

For antenna spacing $d = \lambda/2$, $|u| \leq \pi$ corresponds to the actual space $|\sin(\theta)| \leq 1$, i.e., $-\pi/2 \leq \theta \leq \pi/2$. This range where the pattern appears in actual space is called the "visible region." Other region is called the "invisible region."

For antenna spacing $d = 1.5\ \lambda$, $|u| \leq 3\pi$ becomes the visible region. In this case, in addition to $u = 0$, two peaks at $|u| = 2\pi$ fall on the visible region. These large radiations correspond to the periodic diffraction fringe caused by the optical diffraction grating effect. They are called "grating lobes."

In array antennas, if the spacing between element antennas is expanded, mutual coupling between element antennas is reduced and narrow beams are obtained. If spacing between element antennas is

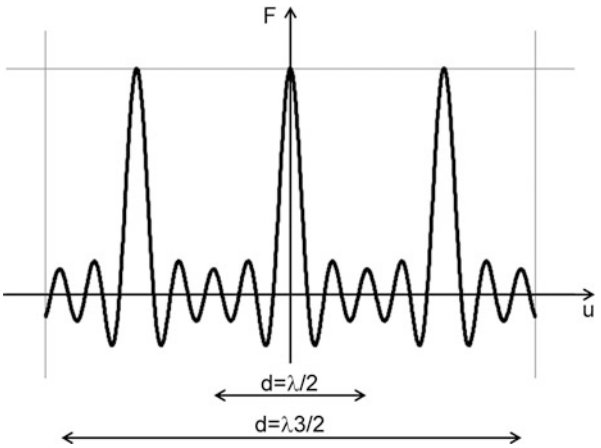


Fig. 3 Array factor of 9 element linear array antenna and visible region for each antenna spacing

DOI 10.1007/978-981-4560-75-7_37-1

expanded within a fixed aperture size, the number of elements is reduced and feeding circuits are accordingly simplified and manufacturing costs are lowered. Thus, antenna designs that offer the widest possible spacing are preferred. Unfortunately, excessively wide antenna spacings trigger grating lobes. Radiation occurs on undesired angles, gain is degraded on the desired angle, and signal-to-noise ratio (S/N) is degraded along with the input impedance of the element antenna. Array antenna designs basically widen the spacing up to the point that grating lobes appear in the visible region.

For a one-dimensional PAA whose main beam is to be scanned, antenna spacing d is calculated as described below. First, the relationship between variable u and main beam angle θ_0 is given by Eq. 6:

$$u = kd(\sin\theta - \sin\theta_0) \tag{6}$$

If $\theta_0 \geq 0$ is assumed, |u| is maximized at $\theta = -\pi/2$, which yields the following equation:

$$|u|_{maz} = \frac{2\pi}{\lambda} d(1 + \sin\theta_0) \tag{10}$$

The first grating lobe appears at $|u| = 2\,\pi$. Grating lobes do not appear in the visible region if the following inequality is satisfied:

$$\frac{2\pi}{\lambda} d(1 + \sin\theta_0) \leq 2\pi \tag{11}$$

From Eq. 11, the upper limit of d without grating lobes can be determined:

$$d = \frac{\lambda}{1 + \sin\theta_0} \tag{12}$$

If d is selected so as to satisfy Eq. 12, where θ_0 is the upper limit of scanning angle, no grating lobe appears within the visible region. If, however, the element spacing d determined from Eq. 12 is used, the peaks of the first grating lobes lie on the edges of the visible region. In order to eliminate the grating lobes completely, including their shoulder, antenna spacings that are 5–10 % smaller than Eq. 12 are generally adopted.

Using Eq. 12, the relationship between antenna spacing and main beam angle with the first grating lobe at $\theta = -\pi/2$ is plotted in Fig. 4. When $d/\lambda = 0.5$ is used, no grating lobe appears within the scan angle $\pm\pi/2$.

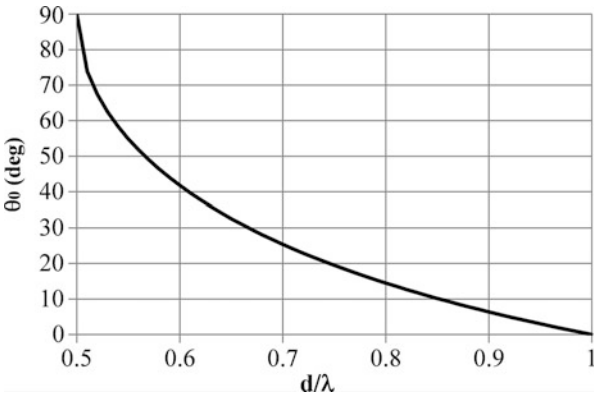


Fig. 4 Relationship between the antenna spacing and main beam angle with the first grating lobes at the visible region edges

DOI 10.1007/978-981-4560-75-7_37-1

If grating lobes in the visible region are permitted, the relationship between grating lobe angle θ_g and main beam angle θ_0 is given by

$$\frac{2\pi}{\lambda} d\left(\sin\theta_g - \sin\theta_0\right) = -2\pi \tag{13}$$

$$\sin\theta_g = \sin\theta_0 - \frac{\lambda}{d} \tag{14}$$

This is illustrated in Fig. 5.

Planar Array Antenna

While linear array antennas have radiation patterns that take the form of a dish or circular cone, planar array antennas have pencil beam-like radiation patterns (Mailloux 2005). Therefore, planar array antennas are widely used as practical array antennas.

Most planar array antennas use either a rectangular grid array or a triangular grid array; see Fig. 6. The rectangular grid array is realized by expanding the linear array to two dimensions. Its feeding circuit is easier to design than that of the triangular grid array antenna. On the other hand, the triangular grid array

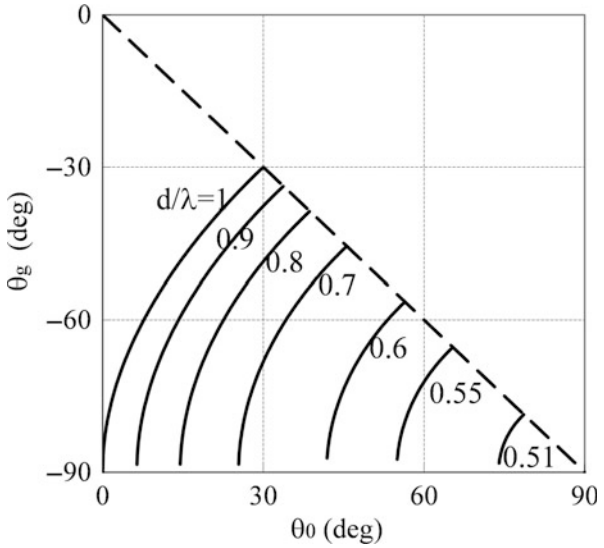


Fig. 5 Relationship between main beam angle and grating lobe angle for different antenna spacings

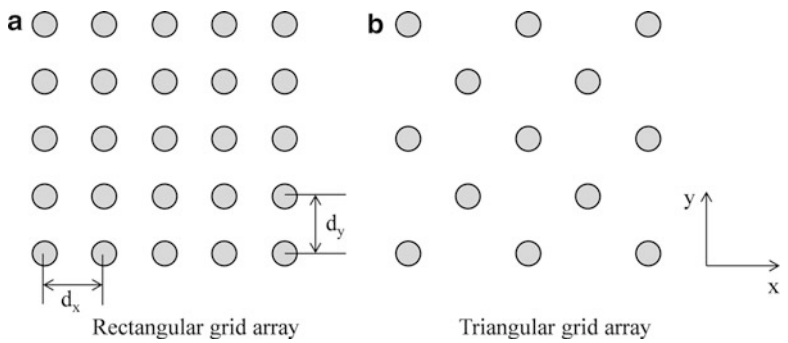


Fig. 6 Rectangular grid array and triangular grid array

DOI 10.1007/978-981-4560-75-7_37-1

allows element antenna spacing to larger than the rectangular grid array. The triangular grid array allows fewer element antennas to be used than the rectangular grid array for the same aperture size and gain and coverage. Additionally, mutual coupling between element antennas is reduced because of wide antenna spacing. So the triangular grid array antennas are widely used.

In the case of the rectangular grid array, the complex amplitude of the element antenna on the nth column (x-direction) and the mth row (y-direction) is defined as A_{mn} ($=a_{mn}$ exp[j(nu + mv)]). Element antenna spacing in x-direction is d_x, and element antenna spacing in y-direction is d_y. The radiation pattern is given by

$$F(\theta, \phi) = \sum_n \sum_m A_{nm} \exp[j(nu + mv)] \tag{15}$$

where

$$u = kd_x \sin\theta \cos\phi \tag{16}$$

$$v = kd_y \sin\theta \sin\phi \tag{17}$$

If the complex amplitude of the element on nth column and mth row, A_{mn} is expressed as the product of x-direction term A_n and y-direction term A_m, that is,

$$A_{nm} = A_n A_m \tag{18}$$

Equation 15 is rewritten as

$$F(\theta, \phi) = \sum_n A_n \exp[jnu] \sum_m A_m \exp[jmv] \tag{19}$$

It means that the radiation pattern of a rectangular grid array antenna is the product of the radiation pattern of the x-directional linear array and the radiation pattern of the y-directional linear array. Basically, rectangular grid array antennas are excited in accordance with Eq. 18. In the case, characteristics of the planar array antenna are those of the linear array antenna. Radiation patterns described by Eq. 19 are called separable patterns. On the other hand, radiation patterns that do not follow Eq. 19 are called non-separable patterns.

The rectangular grid array and the triangular grid array have different grating lobe characteristics. In Fig. 6, the two designs have the same element antenna spacing on the x-axis and on the y-axis, so the incidence of grating lobes on the xz-plane and yz-plane is also the same. However, grating lobe characteristics on the diagonal direction are not the same. A grating lobe chart, based on the reciprocal lattice of crystallography, helps us to analyze the grating lobes of planar array antennas. An example of a grating lobe chart is shown in Fig. 7 for both the rectangular grid array and the triangular grid array; the main beam is directed to the boresight (z-axis).

The grating lobe chart has two axes: axis Tx plots term $\sin(\theta)\cos(\phi)$, while axis Ty plots term $\sin(\theta)\sin(\phi)$; length from the origin is $\sin(\theta)$ and circumferential angle is ϕ. The visible region lies inside the circle of $\sin(\pi/2) = 1$. The grating lobes of the rectangular grid array lie on the reticular pattern of λ/d_x and λ/d_y in the grating lobe chart. The triangular grid array outputs the grating lobes on a triangle lattice in the grating lobe chart. If element antenna spacing is constant, the grating lobes of the triangular grid array are more widely separated than those of the rectangular grid array. Thus, triangular grid array antennas are widely used.

DOI 10.1007/978-981-4560-75-7_37-1

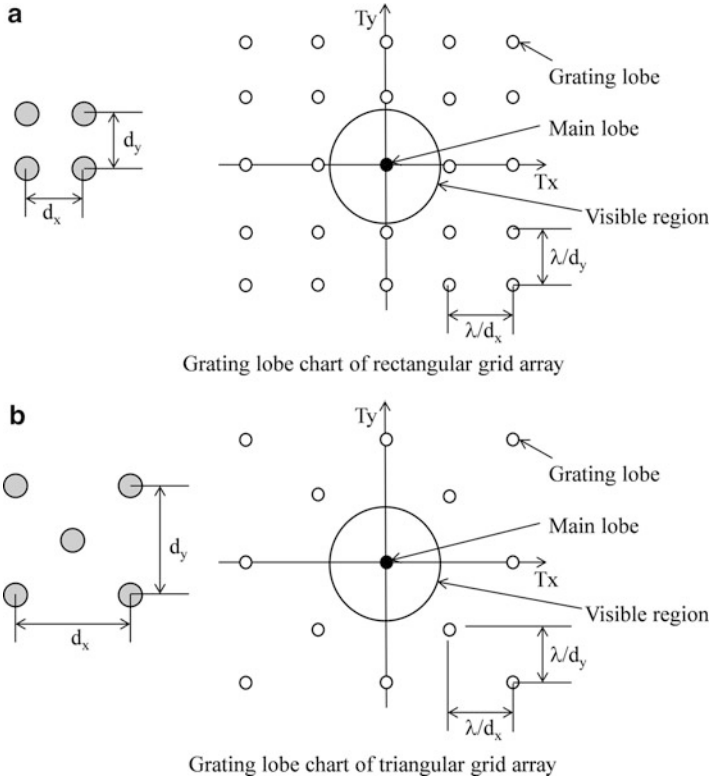


Fig. 7 Grating lobe chart at main beam angle $\theta = 0°$

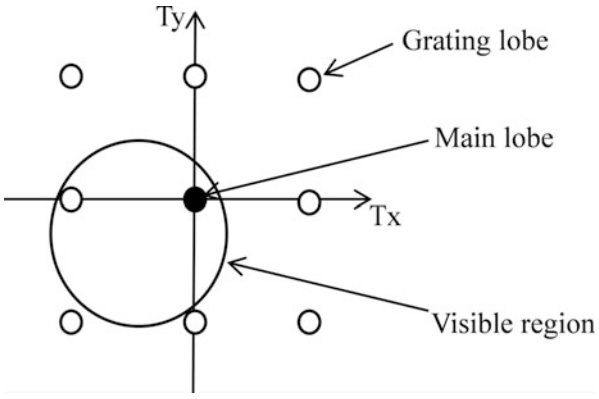


Fig. 8 Grating lobe chart at main beam angle $\theta = 60°$, $\phi = 60°$

The grating lobe chart is helpful in analyzing grating lobes even if the beam is scanned. Figure 8 is an example of the grating lobe chart when the main beam of a rectangular grid array is directed to $\theta = 60°$, $\phi = 30°$.

When the main beam is scanned, its point moves in the grating lobe chart. All grating lobe points move in parallel with that of the main lobe. In the case of Fig. 8, one grating lobe appears in the visible region due to beam scanning.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_37-1

Gain, Beamwidth, and Sidelobe Level

Directivity of array antennas is approximated as follows (Elliott 2003)

Linear Array Antenna When the element antennas in Fig. 2 are point wave sources, their amplitudes are uniform, their number (N) is large, and antenna spacing d is selected so as to exclude grating lobes; the directivity of the resulting linear antenna, D_l, is given by

$$D_l \approx \frac{2Nd}{\lambda} \tag{20}$$

The directivity is determined by the radiation pattern. Loss of feeding circuits and efficiency of element antennas are not included in this expression.

Planar Array Antenna When aperture size is sufficiently larger than wavelength, and element number N is large, the directivity of a planar array, D_p, is given by

$$D_p \approx D_e N \cos\theta_0 \tag{21}$$

$$D_e = \frac{4\pi A_e}{\lambda^2} \tag{22}$$

where θ_0 is main beam angle, D_e is directivity of arrayed element antenna, and A_e is efficient aperture size of the element antenna. If directivity of stand-alone element antenna D_0 is higher than D_e, that is, $D_0 \geq D_e$ and $A_0 \geq A_e$ (A_0: efficient aperture size of stand-alone element antenna), A_e is limited by the area occupied by one element antenna. Thus, $A_e \times N$ is nearly equal to aperture size of total antenna size A. This condition is basically satisfied in the case of array antennas with standard element antenna spacings. Using this approximation, D_p becomes

$$D_p \approx \frac{4\pi A}{\lambda^2} \cos\theta_0 \tag{23}$$

This is equal to the directivity of a planar wave source with aperture size A and with phase distribution so as to direct the main beam to θ_0. If the amplitude distribution of element antennas is unequal to obtain a radiation pattern with low sidelobe levels, the directivity will be reduced since the aperture efficiency is determined by the amplitude distribution.

Actual Gains

The actual gain of a PAA is not equal to the directivity determined from the radiation pattern (Mailloux 2005). The one factor is the impedance mismatch of element antennas and mutual coupling between element antennas. When the number of elements is N, and the aperture efficiency determined from amplitude distribution is η, the actual gain is given by

$$G(\theta_0) = g(\theta_0) N \eta \tag{24}$$

g is the element gain function given by

DOI 10.1007/978-981-4560-75-7_37-1

$$g(\theta_0) = D_e \cos\theta_0 \left(1 - |\Gamma(\theta_0)|^2\right) \tag{25}$$

D_e is the directivity of the arrayed element antenna given by Eq. 22. Γ is reflection coefficient where all elements are excited, that is, active reflection coefficient is defined as

$$\Gamma_n(\theta_0) = \frac{1}{a_n} \sum_{i=1}^{N} S_{i,n} a_i \exp[j(\psi_i(\theta_0) - \psi_n(\theta_0))] \tag{26}$$

Subscript n means element number. $S_{n,n}$ is the self-reflection coefficient of element n. $S_{i,n}$ is mutual coupling from element i to element n. a_i is excitation amplitude of element i. $\psi_i(\theta_0)$ is the excitation phase of element i. Because $\psi_i(\theta_0)$ varies according to main beam angle θ_0, active reflection coefficient Γ is also a function of θ_0.

Beamwidth and Sidelobe Level

From Eq. 8, directivity of a linear array with element number N and uniform amplitude distribution is approximately (Mailloux 2005)

$$F(\theta) = \frac{\sin(u)}{u} \tag{27}$$

where

$$u = \frac{\pi}{\lambda} L(\sin\theta - \sin\theta_0) \tag{28}$$

L is aperture length, $L = Nd$ (d is element antenna spacing). Beamwidth Θ (which is defined as the angle at which gain decreases −3 dB from main lobe peak) of this linear array antenna is directly calculated from Eq. 27. Its approximation is given by

$$\Theta \approx \frac{50}{(L/\lambda)\cos\theta_0} (\text{deg.}) \tag{29}$$

The beamwidth increases with scan angle θ_0.

The directivity of a circular planar array with high enough element number and uniform amplitude distribution is approximately equal to a circular aperture with continuous wave source and uniform amplitude distribution. It is given as follows:

$$F(\theta) = \frac{2J_1(u)}{u} \tag{30}$$

$$u = \frac{\pi}{\lambda} D(\sin\theta - \sin\theta_0) \tag{31}$$

J_1 is the Bessel function order 1. Beamwidth is approximated by

$$\Theta \approx \frac{60}{(D/\lambda)\cos\theta_0} \ (\text{deg.}) \tag{32}$$

DOI 10.1007/978-981-4560-75-7_37-1

From Eqs. 29 and 32, beamwidth of the circular planar array is about 1.2 times greater than that of the linear array or the rectangular planar array.

Sidelobe level can be easily obtained from Eqs. 27 and 30. First sidelobe level of the linear array with uniform amplitude distribution is about −13.3 dB. That of the circular planar array with uniform amplitude distribution is about −17.6 dB.

Mutual Coupling

Mutual coupling between element antennas, a characteristic of array antennas, causes various problems. If the mutual coupling level is not zero, the active reflection coefficient of an arrayed element antenna is not equal to that of a stand-alone element antenna as per Eq. 26. The impedance of an arrayed element antenna where all elements are excited is called the active impedance. From Eq. 26, the active impedance varies according to beam-scanning angle. In large-scale arrays, there is a risk that the active impedance becomes infinite (the active reflection coefficient becomes 1) or the active impedance becomes zero (the active reflection coefficient becomes −1). In such cases, no signal is radiated. This phenomenon is called scan blindness. One example of scan blindness is as follows: element antenna spacing d is larger than $\lambda/2$, and the main beam is widely angled away from the boresight. Scan blindness may arise when the grating lobe appears at a visible region edge. Figure 9 shows a dipole antenna that is $\lambda/4$ above a metal reflector. This antenna is used as the elements of an infinite array antenna. Because antenna spacing in the y-direction is $\lambda/2$, a grating lobe appears at $\phi = 90°$ and $\theta = -90\,°$, when the main beam is directed to $\phi = 90°$ and $\theta = 90°$. Figure 10 is its active impedance plot. When main beam direction is the boresight ($\phi = 90°$, $\theta = 0°$), the impedance is almost matched. When the main beam direction is scanned 90° from the boresight ($\phi = 90°$, $\theta = 90°$), the impedance is rapidly degraded.

Scan blindness can arise without grating lobes. One example is an array antenna from which some elements have been eliminated to create a tapered amplitude distribution, which is explained later. Because some element antennas may not have symmetrical partners, mutual coupling cannot be fully canceled.

One method to estimate the degradation in active impedance is the array-element pattern measurement; the radiation pattern of a single element is measured when only that element is excited and all other elements are terminated. The array-element pattern is the element gain function given by Eq. 25. The gain at the angle at which the scan blindness occurs decreases to zero at worst.

To counter the impedance mismatch due to beam scanning, one approach is to reduce the mutual coupling by changing the element antenna type or the array configuration. Other approach is to add matching circuits to the element antennas. There is no general, deterministic solution to this problem. So computer simulations or experimental techniques with parameterized antenna parameters are used. Note that the edge effect should be considered because element number is finite.

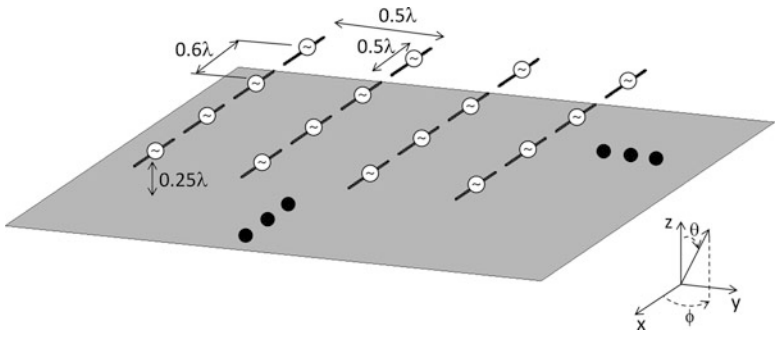


Fig. 9 Antenna configuration

DOI 10.1007/978-981-4560-75-7_37-1

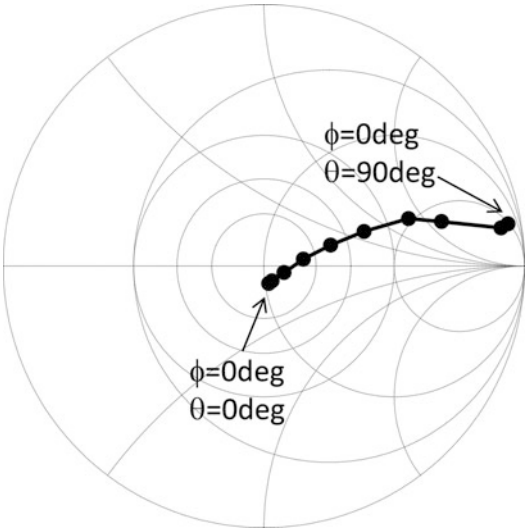


Fig. 10 Active impedance (*circles* are plotted at θ intervals of 10°)

Mutual coupling produces various problems like impedance variation, gain degradation, and radiation pattern degradation. Therefore, it is important for the array antenna designer to determine the antenna type and the array configuration so as to reduce mutual coupling as much as possible.

The Analysis of Infinite Array by Using Periodic Boundary Conditions

When large-scale array antenna is analyzed, simulating the actual array size allows various antenna characteristics, like the impedance of each element antenna, mutual coupling, array-element pattern, and total pattern, to be obtained. Unfortunately, such simulations are usually impossible due to constraints on computer memory resources and calculation time. One approach is to use a subarray, a subset of the full array. However, it is difficult to judge whether the modeled subarray well replicates the characteristics of the full array. Other analysis methods of infinite arrays (Ren et al. 1994; Turner and Christodoulou 1999; Holter and Steyskal 1999) use periodic boundary conditions (Amitay et al. 1972). The analytical region matches the element spacing, but a periodic boundary condition is added. This approach method can simulate one- or two-dimensional infinite arrays by modeling just one element antenna.

Antenna Pattern Synthesis

Amplitude Distribution Yielding Low Sidelobe

In the previous chapter, it was shown that the first sidelobe level of the linear array antenna and the rectangular aperture array antenna is about −13.3 dB, while that of the circular aperture array antenna is about −17.6 dB. In APAA for radar application, lower sidelobe level is required so as to reduce interference and noise from the environment.

It is important to achieve both sidelobe level reduction and gain maintenance. The amplitude distribution yielding the sidelobe level of a defined value (or lower) and minimum beamwidth is the Chebyshev distribution (or Dolph-Chebyshev distribution) (Dolph 1946). The radiation pattern of an array antenna with this amplitude distribution is expressed as the following polynomial:

$$E(u) = T_{N-1}\left(z_0 \cos\frac{u}{2}\right) \tag{33}$$

DOI 10.1007/978-981-4560-75-7_37-1

All sidelobe levels are identical. Z_0 is a constant that defines the sidelobe level. Its value is greater than 1. The ratio of main lobe level to sidelobe level is

$$R = T_{N-1}(z_0) \tag{34}$$

$$z_0 = \cosh\left[\frac{1}{N-1}\cosh^{-1}R\right] \tag{35}$$

The Chebyshev distribution, both ends of which are normalized to 1, is given as

$$a_n = \begin{cases} (n-1)\alpha^2 \sum_{m=0}^{n-1} \dfrac{(n-m)_m (N-n-m-1)_n}{m!(m+1)!}\alpha^{2m} & \left(0 \le n \le \dfrac{N-1}{2}\right) \\ 1 & (n=0) \end{cases} \tag{36}$$

$$\alpha = \frac{\sqrt{z_0^2 - 1}}{z_0} \tag{37}$$

In Eq. 36, $(\beta)_l$ means the following product:

$$\beta_l = \beta(\beta+1)(\beta+2)\cdots(\beta+l-1)$$

$$\beta_0 = 1$$

When the Chebyshev distribution is applied to a large-scale array with element number N, amplitudes of edge elements become discontinuous. The radiation pattern is very sensitive to errors in the amplitude distribution. In such case, the Taylor distribution, which has no such inconvenience, is commonly used.

The Taylor distribution is explained as follows. At first, the Chebyshev distribution of N elements follows the equation below when entire array length $a = (N-1)d$ is kept constant and N approaches infinity:

$$E_1(u) = \cos\left[\pi\sqrt{u^2 - A^2}\right] \tag{38}$$

$$u = \frac{2a}{\lambda}\sin\theta \tag{39}$$

A is a parameter of the sidelobe level, $1/\cosh(\pi A)$. Radiation patterns with Chebyshev distribution have constant sidelobe levels. On the other hand, from Eq. 8, the radiation pattern of a continuous wave source with uniform amplitude distribution is

$$E_2(u) = \frac{\sin(\pi u)}{\pi u} \tag{40}$$

It becomes zero when u is integer. The sidelobe level is reduced by $1/u$.

Here integer number $\overline{n}$ is defined. The Taylor radiation pattern has zero point of E_2 for $u \ge \overline{n}$ and zero point of E_1 for $u \le \overline{n}$ (Taylor 1955).

DOI 10.1007/978-981-4560-75-7_37-1

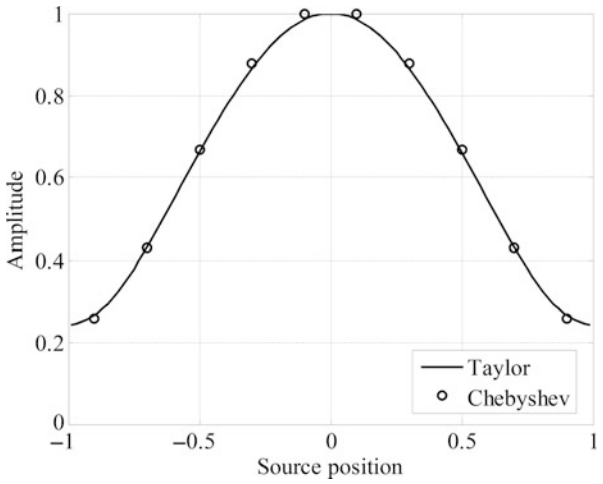


Fig. 11 Taylor distribution (*solid line*) and Chebyshev distribution (*circles*)

The amplitude distribution that yields the Taylor radiation pattern is the Taylor distribution. The Taylor distribution with sidelobe level of -30 dB and $\overline{n}$ =4 is shown in Fig. 11. The Chebyshev distribution with sidelobe level of −30 dB and ten elements is also plotted. Its values are similar to the sampled values of the Taylor distribution.

Discrete values of the Taylor distribution are generally used in designing large-scale array antennas with element number N. Compared to the Chebyshev distribution, the Taylor distribution yields high gain.

Plane-Wave Synthesis

Plane-wave synthesis yields desired radiation pattern by superposition of multiple radiation patterns (Chiba and Mano 1987).

In Fig. 12, the angle of the desired signal is θ_s, and its radiation pattern with main beam angle θ_s is $E_s(\theta)$, which is the initial pattern. The angles forming nulls are θ_m ($m = 1 \cdots M$), and radiation pattern with main beam angle θ_m is $E_m(\theta)$ ($m = 1 \cdots M$). In plane-wave synthesis, null points are formed by superimposing the initial patterns E_s and E_m with multiplication by an appropriate complex factor. When complex amplitudes of element antenna n to form E_s and E_m are A_{sn} and B_{mn}, respectively, from Eq. 1, E_s and E_m are as follows:

$$E_s(\theta) = \sum_{n=1}^{N} A_{sn} \exp(jk\mathbf{r}_n \cdot \mathbf{R}) \tag{41}$$

$$E_m(\theta) = \sum_{n=1}^{N} B_{mn} \exp(jk\mathbf{r}_n \cdot \mathbf{R}) \tag{42}$$

When complex factor multiplied to E_m is a_m, desired radiation pattern yielding desired angle θ_s and null angle θ_m is given as

$$E(\theta) = E_s(\theta) + \sum_{m=1}^{M} a_m E_m(\theta) = \sum_{n=1}^{N} \left(A_{sn} + \sum_{m=1}^{M} a_m B_{mn} \right) \exp(jk\mathbf{r}_n \cdot \mathbf{R}) \tag{43}$$

DOI 10.1007/978-981-4560-75-7_37-1

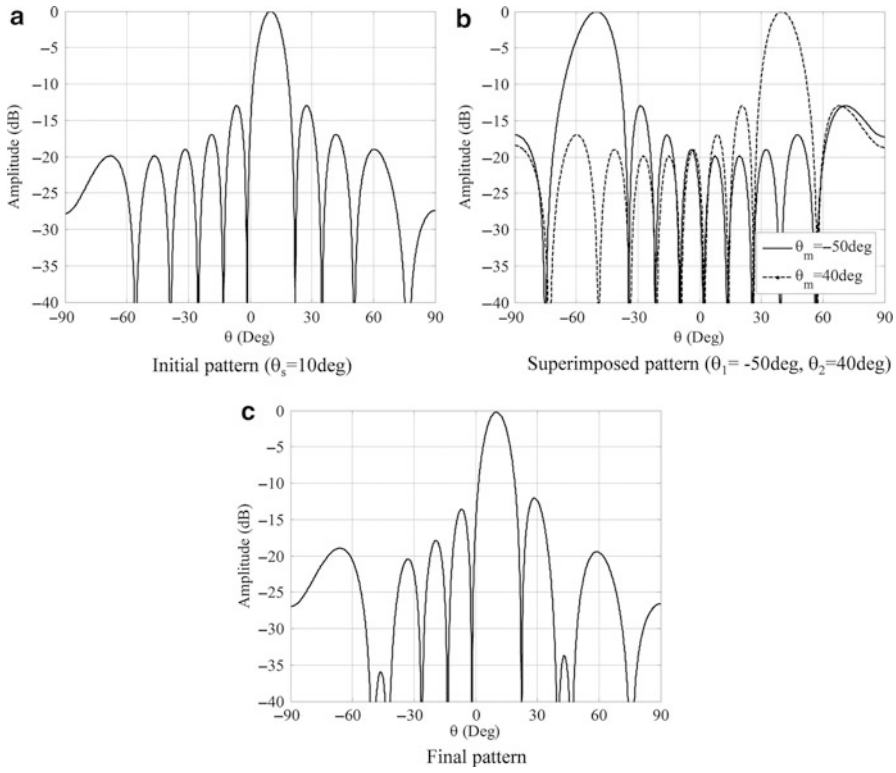


Fig. 12 Plane-wave synthesis

Amplitude distribution A_n to form the desired radiation pattern is expressed by using A_{sn} and B_{mn}:

$$A_n = A_{sn} + \sum_{m=1}^{M} a_m B_{mn} \tag{44}$$

From Eq. 43, the condition that yields nulls on θ_m ($m = 1 \cdots M$) is

$$E(\theta_m) = E_s(\theta_m) + \sum_{m=1}^{M} a_m E_m(\theta) = 0 \tag{45}$$

a_m is the solution of the following simultaneous equation:

$$\mathbf{Ea} = \mathbf{e} \tag{46}$$

$$\mathbf{E} = \begin{bmatrix} E_1(\theta_1) & E_2(\theta_1) & \cdots & E_M(\theta_1) \\ E_1(\theta_2) & E_2(\theta_2) & \cdots & E_M(\theta_2) \\ \vdots & \vdots & \ddots & \vdots \\ E_1(\theta_M) & E_2(\theta_M) & \cdots & E_M(\theta_M) \end{bmatrix} \tag{47}$$

$$\mathbf{a} = [\, a_1 \quad a_2 \quad \cdots \quad a_M \,] \tag{48}$$

$$\mathbf{e} = [\, -E_s(\theta_1) \quad -E_s(\theta_2) \quad \cdots \quad -E_s(\theta_M) \,] \tag{49}$$

DOI 10.1007/978-981-4560-75-7_37-1

If A_{sn} and B_{mn} are array factors for θ_s and θ_m, that is, the conjugate of the electric field from element antenna, A_n becomes the complex amplitude that maximizes gain for θ_s and sets nulls on θ_m. Because A_{sn} and B_{mn} are amplitudes that form a plane wave, this method is called plane-wave synthesis. Figure 12 is an example of plane-wave synthesis when the number of element antennas is 10. In (a), θ_s is 10°. In (b), θ_1 and θ_2 are −50° and 40°, respectively. The resulting pattern, (c), has two nulls at −50° and 40°.

Plane-wave synthesis gives a unique complex value that includes amplitude and phase information for each element. This method requires both amplitude and phase control. Because standard PAA has only phase control, the desired pattern may not be obtained. In such cases, antenna pattern synthesis with only phase control, which demands nonlinear optimization like the minimax method, is used.

Minimax Algorithm

When plane-wave synthesis or other analytical synthesis algorithms are not applicable, or the amplitude distribution cannot be analytically obtained due to the complexity of the desired radiation pattern, nonlinear optimization algorithms are used. One example, the minimax algorithm, addresses array antenna power (Klein 1984). It is effective when the reference point number of the radiation pattern, M, is larger than element antenna number, N.

Parameters are commonly defined as follows:

N: Element antenna number
M: Reference point number of radiation pattern
E_{mn}: Radiated electric field amplitude of element antenna n, angle θ_m
ϕ_{mn}: Radiated electric field phase of element antenna n, angle θ_m
a_n: Excitation amplitude of element antenna n
φ_n: Excitation phase of element antenna n
P_{0m}: Desired power for θ_m
W_m: Weight of each reference point

The evaluation function is defined as

$$F = \sum_{m=1}^{M} W_m \left(\left| \sum_{n=1}^{N} a_n \exp(j\varphi_n) E_{nm} \exp(j\phi_{nm}) \right|^2 - P_{0m} \right)^2 \tag{50}$$

Excitation amplitude and phase yielding desired radiation pattern can be obtained by minimizing F by calculating its gradient.

The minimax algorithm can be applied to any requirement. However, the following important points must be noted:

(1) F has many local minima due to its nonlinearity, and it is difficult to judge if an F value is the absolute or local minimum. Moreover, the iteration number for convergence depends on initial excitation amplitude and phase. In order to ease these inconveniences, plane-wave synthesis can be conducted first and its results used as the initial condition.
(2) In the case of an antenna with fixed radiation pattern, feeding circuits cannot be constructed because calculated values of excitation amplitude and phase may be scattered.

DOI 10.1007/978-981-4560-75-7_37-1

Element Thinning Algorithm

In many APAA designs, transmitting power of each element antenna is constant because high-power amplifiers are driven in their saturated region for efficiency. This prevents excitation amplitude of each element from being controlled to obtain radiation patterns with low sidelobe levels. In order to realize low sidelobe levels in this situation, element density distribution is controlled by thinning element antennas. A deterministic approach to the thinning algorithm that attains the element density distribution yielding desired radiation pattern has been proposed (Numazaki et al. 1987).

It has no ambiguity and does not depend on probability.

When element position is P_i and desired amplitude distribution is f, weight f_i of element at P_i is defined as

$$f_i = Cf(P_i)\,\Delta S_i \tag{51}$$

$f(P_i)$ is the amplitude distribution value at P_i. ΔS_i is the exclusive area of point P_i. C is a coefficient such that the upper limit of f_i is 1. An example of P_i is the grid of a linear array, and f is the Taylor distribution. When the order of element position is determined, the thinning function, $T(P_i)$, is given by the cumulative value of f_i:

$$T(P_i) = \left[\sum_{j=1}^{i} f_j + 0.5\right] - \left[\sum_{j=1}^{i-1} f_j + 0.5\right] = 0 \text{ or } 1 \tag{52}$$

$[\cdot]$ is a floor function. Equation 52 outputs 0 or 1. The value means excitation or non-excitation of each element. The difference between the cumulative value of the weight and that of excitation element number is 0.5 or less. An example of thinning is shown in Fig. 13. The number of elements is 20 and the amplitude distribution follows a Taylor distribution with sidelobe of -30 dB and $\overline{n} = 4$. In the figure, the solid line is the cumulative value of f_i, and circle plots are the cumulative value of $T(P_i)$. No difference exceeds 0.5.

In planar array antennas, the radiation pattern on an observing plane is determined by element number distribution vertically projected onto the plane. Let's consider optimizing the x-plane radiation pattern to yield the desired pattern. While P_i has coordinates of (x, y), order of P_i is given lexicographically, that is, x is sorted in ascending order. If multiple elements have identical position according to x, y is sorted in

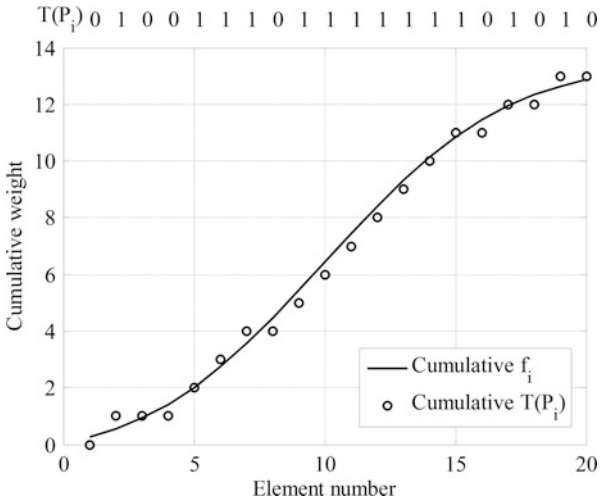


Fig. 13 Element thinning algorithm

DOI 10.1007/978-981-4560-75-7_37-1

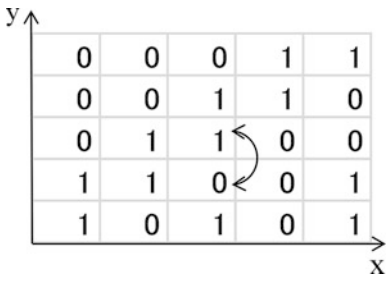


Fig. 14 Interchange of excitation elements

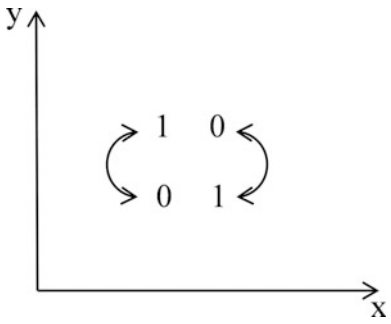


Fig. 15 Interchange of two pairs

ascending order. The thinning function, T_x, for the x-plane pattern is determined from P_i. From T_x, P_{xj}, which is the number of the excitation element at jth position for the x-plane, is calculated. Similarly, T_y and P_{yk} are calculated for the y-plane. When T_x is calculated for x-plane, resulting Q_{yk}, which is the excitation element number at kth position on the y-plane, is also calculated. Difference between ideal distribution P_{yk} and distribution Q_{yk} is calculated as D_k:

$$D_k = Q_{yk} - P_{yk} \tag{53}$$

The summation of D_k must be zero. If all D_k are zero, the desired distribution is obtained for x- and y-plane.

If D_{k1} is positive, D_{k2} having negative value must exist. If the thinning function yields 1 on row k_1 and 0 on row k_2 at position x_1, and 0 on row k_1 and 1 on row k_2 at position x_2, values are interchanged as shown in Fig. 14. D_{k1} and D_{k2} increase or decrease by 1, while P_{xk} is unchanged. This operation enables the thinning function to approach the desired distribution. Though multiple interchangeable pairs may be found, the existence of interchangeable pairs is not assured. When no interchangeable pair is found, two pairs can be interchanged as shown in Fig. 15. After the operation, above changeable pair is searched.

Here one calculated example of the thinning algorithm is shown. A planar array antenna with 10 × 10 elements, rectangular grid, and Taylor distribution with sidelobe level of −25 dB and $\overline{n} = 3$ is assumed. The excitation elements for the x-plane are determined as per Fig. 16a. In the 10 × 10 matrix, "0" means unexcited element. "1" means excited element. P_{xj}, the sum of each column, is the result and has no ambiguity. Resulting Q_{yk}, sum of each row, differs from ideal P_{yk}. Here "0" and "1" are interchanged in each column. Q_{yk} approaches P_{yk} while P_{xj} is unchanged. Interchanged result is shown in Fig. 16b. From the results, the calculated radiation pattern is shown in Fig. 17. Solid line is the result achieved by the thinning algorithm. Broken line is the result gained by an ideal Taylor distribution. The plots are nearly identical.

DOI 10.1007/978-981-4560-75-7_37-1

a

											Q_{yk}	P_{yk}(Ideal)
	0	0	1	1	1	1	1	1	1	1	8	4
	1	1	0	1	1	1	1	1	0	0	7	5
	0	0	1	1	1	1	1	0	1	1	7	7
	1	1	1	1	1	1	1	1	0	0	8	9
	0	0	1	1	1	1	1	1	1	0	7	10
	0	1	0	0	1	1	0	0	0	1	4	10
	1	0	1	1	1	1	1	1	1	0	8	9
	0	1	1	1	1	1	1	1	0	0	7	7
	0	0	0	1	1	1	1	1	1	1	7	5
	1	1	1	1	1	1	1	0	0	0	7	4
P_{xj}	4	5	7	9	10	10	9	7	5	4		

Before interchange

b

											Q_{yk}
	0	0	1	1	1	1	0	0	0	0	4
	0	0	0	1	1	1	1	1	0	0	5
	0	0	1	1	1	1	1	0	1	1	7
	1	1	1	1	1	1	1	1	1	0	9
	1	1	1	1	1	1	1	1	1	1	10
	1	1	1	1	1	1	1	1	1	1	10
	1	0	1	1	1	1	1	1	1	1	9
	0	1	1	1	1	1	1	1	0	0	7
	0	0	0	1	1	1	1	1	0	0	5
	0	1	0	0	1	1	1	0	0	0	4
P_{xj}	4	5	7	9	10	10	9	7	5	4	

After interchange

Fig. 16 Thinned results

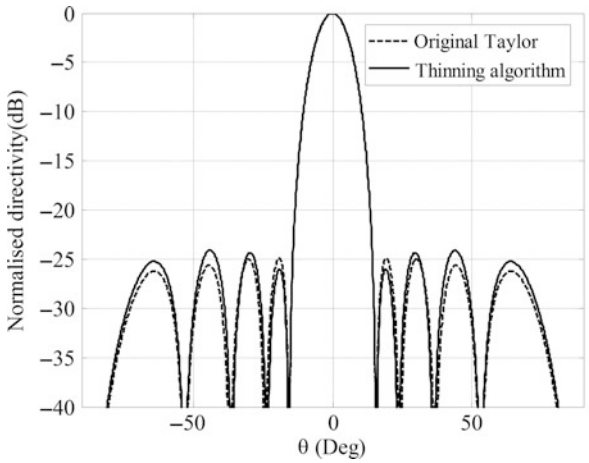


Fig. 17 Radiation pattern

Genetic Algorithm

Introduction of Genetic Algorithm

Though the element thinning algorithm determinately selects elements to excite, there is low flexibility in changing the exciting elements for adjusting the radiation patterns on the two axes. It is not always true that the element thinning algorithm yields the desired pattern. Here the genetic algorithm (GA) (Holland

DOI 10.1007/978-981-4560-75-7_37-1

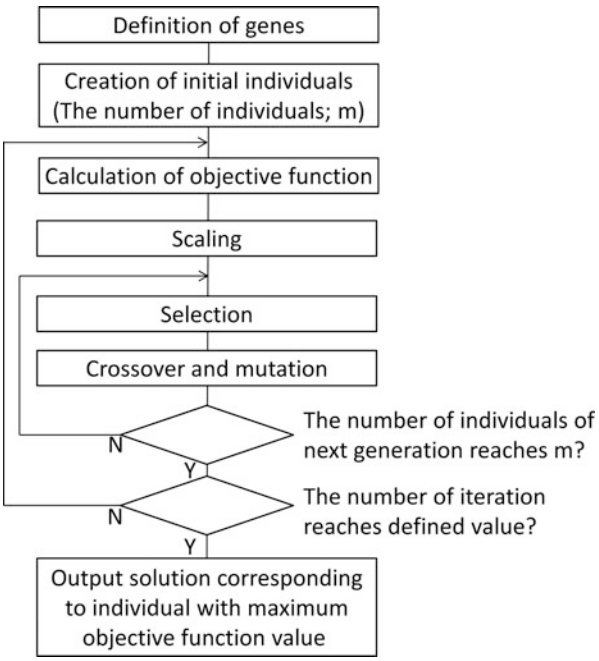


Fig. 18 Flowchart of optimization process of GA

1975), which simulates the process of biological evolution, is a powerful tool for optimization. Features of GA are as follows:

- GA can be applied to many applications by changing genes (input) and objective function (output).
- Full search of space is possible; it is not local solution.
- Multiple-point search is possible.
- Gradient information is not required for optimization.

Non-array antennas have been designed by the GA approach (Altshuler and Linden 1997; Jones and Joines 1997). GA is also applicable to array antennas, element thinning (Haupt 1994), and antenna pattern synthesis (Shimizu 1994; Yan and Lu 1997; Fujita 1999).

A flowchart of the optimization process based on standard GA is shown in Fig. 18. The GA identifies the optimum solution by creating multiple individuals that yield solutions and repeating the GA processes of selection, crossover, and mutation.

Process of GA

Definition of Genes In order to solve a problem by using GA, at first, input variables should be expressed by genes. One set of genes is called an individual. An example of an individual is

$$g_i : 01101010 \tag{54}$$

In this case, the individual consists of 8 genes. In the element thinning problem, each gene corresponds to an element antenna. A value of 0 means a thinned element, and a value of 1 means an excited element.

Initial Individuals A population consisting of multiple individuals is required for GA optimization. The first-generation population is created by random numbers. For example, when each individual has n-genes

DOI 10.1007/978-981-4560-75-7_37-1

and the population consists of m-individuals, $n \times m$ random numbers are generated. If a value close to optimal is already known, the optimal solution can be found more rapidly by using the value and its vicinity as first-generation individuals rather than random numbers.

Objective Function In GA, a new individual, which is closer to the optimal solution, is created as the next generation. In order to determine this criterion quantitatively, an objective function is defined. Though an arbitrary function can be used as the objective function, it outputs larger values as the population more closely approaches the optimal solution.

As one example of the objective function, in order to reduce sidelobe level, the objective function f can be defined as

$$f = -\max(\{S(P_i) \ \ | i = 1 \cdots k\}) \tag{55}$$

where $S(P_i)$ is sidelobe level and P_i is sampling points of angle within the sidelobe region. The solution that maximizes this objective function yields the lowest sidelobe level.

Scaling In GA, two individuals are selected, and new individuals related to them are created as the next generation. Individual selection is based on a probability that corresponds to the value of the objective function (fitness). However, when all individuals have almost the same fitness, individuals will be selected with virtually the same probability. The solution is to transform fitness so as to realize more effective selection. The procedure used for this is scaling. One approach is linear scaling. When original fitness is f and scaled fitness is f', linear scaling is expressed as

$$f' = af + b \tag{56}$$

a and b are constants and are determined so as to satisfy the following equations:

$$f'_{\text{avg}} = f_{\text{avg}} \tag{57}$$

$$f'_{\max} = Cf_{\text{avg}} \tag{58}$$

where f_{avg} is the average value of original fitness, f'_{avg} is average value of scaled fitness, and $f'_{\max}$ is the upper limit of scaled fitness. Equation 57 is introduced to ensure that the next generation includes the average individual.

Selection In GA, an individual superior to others creates many individuals in the next generation. Typical selection methods are explained.

The first one is the roulette rule. Individuals are selected with probability proportional to their fitness. When the number of individuals is n and fitness of the ith individual is f_i, the probability that this individual will be selected is

$$P_i = \frac{f_i}{\sum_{i=1}^{n} f_i} \tag{59}$$

Cumulative probability is calculated from P_i:

DOI 10.1007/978-981-4560-75-7_37-1

$$Q_i = \sum_{j=1}^{i} P_j \quad (i = 1, 2, \cdots, n) \tag{60}$$

By setting a random value r ($0 \leq r \leq 1$), minimum i satisfying $r < Q_i$ is determined, and the ith individual is selected.

The second approach is ranking. Individuals are sorted in descending order of fitness and are selected according to the ranking for the next generation.

Crossover In the crossover process, two individuals swap some of their genes. Gene position is determined in a random manner. In one-point crossover, genes are interchanged at one switching point. In multiple-point crossover, genes are interchanged at multiple switching points. In uniform crossover, all genes are randomly interchanged. Though uniform crossover has the best performance, it is complicated to implement. Note that the crossover process is conducted based on probability. Original genes are passed to the next generation if the crossover is ignored. When original individuals are g_i, g_j, and the individuals that result from crossover are g'_i, g'_j, examples of crossover are as follows.

One-Point Crossover Genes are interchanged at one switching point as per Fig. 19. The switching point is randomly determined.

Multiple-Point Crossover Genes are interchanged at multiple switching points as per Fig. 20. Following is an example of two-point crossover.

Uniform Crossover Random sequence R whose length equals that of the individuals and whose values are 0 or 1 is generated. Genes are interchanged at the positions as per Fig. 21 where the value of the random sequence is 1.

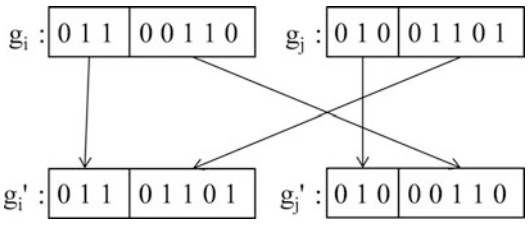


Fig. 19 One-point crossover

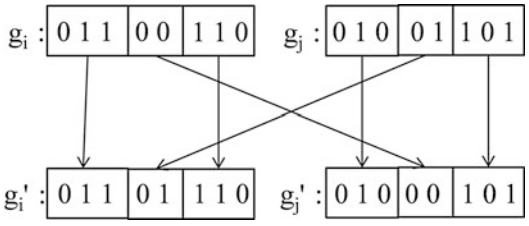


Fig. 20 Multiple-point crossover

DOI 10.1007/978-981-4560-75-7_37-1

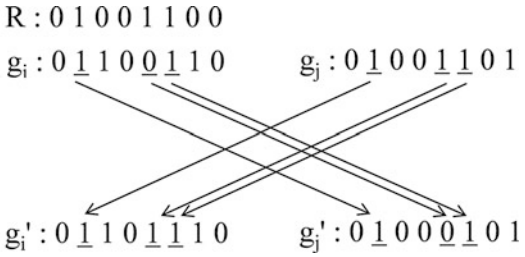


Fig. 21 Uniform-point crossover

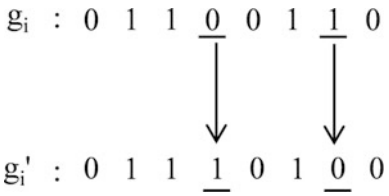


Fig. 22 Mutation

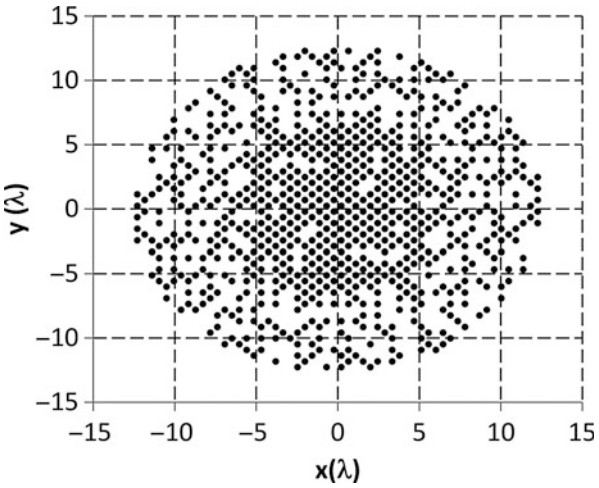


Fig. 23 Thinned aperture by using GA

Mutation The mutation process varies genes at arbitrary positions. Mutation probability per gene is defined. All genes are randomly varied according to the probability. An example of mutation is shown in Fig. 22.

Calculation Example

One example of thinning by using GA is shown. Original aperture is a circle with radius of 12.3λ. The objective is the sidelobe level of −25 dB. Moreover, the algorithm attempts to reduce the sidelobe level as much as possible.

The thinned aperture is shown in Fig. 23. Its radiation patterns are shown in Fig. 24. Radiation patterns with sidelobe level of −28 dB or lower are obtained.

DOI 10.1007/978-981-4560-75-7_37-1

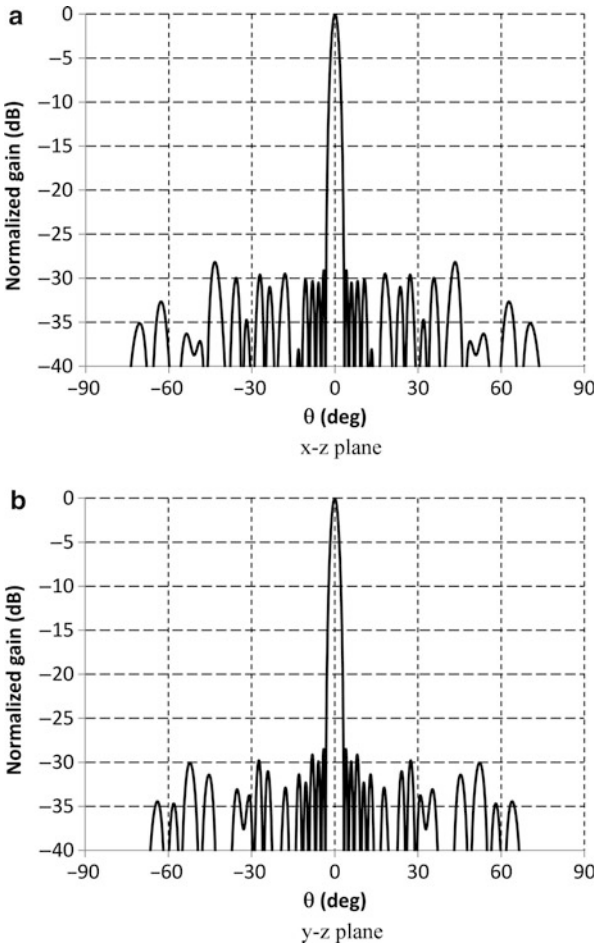


Fig. 24 Radiation pattern

Array Antenna Calibration

This chapter explains the rotating element electric field vector (REV) method, which measures the electric field from each element of an array antenna (Mano and Katagi 1982). The REV method can be adopted for array antenna calibration. It has the following features:

(1) Element electric field can be measured in the state of all elements being driven.
This means that the electric field can be measured in actual environments and so covers manufacturing error, mutual coupling, scattering from surrounding area, etc.
(2) Amplitude and phase of the element electric field can be obtained by measuring the summed power of all elements.
Amplitude and phase of the element electric field can be measured without explicit phase measurement. This feature is very effective in the millimeter wave band.

System Configuration and Measurement Principle

Typical system configuration for REV measurement is shown in Fig. 25. Received element electric field vector of each element and the summed array electric field vector are shown in Fig. 26. Here, the initial amplitude and initial phase of the summed array electric field are defined as E_0 and ϕ_0, respectively. Similarly, the initial amplitude and initial phase of the element electric field of the mth element antenna are

DOI 10.1007/978-981-4560-75-7_37-1

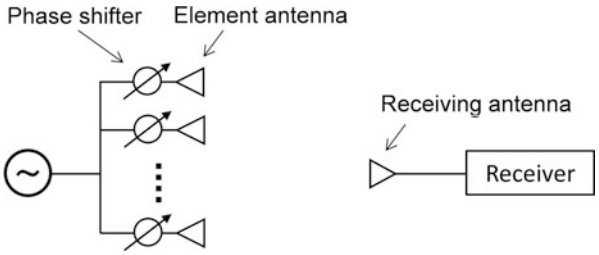


Fig. 25 System configuration for REV measurement

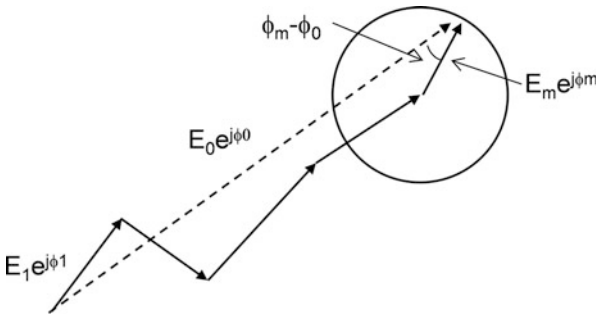


Fig. 26 Summed array electric field and element electric field

E_m and ϕ_m, respectively. Summed array electric field where the phase of the mth element antenna is rotated by Φ_m becomes

$$E = E_0 e^{j\phi 0} - E_m e^{j\phi m} + E_m e^{j(\phi m + \Phi m)} \tag{61}$$

Dividing both sides of Eq. 61 by the initial summed array electric field yields the following equation:

$$\hat{E} = \frac{E}{E_0 e^{j\phi 0}} = 1 - k_m e^{jXm} + k_m e^{j(Xm+\Phi m)} \tag{62}$$

k_m and X_m are relative amplitude and relative phase of mth element from the initial summed array electric field, respectively:

$$k_m = \frac{E_m}{E_0} \tag{63}$$

$$X_m = \phi_m - \phi_0 \tag{64}$$

From Eq. 62, relative value of the summed array electric field becomes

$$f = \left|\hat{E}\right|^2 = \left(Y^2 + {k_m}^2\right) + 2k_m Y \cos\left(\Phi_m + \Phi_{m,0}\right) \tag{65}$$

where

DOI 10.1007/978-981-4560-75-7_37-1

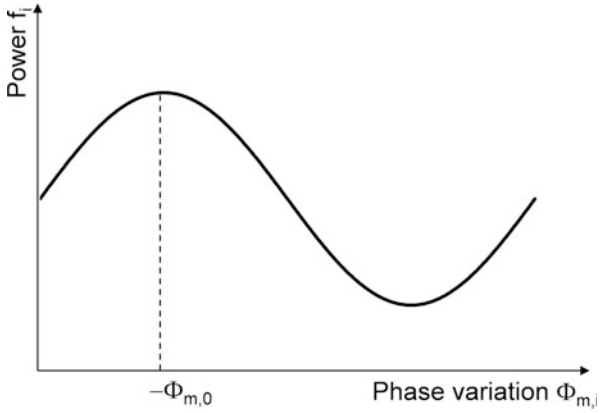


Fig. 27 Power variation of summed array electric field

$$Y^2 = (\cos X_m - k_m)^2 + \sin^2 X_m \tag{66}$$

$$\tan \Phi_{m,0} = \frac{\sin X_m}{\cos X_m - k_m} \tag{67}$$

Therefore, the "phase variation of one element" creates a "cosine-shaped power variation in the summed array electric field." The power variation is shown in Fig. 27. $-\Phi_{m,0}$ in the figure is the phase when the power of the summed array electric field is maximized. In the REV method, relative amplitude k_m and relative phase X_m of the mth element can be estimated from the phase, $-\Phi_{m,0}$, and the ratio of maximum power of the summed array electric field to its minimum, r^2. r^2, is expressed as

$$r^2 = \frac{(Y + k_m)^2}{(Y - k_m)^2} \tag{68}$$

and so r has two solutions:

$$r = \pm \frac{Y + k_m}{Y - k_m} \tag{69}$$

Due to the double sign of Eq. 69, two candidates of relative amplitude k_m and relative phase X_m are obtained:

(A) Solution with plus sign

$$k_m = \frac{\Gamma}{\sqrt{1 + 2\Gamma \cos \Phi_{m,0} + \Gamma^2}} \tag{70}$$

$$\tan X_m = \frac{\sin \Phi_{m,0}}{\cos \Phi_{m,0} + \Gamma} \tag{71}$$

(B) Solution with minus sign

DOI 10.1007/978-981-4560-75-7_37-1

$$k_m = \frac{1}{\sqrt{1 + 2\Gamma \cos \Phi_{m,0} + \Gamma^2}} \tag{72}$$

$$\tan X_m = \frac{\sin \Phi_{m,0}}{\cos \Phi_{m,0} + 1/\Gamma} \tag{73}$$

Γ in Eqs. 70, 71, 72, and 73 is

$$\Gamma = \frac{r-1}{r+1} \tag{74}$$

There is an ambiguity in determining the sign in Eq. 69. The ambiguity is eliminated by obtaining solutions for both signs under two different initial phase distribution states. The solution in the first state and the solution in second state will be identical. They are assumed to be the true solution. This method requires only amplitude measurement. If both amplitude and phase measurement are available, the sign ambiguity be resolved with a single measurement.

Application of REV Method to Actual System

From the previous section, in the REV method, amplitude and phase of electric field of mth element can be estimated from the phase, $-\Phi_{m,0}$, that maximizes the power of the summed array electric field and the ratio r of maximum power of the summed array electric field to its minimum. $-\Phi_{m,0}$ and r are obtained by changing the excitation phase of element antennas. However, it's not always true that the maximum power and minimum power of the summed array electric field can be measured exactly, because actual APAAs use digital phase shifters, which output discrete phases. Additionally, the phase shifter has output phase error for each input phase. Therefore, measured power of the summed array electric field fluctuates from the ideal cosine-shaped curve as per Fig. 28.

A cosine term can be obtained from measured values by using Fourier series expansion. When the excitation phase of element m is $-\Phi_{m,\mathrm{i}}$ and its measured power of the summed array electric field is f_i, the cosine curve yielded by Fourier series expansion is expressed as

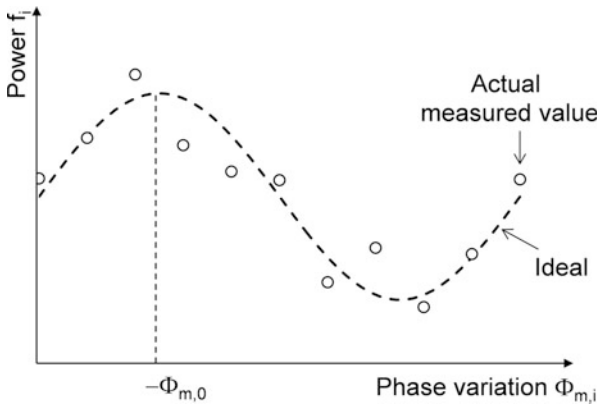


Fig. 28 Example of the power variation of summed array electric field using digital phase shifter

DOI 10.1007/978-981-4560-75-7_37-1

$$f_i = \frac{\alpha}{2} + c\cos\Phi_{m,i} + s\sin\Phi_{m,i} \tag{75}$$

Fourier coefficients are given as follows:

$$\alpha = \sum_{i=1}^{N} f_i \tag{76}$$

$$c = \sum_{i=1}^{N} f_i \cos\Phi_{m,i} \tag{77}$$

$$s = \sum_{i=1}^{N} f_i \sin\Phi_{m,i} \tag{78}$$

N is the number of measurements made while changing excitation phase. N is generally the number of states of the phase shifter. r and $-\Phi_{m,0}$ are calculated as follows:

$$\tan\Phi_{m,0} = -\frac{s}{c} \tag{79}$$

$$r^2 = \frac{\alpha + 2\sqrt{c^2+s^2}}{\alpha - 2\sqrt{c^2+s^2}} \tag{80}$$

Measurement Time Reduction for REV Method

The basic REV method, explained in detail in the previous chapter, has been enhanced in several ways. As one example, measurement time reduction (Takahashi et al. 2008) is introduced. While element phases are changed one at a time in the basic REV method, an enhancement changes the phases of multiple elements with different phase cycles, and the power variation of the summed array electric field is measured. The amplitudes and phases of the elements are estimated by calculation. The time taken to conduct measurements, a significant problem of the basic REV method, is drastically shortened. Though this method does increase measurement error, it is not of practical importance because the error of the original REV is very small. However, this method should be applied with reference to the theoretical error.

True Time Delay Calibration

Previous chapters explained the "phase" calibration method. Array antennas for large aperture and wideband systems suffer from main beam squint. The ideal excitation phase of each element antenna differs with frequency because k in Eq. 4 varies according to frequency (or wavelength). Unfortunately, a standard phase shifter outputs constant phase within the frequency bandwidth. Squint is prevented if signal transmission timing is controlled instead of phase. Such timing control is called true time delay (TTD). Due to cost limitations, TTD is generally applied only to subarrays of element antennas, while phase shifters are set to all element antennas. Such systems require both phase calibration for each element antenna and delay calibration for each subarray. For phase calibration, the REV method explained in the previous chapter can be used. For delay calibration, a method of estimating the delay difference between two subarrays through space has been proposed (Maruyama et al. 2014). Two subarrays transmit a signal, while an already known delay is given to one subarray. When the transmitting frequency is swept, a cosine-shaped response is output. Total delay difference between the subarrays is estimated from the

DOI 10.1007/978-981-4560-75-7_37-1

response. Subtracting the already-known delay from the total delay difference yields the original delay difference. This method does not require phase measurement, while received level variation is referenced. This characteristic is similar to that of the REV method.

Digital Beam Forming (DBF)

Overview and Functions of DBFs

DBF antennas combine array antennas with digital signal processing (Steyskal 1987; Farina 1992; Litva and Lo 1996). Such an antenna can perform a variety of intelligent processes because it uses digital signal processing for beam forming, etc. The major functions that DBFs can provide are listed below:

(1) High-precision beam pointing
(2) Spatial filtering
(3) Spatial division multiplexing

(1) is a function for pointing the main lobe with high precision in the desired signal's direction and a function that ensures low sidelobe levels. DBFs can reproduce signals more easily than APAAs and so can easily realize multibeam forming.

(2) is a function for spatially eliminating an undesired signal by pointing a null in the antenna pattern in the appropriate direction. This results in the receiver achieving a better signal-to-noise ratio (SNR).

(3) refers to transmitting different information from multiple antennas and receiving signals on multiple antennas; this is called the MIMO (multi-input multi-output) technique. This can help expand transmission capacity and improve frequency usage efficiency.

Basic Configuration of DBF Antennas

Figure 29 shows the basic configuration of a DBF antenna. The receiving antenna consists of antenna elements, low-noise amplifiers, down converters, analog-to-digital converters, and digital signal processors.

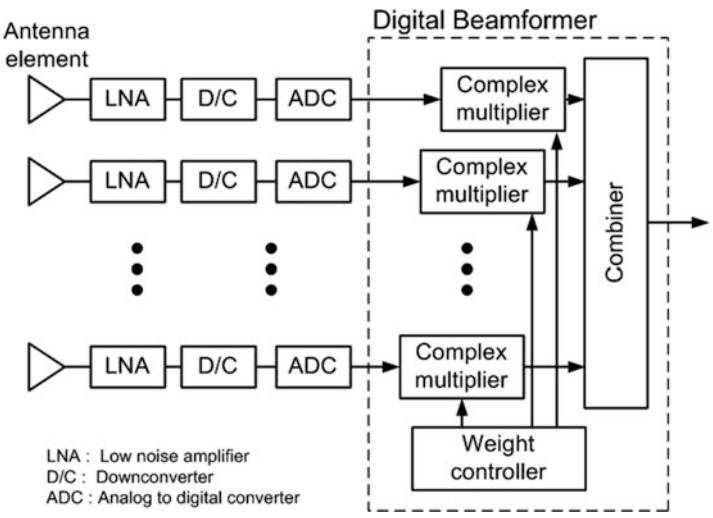


Fig. 29 Configuration of DBF antenna (receiving side)

DOI 10.1007/978-981-4560-75-7_37-1

Antenna Elements

The antenna elements and element arrangement basically follow APAA designs. However, depending on the operating mode (such as whether to achieve high directionality or antenna diversity), element spacing may be about one wavelength.

Low-Noise Amplifiers (LNAs)

LNAs amplify the received signals; they are essential to DBF antennas. They are usually located immediately after the antenna elements. The NF (noise figure), an indicator of input deterioration, and the output signal-to-noise ratio (SNR) determine the performance in the vicinity of the receiving sensitivity point.

Down Converters

A down converter uses local signals generated inside the receiver to convert the received signals in the carrier wave frequency band into intermediate frequencies, etc. A low-pass filter is placed immediately after it because a high-frequency component is generated by frequency conversion.

Analog-to-Digital Converters

An analog-to-digital converter is located at the last stage of an analog circuit. It converts analog signals to digital ones. Sampling frequency and bit precision are important parameters. The analog-to-digital converter adopted is selected based on factors such as the degree of deterioration in receiving performance, based on received signal bandwidth and quantum error.

Digital Signal Processors

The digital signal processor provides two functions. The first is calculating the weight (that is, the amplitude phase adjustment value) to be applied to the signal received on each element. Specifically, this includes calculating the phase value applied to each element in order to make the wave front of the received signal uniform in the main beam direction and calculating the amplitude phase value to point nulls in the direction of arrival of undesired signals. The second function is multiplying the weight to each element's output to get the synthesized output.

There are two ways of implementing the hardware for digital signal processing. The first is the use of an ASIC (application-specific integrated circuit). An ASIC is a dedicated circuit with a particular use. Its wiring is physically fixed, so it cannot be changed once it is produced, but using such a circuit can lower the energy consumption. The other way is to use an FPGA (field-programmable gate array). FPGA circuits are built by hardware programming. The circuits can be rewritten, which makes them more flexible than ASICs. Software-based digital signal processing is also possible by using a DSP (digital signal processor) or a CPU/MPU (central/micro processing unit).

The above has briefly explained the structure of DBF receiving antennas. To make a transmitting DBF antenna, one need only replace the low-noise amplifiers with high-power amplifiers, the down converters with up converters, and the analog-to-digital converters with digital-to-analog converters.

Characteristics of DBF Antennas

One advantage of that DBF antennas is that they can provide a wide variety of functions because they use digital signal processing. They also allow for a simpler feed network and suffer less loss than is the case with APAA technology. Functions can be consolidated into digital devices, enabling a high degree of integration. As digital devices become more advanced, moreover, DBF antennas can also take advantage of lower costs. One disadvantage is that the processing bandwidth is limited by the performance of the analog-to-digital converter or digital-to-analog converter. High-speed analog-to-digital converters have

DOI 10.1007/978-981-4560-75-7_37-1

been developed in recent years, and their communications bands are widening. This is generally costly, however, because each antenna element requires an analog-to-digital or digital-to-analog converter and a down converter or up converter. Because there are so many digital devices and the amount of data processed is growing, digital signal processors are consuming more energy. For that reason, array antennas with a large number of elements often have a subarray structure such that they first synthesize multiple antenna elements at the analog stage and then perform DBF.

Adaptive Arrays

Overview of Adaptive Arrays

If the antenna element can independently find the correct weight and set it automatically, the optimal communications environment becomes possible at all times. This idea was proposed a long ago. Starting in the 1960s, there were many reports, starting with the topic of the sidelobe canceler (SLC) (Howells 1965; Applebaum 1976). At first, the idea was certainly interesting and seemed effective theoretically, but implementing such technology required massive computers and microwave circuits, so in most cases, it did not advance much beyond the theoretical stage.

However, digital technology is advancing rapidly, and the integration of microwave circuits made them less costly and more compact. At the same time, mobile communications were developing, making the radio environment more challenging by the day. These events are highlighting adaptive array technology more and more.

The sidelobe canceler, which might be called the origin of adaptive arrays, was devised by Howells. The sidelobe canceler was intended to enable radar to operate even when there was strong interference. As shown in Fig. 30, it consists of a main, high-gain antenna and K auxiliary antennas. The auxiliary antenna gains are designed to approximate the average sidelobe level of the main antenna gain pattern. The amount of desired signal received by the auxiliary antennas is negligible compared to the desired signal in the main antenna. By controlling the weights on the auxiliary antennas, an interference in the sidelobe of the main antenna may be nulled out.

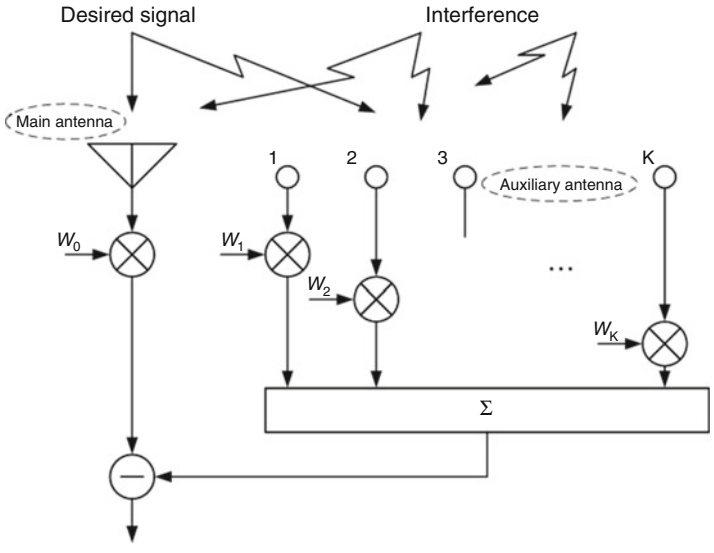


Fig. 30 Configuration of sidelobe canceler

DOI 10.1007/978-981-4560-75-7_37-1

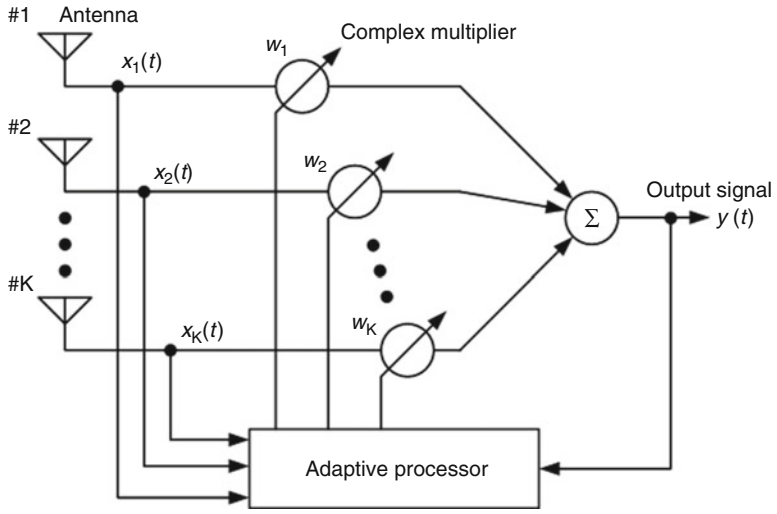


Fig. 31 Configuration of an adaptive array

Functions of Adaptive Arrays

Figure 31 shows the configuration of an adaptive array made up of K elements (Compton 1988; Monzingo and Miller 1980; Widrow et al. 1967; Kikuma and Fujimoto 2003; Ogawa and Ohgane 2001). Weight is first applied to each antenna element's input signal. An adder then synthesizes these into the array output. Each weight is a complex value that can control the amplitude and phase of the antenna elements' input signal.

Vector notation is used in the following formulas, with the desired signal, interference, and thermal noise at kth element labeled $s_k(t)$, $i_k(t)$, and $n_k(t)$, respectively:

$$\begin{aligned} \mathbf{S}(t) &= [s_1(t)\ s_1(t)\ \cdots\ s_K(t)]^T \\ \boldsymbol{I}(t) &= [\mathrm{i}_1(t)\ \mathrm{i}_2(t)\ \cdots\ \mathrm{i}_K(t)]^T \\ \boldsymbol{N}(t) &= [\mathrm{n}_1(t)\ \mathrm{n}_2(t)\ \cdots\ \mathrm{n}_K(t)]^T \end{aligned} \tag{81}$$

Here, subscript t refers to transposition.

Using these formulas, input signal vector $\boldsymbol{X}(t)$, with input signals in vector notation, is given by the following formula:

$$\mathbf{X}(t) = [\,x_1(t) \quad x_2(t) \quad \cdots \quad x_K(t)\,]^T = \mathbf{S}(t) + \mathbf{I}(t) + \mathbf{N}(t) \tag{82}$$

Similarly, given weight w_k, the vector notation yields the following formula:

$$\mathbf{W} = [\,w_1 \quad w_2 \quad \cdots \quad w_K\,]^T \tag{83}$$

Array output $y(t)$ is therefore expressed as the following formula:

$$\mathrm{y}(t) = \mathbf{W}^{\mathrm{H}}\mathbf{X}(t) \tag{84}$$

Subscript H represents complex conjugate transposition. Output power P_{out} is found as in the following formula:

DOI 10.1007/978-981-4560-75-7_37-1

$$P_{out} = \frac{1}{2}E\left[|y(t)|^2\right] = \frac{1}{2}\mathbf{W}^{H}R_{xx}\boldsymbol{W} \tag{85}$$

$\mathbf{R}_{xx}$ is the correlation matrix of the input signal vector and is defined by the following formula:

$$\mathbf{R}_{xx} = E\left[\mathbf{X}(t)\mathbf{X}^{\boldsymbol{H}}(t)\right] \tag{86}$$

$E[]$ represents the expected value.

The output SINR (signal-to-interference-plus-noise ratio) is defined by the following formula. This is normally used as a metric for evaluating an array's output characteristics. The higher the value, the better its characteristics:

$$SINR = \frac{\text{Desired signal power}}{\text{Interference power} + \text{Noise power}} \tag{87}$$

The functions of an adaptive array can be classified into the following two, depending on the objective:

(1) Adaptive beam forming
(2) Adaptive null steering

The first enables automatic tracking of the array's main beam, even if the direction of arrival of the received signal (desired signal) is unknown or changes with time. The second automatically forms the antenna pattern's null points in the direction of interference. There have been many examples of research into null steering. The sidelobe canceler mentioned earlier was the result of some of the earliest research.

Optimization Algorithm

An adaptive array changes directivity and frequency characteristics to conform to the radio environment as it acquires information about that environment. Therefore, it is not necessary to know the undesired signals (interference) in advance. However, extracting just the desired signal from a poor radio environment that also contains undesired signals and noise does require some a priori information of the desired signal. Data needed include the signal's center frequency, direction of arrival, modulation method, polarization used, and training signals used. The following is an explanation of the main optimization algorithms.

Maximal Ratio Combining

Maximal ratio combining (MRC) detects the envelope level and phase information of the signal received on each antenna element (Miura et al. 1999; White 1976). Each received signal is set to the same phase and weighted according to its envelope level. This maximizes the SNR after combining. This was developed as a diversity reception technology to mitigate the effect of fading. It can scan for the main beam to track incoming waves, but it does not suppress interference. Several methods proposed such as closed loop control using a receiving level monitor, control using a direction sensor, and phase control using a feedback loop (PLL system). Moreover, it is important to achieve stable operations in environments with low SNR (Kihira et al. 2006).

MMSE

The development of adaptive arrays based on the minimum mean square error (MMSE) method started when Widrow proposed an adaptive filter in the 1960s. Widrow applied adaptive array technology and

DOI 10.1007/978-981-4560-75-7_37-1

established the least mean square (LMS) algorithm (Widrow et al. 1967). MMSE adaptive arrays work to minimize the error signal, that is, the difference between a reference signal and the actual array output signal. This system can perform the two major functions of adaptive arrays (adaptive beam forming and adaptive null steering) simultaneously and is not limited by the element arrangement. Strictly speaking, the desired signal itself is required as a reference signal, but in actuality, this system works by using something, such as a known training symbol that is inserted, for example, in front of the data.

The error signal of reference signal $r(t)$ (desired time response) and array output signal $y(t)$ can be written as the following formula:

$$\begin{aligned} e(t) &= r(t) - y(t) \\ &= r(t) - \mathbf{W}^H \mathbf{X}(t) \end{aligned} \tag{88}$$

The expected value of the square of this error signal (mean square error) is given by the following formula:

$$\begin{aligned} E\left[|e(t)|^2\right] &= E\left[|r(t) - y(t)|^2\right] = E\left[\left|r(t) - \mathbf{W}^{\mathrm{H}}\mathbf{X}(t)\right|^2\right] \\ &= E\left[|r(t)|^2\right] - \mathbf{W}^T \mathbf{r}^*_{xr} - \mathbf{W}^{\mathrm{H}}\mathbf{r}_{xr} + \mathbf{W}^{\mathrm{H}}\mathbf{R}_{\mathbf{xx}}\mathbf{W} \end{aligned} \tag{89}$$

$\boldsymbol{r}_{xr}$ is the correlation vector of the reference signal and input signal vector and is defined by the following formula:

$$\mathbf{r}_{xr} = E[\mathbf{X}(t) r^*(t)] \tag{90}$$

In this formula, the weight vector is a quadratic function, and to minimize its mean square error, it is necessary to satisfy the following formula:

$$\frac{\partial}{\partial \mathbf{W}} E\left[|e(t)|^2\right] = -2\mathbf{r}_{xr} + 2R_{xx}\mathbf{W} = \mathbf{0} \tag{91}$$

Therefore, the optimal weight is given by the following formula:

$$\mathbf{W}_{opt} = \mathbf{R}_{\mathbf{xx}}{}^{-1}\mathbf{r}_{xr} \tag{92}$$

Several optimization techniques using digital control have been proposed, such as the LMS algorithm based on the steepest descent method (Haykin 2002); sample matrix inversion (SMI), which uses sample values to find a solution directly; and recursive least squares (RLS) (Compton 1988).

Figure 32 shows an example of an antenna pattern after convergence in a four-element, two-incoming wave model. The main beam is formed in the direction of arrival of the desired signal (0°), and a deep null is pointed at the direction of arrival of interference (60°). Figure 33 shows the output power after array synthesis of each signal in the LMS algorithm. The horizontal axis is the number of weight updates (iterations). The results clearly show that the desired signal's output power is maintained, while the interference's output power is diminished.

Having a reference signal is a precondition for the MMSE adaptive array, so it is more effective as a means for achieving a spatially optimized filter in a communication system than in radar. A training symbol inserted into a data signal in a communications system using TDMA can be used as a reference signal. This potential is gaining MMSE attention for applications in mobile communications.

DOI 10.1007/978-981-4560-75-7_37-1

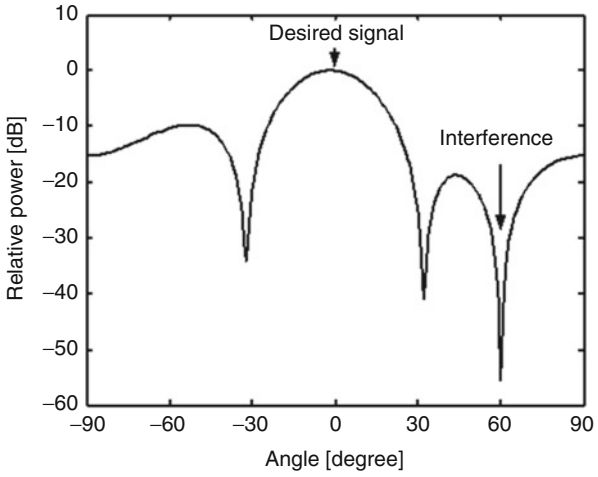


Fig. 32 Antenna pattern after convergence

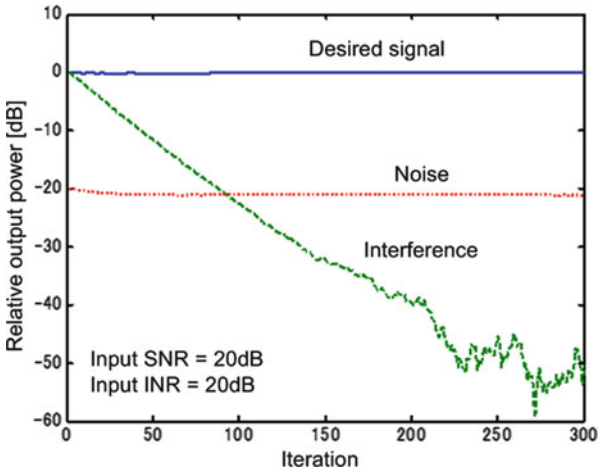


Fig. 33 Output power after array synthesis

MSN

Applebaum took the sidelobe canceler developed by Howells and developed it into the MSN (maximum SNR method) adaptive array; this is also referred to as the Howells-Applebaum (HA) adaptive array (Applebaum 1976). Because MSN adaptive arrays were applied to radar systems, a precondition is that the direction of arrival of the desired signal must already be known.

This algorithm finds the weight that maximizes output SNR. The desired signal power, interference power, and thermal noise power in the array output are expressed in the following formula:

$$\begin{aligned} P_{Sout} &= \frac{1}{2}E\left[|y_s(t)|^2\right] = \frac{1}{2}\mathbf{W}^H\mathbf{R}_{SS}\mathbf{W} \\ P_{Iout} &= \frac{1}{2}E\left[|y_i(t)|^2\right] = \frac{1}{2}\mathbf{W}^H\mathbf{R}_{ii}\mathbf{W} \\ P_{Nout} &= \frac{1}{2}E\left[|y_n(t)|^2\right] = \frac{1}{2}P_n\mathbf{W}^H\mathbf{W} \end{aligned} \tag{93}$$

DOI 10.1007/978-981-4560-75-7_37-1

Here, $\mathbf{R}_{ss}$ and $\mathbf{R}_{ii}$ are the correlation matrices for the desired signal and interference, respectively, while P_n is the thermal noise power per antenna element. If the direction of arrival of the desired signal is known, the following formula can be established if it is assumed that the desired signal's bandwidth is narrow enough:

$$\begin{aligned}\mathbf{S}(t) &= s(t)\mathbf{V}_s \\ \mathbf{V}_s &= \left[v_{s1}\ v_{s2}\cdots v_{sK}\right]^T \\ \mathbf{R}_{ss} &= E\left[\mathbf{S}(t)\mathbf{S}^H(t)\right] = P_s\mathbf{V}_s\mathbf{V}_s^H\end{aligned} \tag{94}$$

Here, P_s is the desired signal power per element.

Therefore, the output SINR can be found using the following formula:

$$SINR = \frac{P_{Sout}}{P_{Iout} + P_{Nout}} = \frac{\mathbf{W}^H\mathbf{R}_{SS}\mathbf{W}}{\mathbf{W}^H\mathbf{R}_{nn}\mathbf{W}} \tag{95}$$

Here, R_{nn} is the correlation matrix for the undesired signal component as expressed below:

$$\mathbf{R}_{nn} = \mathbf{R}_{ii} + \mathbf{P}_n\mathbf{I} \tag{96}$$

Note that $\mathbf{I}$ is the identity matrix of dimension K.

The optimal weight, W_{opt}, that maximizes the SINR satisfies the following equation:

$$\nabla\left(\frac{\mathbf{W}^H\mathbf{R}_{ss}\mathbf{W}}{\mathbf{W}^H\mathbf{R}_{nn}\mathbf{W}}\right) = \mathbf{0} \tag{97}$$

Solving this yields the following formula:

$$\mathbf{W}_{opt} = \alpha\mathbf{R}_{nn}^{-1}\mathbf{V}_s \tag{98}$$

Here, α is a constant.

Finding the optimal weight requires separating the desired signal from the interference and then finding their correlation matrix $\mathbf{R}_{\mathrm{nn}}$. With radar, assuming that interference is much greater than the desired signal, one can take as a substitute the correlation matrix of the received signal (a signal that contains both the desired signal and interference).

As indicated above, MSN adaptive arrays require setting the desired signal's direction of arrival in advance, and so one important question is how to get correct direction information for the desired signal. Techniques have been proposed in recent years that estimate the direction of arrival with high precision; examples include MUSIC (multiple signal classification) (Schmidt 1986) and ESPRIT (estimation of signal parameters via rotational invariance techniques) (Roy et al. 1986). These estimates can be used for preprocessing MSN control. This, it is believed, would make it possible to use MSN adaptive arrays even in such applications as mobile communications. However, the direction of arrival estimation errors causes the performance of MSN adaptive arrays to degrade sharply. The so-called tamed adaptive antenna, which adds "pseudo"-noise to the correlation matrix, offers greater resistance to these errors (Takao and Kikuma 1986).

DCMP

This algorithm was proposed by Frost (Frost 1972) and further expanded by Takao et al. (Takao et al. 1972). If the direction of arrival, frequency, and array arrangement is decided, then the array

DOI 10.1007/978-981-4560-75-7_37-1

response vector is decided. In DCMP, this is treated as constraint vector **C**, and the following formula's constraint conditions are imposed.

$$\mathbf{C}^T\mathbf{W}^* = H \tag{99}$$

Here, H is called the constraint response value; it is a constant that determines the array response value in direction θ_s.

The DCMP adaptive array determines the weight vector that minimizes output power under these conditions. In other words, if the array response in the direction of arrival of the desired signal is constrained and the output power is minimized, then interference is suppressed. The weight vector that satisfies this is given by the following formula:

$$\mathbf{W}_{opt} = \mathbf{R}_{xx}^{-1}\mathbf{C}\left(\mathbf{C}^H R_{xx}^{-1}\mathbf{C}\right)^{-1}H^* \tag{100}$$

Moreover, if constraint conditions must be imposed in multiple directions, this can be handled by using a constraint matrix in place of a constraint vector and by rendering the constraint response value into vector form.

DCMP adaptive arrays work with the direction of arrival of the desired signal already known, but in reality, there are cases when there is some variance between the designated direction of arrival and the actual angle of arrival, for example, because of the error introduced when the antenna was installed or because of the movement in the antenna itself. Similar problems can also be caused by mutual coupling between antenna elements. It is possible to suppress this problem by using a tamed adaptive antenna (Takao and Kikuma 1986), which, as mentioned above, adds pseudo noise to the correlation matrix.

Power Inversion

Conditions for applying a power inversion adaptive array (PIAA) (Compton 1979) are as follows:

- The array's degrees of freedom (one less than the number of elements) must be the same as the number of interference signals.
- The desired signal's output power must be (sufficiently) smaller than the interference power.

Power inversion can be used if the above conditions are satisfied. In other words, the desired signal and interference are distinguished from each other merely from the difference in their input powers, so a priori information, such as the reference signal needed in MMSE, is not required.

Under the above conditions, minimizing the output power is a direct way of suppressing the unnecessary component. As is, however, this means that all weights are zero, which does not allow the antenna to receive even the desired signal. Under PIAA, the simplest conditions are imposed in order to eliminate trivial solutions. Under this rule, in other words, the minimal output power is achieved when one antenna element's weight is set at a fixed value. Thus, PIAA is a kind of output power minimization method with constraints. Following this principle, the greater the power of the interference is, the more strongly it is suppressed.

The optimal weight when the weight of element 1 is fixed is given by

$$\begin{aligned}\mathbf{W}_{opt} &= \mathbf{R}_{xx}^{-1}\overline{\mathbf{S}} \\ \overline{\mathbf{S}} &= [1,\, 0, \cdots,\, 0]^T\end{aligned} \tag{101}$$

DOI 10.1007/978-981-4560-75-7_37-1

Here, $\overline{\mathbf{S}}$ is a constraint vector.

Its application in mobile communications is also being studied because PIAA is a blind algorithm that does not use a priori information, which allows simpler configurations and processing (Kihira et al. 2003).

CMA

Whereas MMSE adaptive arrays require a reference signal, phase modulation signals like PSK, which are currently widely used for mobile communications, basically have a constant envelope characteristic. Taking advantage of this feature, CMA controls antenna weights by making the array output signal's envelope a desired fixed value (Treichler and Agee 1983). Therefore, it is possible to obtain the optimal weight by minimizing the evaluation function expressed by the following formula:

$$Q(\mathbf{W}) = E[|\,|y(t)|^p - \sigma^p\,|^q] \tag{102}$$

Here, σ is the desired envelope value, and p and q are positive integers, usually 1 or 2. The solution $\mathbf{W}$ that minimizes evaluation function Q cannot be found analytically, and so repeated operations are needed using a nonlinear optimization method. If the steepest descent method is used as the operation algorithm, weight updates follow the following formula ($p = q = 2$):

$$\mathbf{W}(m+1) = \mathbf{W}(m) - \mu 4\mathbf{X}(m)y^*(m)\left(|y(m)|^2 - \sigma^2\right) \tag{103}$$

Here, μ is a constant that expresses the feedback step size, and m is the number of times (iterations) a weight is updated.

CMA adaptive arrays allow blind processing; no a priori information is needed. Therefore, many studies have been performed on applications for mobile communications.

MIMO Technology

MIMO technology uses multiple antennas (that is, an array antenna) for both the transmitter and receiver (Foschini and Gans 1998; Telatar 1999). This makes it possible to achieve either (i) faster transmission speed or (ii) greater reliability or both (Paulaj et al. 2003). In the case of (i), the technology is described as spatial division multiplexing (SDM). In the case of (ii), it is described as spatial diversity (SD). Particularly in the case of (i) SDM, multiple different signal data streams are transmitted from multiple antennas, meaning that they are spatially multiplexed. At the receiving side, the multiple signals are separated by signal processing technology. With SDM, as the number of antennas at both the transmitting and receiving sides is increased, the frequency usage efficiency increases at a rate that is proportional to the number of antennas. These results have been demonstrated experimentally (Nishimori et al. 2005; Taoka et al. 2007). MIMO technology is now indispensable to wireless communication systems.

In recent years, furthermore, there has been much study of multiuser MIMO (MU-MIMO) technology (Spencer et al. 2004). This technology uses many antennas at the base station to enable MIMO communication among multiple users and a base station. MU-MIMO is being studied in terms of IEEE 802.11 ac (which is being studied as the next-generation LAN standard) and LTE-Advanced (which is being studied as the next-generation mobile communications standard), among others. For MU-MIMO comparison purposes, a scenario in which one user communicates by MIMO with a base station is called single-user MIMO (SU-MIMO).

DOI 10.1007/978-981-4560-75-7_37-1

MIMO technology, originally developed for communication, is now also being studied for applications in radar (Robey et al. 2004). This research into MIMO radar is also gaining much attention.

Conclusion

This chapter has introduced the fundamental points of PAAs. One of the important features of PAAs is pattern synthesis. Some methods of pattern synthesis were explained. In order to operate PAAs correctly, element antenna calibration is an essential procedure. Array calibration methods were also explained. In the later part of this chapter, DBF including signal processing and MIMO techniques were introduced. All types of devices have advanced in recent years, and the various DBF prototypes and DBF use in practical fields have earned high praise. This has created fertile ground for practical applications of DBF. Very recently, products have appeared that apply MIMO technology. This is one example of how research and development for the large-scale diffusion of this technology is speeding up.

Cross-References

► Multibeam Antenna Arrays
► Reflectarray Antennas

References

Altshuler EE, Linden DS (1997) Wire-antenna designs using genetic algorithms. IEEE Antennas Propag Mag 39(2):33–43
Amitay N, Galindo Y, Wu C (1972) Theory and analysis of phased array antennas. Wiley, New York
Applebaum SP (1976) Adaptive arrays. IEEE Trans Antennas Propag AP-24(5):585–598
Chiba I, Mano S (1987) Null forming method by phase control of selected array elements using plane-wave synthesis. In: Proceeding of IEEE AP-S,Blacksburg, VA, pp 70–73
Chiba I, Miura R, Tanaka T, Karasawa Y (1997) Digital beam forming (DBF) antenna system for mobile communications. IEEE Aerosp Electron Syst Mag, Blacksburg, VA 12(9):31
Compton RT Jr (1979) The power inversion adaptive array: concept and performance. IEEE Trans Aerosp Electron Syst AES-15(6):803–814
Compton RT Jr (1988) Adaptive antennas: concepts and performance. Prentice Hall, Englewood Cliffs
Dolph CL (1946) A current distribution for broadside arrays which optimizes the relationship between beam width and sidelobe level. Proc IRE 34:335–348
Elliott RS (2003) Antenna theory and design revised edition. Wiley, Hoboken
Farina A (1992) Antenna based signal processing techniques for radar systems. Artech House, Boston
Foschini GJ, Gans MJ (1998) On limits of wireless communications in a fading environment when using multiple antennas. Wirel Pers Commun 6:311–335
Frost OL III (1972) An algorithm for linearly constrained adaptive array processing. Proc IEEE 60(8):926–935
Fujita M (1999) A trinary-phased array. IEICE Trans Commun E82-B:564–566
Hansen RC (1966) Microwave scanning antennas II, Chap. 1, 3, 4. Academic, New York
Haupt RL (1994) Thinned arrays using genetic algorithms. IEEE Trans Antennas Propag AP-42:993–999
Haykin S (2002) Adaptive filter theory, 4th edn. Prentice Hall, Upper Saddle River

Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor
Holter H, Steyskal H (1999) Infinite phased-array analysis using FDTD periodic boundary conditions-pulse scanning in oblique direction. IEEE Trans Antennas Propag 47(10):1508–1514
Howells PW (1965) Intermediate frequency sidelobe canceller. US Patent 3,202,990
Jones EA, Joines WT (1997) Design of Yagi-Uda antennas using genetic algorithms. IEEE Trans Antennas Propag 45(9):1386–1392
Kihira K, Yonezawa R, Chiba I (2003) A simple configuration of adaptive array antenna for DS-CDMA systems. IEICE Trans Commun E86-B(3):117–1124
Kihira K, Chiba I, Konishi Y, Masuda H, Sato H, Ikematsu H, Miyashita H, Makino S (2006) A Ka-band DBF array antenna for land mobile satellite communications. IEICE Trans Commun J89-B (9):1705–1716, (in Japanese)
Kikuma N, Fujimoto M (2003) Adaptive antennas. IEICE Trans Commun E86-B(3):968–979
Klein C (1984) Design of shaped-beam antennas through minimax gain optimization. IEEE Trans Antennas Propag 32(9):963–968
Litva J, Lo KY (1996) Digital beamforming in wireless communications. Artech House, Boston
Mailloux RJ (2005) Phased array antenna handbook, 2nd edn. Artech House, Boston
Mano S, Katagi T (1982) A method for measuring amplitude and phase of each radiating element of a phased array antenna. Electron Commun Jpn 65(5):58–64
Marpaung D, Zhuang L, Burla M, Roeloffzen C, Verpoorte J, Schippers H, Hulzinga A, Jorna P, Beeker WP, Leinse A, Heideman R, Noharet B, Wang Q, Sanadgol B, Baggen R (2011) Towards a broadband and squint-free ku-band phased array antenna system for airborne satellite communications. In: Proceedings of 5th European Conference on Antennas and Propagation (EUCAP), Rome, Italy, pp 2623–2627
Maruyama T, Yamaguchi S, Takahashi T, Otsuka M, Miyashita H (2014) Estimation of delay time difference through space for phased array antennas with true time delay. IEICE Commun Express 3(6):200–205
Miura R, Tanaka T, Horie A, Karasawa Y (1999) A DBF self-beam steering array antenna for mobile satellite applications using beam-space maximal-ratio combination. IEEE Trans Veh Technol 48(3):665–675
Monzingo RA, Miller TW (1980) Introduction to adaptive arrays. Wiley, New York
Nishimori K, Kudo R, Takatori Y, Tsunekawa K (2005) Evaluation of 8 × 4 eigenmode SDM transmission in broadband MIMO-OFDM systems. NTT Tech Rev 3(9):50–59
Numazaki T, Mano S, Katagi T, Mizusawa M (1987) An improved thinning method for density tapering of planar array antennas. IEEE Trans Antennas Propag AP-35(9):1066–1070
Ogawa Y, Ohgane T (2001) Advances in adaptive antenna technologies in Japan. IEICE Trans Commun E84-B(7):1704–1712
Paulaj A, Nabar R, Gore D (2003) Introduction to space-time wireless communications. Cambridge University Press, Cambridge, United Kingdom
Ren J, Gandhi OP, Walker LR, Fraschilla J, Boerman CR (1994) Floquet-based FDTD analysis of two-dimensional phased array antennas. IEEE Trans Antennas Propag 4(4):109–111
Robey FC, Coutts S, Weikle D, Mcharg JC, Cuomo K (2004) MIMO radar theory and experimental results. In: Conference record of the 38th asilomar conference on signals, systems and computars, Pacific grove, CA, vol 1, pp 300–304
Roy R, Paulraj A, Kailath T (1986) ESPRIT – a subspace rotation approach to estimation of parameters of cisoids in noise. IEEE Trans Acoust Speech Signal Process ASSP-34(4):1340–1342
Schmidt RO (1986) Multiple emitter location and signal parameter estimation. IEEE Trans Antennas Propag AP-34(3):276–280

Shimizu M (1994) Determining the excitation coefficients of an array using genetic algorithms. In: Proceedings of IEEE AP-S, Seattle, WA, pp 530–533

Spencer QH, Peel CB, Swindlehurst AL, Haardt M (2004) An introduction to the multi-user MIMO downlink. IEEE Commun Mag 42(10):60–67

Steyskal H (1987) Digital beamforming antennas - an introduction. Microw J 30(1):107–124

Steyskal H (1988) Digital beamforming. In: European microwave conference, Stockholm, Sweden, pp 49–57

Sumiyoshi H, Nagase M, Iguchi T, Owada A, Akiyama T, Takahashi T, Aoki T, Sato M, Shoji Y, Suzuki R, Fujino Y, Akaishi A (2010) Optically controlled phased array antenna using spatial light modulator. In: IEEE international symposium on phased array systems and technology, Waltham, MA, pp 953–958

Takahashi T, Konishi Y, Makino S, Ohmine H, Nakaguro H (2008) Fast measurement technique for phased array calibration. IEEE Trans Antennas Propag 56(7):1888–1899

Takao K, Kikuma N (1986) Tamed adaptive antenna array. IEEE Trans Antennas Propag AP-34(3):388–394

Takao K, Fujita M, Nishi T (1972) An adaptive antenna array under directional constraint. IEEE Trans Antennas Propag AP-24(5):662–669

Taoka H, Dai K, Higuchi K and Sawahashi M (2007) Field experiments on ultimate frequency efficiency exceeding 30 bit/second/Hz using MLD signal detection in MIMO-OFDM broadband packet radio access. In: Proceedings of IEEE VTC2007-Spring, Dublin, Ireland, pp 2129–2134

Taylor TT (1955) Design of line-source antennas for narrow beamwidth and low side lobes. IRE Trans AP-3(1):16–28

Telatar IE (1999) Capacity of multiantenna Gaussian channels. Eur Trans Telecommun 1(6):585

Treichler JR, Agee BG (1983) A new approach to multipath correction of constant modulus signals. IEEE Trans Acoust Speech Signal Process ASSP-31(2):459–472

Turner GM, Christodoulou C (1999) FDTD analysis of phased array antennas. IEEE Trans Antennas Propag 47(4):661–667

White WD (1976) Cascade preprocessors of adaptive antennas. IEEE Trans Antennas Propag AP-24(5):670–684

Widrow B, Mantey PE, Griffiths LJ and Goode BB (1967) Adaptive antenna systems. Proc IEEE 55(12):2143–2159

Yan KK, Lu Y (1997) Sidelobe reduction in array-pattern synthesis using genetic algorithms. IEEE Trans Antennas Propag AP-45:1117–1122

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_40-1

Dielectric Lens Antennas

Carlos A. Fernandes[a]*, Eduardo B. Lima[a] and Jorge R. Costa[b]
[a]Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
[b]Instituto de Telecomunicações, Instituto Universitário de Lisboa (ISCTE-IUL), Lisbon, Portugal

Abstract

Dielectric lens antennas are attracting a renewed interest for millimeter- and submillimeter-wave applications where they become compact, especially for configurations with integrated feeds usually referred as integrated lens antennas. Lenses are very flexible and simple to design and fabricate, being a reliable alternative at these frequencies to reflector antennas. Lens target output can range from a simple collimated beam (increasing the feed directivity) to more complex multi-objective specifications.

This chapter presents a review of different types of dielectric lens antennas and lens design methods. Representative lens antenna design examples are described in detail, with emphasis on homogeneous integrated lenses. A review of the different lens analysis methods is performed, followed by the discussion of relevant lens antenna implementation issues like feeding options, dielectric material characteristics, fabrication methods, and a few dedicated measurement techniques. The chapter ends with a detailed presentation of some recent application examples involving dielectric lens antennas.

Keywords

Lens antennas; Geometrical optics; Physical optics; Lens feeds; Dielectric materials; Lens manufacturing; Lens profile design; Optimization

Introduction

The use of a dielectric lens as part of an antenna is almost as old as the demonstration of electromagnetic waves by Hertz. In fact, in 1888, Oliver Lodge used a dielectric lens in his experiments at 1-m wavelength (Lodge and Howard 1888). However, it was not until World War II that research on lens antennas has progressed further. Lenses were used to transform the radiation pattern of the primary feed into some high-gain radiation pattern, either for fixed or scanning beam applications. But at that time, they were supplanted by reflector antennas, less bulky and lighter at microwaves.

With the fast advances on the millimeter and submillimeter waves' circuit technology for the past two decades, there has been a renewed interest on lens antennas, which present a more acceptable size at these frequencies. Lenses are being explored for imaging applications, for fixed and mobile broadband communications, and for automotive radar applications, among others. In most cases, the target radiation pattern continues to be the collimated beam type (plane wave output), either fixed or scanning. But interesting highly shaped beams conforming to demanding amplitude templates have also been explored.

Lenses can be used to modify the phase or the amplitude (or both) of the primary feed radiation pattern in order to transform it into a prescribed output radiation pattern. In this sense, lenses are equivalent to reflectors. However, instead of reflection, the lens operation principle is based on the refraction of

*Email: carlos.fernandes@lx.it.pt

DOI 10.1007/978-981-4560-75-7_40-1

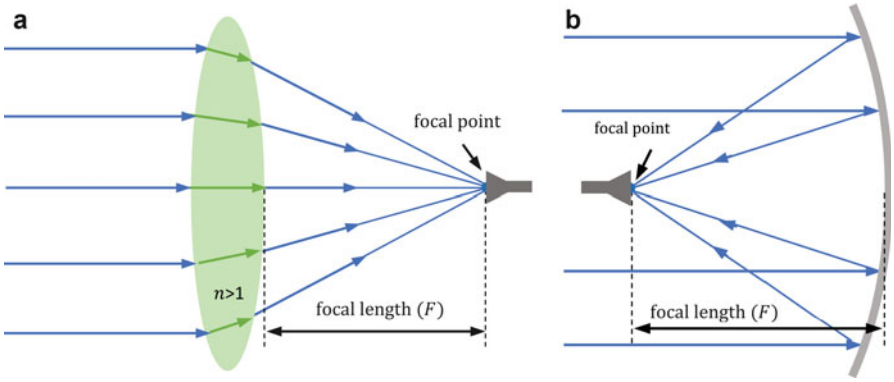


Fig. 1 Focusing antennas: (**a**) lens; (**b**) reflector

electromagnetic waves at the lens surfaces (in the case of isotropic homogenous lenses) or within the lens dielectric material in the case of nonuniform refractive index lenses. For instance, in one of its most basic configurations (Fig. 1a), parallel rays of an incident plane wave are refracted at the lens surfaces in such a way that all output rays intersect at a point, the lens focal point. All these rays have the same electrical path length, that is, they arrive in phase at the focal point (Fermat's principle), despite their different physical lengths which are compensated by a slower phase velocity ($v = c/n$) in different portions of the lens. In most designs, lens dimensions are large in terms of wavelength, enabling the use of quasi-optical design methods. Large lenses share with reflectors its inherent large bandwidth, which is limited only by the feed bandwidth.

One main advantage of lenses over reflectors is that the feed and its supporting structure do not block the antenna aperture. This back-feeding feature was key for the development of the millimeter- and submillimeter-wave integrated lens antenna concept where the lens base is positioned directly in contact with the feed, such as an integrated circuit front end, to produce a directive radiation pattern, either single beam or multiple beam. The integrated lens structure is very flexible to accommodate demanding output radiation pattern specifications, while multiple shells can be added, for instance, to increase the design degrees of freedom maintaining a compact structure. This contrasts with multi-reflector systems where blockage issues force large complex structures.

The existence of powerful software simulation tools, numerically controlled machines, 3D additive manufacturing technologies, low-loss dielectric materials, and an exploding knowledge on artificial dielectrics is favoring the development and fabrication of very sophisticated high-performance lens antennas, namely, integrated lens antennas, that are easily accessible to most laboratories or companies and affordable for mass-market products.

Lens Theory

This section starts with a brief overview of different types of known lenses and presents a possible classification. This is followed by some basic concepts applicable for quasi-optical lens design, assuming that the lens dimensions and radius of curvature at every point of the surfaces are large compared to the wavelength. Finally this section presents some alternative lens design approaches, involving a combination of iterative algorithms and lens analysis methods.

DOI 10.1007/978-981-4560-75-7_40-1

Table 1 Lens classification based on physical characteristics

1. Off-body fed	1.1 Homogeneous	A. Single refraction
		B. Multiple refraction
2. Integrated	2.1 Homogeneous	C. Single refraction
		D. Multiple refraction
	2.2 Nonuniform index	E. Continuous refraction
		F. Multiple refraction

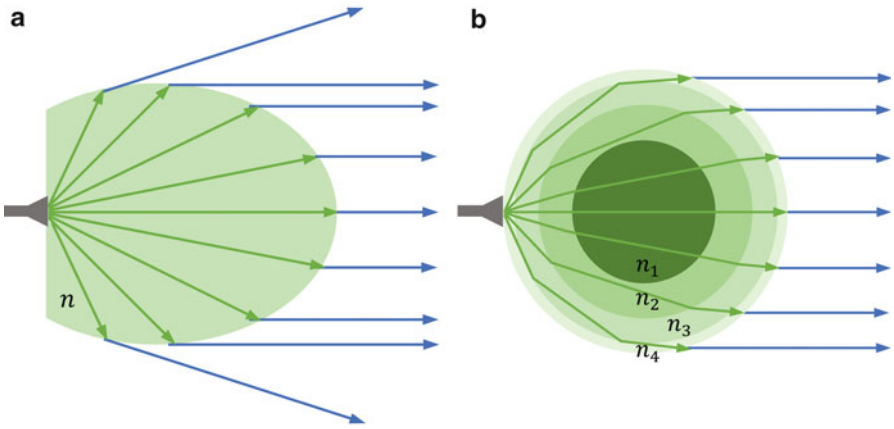


Fig. 2 Examples of integrated lens antennas

Lens Types

The lens classification adopted in this chapter is based on three different physical characteristics: the feed position relative to the lens body (far from the lens or in direct contact), refractive index profile (constant/stepped or nonuniform), and number of refraction surfaces (see Table 1). For each of these categories, the lenses can be further classified according to the type of output radiation pattern: fixed beam (collimated or shaped) or scanning beam (usually collimated).

In early lenses and in a few current designs, the focal point is located well away from the lens, at a distance comparable with its diameter as in Fig. 1a. These lenses are named in this chapter the off-body fed lenses and all examples found in the literature are axial-symmetric.

Lenses can alternatively be designed to have the feed in direct contact with the lens body (or within the body or at a fraction of the wavelength away). These are referenced in the literature as integrated lens antennas (ILA). These lenses can be made with one or more shells although the most common designs use a single layer (Fig. 2).

The concept of integrated lenses started with single-material hemispherical lenses added on top of integrated circuit antennas to eliminate substrate modes and increase radiation efficiency (Rebeiz 1992). This has evolved to the use of other fixed canonical shapes like the elliptical or extended hemispherical to further enhance gain, producing collimated output beams (Filipovic et al. 1993).

But the integrated lens configuration is especially flexible to satisfy more elaborate output beam specifications like secant square type of radiation patterns using more complex lens surfaces (Fernandes 1999), which can have an arbitrary 3D shape in the general case to produce nonsymmetrical radiation patterns (Bares and Sauleau 2007). In Fernandes and Anunciada (2001), the lens shape is adapted to transform the radiation of the circular-symmetrical source into an output shaped beam with square or rectangular footprint when pointed at the ground.

DOI 10.1007/978-981-4560-75-7_40-1

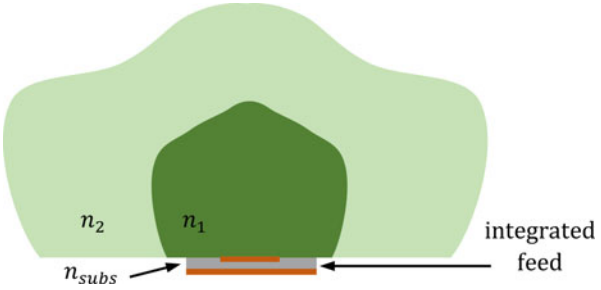


Fig. 3 Shaped integrated dielectric lens antenna

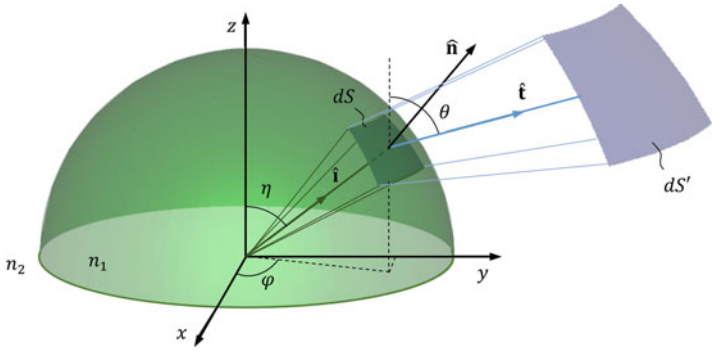


Fig. 4 Geometry of the lens and ray tube for GO formulation

Adding a further shell to the integrated lens adds a second refracting surface offering a supplementary degree of freedom to impose another design condition without sacrificing too much the lens compactness (see Fig. 3). For instance, in Costa et al. (2008a), a double-shell lens is designed to satisfy both a beam-scanning condition and maximum power transfer across the lens surfaces.

The Luneburg lens is a special case of integrated lenses, with nonuniform refractive index, where ideally an incident plane wave is focused on the antipodal point on the lens surface (see Fig. 2b).

Geometrical Optics for Lens Design

Geometrical optics (GO) is a very convenient formulation for lens (or reflector) design. It derives from the asymptotic solution of Maxwell's equations in the high-frequency limit (Kay 1965). As long as the overall lens dimensions and surface radius of curvature at any point are much larger than the wavelength, wave propagation inside a homogenous isotropic lens may be conveniently modeled in terms of elementary ray tubes. These emanate from the phase center of the source (see Fig. 4) along straight lines, with the amplitude weighted by the radiation pattern of the source and decaying with path length in the inverse proportion of the square root of the tube cross section and with phase given by the electrical path length. Reflection and transmission at an interface are made according to Snell's law (which derive from the Fermat's principle), and ray amplitude is affected by Fresnel coefficients and a divergence factor.

Assuming that the interface between two dielectric media can be considered locally plane, the reflection of an incident plane wave occurs in the same medium, with equal incident and reflected angles – Snell's law for reflection. The refraction is governed by the following Snell's law:

$$n_1 \sin(\theta_{\text{inc}}) = n_2 \sin(\theta_{\text{trans}}) \tag{1}$$

DOI 10.1007/978-981-4560-75-7_40-1

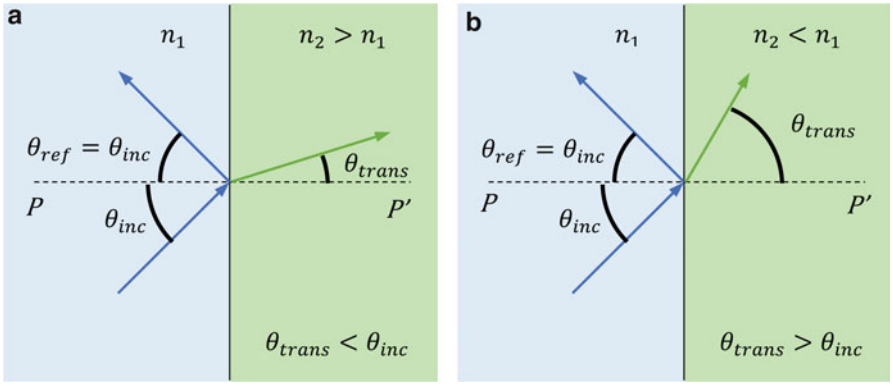


Fig. 5 Plane wave incident at an interface between two dielectrics; (**a**) incidence from the lower index medium; (**b**) incidence from the higher index media

where n_1 and n_2 are the refraction indexes of each medium and θ_{inc} and θ_{trans} are the incidence and transmission angles defined with respect to the interface normal (see Fig. 5). If the two media present the same magnetic permeability of air (most of the usual lens materials do), then $n_1 = \sqrt{\varepsilon_{r1}}$ and $n_2 = \sqrt{\varepsilon_{r2}}$ where ε_{r1} and ε_{r2} are the relative electric permittivity of each media. The refracted wave is bent toward the surface normal if the wave enters a medium with higher dielectric constant (see Fig. 5a), and it is bent away from the normal when exiting a medium with higher dielectric constant (see Fig. 5b).

In the general case, the lens surface has an arbitrary curved shape, although with large curvature radius (compared to the wavelength) at any point, and so it is convenient to present Eq. 2 in a more general form:

$$\left(n_1\,\hat{\mathbf{i}} \;-\; n_2\,\hat{\mathbf{t}}\right) \;\times\; \hat{\mathbf{n}} \;=\; 0 \tag{2}$$

where $\hat{\mathbf{i}}$ and $\hat{\mathbf{t}}$ represent the incident and refracted wave number vectors, respectively (see Fig. 4), and $\hat{\mathbf{n}}$ represents the surface normal. By expressing these three unit vectors in spherical coordinates centered at the feed phase center, Eq. 2 for an axial-symmetric lens takes the form:

$$\frac{dr}{d\eta} = \frac{n_2\, r \sin\left(\theta - \eta\right)}{n_1 \;- n_2\, \cos\left(\theta - \eta\right)} \tag{3}$$

where $r(\eta)$ represents the unknown lens profile while the output angle function $\theta(\eta)$ is the other unknown to be specified by some design condition involving phase or amplitude or eventually polarization. This condition may be given by an algebraic equation or by a differential equation. The simultaneous equations must be integrated in the interval $\eta \in [0,\ \eta_{max}]$ with applicable initial conditions $r(0)$ and $\theta(0)$ and final value η_{max}. This integration is usually performed numerically, taking only a few seconds to provide the final lens shape.

If other dielectric interfaces are involved in the ray paths, each one must conform to its Snell's equation equivalent to Eq. 3. Additional interfaces allow the enforcement of additional radiation pattern specifications which translate into corresponding equations involving the radius $r_n(\eta)$ of the n^{th} interface and the ray exit angle $\theta_n(\eta)$. The system of all the involved equations is solved by generalizing the procedure described above for the single interface. It is noted however that the formulation is different for spherical lenses like the Luneburg or Maxwell fish-eye lenses where the unknown function is no longer $r(\eta)$. The refractive index profile $n(r)$ is the unknown function in these lenses. These types of lenses are not the main focus of the present chapter and are treated as specific cases ahead.

DOI 10.1007/978-981-4560-75-7_40-1

The characterization of multiple internal reflections becomes intricate with the increase in the number of lens shells, but in general, there is no significant advantage in including its effect in the lens synthesis process. The design formulation must include however the transmission coefficients, divergence factor, and material dissipation losses where applicable. Internal reflection analysis can be performed a posteriori when evaluating the lens performance.

Because of lenses' large dimensions in terms of wavelength, dissipation losses are also an important aspect to mind. The material permittivity is a complex-valued function of the form $\varepsilon_r(1 - j\ \tan\delta)$ where the imaginary part inside the parenthesis is the loss tangent. A spherical wave propagating inside the lens material has the form:

$$E_0(\theta,\varphi)\frac{e^{-j2\pi\sqrt{\varepsilon_r(1-j\tan\delta)}\frac{r}{\lambda}}}{r} \tag{4}$$

So the total dissipation loss depends on lens dimensions, its shape, and feed illumination function. For low-loss materials, loose bounds for the dissipation loss in a homogeneous lens can be calculated for the minimum and maximum radius r of the lens using

$$L = 27.3\sqrt{\varepsilon_r}\frac{r}{\lambda}\tan\delta\ [dB] \tag{5}$$

Considering that typical lens radius that copes with the geometrical optics approximation ranges from 10 to 30 λ, a value of tan $\delta = 10^{-3}$ corresponds to dissipation losses ranging from about 0.4 to 1.3 dB.

A simple procedure to verify the lens design is to perform a ray tracing, by shooting rays from the feed point at equal angular intervals and using Snell's law to trace the rays' propagation inside and outside the lens (see Fig. 6). This can reveal, for instance, the shape of the output phase front, the phase center position, or the existence of caustics. The lens radiation pattern in the optical limit can be easily obtained by representing throughout the solid angle the ratio dS/dS' of the ray tubes, defined in Fig. 4, weighted by the feed radiation pattern.

GO-based lens synthesis has been used in the literature for different design specifications ranging from simple phase correction problems (Olver et al. 1994) or with additional aperture taper specification to multi-beam or beam-scanning problems (Kelleher 1961) or to stringent beam-shaping problems (Fernandes 1999). It has been applied for axial-symmetric lens synthesis as well as for any arbitrarily shaped lens (Salema et al. 1998; Fernandes and Anunciada 2001; Sauleau and Bares 2006; Bares and

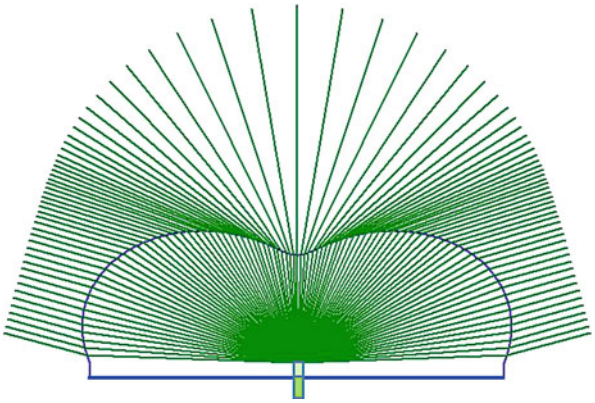


Fig. 6 Ray tracing in a secant-squared lens antenna example with embedded feed

DOI 10.1007/978-981-4560-75-7_40-1

Sauleau 2007), for multiple-shell lenses (Silveirinha and Fernandes 2000), for nonuniform refractive index lenses (Cornbleet 1994), or, in fact, for any combination of the previous cases.

The exact solution for the generic 3D structures using GO is numerically involved but feasible (Salema et al. 1998; Sauleau and Bares 2006; Bares and Sauleau 2007). Perturbation methods like the one proposed in Fernandes and Anunciada (2001) may be simpler to implement for specific types of nonsymmetrical target radiation patterns, and it allows to find the appropriate shape for the required nonsymmetrical lens. In this reference, the lens shape is adapted to transform the radiation of the circular-symmetrical source into an output shaped beam with square or rectangular cross section. An adaptation of the same principle can be used to adjust the lens shape to produce an axial-symmetric output beam.

Other Lens Design Methods

The advantage of the direct GO synthesis method presented in the previous section is that it provides the lens shape that satisfies the design requirements after elementary numerical evaluation of closed-form analytical expressions, without the need for any trial and error iterations. Required memory and CPU resources are really modest in the case of axial-symmetric lenses. Although the direct GO synthesis is enough for a large number of applications, the method is asymptotic, valid in the optical limit, thus neglecting diffraction effects that become important as the lens size reduces.

Alternative lens design methods can be considered when a more accurate lens solution is required for given specifications. They are typically based on a trial and error process involving a parameterized lens model and a lens performance analysis method. The accuracy of the design depends on the accuracy of the numerical modeling and a number of required iterations, conditioned by available computational resources. The efficiency of the design process depends very much on how smart the optimization procedure is, due to lens-specific issues which are detailed ahead. Using the direct GO synthesis lens as the first guess for the iterative process may enhance its efficiency.

Before describing the alternative iterative lens design method, it is worth summarizing the following concepts:

- A *closed-form lens synthesis* method (like the GO method described in the previous section) starts from a list of input parameters and design specifications and directly provides an appropriate lens shape based on a closed-form formulation, without the need for trial and error iterations. In this sense, this can be further classified as a direct synthesis method. In general the verification of the lens performance requires the subsequent single-time use of a lens analysis method.
- In *iterative lens synthesis* methods, the lens shape is described by some analytical or numerical representation with unknown coefficients which are determined inside an iterative optimization cycle that tests each generated lens using an appropriate lens analysis method until the target specifications are met by trial and error.
- A *lens analysis method* is intended to evaluate the performance of an existing lens either using an approximate method or a full-wave solver. Its output is the lens performance, not the lens shape.

Unlike direct synthesis methods, the list of available analysis methods is considerably larger. The most used are presented in Table 2, grouped according to the type of electromagnetic modeling. A brief description of these methods is provided in section "Lens Design, Fabrication, and Test."

Figure 7 presents a schematic diagram of the two mentioned approaches for lens design – the direct synthesis method on the left and the iterative synthesis methods on the right. Naturally both start from the definition of the target specifications, lens material, and primary feed characteristics. The path on the left provides a fast first guess of the solution, which can be enough for certain problems or can be fed for refinement through the iterative process on the right.

DOI 10.1007/978-981-4560-75-7_40-1

Table 2 Possible lens antenna analysis methods

Approximate	Geometrical optics/physical optics (GO/PO)
	Physical optics/physical optics (PO/PO)
	Spectral domain method (SDM)
Full wave	Spherical wave modal method
	Finite element method (FEM)
	Method of moments (MoM)
	Finite-difference time-domain (FDTD)

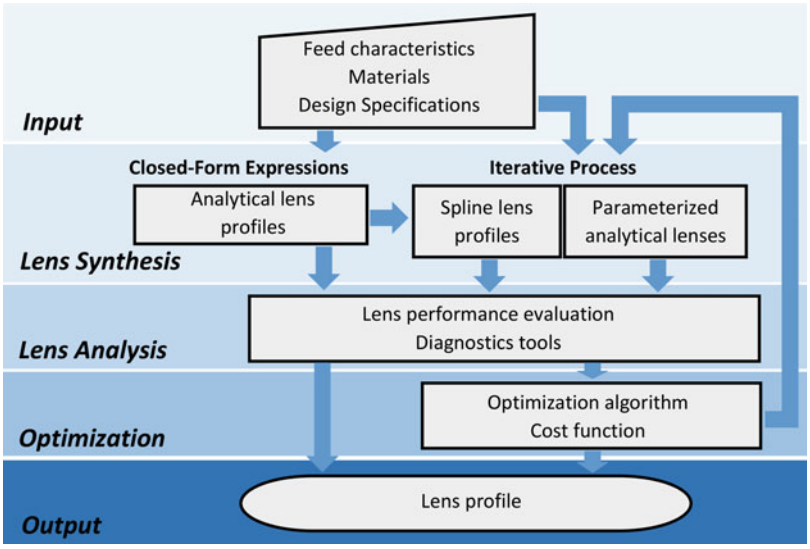


Fig. 7 Block diagram of lens design steps

Two alternative approaches are considered for the parametric modeling of the lens. A polynomial-type representation can be adopted, where the coefficients are the unknowns to be optimized in the loop instead of the large collection of the lens surface coordinates. This solution is very flexible allowing the representation of arbitrary shapes. The disadvantage is that the optimization algorithm may generate an unnecessary number of useless lenses, not because the geometry is impossible (it can be checked before starting the lens analysis and removed from the test population) but because randomly generated lenses can easily originate total internal reflection, surface wave modes, and caustics, especially for integrated multiple-shell lenses which can only be detected after the analysis process.

In the second approach (Lima et al. 2008), the search space can be narrowed to specific classes of lenses by using the analytical lens profiles from the GO synthesis method but letting the involved parameters be managed by the optimization algorithm. The analytical-based solutions ensure the electromagnetic viability of all the solutions tested in the iterative process. There is a convergence time versus design flexibility trade-off between these two approaches.

The described procedures and the workflow presented in Fig. 7 are implemented in a freeware lens design, analysis, and optimization tool – the ILASH software tool (Lima et al. 2008). It was developed for circular-symmetric-shaped integrated lens antennas with single or double layer and can handle multiple target specification definitions. The lens analysis method is based on the GO/PO, and the optimization is based on genetic algorithms. A screenshot of the ILASH user interface is presented in Fig. 8. It allows simple interaction with the kernel, to generate and manipulate lens design data, to fully characterize the

DOI 10.1007/978-981-4560-75-7_40-1

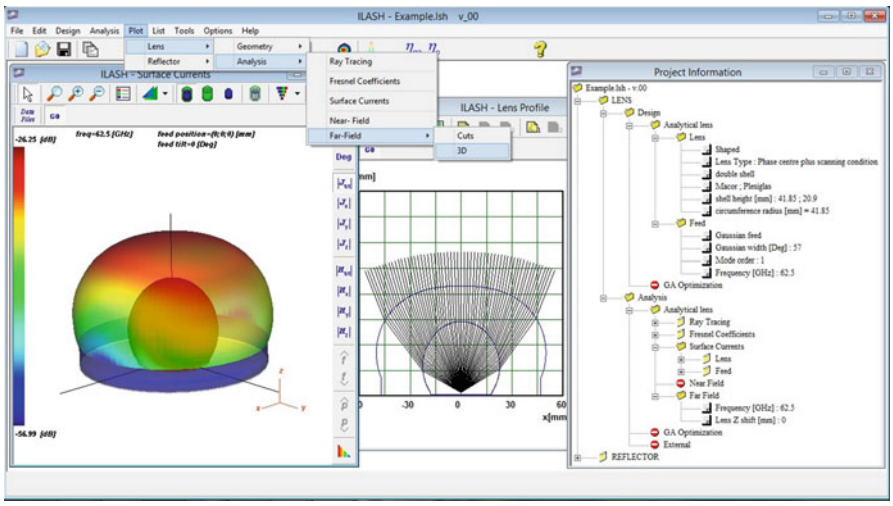


Fig. 8 Screenshot of the user interface of the freeware ILASH software

lens performance, and to export and import results. It is also possible to monitor in real time several aspects of the optimization process like the cost function value, lens parameter evolution, and convergence. ILASH was used to design the lens examples described in section "Applications."

Lens Design, Fabrication, and Test

This section presents a few representative examples of lens design, organized according to the lens classification introduced in section "Lens Theory." In most of the cases, the design is based on GO formulation. The section includes also a discussion of the steps that follow the lens design toward fabrication and test.

Off-Body Fed Lens Design Examples

Early works on dielectric lens antennas were based upon the concepts of optical lenses. Most of the late 1800 and early 1900 tests were conducted with off-body fed lenses which were used to collimate the radiation of a plane wave into the feed placed at the focal point in the opposite side of the lens. Also most off-body fed lenses have an axial-symmetric shape, and, at least during design, the feed radiation pattern is considered also circularly symmetric. These assumptions remove the dependence of the lens with the φ angle, the angle of rotation about the antenna axis of symmetry. Off-body fed lens antennas are usually made of a single material and can have either one or two refracting surfaces.

Single Refraction Lenses

The simpler lens configurations with a single refracting surface are the elliptical and the hyperbolic lenses (see Fig. 9). In an elliptical lens, the surface closer to the feed (inner surface) has a spherical shape and does not refract the rays. The collimation of the rays is achieved from the elliptical surface farther away from the feed (outer surface). In a hyperbolic lens, the refraction occurs in the hyperbolic lens surface closer to the feed. In this configuration, the outer lens surface is planar and does not refract the rays.

The shape of the outer surface of the elliptic lens in polar coordinates can be obtained by imposing the path length collimation condition:

$$r_1 + n\, l(\eta) + \; s(\eta) = r_1 + n\, T \tag{6}$$

DOI 10.1007/978-981-4560-75-7_40-1

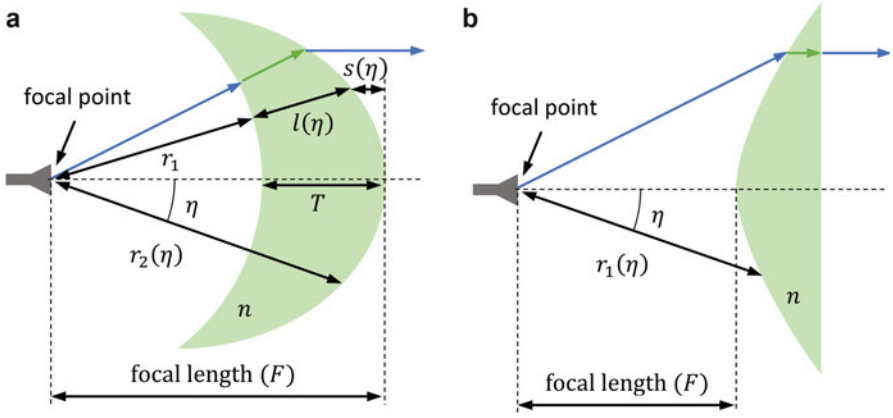


Fig. 9 Off-body fed lens examples. (**a**) Elliptical lens; (**b**) hyperbolic lens

and the following physical length condition:

$$[r_1 + l(\eta)]\cos(\eta) + s(\eta) = r_1 + T \tag{7}$$

where T is the lens selected thickness at the axis. Subtracting the two equations and noting that $r_1 + l(\eta) = r_2(\eta)$ and that $r_1 + T = F$ where F represents the lens focal distance, the outer surface profile is expressed as

$$r_2(\eta) = \frac{F(n-1)}{n - \cos(\eta)} \tag{8}$$

where n is the refraction index of the lens material and F is the lens focal distance. Snell's equations are implicitly satisfied by the collimation condition (which translates Fermat's principle).

The inner shape of the hyperbolic lens can be obtained using a similar set of equations and is defined by

$$r_1(\eta) = \frac{F(n-1)}{\cos(\eta) - 1} \tag{9}$$

Note that the design is based solely on phase considerations, assuming a point source. In the ideal case when both lens types with identical diameters are fed with a point source, the elliptical lens presents higher directivity and lower side lobe levels. In fact, all the inner surface points of the elliptical lens are at the same distance from the source and are illuminated with identical amplitude. The main disadvantage of the elliptical lens is that the reflection at the inner surface of the lens, due to the contrast between the air and the dielectric permittivity, is normal to the spherical surface and is reflected back toward the feed (Piksa et al. 2011). In the hyperbolic lens, the diffraction effects at the edges of the lens tend to influence more the main beam of the radiation pattern and increase the side lobe level (Piksa et al. 2011). The hyperbolic lens however may be simpler to manufacture since it presents one planar surface.

Double Refraction Lenses

Lenses with two refracting surfaces allow more control of the radiation pattern characteristics as discussed in section "Geometrical Optics for Lens Design." In fact, based upon geometrical optics, a set of differential and linear equations can be obtained to determine the coordinates of one of the surfaces

DOI 10.1007/978-981-4560-75-7_40-1

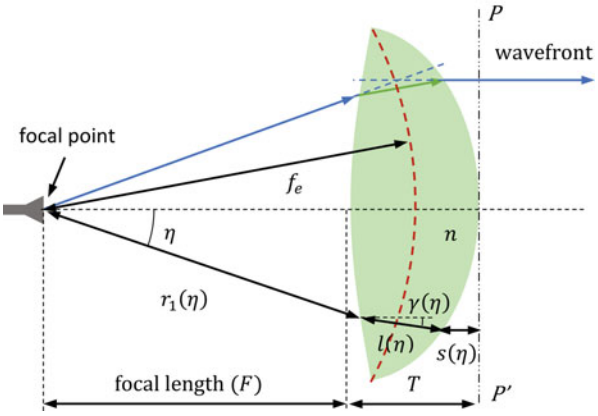


Fig. 10 Geometry of a two refracting surface lens for Abbe sine condition enforcing

(Silver 1984). However, those equations are not sufficient to determine uniquely both surfaces, so another design condition must be imposed.

One interesting use for the extra degree of freedom provided by the second refracting surface is in scanning beam applications. If the feed in the previous elliptical or hyperbolic lenses is shifted off the focus on a plane normal to the lens axis, the beam tilts linearly for very small shifts and with a more complex dependence for larger shifts along with increasing beam deformations. In fact the phase front of the output beam is no longer plane for off-axis feed positions. Expanding it in Taylor series reveals a linear term responsible for the linear tilt and nonlinear terms responsible for different types of superimposed beam deformations (Born and Wolf 1959).

It is possible to extend the linear tilt angle range for scanning beam applications by imposing the so-called Abbe sine condition for the design of the second available lens-refracting surface. This enables designing a collimating lens which is free from comma aberration when the feed is transversely displaced away from the lens axis (Born and Wolf 1959). The inner surface of the lens can be defined by the unknown function $r_1(\eta)$ and the outer surface by the unknown length $l(\eta)$ and angle $\gamma(\eta)$ represented in Fig. 10. The lens focal length is F and the axial thickness is T. The Abbe sine condition is verified when the intersections of the extended $r_1(\eta)$ rays departing from the lens focal point and the corresponding extended transmitted $s(\eta)$ rays all lie over an arc of circumference with a certain radius f_c centered at the focal point. This is represented by the thick dashed line in Fig. 10. In view of the geometry presented in Fig. 10, the Abbe sine condition can be written as

$$(f_e - r_1)\sin(\eta) = l\sin(\gamma) \tag{10}$$

Snell's law at the inner surface of the lens implies that (Eq. 3)

$$\frac{dr_1}{d\eta} = \frac{n\, r_1 \sin(\gamma - \eta)}{1 - n\,\cos(\gamma - \eta)} \tag{11}$$

In order that the electrical path length of every ray is the same at the exiting wavefront, it is required that

$$r_1 + nl + s = F + nT, \tag{12}$$

$$s + r_1\cos(\eta) + l\cos(\gamma) = F + T \tag{13}$$

DOI 10.1007/978-981-4560-75-7_40-1

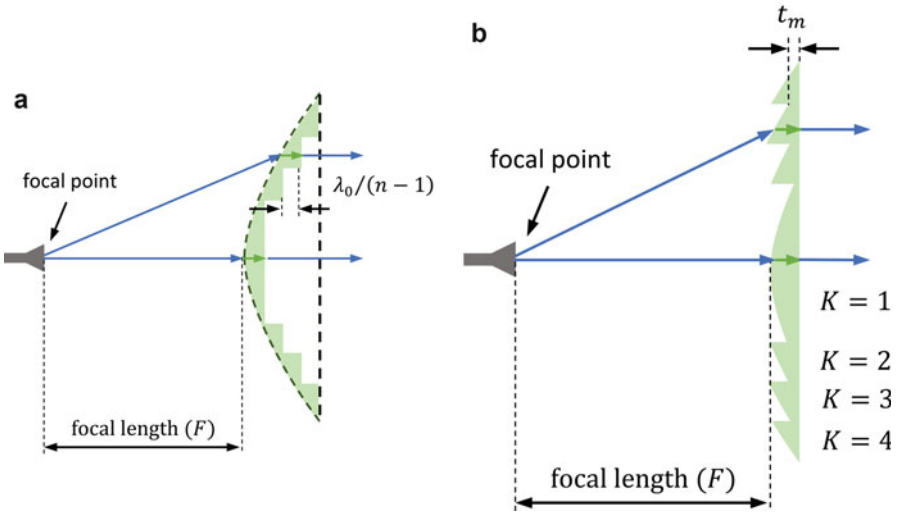


Fig. 11 Zoning of a hyperbolic lens. (**a**) Original outer hyperbolic shape; (**b**) corresponding zoned lens

Equations 10, 11, 12, and 13 can be solved simultaneously for $r_1(\eta)$, $l(\eta)$, $s(\eta)$, and $\gamma(\eta)$ taking η as the independent variable and the initial conditions $r_1(0) = F$, $l(0) = T$, $s(0) = 0$, and $\gamma(\eta) = 0$. The outer surface of the lens can be calculated from the previous four functions.

Zoning

Usually lenses are several wavelengths thick, particularly near the lens axis. For microwave applications, this may lead to a very bulky antenna and originate non-negligible dissipation losses in the dielectric. To attenuate those effects, rings of material with a thickness equal to an integer multiple of the wavelength can be removed from the lens, a process called zoning. Figure 11 shows an example for the hyperbolic lens discussed in section "Single Refraction Lenses" with four zones ($K = 4$). A minimum physical lens thickness t_m needs to be maintained in the zoned lens to provide structural support, as indicated in Fig. 11. The thickness of the zoned lens is independent of the number of zones K, and it is given by $t_m + \lambda_0/(n-1)$, which for the example in Fig. 11 represents about a quarter of the original lens thickness.

The downside of zoned lens is the associated frequency dependence which becomes more important as the lens diameter increases and more zones are used. A criterion to estimate the bandwidth of zoned lenses is given by Silver (1984): under uniform aperture distribution and lens aperture phase error smaller than $\pi/4$

$$\text{Bandwidth} \approx \frac{25\%}{K-1} \tag{14}$$

If the zoning is performed in the non-refracting surface of the lens, shadowing losses can appear at the transition regions, and this effect is stronger for lenses with larger ratios between the focal length and the lens diameter F/D (Petosa and Ittipiboon 2000).

Integrated Lens Design Examples

This section addresses briefly two classical examples of nonuniform refractive index lenses – the Luneburg and the Maxwell fish-eye – and focuses more on the homogeneous shaped lenses. Other nonuniform refractive index lenses have also gained interest with the recent advances in the fields of transformation optics and metamaterials, but these exceed the scope of this chapter.

Homogeneous integrated lenses are very effective for integration with low-cost millimeter-wave circuits, allowing low form-factor solutions with sophisticated radiation pattern characteristics owing to

DOI 10.1007/978-981-4560-75-7_40-1

the great design flexibility this lens configuration provides. However, internal reflections may critically influence the lens performance and must be properly analyzed in each design. This section presents some representative design examples of canonical lenses as well as some more recent shaped lenses based on GO.

Nonuniform Index Spherical Lenses

In the classical nonuniform spherical lenses, the dielectric constant of the material has only radial variation. Therefore, the lens is symmetric in relation to any axis passing the center of the lens. As a consequence, these spherical lenses do not present a unique focal point but rather a spherical surface concentric with the lens where the focal point can be located depending on the direction of the incident wave.

The most known of the spherical lenses is the Luneburg lens (Luneburg 1943) where the lens material permittivity profile varies with the square of the r distance from the lens center:

$$\varepsilon_r(r) = 2 - \left(\frac{r}{R}\right)^2 \tag{15}$$

R being the outer radius of the lens. This formulation results in a focal region located at the outer surface $r = R$ of the lens. Therefore, a point source located at any point of the lens surface originates a collimated beam in the opposite direction (Fig. 12). This property is independent of the lens diameter. The Luneburg lens is particularly adequate for multi-beam applications since the symmetry ensures that all beams are equally independent of the feed position on the lens surface.

Luneburg did not have the opportunity to demonstrate such antenna experimentally since no suitable lens materials were available at the time. Nowadays, Luneburg lenses are formed by a discrete number of concentric dielectric layers that approximate the ideal permittivity law of Eq. 15. It is expected that as the number of discrete layers increases, the performance of the lens improves; however, the manufacturing complexity increases as well as the influence of possible air gaps between the layers (Kim and Rahmat-Samii 1998). It has been demonstrated that a quite small number of shells is enough to obtain a directivity and side lobe level close to the ideal case (Fuchs et al. 2007a). Nevertheless, the recommended number of discrete layers increases with the lens diameter compared to the wavelength. For example, just six layers were required to achieve adequate performance with an $8\lambda_0$-diameter Luneburg lens (Fuchs et al. 2007a). Even lesser number of layers is possible by optimizing both the permittivity and thickness of the layers (Mosallaei and Rahmat-Samii 2001; Boriskin et al. 2011).

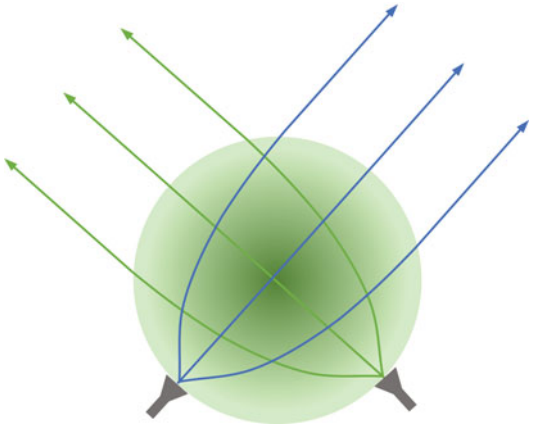


Fig. 12 Luneburg lens with darker color representing increasing permittivity

DOI 10.1007/978-981-4560-75-7_40-1

There are also examples of Luneburg lenses fabricated from a single material with controllable dielectric constant. For instance, in Min et al. (2014), a 3D rapid prototyping machine was used to achieve the desired permittivity value by controlling the filling ratio of a polymer/air-based unit cell. Alternatively, it is possible to use foam material and press it to obtain the desired dielectric constant. In fact, when the pressure increases and the foam is compacted, the quantity of air in the material decreases, increasing the permittivity. This technique has been tested for the fabrication of a Luneburg lens in Bor et al. (2014). It is also possible to control the dielectric constant in the Luneburg lens by drilling holes into the material and varying either the diameter of the holes or their density (Sato and Ujiie 2002; Xue and Fusco 2007).

One of the major disadvantages of the Luneburg lens is its high profile. Transformation optics (Do-Hoon and Werner 2010) can be used to change the shape of the lens into a much lower cylindrical profile. However, common solutions end up resulting in materials that are anisotropic and/or must have relative permeability different than unity forcing the use of metamaterials that are usually narrowband, lossy, dispersive, and sometimes cumbersome to manufacture. A reasonable performance has been achieved with a compacted Luneburg lens with cylindrical geometry, formed by discrete layers with step permittivity variation along the radius and along the height (Mateo-Segura et al. 2014). Although a height reduction was obtained, permittivity values as high as 12 are required, and the structure presents some scanning loss.

A commonly used solution to reduce the profile of the Luneburg lens by a factor of 2 is to combine half of the lens with a flat ground plane (Fig. 13). The ground plane creates an image of the upper hemisphere and simulates the complete Luneburg lens. The size of the ground plane must be adequate to produce the required image, and the necessary dimensions are the function of the elevation angle of the beam in relation to the ground plane. An insufficient ground plane size will reduce the directivity of the beam and cause scanning loss. This hemispheric lens solution also suffers from feed blockage for high elevation angles. One advantage of this configuration is that it is easier to mechanically stabilize than the complete spherical lens. Recently a quarter Luneburg lens with a 90° corner ground plane has been demonstrated (Nikolic et al. 2012).

A similar configuration to the Luneburg spherical lens is the cylindrical one (Fig. 14) where the variation of the dielectric constant occurs only in the radial direction. In this solution, the beam collimation is obtained only in one plane, leading to fan-shaped beams instead of pencillike beams (Komljenovic et al. 2010).

Another classical nonuniform index spherical lens is the Maxwell fish-eye lens, which predates the Luneburg lens (Maxwell 1860). The dielectric constant of the lens material is given by

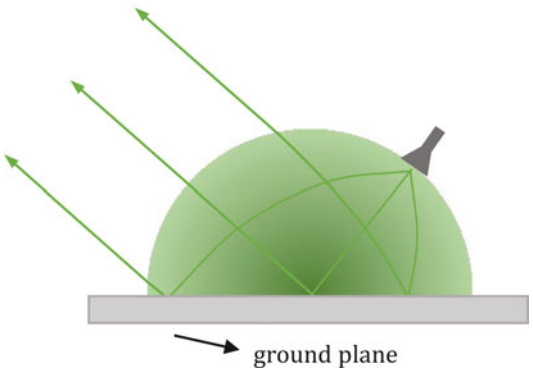


Fig. 13 Half hemisphere Luneburg lens

DOI 10.1007/978-981-4560-75-7_40-1

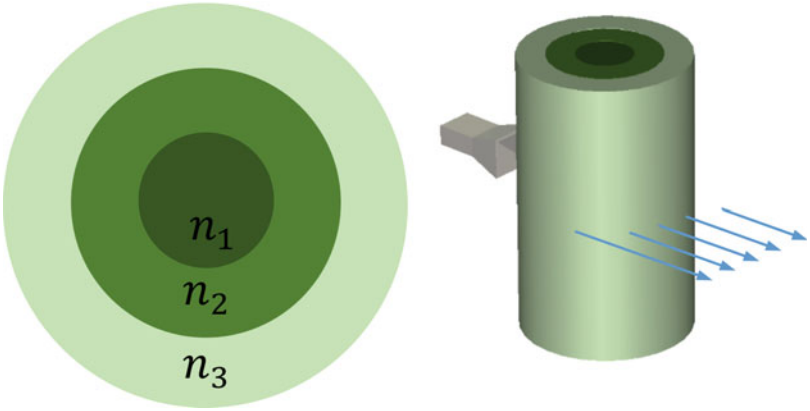


Fig. 14 Cylindrical Luneburg lens

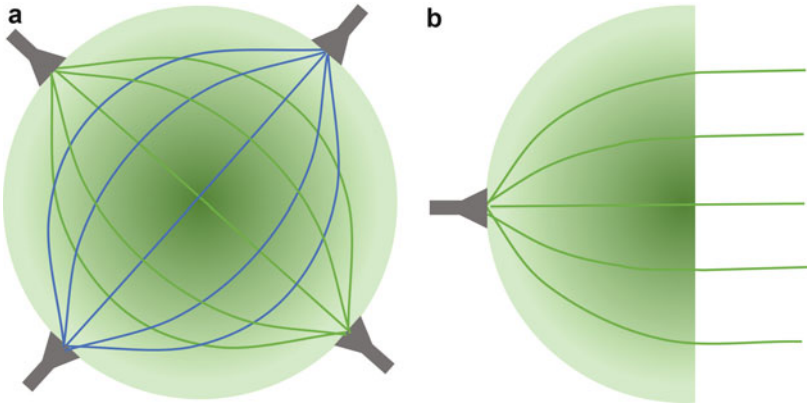


Fig. 15 (**a**) Maxwell fish-eye lens; (**b**) half Maxwell fish-eye lens. The darker color represents higher values of permittivity

$$\varepsilon_r(r) = \left(\frac{4}{\left[1 + \left(\frac{r}{R}\right)^2\right]^2} \right) \tag{16}$$

where R is the outer radius of the lens. When a point source is placed on the surface of the lens, the radiation is focused at the antipodal point of the lens as represented in Fig. 15a. Due to the lens symmetry, the source radiation is converted into a local plane wave at the center of the lens. Therefore, if the Maxwell fish-eye lens is cut in half, it can be used as a collimating lens that focuses a plane wave into a focal point on the surface of the lens, which becomes unique in this case (Fig. 15b). The half Maxwell fish-eye (HMFE) lens has lower profile than the Luneburg lens and is easier to flush mount. Practical implementation of the HMFE has been done using a few discrete layers (Fuchs et al. 2006), and the performance was shown to be comparable with the Luneburg lens when the feed is centered with the lens (Fuchs et al. 2008a).

Unlike the Luneburg lens, the HMFE lens loses interest for scanning beam applications since scanning losses appear as the feed is moved over the lens surface (Fuchs et al. 2007b).

Elliptical and Hemispherical Lenses

The homogeneous elliptical integrated lens, just like its off-body fed counterpart treated in section "Single Refraction Lenses," is used to transform the radiation pattern of a feed placed at the focal point of the lens

DOI 10.1007/978-981-4560-75-7_40-1

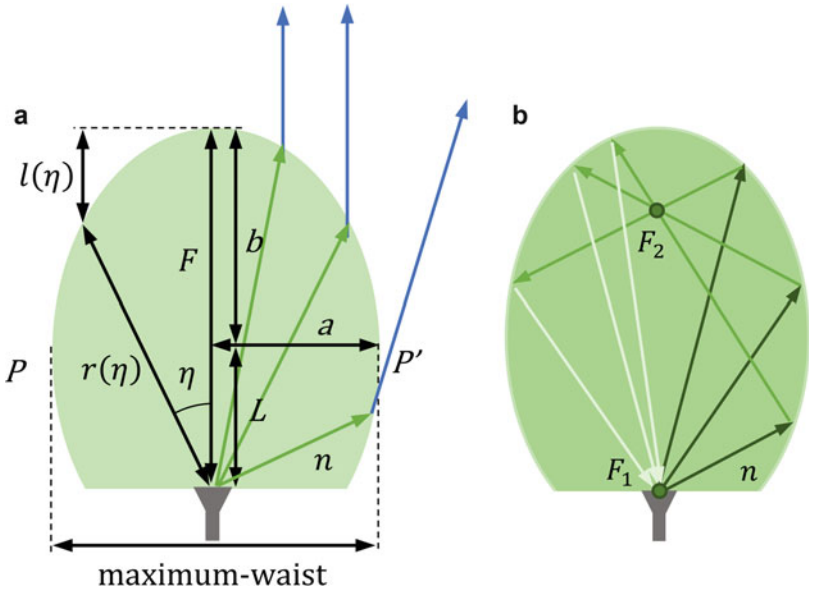


Fig. 16 (**a**) Design variables of an elliptical integrated lens antenna; (**b**) rays' reflection inside the lens

into a plane wave in the air medium, propagating along the lens axis (Fig. 16). It can be viewed as the limit when r_1 tends to zero in Eqs. 6 and 7 from section "Single Refraction Lenses":

$$n\, r(\eta) + l(\eta) = n\, F \tag{17}$$

Also the following physical length condition can be imposed:

$$r(\eta)\cos(\eta) + l(\eta) = F \tag{18}$$

Eliminating $l(\eta)$ between the two equations leads to the same elliptical lens profile found in section "Single Refraction Lenses":

$$r(\eta) = \frac{F(n-1)}{n - \cos(\eta)} \tag{19}$$

It is a simple matter to show that this corresponds in rectangular coordinates to

$$\left(\frac{x}{a}\right)^2 + \left(\frac{z-L}{b}\right)^2 = 1 \tag{20}$$

where a is the radius along the x-axis, b is the radius along the z-axis, L is the position of the focal point in relation to the center of the lens, and $b + L = F$ (Fig. 16a). The eccentricity of the elliptical lens is related to the dielectric constant of the material. The following relations hold (Filipovic et al. 1993):

$$b = \frac{a}{\sqrt{\frac{1-1}{n^2}}} \tag{21}$$

DOI 10.1007/978-981-4560-75-7_40-1

$$L = \frac{b}{n} \tag{22}$$

The value of a is chosen to define the lens size and consequently the aperture size required to achieve a certain directivity of the beam.

Only the rays that hit the lens surface above the plane of maximum waist PP' are focused (Fig. 16b). The feed radiations intersecting the lens below the maximum waist are not collimated but rather propagate along undesired directions or excite surface wave modes (Pasqualini and Maci 2004) giving rise to side lobes or other perturbations in the lens radiation pattern. For this reason, proper feed configurations should be used to minimize the illumination of the lower part of the lens surface.

Internal reflections may be especially critical in integrated lens antennas. For instance, if the feed is at the center of a hemispherical lens, all reflected rays travel back along the same path of the corresponding incident rays, concentrating at the focal point, usually causing a substantial mismatch at the feed impedance. A similar effect occurs in the elliptical lens; all reflected rays concentrate at the feed point after traveling through the second focal point of the elliptical lens, originating a second-order reflection at another point of the lens surface (Fig. 16b) (Neto et al. 1998, 1999; Van Der Vorst et al. 1999, 2001). Again, this causes feed impedance mismatch. Part of the second-order ray is transmitted to the air at an undesired direction causing side lobes in the lens radiation pattern. These undesired effects increase with the contrast between the material and the air refraction indexes. For other integrated lens shapes, the reflected rays may not be all reflected back into the feed; nevertheless, they may be responsible for transmission of higher than first-order rays along undesired directions, reducing the main beam efficiency or causing ripple.

The problem of internal reflections can be tackled, at least within a limited bandwidth. For normal incidence on a planar interface between two dielectric media with refraction index n_1 and n_2, it is possible to cancel the reflected ray by adding an appropriate intermediate layer, Fig. 17a. Its dielectric refraction index n^{match} must be

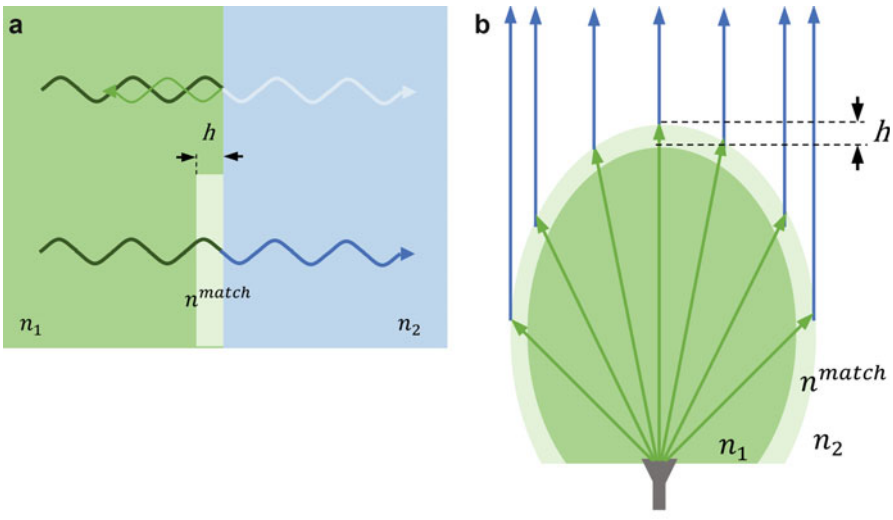


Fig. 17 Matching layer (**a**) in a planar dielectric-air interface. The intensity of ray color is proportional to the wave power density; (**b**) in an elliptical integrated lens antenna

$$n^{\text{match}} = \sqrt{n_1 n_2} \tag{23}$$

and a thickness h

$$h = \frac{\lambda_0}{4n^{\text{match}}} \tag{24}$$

where λ_0 is the free-space wavelength. This intermediate layer is usually called the matching layer.

By adding a matching layer to the lens surface, it is possible to alleviate the effect of internal reflections (Fig. 17b). It is a fact that the rays do not hit the lens surface along the normal direction, and therefore the thickness of the matching layer should not be given by Eq. 24. Instead it should vary along the surface of the lens according to the ray incident angle. In a study presented in Van Der Vorst et al. (1999) for an elliptical integrated lens of silicon ($\varepsilon_r = 11.7$), it was demonstrated that there is no major improvement on the lens radiation performance by using the optimum-thickness matching layer. Therefore, a constant thickness layer given by Eq. 24 is generally used, facilitating the manufacturing process. Because it is not easy to find natural materials with specific permittivity values, it is common to synthesize an effective medium layer by periodically removing a fraction of the dielectric material from the lens surface, such as drilling holes or cutting groves (Ngoc Tinh et al. 2010).

The matching layer thickness is frequency dependent and, therefore, reduces the bandwidth of the lens antenna. To improve the bandwidth, multiple consecutive matching layers can be used, performing a gradual transition between the two dielectric constants at the interface (Ngoc Tinh et al. 2009).

The extended hemispherical lens is another classical configuration of integrated lens, owing to its simple shape for fabrication. It consists of a half sphere of radius R with a cylindrical extension of height L. The feed is located at the base of the lens (Fig. 18b).

If the length of the cylindrical extension is chosen to be

$$L = \frac{R}{n-1} \tag{25}$$

most of the spherical part of the hyper-hemispherical lens coincides with an elliptical lens (Fig. 18a). This lens is usually called synthesized elliptical lens (Filipovic et al. 1993) although it tends to present a slightly lower directivity when compared to a true elliptical lens with the same diameter.

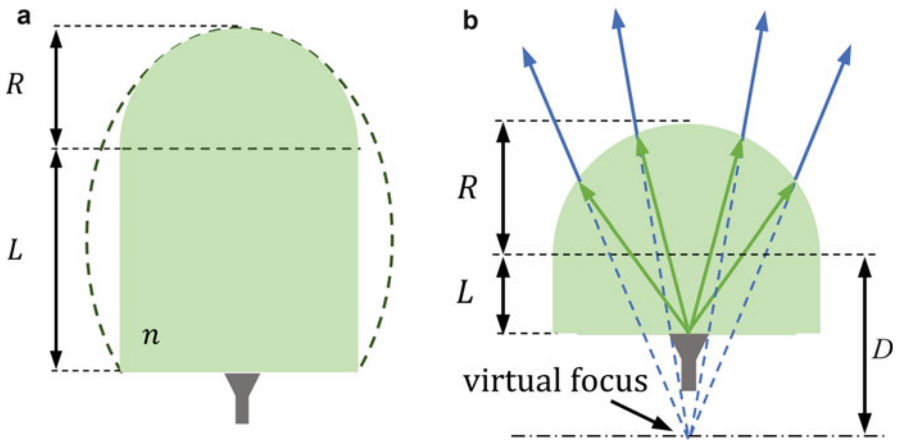


Fig. 18 Extended hemispherical integrated lens antenna: (**a**) synthesized elliptical; (**b**) hyper-hemispherical

DOI 10.1007/978-981-4560-75-7_40-1

A second type of extended hemispherical lens is the hyper-hemispherical (Rebeiz 1992) where the cylindrical extension length is

$$L = \frac{R}{n} \tag{26}$$

For this particular type of lens, the output beam is not collimated, presenting a much broader (and sometimes multilobed) radiation pattern when compared to an elliptical lens with the same radius. Nevertheless, the hyper-hemispherical lens bends the rays radiated by the feed toward the axis of the lens (Fig. 18b). The lens sharpens the radiation pattern, effectively increasing the gain of the feed by a factor n^2 (Rebeiz 1992). However, unlike collimated lenses, the directivity of this lens does not increase with lens size (or aperture size).

The hyper-hemispherical lens satisfies the Abbe sine condition (Born and Wolf 1959). Therefore, this lens is free from comma aberration when the feed is transversely displaced within certain limits away from the lens axis. This type of lens is particularly useful for beam-scanning applications. Another interesting characteristic of hyper-hemispherical lenses is that all the rays transmitted into the air focus in a virtual point behind the lens at a distance:

$$D = Rn \tag{27}$$

from the center of the spherical part of the lens. This means that the radiation pattern of this lens presents a very stable phase center position coincident with this virtual focus and its position does not shift with the frequency (in the optical limit).

There are several other types of shaped dielectric integrated lens antennas with profiles not given by canonical expressions like the elliptical and extended hemispherical lenses. In section "Applications" of this chapter, several design and implementation examples of noncanonical shaped lenses will be given in more detail.

Lens Designed to Match an Output Power Template

Integrated lenses can also be used to conform the output radiation pattern to some far-field power template $G(\theta)$. The formalism of a GO direct synthesis method is described for an axial-symmetric single-material lens with refraction index n, fed on-axis at the base of the lens as shown in Fig. 19 (Fernandes 1999).

The lens design requires the knowledge of the feed power pattern $U(\eta)$ when the feed is inside the lens material. This can be obtained using a full-wave analysis of the feed radiating into an unbounded medium with the same permittivity as the lens. Alternatively it can be obtained experimentally using the procedure described in section "Lens Measurements."

The unknown lens profile is represented by the $r(\eta)$ function. As before, Snell's equation for refraction at the dielectric interface is written as

$$\frac{\partial r}{\partial \eta} = \frac{r(\eta)\sin(\theta - \eta)}{n - \cos(\theta - \eta)} \tag{28}$$

In order to enforce the output power pattern template, the elementary ray tube concept introduced in section "Geometrical Optics for Lens Design" is used (Fig. 4). Power conservation in the elementary ray tube is expressed as

$$U(\eta)T\sin(\eta)d\eta = KG(\theta)\sin(\theta)d\theta \tag{29}$$

DOI 10.1007/978-981-4560-75-7_40-1

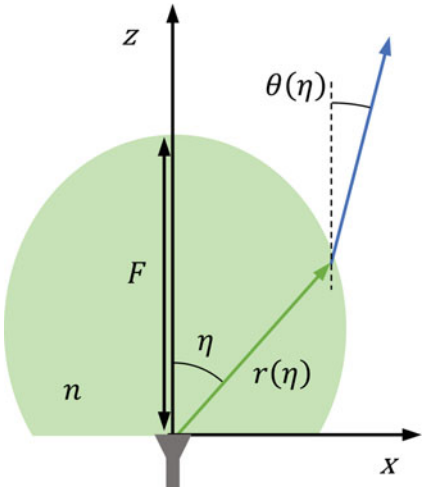


Fig. 19 Geometry of a single-shell lens

or rearranging

$$\frac{d\theta}{d\eta} = \frac{T}{K}\frac{U(\eta)}{G(\theta)}\frac{\sin(\eta)}{\sin(\theta)} \tag{30}$$

where $T(\eta)$ represents the transmissivity, that is, the ratio of the power P_t transmitted across dS to the incident power P_i (Fig. 4).

$$T = \frac{U_{\parallel}\left|t_{\parallel}^2\right| + U_{\perp}\left|t_{\perp}^2\right|}{U}\frac{1}{n}\frac{\cos(\beta)}{\cos(\alpha)}T \tag{31}$$

$$\cos(\alpha) = \hat{\mathbf{i}} \cdot \hat{\mathbf{n}} \tag{32}$$

$$\cos(\beta) = \hat{\mathbf{t}} \cdot \hat{\mathbf{n}} \tag{33}$$

where $t_{//}$ and $t_{\perp}$ represent the Fresnel transmission coefficients for parallel and perpendicular polarization, respectively. K is a normalization constant to be determined from the balance between the total power inside the lens and total power outside the lens:

$$K = \frac{\int_0^{\eta_{\max}} T(\eta)U(\eta)\sin(\eta)d\eta}{\int_0^{\theta_{\max}} G(\theta)\sin(\theta)d\theta} \tag{34}$$

$\eta_{\max}$ being the maximum feed aperture and $\theta_{\max}$ the maximum output angle.

The unknowns $r(\eta)$ and $\theta(\eta)$ are obtained by integrating the system of equations formed by Eqs. 28 and 30 from $\eta = 0$ to $\eta_{\max}$ (typically 90°) using the initial condition $r(0) = F$ and $\theta(0) = 0$. F acts as a scaling factor not affecting the lens shape. However, the larger is this value, the larger is the lens size. Inherent to the design, increasing the size also improves the lens radiation pattern compliance with the target $G(\theta)$.

DOI 10.1007/978-981-4560-75-7_40-1

Function $T(\eta)$ in Eqs. 28, 29, 30, and 31 depends indirectly on the unknown function $r(\eta)$. An iterative process can be adopted, repeating successively the integration of Eqs. 28 and 30 considering $T(\eta) =$ constant in the first step. The obtained solution $r(\eta)$ can then be used to calculate $T(\eta)$ for the next evaluation of Eqs. 28, 29, 30, and 31, and the process is repeated until convergence; usually two or three iterations are enough.

The presented formulation assumes φ-independent feed and target template power patterns $U(\eta)$ and $G(\eta)$. In most cases, the feed radiation pattern is not axial-symmetric, but an approximation can be generated as an average of the co-components in the main planes for the lens synthesis purpose.

Frequency-Stable Radiation Pattern and Phase Center Position

This section presents the design of a double-shell axial-symmetric lens that complies with two design conditions: a well-defined phase center located behind the lens (outside its body) and a target far-field amplitude template (Fernandes et al. 2010). This type of virtual focus lens can be useful as a reflector primary feed as it will be discussed in the "Applications" section ahead. Lens dimensions are assumed to be large relative to the wavelength, so direct GO synthesis will be used for the lens design.

Considering the geometry from Fig. 20, n_1 and n_2 are the refraction indexes from the inner and outer lens shells, respectively, and S and F are the corresponding axial depths. The lower value of the refraction index is used at the outer shell to favor lower reflections at the air interface. The two lens surfaces are defined by the unknown functions $r_1(\eta)$, $\theta(\eta)$, and $R(\eta)$ and are obtained by solving a system of three differential equations as explained next.

One of the equations derives from the power conservation condition in an elementary ray tube crossing the lens system, the same discussed in section "Lens Designed to Match an Output Power Template":

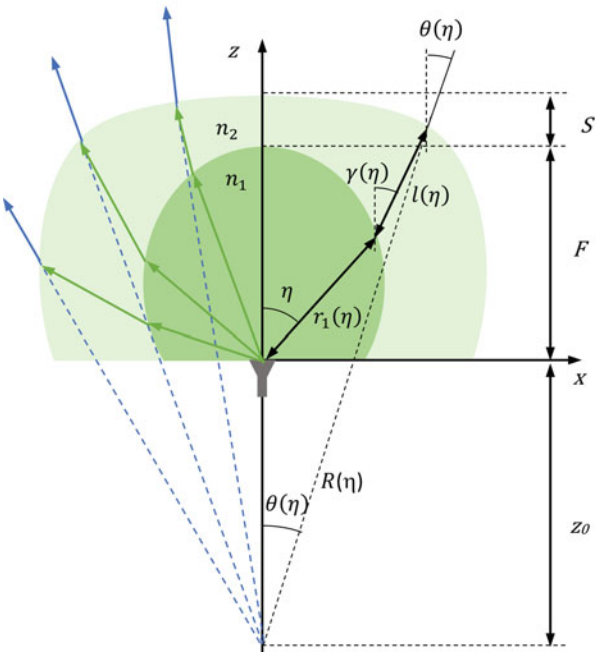


Fig. 20 Geometry for the lens design

DOI 10.1007/978-981-4560-75-7_40-1

$$\frac{d\theta}{d\eta} = \frac{T(\eta)U(\eta)\sin(\eta)}{KG(\theta)\sin(\theta)} \tag{35}$$

where $U(\eta)$, $G(\theta)$, $T(\eta)$, and K maintain the previous definition. Further imposing Snell's law for refraction at the inner interface results in (Eq. 3)

$$\frac{dr_1}{d\eta} = \frac{r_1(\eta)\sin(\gamma-\eta)}{\frac{n_1}{n_2} - \cos(\gamma-\eta)} \tag{36}$$

while imposing Snell's law at the outer interface results in

$$\frac{dR}{d\eta} = \frac{dR}{d\theta}\frac{d\theta}{d\eta} = \frac{Rn_2\sin(\gamma-\theta)}{1-n_2\cos(\gamma-\theta)}\frac{T(\eta)U(\eta)\sin(\eta)}{KG(\theta)\sin(\theta)} \tag{37}$$

Finally, the following path length conditions are imposed:

$$\begin{cases} r_1 n_1 + l n_2 = R \\ R\cos(\theta) = r_1\cos(\eta) + l\cos(\gamma) + z_0 \end{cases} \tag{38}$$

The lens profile is obtained by solving the system of three differential equations, Eqs. 35, 36, and 37, with respect to η integrating from $\eta = 0$ to $\eta = \eta_{max}$. The γ angle required in Eqs. 36 and 37 is obtained from the following explicit expression derived from Eq. 38:

$$\gamma = \cos^{-1}\left[\frac{n_2(R\cos(\theta) - r_1\cos(\eta) - z_0)}{R - n_1 r_1}\right] \tag{39}$$

The initial conditions at $\eta = 0$ are

$$\begin{cases} \theta = 0 \\ r_1 = F \\ \gamma = 0 \\ l = S \\ R = n_1 F + n_2 S \end{cases} \tag{40}$$

The distance from the phase center to the base of the lens z_0 is settled once the lens shells' refraction index and axial depths are chosen. In fact, from Eq. 40, one has

$$z_0 = R(\eta = 0) - F - S = F(n_1 - 1) + S(n_2 - 1) \tag{41}$$

Equation 35 is numerically undetermined at $\eta = 0$. To obtain its value at $\eta = 0$, the following relation is written from

$$\int_0^{\theta} KG(\theta)\sin(\theta)d\theta = \int_0^{\eta} T(\eta)U(\eta)\sin(\eta)d\eta \tag{42}$$

DOI 10.1007/978-981-4560-75-7_40-1

and evaluated in the limit as $\theta \to 0$ and $\eta \to 0$. The result is

$$\theta = \sqrt{\frac{T(\eta)U(\eta)}{KG(\theta)}}\eta \tag{43}$$

where $T(\eta)$, $U(\eta)$, and $G(\theta)$ are assumed constant near $\eta = 0$ and $\theta = 0$. Finally, from the above equation, the corresponding limit value for Eq. 35 is given by

$$\frac{d\theta}{d\eta} = \sqrt{\frac{T(\eta)U(\eta)}{KG(\theta)}} \tag{44}$$

This result is also used in the related part of Eq. 37 for $\eta = 0$.

The GO lens synthesis is inherently broadband as long as lens dimensions are large compared to the wavelength and as long as the feed radiation pattern remains constant and coincident with the used $U(\eta)$ template.

Multi-beam Lens

An integrated lens configuration, formed by two embedded shaped shells, can be designed for multi-beam or scanning applications (Costa et al. 2008a). The possibility to shape two lens surfaces enables that two independent design goals can be specified (in a single-material lens, only one design goal can be defined):

- A beam collimation condition, i.e., the output rays emerging from the lens are required to be parallel to each other
- A condition that minimizes aberrations of the output beam for off-axis feed positions

The lens geometry is presented in Fig. 21. The structure is axial-symmetric, formed by two embedded shells of different materials; the inner shell presents the higher refraction index value n_1 and the outer shell the lower value n_2. An array of feeds is distributed at the planar base of the lens within a small area centered with the lens axis.

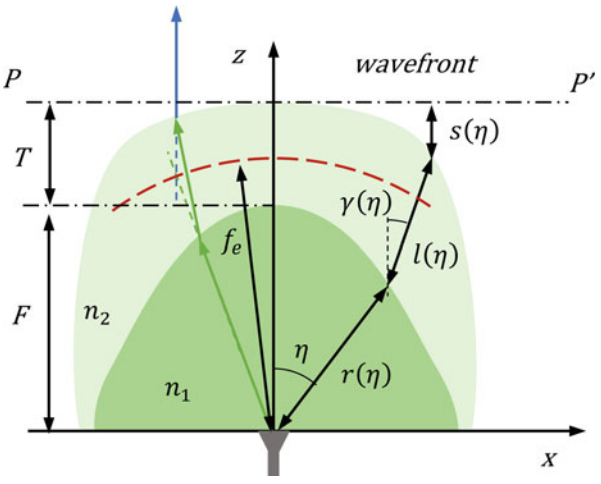


Fig. 21 Geometry for the lens design

DOI 10.1007/978-981-4560-75-7_40-1

The inner lens surface is defined by the (unknown) function $r(\eta)$, and the outer lens surface is defined by the (unknown) length $l(\eta)$ and angle $\gamma(\eta)$ (Fig. 21). The lens axial thickness is F and T for the inner and outer shells, respectively.

It is well known that a lens satisfying the so-called Abbe sine condition is free from comma aberration for a small off-axis transversal displacement of the feed (Born and Wolf 1959). The Abbe sine condition is verified when the intersection points of the extended $r(\eta)$ rays departing from the on-axis feed and the corresponding extended transmitted $s(\eta)$ rays all lie over an arc of circumference with a certain radius f_e centered at the sensor (Born and Wolf 1959). This is represented by the thick dashed arc in Fig. 21. In view of the geometry of Fig. 21, the Abbe sine condition can be written as

$$(f_e - r)\sin\eta = l\sin\gamma \tag{45}$$

Snell's law at the inner interface implies that

$$\frac{dr}{d\eta} = \frac{r(\eta)\sin(\gamma-\eta)}{\frac{n_1}{n_2} - \cos(\gamma-\eta)} \tag{46}$$

In order that the optical path length of every ray is the same at the exiting wavefront, it is required that

$$n_1 r + n_2 l + s = n_1 F + n_2 T \tag{47}$$

$$s = F + T - r\cos(\eta) - l\cos(\gamma) \tag{48}$$

Equations 45, 46, 47, and 48 can be solved simultaneously to determine both the inner and the outer shell profiles, taking η as the independent variable. The initial condition for $\eta = 0$ is $r = F$, $l = T$, and $\gamma = 0$, and the integration is extended up to $\eta = \eta_{edge}$, where generally $\eta_{edge} < \pi/2$. The calculation is stopped at the point where the outer shell intersects the f_e circle or where the total internal reflection condition is reached.

For a given combination of n_1, n_2, F, and T values, the f_e parameter can be adjusted between F and $F + T$ to control the shape of the lens surfaces (and indirectly the lens scanning characteristics). In the optical limit, the lens shape obtained with the above GO-based formulation is independent of the absolute dimensions and thus, for instance, F, can be taken as a scaling factor.

It is noteworthy that, in line with the classical Abbe formulation for single-material lenses (Born and Wolf 1959), the above design equations consider only the central feed. The imposed Abbe condition implicitly determines the scanning behavior for off-axis feeds.

Beam-Steering Lens

In the traditional approach for mechanical beam steering, the primary feed is displaced over the focal arch of the focusing element (lens or reflector) originating a corresponding beam tilt. Alternatively the feed can be fixed and the lens (or reflector) tilted in such a way that its focal arch passes always through the feed. Off-axis feed displacement leads to increased phase error at the aperture, originating progressive beam degradation. These become unacceptable typically beyond 25° beam tilt (notable exception is the Luneburg lens; section "Nonuniform Index Spherical Lenses").

This section presents an alternative approach where the axis for lens tilting is coincident with the lens focus and consequently coincident with the feed phase center (Fig. 22) (Costa et al. 2009). In this way, phase errors associated with the lens tilting are eliminated, and in principle, a larger beam-scanning range can be reached, provided that proper illumination of the lens is maintained for all lens tilt angles. The feed

DOI 10.1007/978-981-4560-75-7_40-1

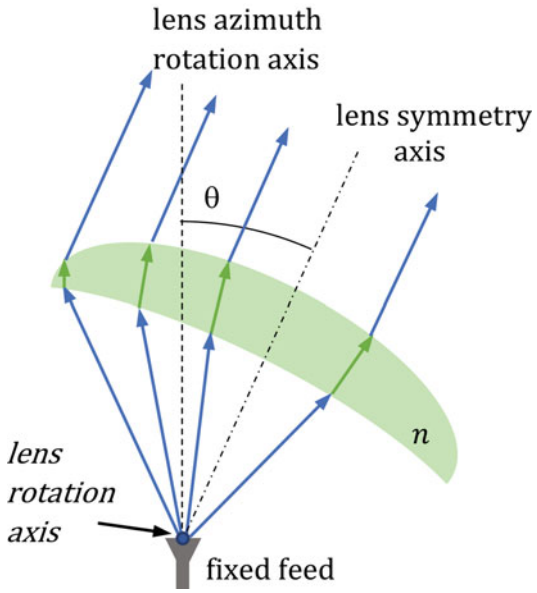


Fig. 22 Working principle of the beam-steering lens

remains stationary. An application example of this lens is presented in section "Mechanical Beam-Steering Lenses."

A collimated beam lens is used for this concept, so, by definition, all output rays emerge always parallel to the lens symmetry axis, making the θ beam tilt angle coincident with the lens tilt angle. The azimuth beam scanning is obtained by simultaneous rotation of the lens about the feed axis.

Of course in this case the feed cannot be in contact with the lens. But the system is designed to have the feed as close to the lens base as possible (in the order of one wavelength) to favor proper lens illumination for all lens tilts. Issues to consider when designing the collimated beam lens are the maximum achievable beam tilt angle, the maximum gain, and the minimum gain scan loss. These characteristics are determined by the lens profile, but they are in part limited by reflections at the dielectric interfaces and by feed illumination spillover as lens tilt increases. An appropriate feed must be designed for proper lens illumination.

This is a case where numerical optimization is required for the lens design. However, instead of brute force optimization of a spline representation of the lens, an alternative hybrid approach is considered (section "Other Lens Design Methods") where GO design equations are combined with a parametric representation of the lens base surface to narrow the search space. The two refraction surfaces of the lens must be designed maximizing the portion of the output lens surface that is able to collimate the feed's radiation. This is equivalent to maximize the lens η_{max} value defined in Fig. 23. The two lens surfaces play a role to broaden this maximum angle.

The lens geometry is shown in Fig. 23. Using Snell's refraction law at the bottom lens interface leads to

$$\frac{\partial r(\eta)}{\partial \eta} = \frac{r(\eta)\, n\, \sin(\gamma - \eta)}{1 - n\, \cos(\gamma - \eta)} \tag{49}$$

where η is the independent variable. On the other hand, by imposing an electrical path length condition, one gets

$$r + n\, l + s = F + n\, T \tag{50}$$

DOI 10.1007/978-981-4560-75-7_40-1

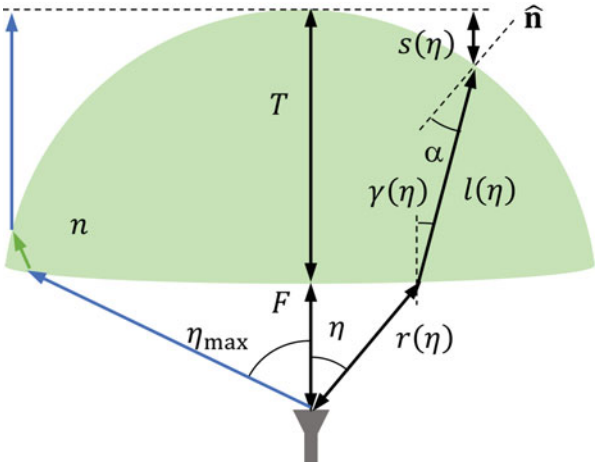


Fig. 23 Geometry for the design of the beam-steering lens

where

$$s = F + T - r\cos(\eta) - l\cos(\gamma) \tag{51}$$

F and T are input constants, whereas $r(\eta)$, $l(\eta)$, and $\gamma(\eta)$ are unknown functions. A third design condition is required to define a unique solution. For that, $r(\eta)$ is analytically written as a Taylor series expansion in η:

$$r(\eta) = \sum_{n=0}^{8} C_n \eta^n \tag{52}$$

So the left-hand side of Eq. 49 can also be written analytically. The following values are set $C_0 = F$ and $C_1 = 0$ in order to impose $\partial r/\partial \eta = 0$ at $\eta = 0$. This ensures null refraction for the central ray. Coefficients C_2 to C_8 are generated by using the genetic algorithm (GA) optimization method. Setting the C_n coefficients defines the $r(\eta)$ function in Eq. 52 so $\gamma(\eta)$ can be calculated from Eq. 49 and then $l(\eta)$ can be calculated from Eqs. 50 and 51. The latter functions define the lens upper surface.

As mentioned, the above formulation is integrated with a GA loop to test different shapes of the bottom lens surface $r(\eta)$, with the goal to maximize the η_{max} angle of the lens, subject to the following constraints:

(a) $r(\eta)$ must be large enough to ensure that the edges of the feed never touch the bottom lens surface when the lens is tilted.
(b) The bottom lens surface cannot cross the upper surface except at the edge of the lens.
(c) Ray incidence angle at the upper lens interface must be below 95 % of the critical angle α_c.

This latter constraint minimizes the excitation of a lateral wave (Pasqualini and Maci 2004) along the lens upper surface. This can happen when ray's incidence angle α, measured with respect to lens local normal $\hat{\mathbf{n}}$, approaches the total reflection condition:

$$\alpha_c = asin\left(\frac{1}{n}\right) \tag{53}$$

DOI 10.1007/978-981-4560-75-7_40-1

As previously referred, the surface wave tends to deflect part of the lens radiation away from the main beam direction reducing the directivity.

Lens Analysis Methods

Geometrical Optics/Physical Optics (GO/PO) Analysis Method

The hybrid GO/PO method is certainly the most used approach for lens (reflector and other open structures) analysis. It takes as input the lens shape and material permittivity, the feed position, and the far-field pattern radiated by the feed when immersed in an unbounded media that has the same permittivity as the lens. The GO/PO procedure involves two steps, as implied by the acronym.

In the first step, GO formulation, as explained in section "Geometrical Optics for Lens Design," is used to compute the field distribution at the inner face of the lens interface; Fresnel coefficients are then used to compute the fields at the outer face of the lens. When one or more dielectric interfaces are crossed by the ray tubes originated at the feed phase center, appropriate Fresnel coefficients and divergence factors must be used.

From the field distribution obtained in the first step, equivalent currents are calculated over the outer face of the lens, and these are Kirchhoff-Huygens (KH) integrated over the lens aperture S to provide the lens far-field radiation pattern:

$$\mathbf{E}(P) = \frac{je^{-jkR}}{2\lambda r}\int_S [Z(\hat{\mathbf{n}} \times \mathbf{H}(P')) \times \mathbf{R}_1 + (\hat{\mathbf{n}} \times \mathbf{E}(P'))] \times \mathbf{R}_1 e^{jk\boldsymbol{\rho}\cdot\mathbf{R}_1} dS \qquad (54)$$

$\mathbf{E}(P')$ and $\mathbf{H}(P')$ represent the field produced by the feed over the lens outer surface, calculated in the first step. $\mathbf{R}_1$ is a unit vector directed from the origin toward the observation point P, $\boldsymbol{\rho}$ is a vector directed from the origin toward the integration point P' on the lens surface, and $\hat{\mathbf{n}}$ is the outward normal to the lens surface. This second step is referred in the literature as the physical optics (PO) integration (Fig. 24).

The GO/PO approach provides very good results for most large aperture antenna problems, provided that the accuracy of the aperture field description is also good. While in the GO direct synthesis approach discussed in the previous section it is enough to consider only the forward propagating rays to define the aperture fields (as already discussed), in the analysis process all the multiple reflected and transmitted rays must be properly accounted. Ling et al. (1989) propose a general procedure for tracing the rays in complex arbitrarily shaped structures in the context of radar cross-section problems, which is referred as the shoot and bounce ray (SBR) method. The SBR method was used for computation of the internally reflected rays in an integrated elliptical lens (Neto et al. 1998) or in an off-axis feed extended hemispherical lens (Pavacic et al. 2006).

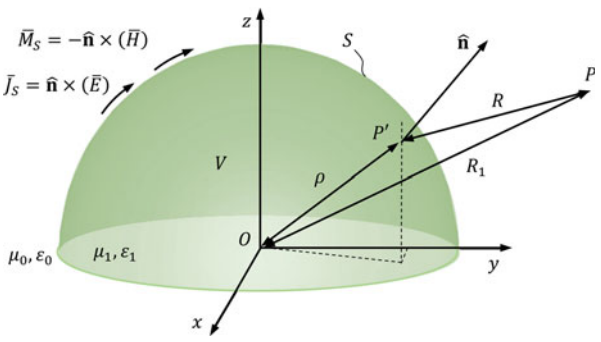


Fig. 24 Geometry for the PO aperture integration

DOI 10.1007/978-981-4560-75-7_40-1

The computation time for the GO/PO method becomes really insignificant when structures are axial-symmetric. Even if the source fields are not axial-symmetric, an appropriate decomposition of those source fields in terms of a series of azimuthal harmonics may transform the nonsymmetrical problem into a superposition of solutions for axial-symmetrical problems.

Memory requirements for the GO part of the procedure are reasonably modest, especially for circular-symmetric structures. A similar comment is valid in general also for the PO part of the procedure, having in mind that in some approaches the far-field pattern can be calculated as the superposition of the sequentially calculated closed-form KH integrations of all exit ray tube fields.

Physical Optics/Physical Optics (PO/PO) Analysis Method

This is again a two-step analysis method, like the previously studied GO/PO, but now the calculation of the aperture fields in the first step is based on the PO formulation. This allows circumventing two GO limitations:

- GO cannot be used in the first step for small lenses, where the feed can no longer be accurately represented by a point and by its far-field radiation pattern.
- GO fails to predict the fields near the caustics. This aspect becomes critical when caustics approach the edge of the lens affecting the correct evaluation of edge diffraction effects or when the caustics approach the feed region, thus affecting significantly the feed impedance (Neto et al. 1998).

With the a priori knowledge of the source currents, a free-space dyadic Green's function can be used to calculate the fields in an unbounded media with the same permittivity as the lens. The fields are calculated at a fictitious surface S with the shape of the lens to represent the incident field ($\mathbf{E}_{inc}$, $\mathbf{H}_{inc}$) at a lens dielectric-air interface. Fresnel coefficients are used to calculate the reflected and transmitted fields at the interface, and equivalent electric and magnetic surface currents are then calculated on both sides of the interface. PO integration of the equivalent currents at the inner face of the lens allows calculating the reflected fields everywhere inside the lens. The far-field radiation pattern of the lens is determined from PO integration of the equivalent surface currents defined outside the interface from the fields obtained in the previous steps (Fig. 25).

The asymptotic expressions that are commonly used for the PO evaluation of the far field cannot be used in general for the evaluation of the interior fields, since the observation point may lie close to the aperture. The use of the general PO formulation valid for near field is a little more cumbersome and more

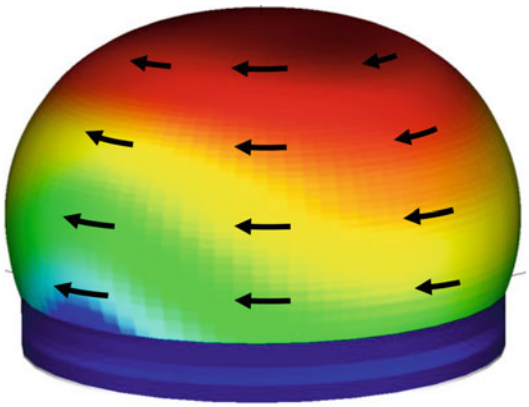


Fig. 25 Equivalent currents on the lens surface

DOI 10.1007/978-981-4560-75-7_40-1

CPU time-consuming. This point aggravates for more complex structures like multiple-shell lenses where more than one PO integration would have to be performed for complete evaluation of the internal fields.

This is clearly a less flexible method and more CPU-intensive than GO/PO. The insertion into an optimization loop is similar to what was discussed before for the GO/PO method, but each iteration takes longer.

Spectral Domain Methods (SDMs)

Spectral domain methods (SDMs) are a possible alternative to GO/PO for the analysis of integrated lens antenna. They are particularly interesting for small lens with only a few wavelengths in size where the classic geometrical optics approach fails.

In SDMs, the lens far field is determined from the decomposition of the feed aperture fields at the lens base into given basis functions. These functions can be, for example, plane waves or Gaussian beams (Maciel and Felsen 1989). The main advantage of Gaussian beams is that they are both spatially and spectrally limited. The estimation of the near field at the lens base requires the use of another method, like the method of moments. The Gaussian decomposition presents several beam parameters that can set the beam width, tilt, and spatial and spectral separation between basis functions.

In Hailu et al. (2009, 2011), a spectral domain decomposition combined with ray tracing is used to analyze an extended hemispherical lens. Results are compared against a full-wave commercial software tool and measurements.

Spherical Wave Modal Method

The spherical wave modal method is based on the discretization of the electromagnetic field in a set of base functions, which are solutions of the wave equation in spherical coordinates. An example of this approach is found, for instance, in Sanford (1994) and Fuchs et al. (2008b) for the analysis of a spherical homogenous or stratified Luneburg lens.

For each layer of the lens, the interior and exterior fields are discretized in spherical modes. A mode-matching technique is used to obtain the expansion coefficients. A priori, the only known coefficients are those of the incident wave at the exterior layer of the lens. All the other coefficients have to be determined from imposing the boundary conditions at the layer interfaces. The application of the boundary conditions creates a series of linear equations that once solved provide the coefficients of all the fields in each region.

To apply the modal method to an integrated lens antenna, the first step is to decompose the radiated field from the planar feed. First, the feed is replaced by its planar representation in terms of equivalent electric and magnetic surface currents. Then that plane is meshed, and each segment is considered as an independent elementary radiator with a known discretization in spherical modes. Therefore, an initial knowledge of the feed's current is needed. Since the modal method is a full-wave method, it implicitly takes into account the internal reflections at the lens interfaces.

The spherical wave mode approach is especially adequate for perfect spherical shapes. For arbitrarily shaped lenses, the boundary conditions of the fields are much more difficult to express in the spherical coordinate system. It is also expected that the convergence of the series becomes unacceptably slow especially when the feed off-axis distance increases.

Method of Moments (MoM)

One of the most popular methods for the full-wave analysis of arbitrary electromagnetic structures is the method of moments (MoM) (Peterson et al. 1998). This method is intended for the evaluation of electric and magnetic field integral equations (EFIE and MFIE, respectively) that describe the EM problems. Those equations may be written in the form $Lf = g$, where L is the integral operator, f the unknown, and g the excitation. An approximate solution may be obtained by MoM in the form:

DOI 10.1007/978-981-4560-75-7_40-1

$$f \cong \sum_{n=1}^{N} \alpha_n B_n \tag{55}$$

where functions B_n are known basis functions defined on the domain of L and the scalars α_n are unknown coefficients to be determined. Equation 55 is then substituted into $Lf = g$, and a system of linear equations is obtained by forcing the residual

$$L\left(\sum_{n=1}^{N} \alpha_n B_n\right) - g = \sum_{n=1}^{N} \alpha_n L B_n - g \tag{56}$$

to be orthogonal to a set of test functions $\{T_1, T_2, \ldots, T_N\}$. The result will be the matrix equation $L\alpha = \beta$ with the entries

$$l_{mn} = \langle T_m, LB_n \rangle \tag{57}$$

and

$$\beta_m = \langle T_m, g \rangle \tag{58}$$

where $\langle , \rangle$ represents the inner product operator. The unknown coefficients can be determined by inverting the matrix (Peterson et al. 1998).

The size of this matrix may become prohibitive for large electromagnetic objects. As previously referred, lenses are usually large in terms of wavelength, so there is not much work reported in the literature about MoM analysis of dielectric lenses.

Finite-Difference Time-Domain (FDTD)

The FDTD technique utilizes a second-order central-difference scheme to approximate Maxwell's curl equations on a representative grid of discrete time and space. The electric and magnetic fields' components are successively updated by repeated implementation of the finite-difference equivalents of the curl equations until the desired transient or steady-state response is achieved. This process is known as leapfrog time marching. Because the solution in the FDTD method is built up through time, it alleviates the need to form a large system of equations and thereby turns an otherwise intractable problem into one of patience.

As with MoM, the use of traditional 3D FDTD as an iterative analysis algorithm is restricted to lenses with dimensions similar to the wavelength. In fact, the number of variables of the algorithm increases cubically with the radius of the lens leading to impractical computation times for most common optimization problems. However, for structures with axial symmetry, it is possible to reduce the 3D problem to a simpler 2D one and implement a faster algorithm called body of revolution (BOR)-FDTD. BOR-FDTD can be used even if the symmetric lens is fed by an asymmetric source (Van Der Vorst and De Maagt 2002). In that case, the feed's current or its radiated field may be discretized into a Fourier expansion. Each element of the expansion is separately analyzed by the BOR-FDTD algorithm, and the final result is obtained by the vector sum of all the partial BOR-FDTD analysis.

BOR-FDTD can be implemented for small lenses where, for instance, the GO/PO method cannot be used. In fact, the smaller the lens, the faster the BOR-FDTD will converge for the solution since the number of space grid elements reduces with the size of the lens. BOR-FDTD can easily handle internal

DOI 10.1007/978-981-4560-75-7_40-1

reflections since it is a full-wave method and can also be implemented for multiple-layer or varying dielectric constant lenses.

Lens Materials

There are a few aspects to be taken into consideration when selecting dielectric materials for lens fabrication. Usually the materials are selected based upon their dielectric constant or relative permittivity, and the usual values range from 1.2 to 11. A second important parameter is the dielectric loss tangent (tan δ) which measures the material dissipation loss. The lower the loss tangent value, the lower the material losses per wavelength. Acceptable values of the loss tangent depend upon the application; nevertheless, in most cases, a value lower than 10^{-3} is usually considered adequate. It is also important to know if the dielectric materials are isotropic and homogenous. There are also mechanical considerations like material hardness, fracture toughness, or melting temperature which may be relevant for the selection of the lens fabrication technology and thermal expansion coefficients which may be relevant for multi-shell lenses.

The electromagnetic properties of dielectric materials at millimeter wave may vary between manufacturers and even between material batches. For demanding specifications, it is advisable to accurately measure the dielectric properties of a sample of the material that will be used for the lens fabrication. Evaluation of the complex permittivity of dielectric materials usually involves vector network analyzer measurement of the frequency response of a test millimeter-wave circuit (it can be a resonator or not) when loaded with a dielectric sample and when empty (Chen et al. 2005). Two classical methods are described next: the waveguide method and the open Fabry-Perot resonator method.

Parallelepiped samples of the dielectric material are required for the waveguide method, precisely cut with the same length of the short-circuited waveguide sample holder and filling completely its cross section (Chen et al. 2005). These waveguide sections operate in single-mode TE_{10} conditions. The complex permittivity of the material is deembedded from the measured s_{11} frequency response at the open port of the sample holder, both filled with the dielectric sample and when empty (Silveirinha et al. 2014). Two sample holders with different lengths may be used for redundant permittivity determination and detection of eventual measurement uncertainties (Fig. 26).

One advantage of this method is that it is quite precise, provided that the air gap between the dielectric sample and the involved five walls of the waveguide sample holder is negligible. Another advantage is that it requires very small samples of the material (typically $3.9 \times 1.9 \times 5.0$ mm^3 in the V-band) which enables to detect material inhomogeneity or anisotropy by cutting samples from different parts and different orientations of the bulk material. Besides, it can be shown that some fortuitous measurement errors or the appearance of higher-order modes can be easily perceived during the measurements, indicated by abnormal irregularities in the measured s_{11} frequency response

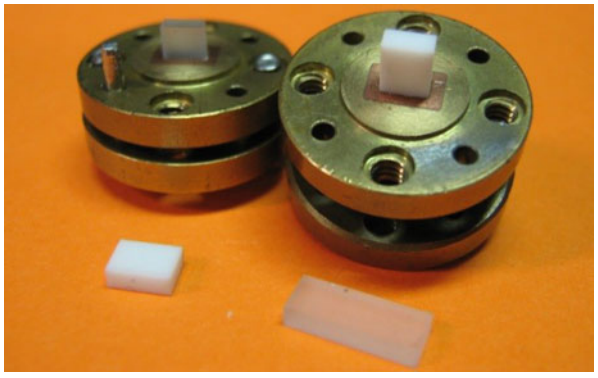


Fig. 26 Example of waveguide short-circuited sample holder for V-band and a few dielectric samples

DOI 10.1007/978-981-4560-75-7_40-1

(Silveirinha et al. 2014). Another interesting point with this method is that it allows measuring loss tangents in the range up to 10^{-2}.

The method presents however limitations for high dielectric constant values (typically larger than 6), associated with the excitation of higher-order modes in the dielectric-filled sample holder. This introduces significant error in the complex permittivity evaluation, especially in the loss factor.

The Fabry-Perot resonator method does not present the previous limitation for high permittivity samples. Its theory and operation principles are very well known (Afsar et al. 1990; Komiyama et al. 1991; Hirvonen et al. 1996). There are different possible configurations for the resonator. The plane-concave configuration is commonly used since it involves only one spherical mirror and allows precise positioning of the dielectric samples at the Gaussian beam waist without the need for perturbing holders. The material samples are typically cut as disks with a diameter typically three times larger than the Gaussian beam waist and thickness of the order of half the wavelength in the material under test (Hirvonen et al. 1996).

In the example shown in Fig. 27 for the V-band (Fernandes and Costa 2009), the spherical mirror is made of aluminum, with 160.3-mm curvature radius and 240-mm projected diameter. A linear translation is allowed for the planar mirror, and a 10-μm precision gauge is associated with this translation. When the central distance between mirrors is around 157.3 mm, this enables a good compromise at 63 GHz between a narrow beam waist of the fundamental Gaussian mode at the planar mirror ($w_0 = 5.7$ mm) and the Gaussian beam width at the spherical mirror ($w_z = 42.9$ mm). This choice makes it possible to measure material samples with diameters as small as 20 mm. Increasing the distance between mirrors reduces w_0 but increases w_z with consequent risk of spillover at the edges of the spherical mirror and consequent reduction of the cavity quality factor.

Measurements can be performed using the fixed mirror distance approach. A vector network analyzer sweeps the frequency within the V-band and acquires the s_{12} transfer function between the two waveguide ports connected to the back of the spherical mirror. In a first run, the resonator is measured without any sample. Then a second measurement is performed with a dielectric disk sample placed at the center of the plane mirror. The dielectric constant of the sample is determined from the frequency shift in the resonator modes, and the loss tangent is obtained from the resonance bandwidth of a given mode (Komiyama et al. 1991).

This method also allows evaluating the anisotropy of the material by performing successive measurements where the sample is rotated in relation to the spherical mirror axis of symmetry. In fact, since the cavity is operated with linearly polarized modes, the s_{12} transfer functions depend on the material

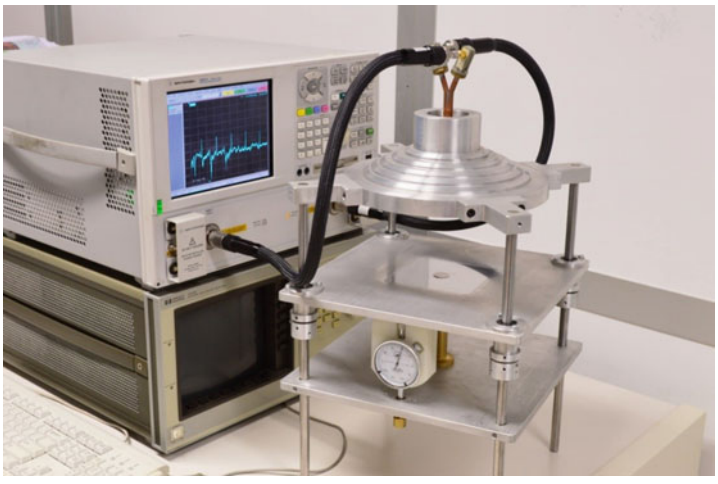


Fig. 27 Fabry-Perot open resonator for complex permittivity measurements in V-band

DOI 10.1007/978-981-4560-75-7_40-1

Table 3 Measured permittivity values of different materials at 60 GHz

Material	Dielectric constant	Loss tangent
ABS-M30 (3D printed)	2.48	0.008
Acrylic glass	2.5	0.0118
Alumina	9.3	0.0013
Fused quartz	3.8	0.0015
MACOR	5.5	0.0118
Polyethylene	2.3	0.0003
Polypropylene	2.2	0.0005
Polystyrene	2.5	0.0004
Teflon	2.2	0.0002

Fig. 28 Lens fabrication by CNC milling machine technique

anisotropy axis orientation in the disk plane (Fernandes and Costa 2009), providing different permittivity values depending on the sample rotation.

Table 3 presents a list of measured electromagnetic characteristics at 60 GHz of some commonly used dielectric lens materials.

Lens Fabrication

Several techniques exist to manufacture a dielectric lens antenna. The most common approaches use either computer numerical control (CNC) milling machine, molding, or 3D additive manufacturing.

CNC enables to produce physical objects from its digitized 3D description, by automatically excavating the object shape out from a block of raw material (Fig. 28). This type of fabrication process is especially appropriate for antennas with complex shapes, tight dimensional tolerances, and good surface finish as happens, for instance, with shaped lens antennas at millimeter waves. It is more indicated for small series production or for lab prototyping as opposed, for instance, to the injection molding process, which is cost-effective for mass production. CNC milling may, however, be used for mold production for the latter technology.

There are no special limitations on the materials that can be used for CNC milling, except that they must be machineable. Some examples of commercially available machineable dielectrics are listed in Table 3. The cutting tool and its rotation speed, cutting speed, feed rate, and cutting depth must be appropriate for

DOI 10.1007/978-981-4560-75-7_40-1

Fig. 29 Lens fabrication by molding technique

the selected material. The specific values for these cutting parameters can be found in manufacturers' material data sheets. Depending on the dielectric material and CNC milling machine characteristics, 50-μm fabrication accuracies can be achieved. For some types of materials, it is possible to fabricate the lens using a molding technique (Fig. 29). In this case, a mold with the shape of the lens has to be manufactured where the dielectric, in liquid phase, is poured or injected into. Usually, the assembly must be left for a few hours before the mold can be removed. Depending on the material, the lens dimensions may change marginally after material curing; this has to be considered when designing the mold. Molding is generally considered a cost-effective technique for mass production, but it is of limited advantage for lens lab prototyping due to the mold manufacturing cost. Furthermore, it is not easy to avoid trapped gas bubbles or to control the homogeneity of the mixture as it becomes solid without an appropriate facility.

In the case of multilayer lenses, particular attention has to be given to the matching and alignment of the layers produced by milling or molding technologies. Air gaps between the layers should be avoided when assembling the lens. The presence of air gaps as small as 0.03 wavelengths in multilayer lenses can reduce the antenna gain and increase side lobe level of the radiation pattern. This effect is more intense with the increase of the number of layers (Kim and Rahmat-Samii 1998) or with higher-permittivity materials (Nguyen et al. 2010). A glue can be used to mitigate the effect of the air gaps as long as it has approximately the same dielectric constant of one of the adjacent layers.

3D printing is becoming an attractive technique to fabricate elaborate dielectric prototypes at relatively low cost. As with CNC, it starts with a 3D digitized model of the object, but the rendering is based on an additive process where successive layers of a material are laid down dot by dot to construct the desired shape. Various materials can be considered, but thermoplastic-based 3D printers are the most popular. Although the commonly used materials tend to have moderately high losses and the fabrication tolerance of low-end printers is presently of the order of 200 μm, the viability of a 3D-printed lens antenna made of ABS ($\tan \delta = 0.008$) for short-range indoor wireless link at 60 GHz has been recently demonstrated (Bisognin et al. 2014). A similar technique can be used to manufacture lenses using higher-permittivity lower-loss materials like alumina (Ngoc Tinh et al. 2010).

Lens Feeds

The main requirements for a lens feed are a reasonably uniform radiation pattern within the lens aperture subtended angle, minimum energy spillover at the lens edges, and a well-defined phase center coincident with the lens focal point. The most common families of lens feeds are either patches, slots, or horns/waveguides. Patches and slots have the advantage of low profile and in some cases can be made conformal with the lens surface. However, they tend to have a broad radiation pattern which may lead to excessive spillover losses particularly in off-body fed lenses. Patches and slots are thus more appropriate for integrated lens feeding. Waveguides, although bulky, may present some advantages for lab testing of

DOI 10.1007/978-981-4560-75-7_40-1

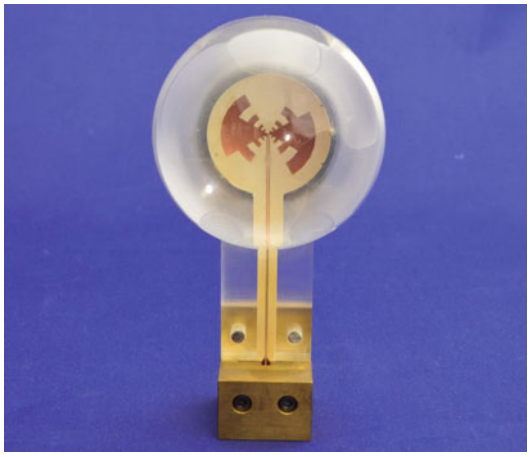


Fig. 30 Integrated lens antenna fed by a planar self-complementary log-periodic feed

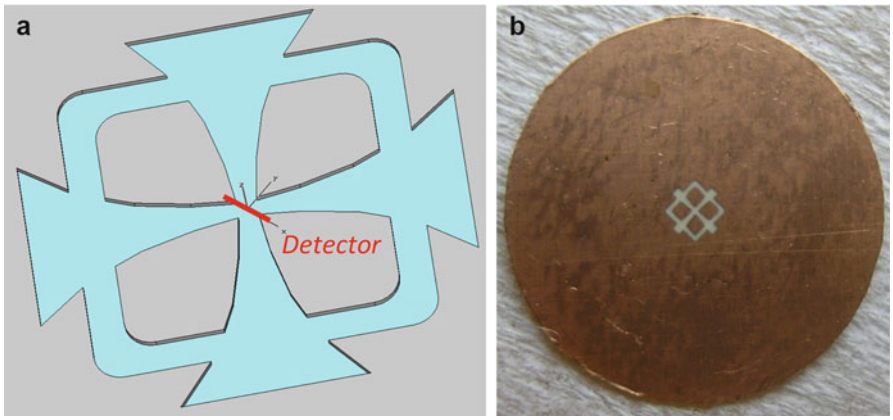


Fig. 31 (**a**) Configuration of a crossed exponentially tapered slot (XETS) antenna; (**b**) prototype of the XETS for operation from 40 to 75 GHz

integrated lens prototypes as they allow reusing the same feed in several lens prototypes while ensuring measurement repeatability. Horns tend to have a much more directive beam than all the previous types of feeds and are more common as off-body lens feeds.

The earlier works on integrated lens antennas at submillimeter waves used a double-slot antenna as the lens feed (Filipovic et al. 1993, 1997). This type of feed is particularly indicated for integration, for instance, with bolometers, in applications where the lens is used only as an amplitude receptor like in radio astronomy. However, the double-slot antenna is narrowband and therefore limits the inherent wideband nature of the lens. Wideband feeds like the sinuous antenna (Edwards et al. 2012) and log-spiral or log-periodic antennas (Semenov et al. 2007) have been used for radio astronomy lens applications (Fig. 30).

One problem with the previous wideband feeds is that the polarization is not stable with frequency. An alternative wideband planar feed that solves that problem is the crossed exponentially tapered slot (XETS) antenna (Costa and Fernandes 2007a). It presents a frequency-stable linearly polarized radiation pattern versus frequency with stable phase center position at the center of the antenna. The operation bandwidth is 1:3 (or 100 %). The antenna geometry is presented in Fig. 31: it is formed by a single metal layer with two crossed exponentially tapered slots, intersected by a square (or star)-like slot. The antenna is fed between two opposite petals, as shown in Fig. 31a, which define the antenna E-plane. The antenna perfect

DOI 10.1007/978-981-4560-75-7_40-1

symmetry with respect to the feed point ensures perfect symmetry of the antenna currents with respect to the E-plane and perfect anti-symmetry with respect to the H-plane, irrespective to frequency. Consequently, pure linear polarization is obtained versus frequency at least in the main planes.

A slot feed with an extremely wide band of more than 1:10 has been presented in Neto (2010) and Neto et al. (2010). It is composed by a long leaky slot line fed at the middle by an orthogonal microstrip line printed in the other face of the substrate. An air gap smaller than 1/16 of the wavelength is placed between the slot and the base of the lens in order to improve the feed radiation performance. This concept has recently been extended to a square array of leaky slots (Yurduseven et al. 2014).

All of the above single-layer antennas present a bidirectional radiation pattern due to the absence of a ground plane. This is attractive for THz applications to ease the feed fabrication. However, when attached to the base of a dielectric lens, most of the feed radiation is pulled into the lens body. In fact, the fraction of power that is coupled to the lens increases with the permittivity as $\varepsilon_r^{3/2}$ (Rutledge et al. 1983). Thus, the lens material permittivity tends to be high to operate with these uniplanar feeds. Higher reflection is thus expected at the air/dielectric interface, but it can be overcome with the matching layer approach or with the double-shell lens configurations discussed in section "Integrated Lens Design Examples."

Classical patch antennas are also used as integrated feeds. For example, an integrated lens with an array of aperture-coupled rectangular patches is fed by a microstrip feeding network at 28 GHz (Nguyen et al. 2011) or at 60 GHz (Artemenko et al. 2013a, b).

An open-end waveguide aperture is sometimes preferable for integrated lens antenna feeding, as it is easier to couple the full power to the lens without the need to increase the lens permittivity as it happens with uniplanar printed antennas. However, the standard waveguide open end has to be modified to avoid excessive return loss and asymmetrical radiation pattern between E- and H-planes. The inclusion of an edge-tapered dielectric filling with the same permittivity of the lens material can mitigate the return loss problem (Ngoc Tinh et al. 2010); however, it may originate higher-order modes that affect the lens performance. To avoid this difficulty, the cross-section dimensions of the dielectric-filled waveguide must be conveniently tapered to enable single-mode operation (Fernandes 2002; Fernandes et al. 2011). When appropriately designed, the waveguide open end itself is enough to reproduce the radiation pattern inside the lens body comparable to that of a printed or slot broadband antenna. Although the operation band of the waveguide feed may not be wide enough, the broadband characteristic of the lens can still be tested using separate waveguide feeds, each one designed for partial bands. Figure 32 presents an example of the layout and photograph of two similar waveguide feeds for Q- and V-band (Fernandes et al. 2011). Due to the outer thread, these feeds can be easily attached or removed from the lens and reused in other lenses ensuring repeatable measurements.

For off-body fed lenses, it is common to use horn feeds. The requirements and the used configurations are exactly the same as for reflectors, so ample literature exists (Olver et al. 1994). In order to avoid excessive spillover losses, usually it is required that the illumination at the edge of the lens is 10 dB below the power level at the center of the lens.

Lens Measurements

Most of the lens antenna measurement techniques are no different from those used for reflector antennas or for any other large aperture antennas. So the most common measurement procedures are not repeated here. This section addresses only two specific issues which are relevant for integrated lens antenna measurements.

Feed Radiation Pattern in Unbounded Dielectric Media

Integrated lens antenna design requires the knowledge of the actual radiation pattern of the feed when it is embedded in the lens material. This is of course different from the radiation pattern of the same feed in air.

DOI 10.1007/978-981-4560-75-7_40-1

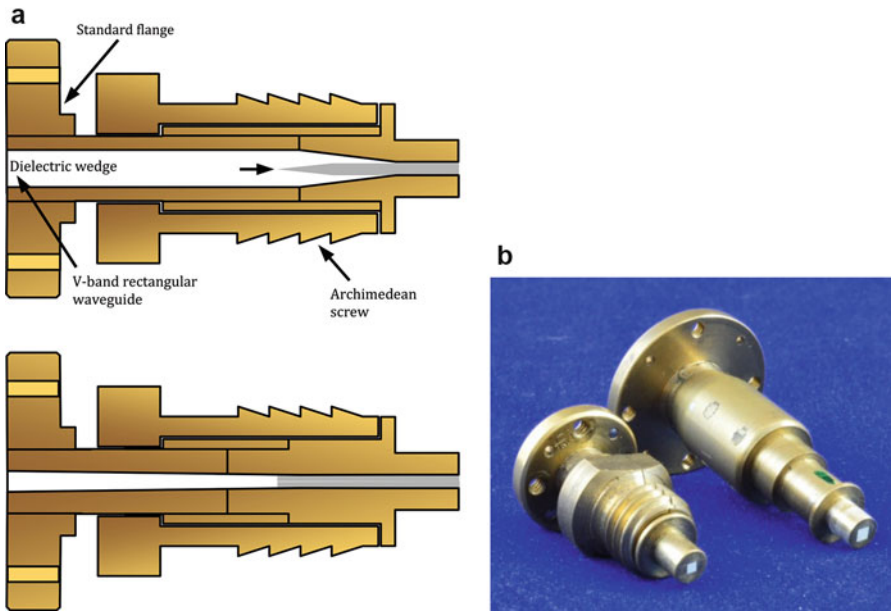


Fig. 32 (**a**) Schematic view of the feed prototype (not to scale). *Above*, longitudinal cut in the H-plane; *below*, E-plane; (**b**) photograph of the fabricated feeds. *Left*, for V-band operation; *right*, for Q-band operation

Although the feed radiation pattern in unbounded dielectric medium can be obtained from some full-wave EM solvers, the direct measurement is not possible.

One practical solution that provides a reasonably good approximation is to use a spherical lens of the same dielectric material and measure at its surface the tangential fields produced by the centered embedded feed. The lens radius must be large enough to ensure the predominance of the transverse field components of the feed radiation at the lens surface. Noting that the tangential components are continuous across the dielectric-air interface and that there is no refraction in this configuration, the measured tangential fields can be considered a fair estimation of the feed radiation pattern inside the dielectric. Internal reflections may generate some ripple but generally do not preclude the procedure.

Figure 33 shows the application of the method for the waveguide feed presented in the previous section. The feed is embedded in a MACOR hemispherical lens, and the near-field scanning probe is also a waveguide open end (Fig. 33a). The near-field measured result is presented in Fig. 33b, superimposed on the simulated far field of the feed in the dielectric, computed with WIPL-D EM solver (Kolundzija and Djordjevic 2002). The agreement is quite acceptable.

Phase Center Evaluation

Phase center knowledge is required when integrated lenses are used, for instance, as primary feeds in focusing systems. Unlike horns, where the phase center position is normally located close to the aperture plane, the phase center of integrated lenses is very dependent upon the lens shape and may even be far out from the lens body. It is a simple matter to determine it when the radiation pattern phase $\phi(\theta, \varphi)$ can be measured. Considering the geometry of Fig. 34a, it can be shown that the distance d between the phase center and the rotation axis in a constant ϕ-plane cut is given by

$$d = -\frac{\lambda}{2\pi} \frac{\partial \phi}{\partial(\cos(\theta))} \tag{59}$$

DOI 10.1007/978-981-4560-75-7_40-1

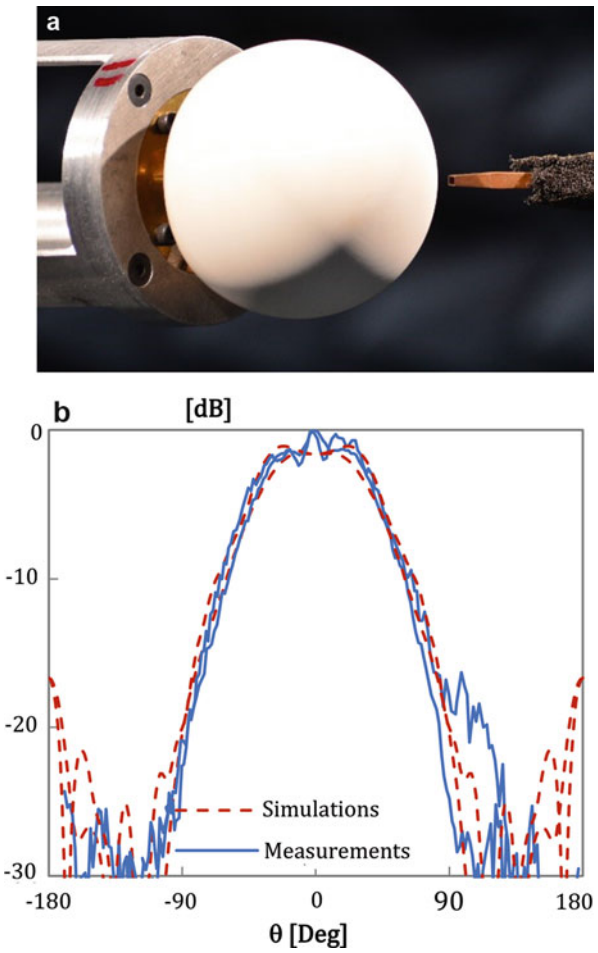


Fig. 33 Near-field scan of the tangential fields of a hemispherical MACOR lens fed by a centered waveguide feed. (**a**) Setup; (**b**) simulated (*red dashed curves*) and measured (*blue solid curve*) E- and H-planes

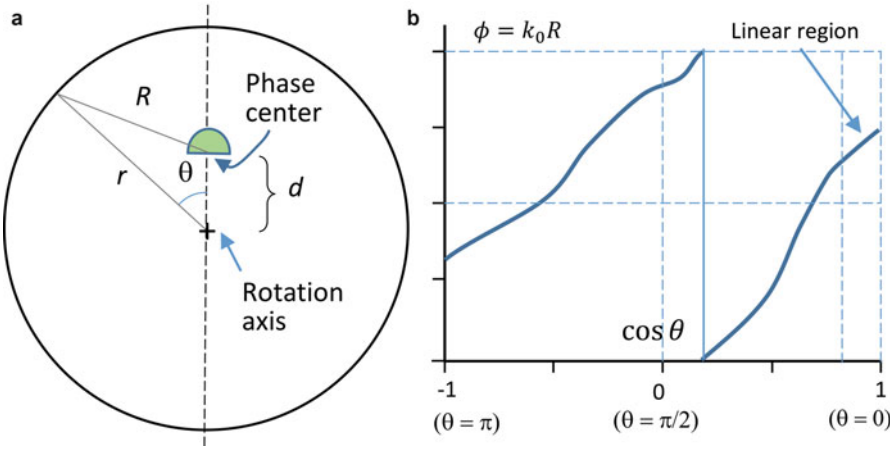


Fig. 34 (**a**) Geometry for phase center determination; (**b**) representation of the radiation patter phase ϕ versus cos θ

DOI 10.1007/978-981-4560-75-7_40-1

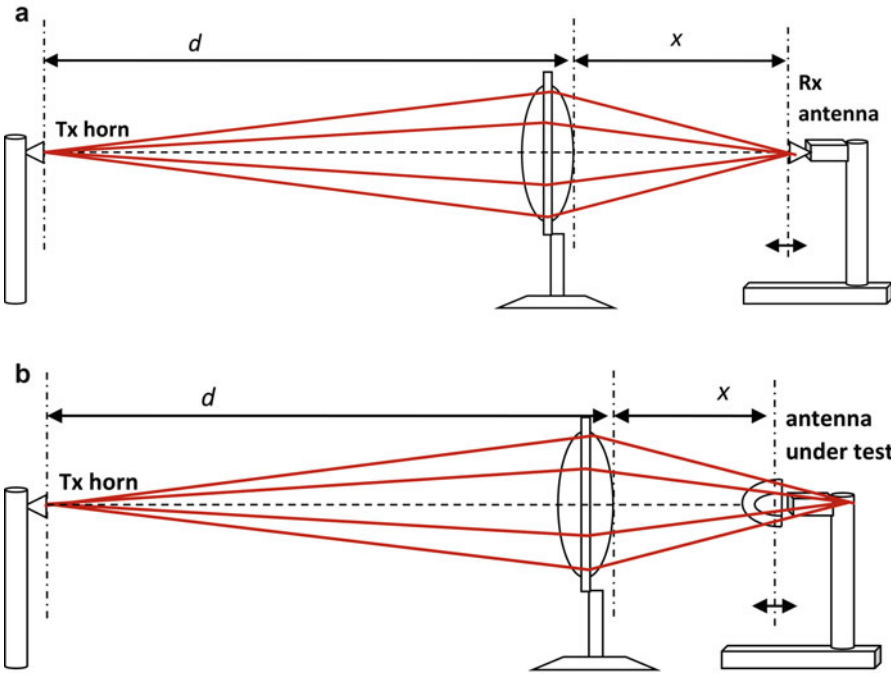


Fig. 35 Schematic setup for phase center determination based on power measurements: (**a**) reference measurement; (**b**) antenna under test measurement

assuming that $d \ll r$. So, when the phase center exists, d is proportional to the slope of the linear variation of ϕ versus cos (θ). This allows calculating d using just a single radiation pattern measurement.

However, when bolometers or other detectors are used as integrated lens feeds (Filipovic et al. 1993), only amplitude radiation patterns can be measured, preventing the application of the previous phase center determination method. The obvious alternative is to place the lens in the focal plane of a simple focusing system. An auxiliary convergent lens (ACL) can be interposed at a fixed position between the transmitting antenna and the lens under test (LUT), while the latter is translated along the link axis until the received power is maximized. That would correspond to coincident ACL and LUT phase centers. An additional reference measurement is required with the lens replaced by an antenna with known phase center to determine the absolute phase center of the LUT (Costa et al. 2010; Fig. 35).

The advantage of using a lens (instead of a reflector) is that it does not need to be very large. An example is presented below where the ACL lens diameter is only three times larger than the LUT diameter, so the procedure is very convenient for small anechoic chambers. If such a small reflector was used instead, the LUT would interpose between the Tx horn and the reflector producing severe aperture blockage and consequent perturbation of the measurement.

Figure 36 shows an example of the setup where the ALC, made of Teflon ($\varepsilon_r = 1.96$ at 62.5 GHz), has a 300-mm diameter and its focal points are located 350 mm to the Rx side and d = 4.550 mm on the Tx side. The LUT is a double-shell lens fed by a waveguide (Fernandes et al. 2010) to allow verification of the amplitude-based method against the phase-based method. The reference antenna is a waveguide probe with the phase center practically coincident with its open-end aperture. The difference between values from phase measurements and power measurements is of the order of 2 mm at 62.5 GHz which is much smaller than the depth of field of practical focusing systems at this frequency.

DOI 10.1007/978-981-4560-75-7_40-1

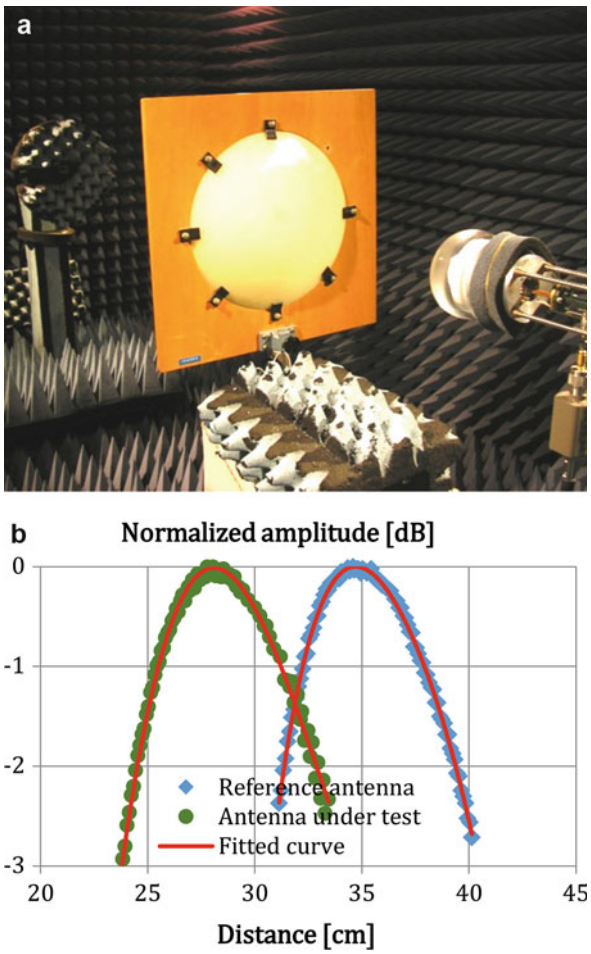


Fig. 36 Practical example. (**a**) Setup; (**b**) measured normalized received power vs. LUT translation compared with reference antenna translation

Applications

Overview of Applications

This section focuses on the recent trends of dielectric lens applications, mostly involving the integrated lens configuration for microwave and millimeter-wave applications where the lens size and weight are reduced. Some of the examples listed below and associated lenses are detailed in the subsequent sections.

The most common function of an integrated lens is to focus the radiation of the feed and increase the overall gain while keeping with a lower profile than what can be achieved with a reflector. This is particularly important in applications at millimeter waves like high bit rate communications at 60 GHz (Bisognin et al. 2014) or automotive anti-collusion radar at 77 GHz (Ka Fai et al. 2014). Alternatively, the integrated lens can be used to shape the radiation pattern of the feed into a desired pattern. For example, for 60-GHz indoor wireless networks, it is required to have base stations with adequately shaped radiation patterns like flat-top (Ngoc Tinh et al. 2011; Rolland et al. 2011) or secant-squared (Fernandes and Anunciada 2001; Bares and Sauleau 2007).

Lenses are inherently wideband so they are interesting feed candidates for some reflector focusing systems in radio telescope applications, where a wideband frequency-stable radiation pattern and phase center position are required for reflector illumination. Integrated lenses have been proposed for stable

DOI 10.1007/978-981-4560-75-7_40-1

performance with either multilayer (Fernandes et al. 2010) or dome (Ngoc Tinh et al. 2013) configurations. In both cases, noncanonical lens shapes were obtained after proper design and/or optimization.

Lenses can be used in imaging applications, where the image of the target is focused through the lens into an array of sensors (or detectors) at the other side of the lens (Filipovic et al. 1997). The spacing between sensors should satisfy the Nyquist sampling criterion (typically two detectors per wavelength) in order to recover the original target image. Security screening and medical imaging are examples of applications using lens antennas. For example, in Trichopoulos et al. (2010, 2013), a hemispherical lens is used with 31 by 31 detectors at its base to detect the object image. Alternatively it is possible to use an array of closely spaced lenses with one detector per lens (Llombart et al. 2013; Naruse et al. 2013). However, in this case, the Nyquist criterion is not fulfilled. There are also some applications requiring multi-beam antennas but with a lower density of detectors like for a lens with an electrically reconfigurable beam shape (Nguyen et al. 2011) or some degree of electrically controlled discrete beam steering (Costa et al. 2008a; Artemenko et al. 2013a, b).

Communications between a movable vehicle and high-altitude platform or satellites require a ground terminal antenna with a beam-steerable capability. Some of the few examples in the literature that use lenses for satellite-on-the-move (SOTM) applications adopt Luneburg-based configurations with mechanical beam-steering capability. For example, companies like Lun'tech in France are selling commercial solutions. Alternatively, integrated lens antennas can also be used for mechanical beam steering in SOTM applications (Costa et al. 2008b). In those configurations, the inner shell of the lens is air, allowing the lens to move freely above the fixed feed to steer the direction of the main beam.

Lens Application Examples

Constant Flux and Flat-Top Lenses for Millimeter-Wave WLAN

Pioneering work on millimeter-wave cellular mobile communications started in the early 1990s (Fernandes 1995) at that time aiming at 150 Mbit/s per channel. Oxygen attenuation is extremely high in this band (15 dB/km), free-space attenuation is also very high, and the same happens with obstacle attenuation. Altogether, these factors intrinsically confine radio coverage to very small cells, so unlicensed spectrum is available for emerging broadband applications.

The 60-GHz band is recently being explored for short-range multi-gigabit wireless local area networks. In indoor scenarios (Fig. 37a), the difference in free-space attenuation for different locations in the same room may represent a high dynamic range which can be compensated with an appropriately shaped antenna radiation pattern. A shaped dielectric lens antenna can be used at the base station (BST) hanging from the ceiling that produces a secant-squared ($\sec^2\theta$) type of radiation pattern in the elevation plane in order to compensate for the free-space attenuation at each direction. The target radiation pattern can be expressed as

$$G(\theta) = \begin{cases} G(\theta_{\max}) \frac{\sec(\theta)}{\sec(\theta_{\max})} & \theta \le \theta_{\max} \\ 0 & \theta > \theta_{\max} \end{cases} \tag{60}$$

The maximum elevation angle $\theta_{\max}$ is selected according to the desired cell radius. A sharp fall of the radiation can be imposed for $\theta > \theta_{\max}$ to conserve power and to prevent excessive reflections from the walls. An estimation of $G(\theta_{\max})$ can be obtained from Fig. 37b which presents the directivity of an ideal radiation $\sec^2$ pattern versus the maximum illumination angle $\theta_{\max}$ (Fernandes 1999).

The BST antenna must be paired with a flat-top radiation pattern antenna at the mobile terminal (MT) so that the average received power remains reasonably constant for all positions of the mobile or portable terminal within the cell. Many antenna technologies can be used to obtain such type of radiation pattern.

DOI 10.1007/978-981-4560-75-7_40-1

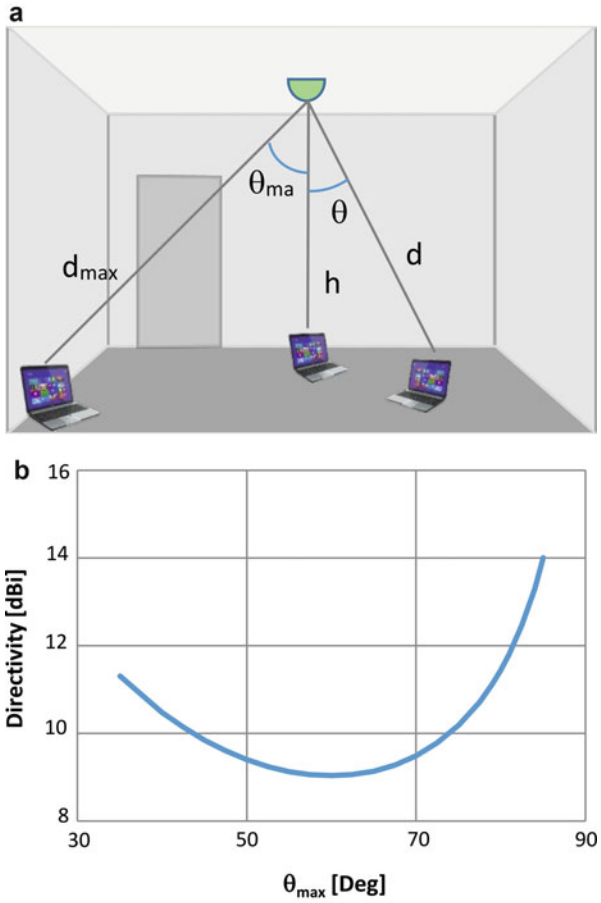


Fig. 37 (**a**) Millimeter-wave wireless LAN scenario; (**b**) directivity of an ideal $\sec^2$ elevation radiation pattern with azimuth symmetry versus θ_{max}

Here a shaped lens is demonstrated, due to its possibility to reduce sharply the radiation beyond θ_{max} and thus reduce multipath pickup. Although the radiation patterns of the BST and MT lens antennas are circular symmetric in the present example, a square coverage footprint could also be obtained for the BST with a 3D-shaped lens, if required (Fernandes and Anunciada 2001).

The design of these axial-symmetric amplitude shaping lenses followed the GO formulation presented in section "Lens Designed to Match an Output Power Template." Plexiglas material ($\varepsilon_r = 2.53$, $\tan\delta = 0.012$) was used for both lenses, although other lower-loss commercially available materials like polystyrene ($\varepsilon_r = 2.53$, $\tan\delta = 10^{-4}$) might also be possible. Both lenses were fed by the aperture of circular metallic waveguides embedded in the lens body, carrying a circular polarized TE_{11} mode. The lens radiation pattern dropout was fixed at $\theta_{max} = 75°$ for both lenses. This corresponds roughly to 11-mm cell radius for h = 3 m BST antenna height with respect to MT.

Figure 38a shows the fabricated prototype of the BST $\sec^2$ lens, with 66-mm radius. Its measured radiation pattern at 62.5 GHz is presented in Fig. 38b. The nadir (the floor in this case) corresponds to $\theta = 0$, and $\theta_{max} = 75°$ corresponds to the maximum gain direction.

The fabricated flat-top lens is shown in Fig. 39. In this case, the lens radius is 35 mm, considerably smaller then for the $\sec^2$ lens where the sharp features of the target pattern were more pronounced. The MT elevation pattern is presented in Fig. 40b, showing a good coincidence with the desired flat-top characteristic. The zenith (ceiling in this case) corresponds to $\theta = 0$. This flat-top characteristic and the pattern circular symmetry favor free movement of the MT within the cell limits and even some tilting.

DOI 10.1007/978-981-4560-75-7_40-1

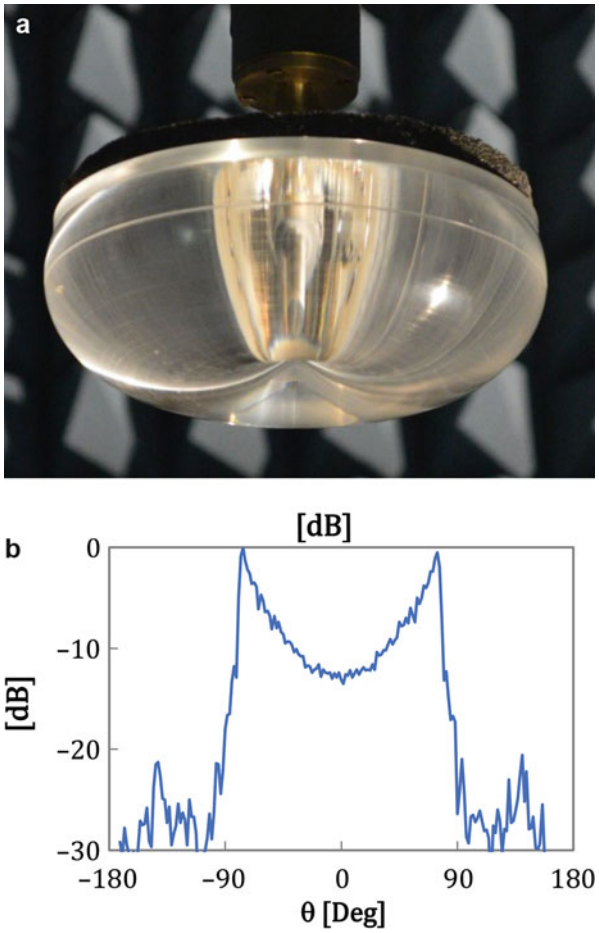


Fig. 38 (**a**) Prototype of the $\sec^2$ lens; (**b**) measured circular polarization radiation patterns at 62.5 GHz

Considering that the link budget is affected by the product of the two antenna gain functions, this lens combination further provides very sharp cell boundaries at θ_{max}, with negligible radiation outside the cell limits. A remarkable characteristic of $\sec^2$ patterns is that cell dimensions are scaled to the antenna height. This provides a simple means to control the illumination of the walls at the cell edges and maintaining an adequate compromise between multipath effects and the need for alternative paths in case of line of sight (LOS) blockage.

Broadband Lens for Constant Illumination of a Reflector

The most common reflector feeds are based on horn antennas. Despite many variants that horns may have, they all share the same feature of aperture antennas, which is a radiation pattern that increases its directivity with the frequency. When large bandwidths are involved, this may represent an important reduction of the reflector aperture illumination efficiency. This problem is often circumvented by splitting the wide band into multiple sub-bands, each with its own feed horn; but then a complex setup is required to minimize aberrations due to off-axis positioning of the multiple horns.

A challenging specification for a reflector feed is the ability to maintain a constant radiation pattern beam width with stable phase center over a 1:3 bandwidth. A double-shell shaped dielectric lens can be designed to meet such specification using the formulation from section "Frequency-Stable Radiation Pattern and Phase Center Position" (Fernandes et al. 2010). The lens is intended to feed a 90° offset

DOI 10.1007/978-981-4560-75-7_40-1

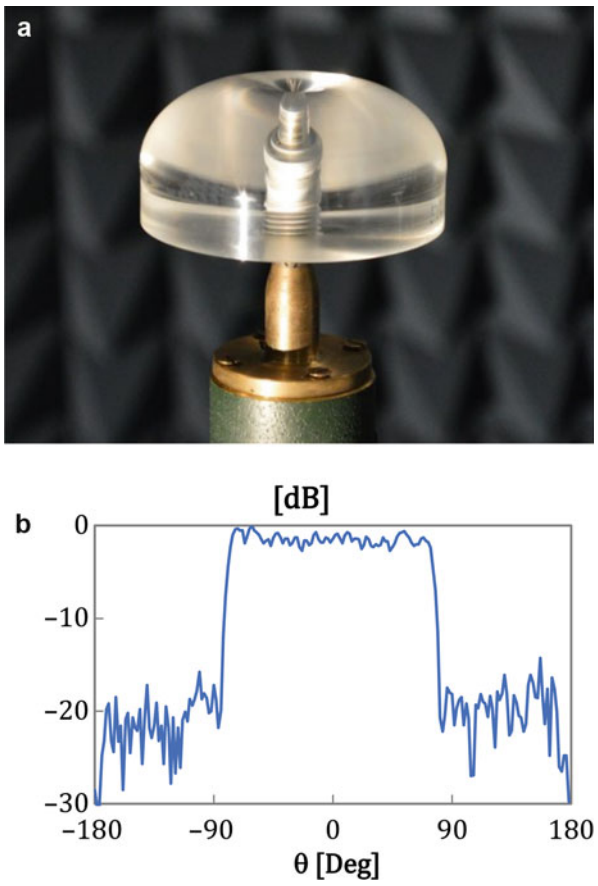


Fig. 39 (**a**) Prototype of the flat-top lens; (**b**) measured circular polarization radiation patterns at 62.5 GHz

reflector. The lens output beam produces a virtual focus located far below the lens and close to the corresponding reflector focal point (Fig. 40a).

A 20 λ diameter lens prototype was designed and fabricated using MACOR and acrylic (Fig. 40b). The optical transparency of the acrylic reveals the inner shell, which appears distorted due to refraction effects. The lens is fed by the waveguide device described in section "Lens Feeds" (Fig. 32). The measured lens radiation patterns are shown in Fig. 41a for two distinct frequencies in the Q- and V-band (40 GHz and 62.5 GHz). The measured lens radiation pattern agrees very well with the Gaussian template with $\alpha_0 = 23^\circ$ Gaussian width (dashed curve). It is possible to see that the shape and beam width of the lens radiation pattern are identical in both bands as desired. The same behavior was demonstrated by full-wave simulation between 30 and 90 GHz (Fernandes et al. 2010). The lens output beam Gaussicity is better than 94 % over the band. The measured phase center position of the lens agrees with the value imposed during the design ($z_0 = 68.7$ mm), and it is identical in both frequencies. Full-wave simulations confirm phase center stability over the 1:3 bandwidth.

The performance of the lens-reflector assembly was simulated using the ILASH software (Lima et al. 2008). Figure 41b shows that the aperture efficiency is practically constant versus frequency as required, and, in agreement with that property, the reflector directivity in dB increases almost linearly with logarithm of the frequency.

Multi-beam Lens

Two examples of compact integrated lens antennas are presented here for multi-beam applications operating over a 40 % bandwidth from 40 to 60 GHz. The first example is intended to demonstrate the

DOI 10.1007/978-981-4560-75-7_40-1

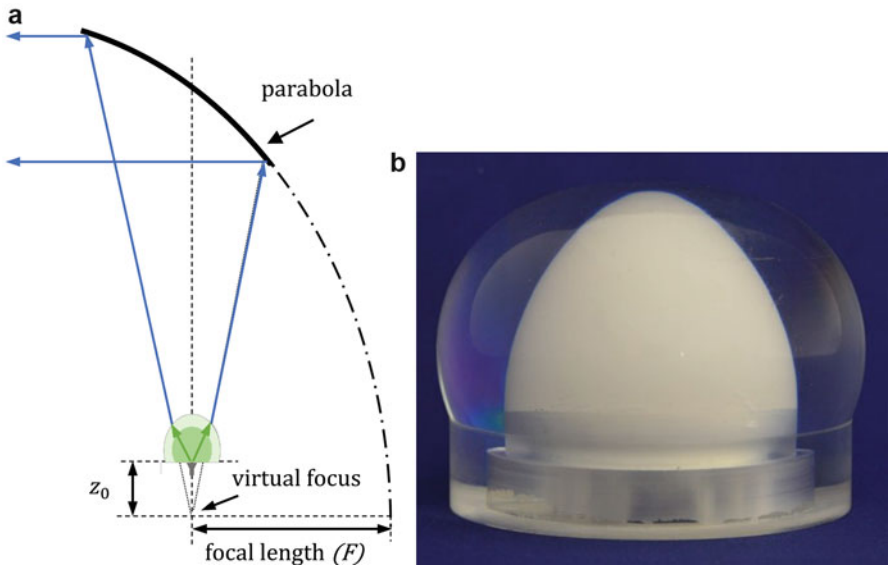


Fig. 40 (**a**) Problem geometry; (**b**) fabricated lens prototype with a MACOR^{TM} inner shell and an acrylic outer shell

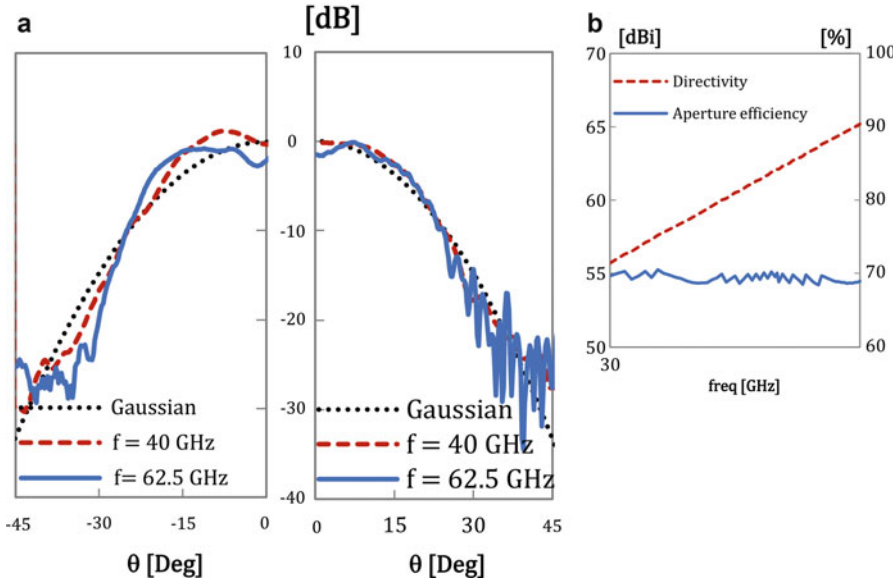


Fig. 41 (**a**) Measured radiation patterns of the lens at 40 and 62.5 GHz; (**b**) calculated directivity and aperture efficiency of the reflector with frequency plotted on log scale

potential of a double-shell lens to produce 11 beams within ±20° scan intervals with 6° beam width at 60 GHz, gain scan loss lower than 1 dB, and better than 95 % Gaussissity throughout the 40 % bandwidth. In this example, the lens is fed by a waveguide. The main purpose of the second example is to demonstrate the viability of the broadband XETS antenna introduced in section "Lens Feeds" as an effective uniplanar linear polarization lens feed for multi-beam applications. Its configuration allows very tight antenna packing at the lens base for proper adjacent beam overlap. The tight packing does not preclude the integration of each XETS element with its Schottky diode for operation as a mixer. A canonical integrated lens is used for this feed demonstration.

DOI 10.1007/978-981-4560-75-7_40-1

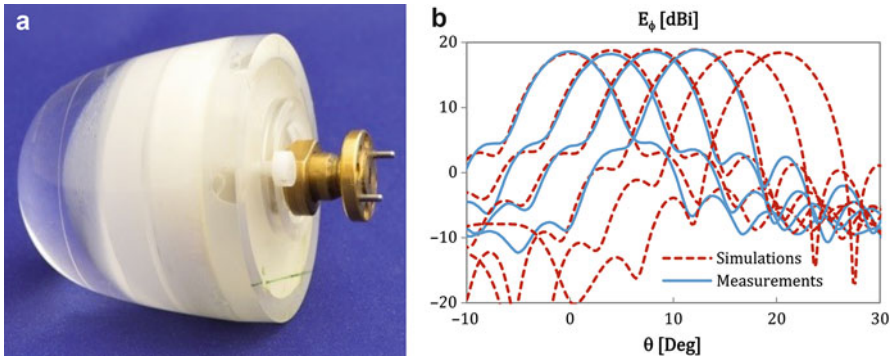


Fig. 42 (**a**) Fabricated MACOR™/acrylic lens prototype, showing the attached 60-GHz band waveguide feed; (**b**) simulated and measured gain radiation patterns at 62.5 GHz, in the H-plane, with feed at 0, 1.1, 2.2, 3.3, 4.4, and 5.5 mm from the lens axis

The scarce integrated lens configurations reported in the literature for scanning or multi-beam applications are based only on extended hemispherical or elliptical lens configurations (Filipovic et al. 1997; Wu et al. 2001). In these canonical structures, the lens shape is fixed, and so the available degree of freedom for design is just the feed distance to the lens radiating surface which does not allow enforcing any condition to reduce the aperture phase errors for off-axis feed positions. A double-shell configuration was adopted in this example to allow imposing the Abbe sine condition. The selected lens materials were the MACOR^{TM}/acrylic combination (permittivity 5.5/2.53), and the selected feed was the waveguide device described in section "Lens Feeds" for 40-GHz and 62.5-GHz operation; see Fig. 32.

Figure 42a shows the fabricated lens antenna prototype (Costa et al. 2008a) which was designed using the GO formulation from section "Multi-beam Lens" implemented in the ILASH software (Lima et al. 2008). The lens dimensions are 60 mm in diameter at the base and 37 mm in height. Figure 42b shows the measured and simulated radiation patterns for different feed positions at 1.1-mm steps along the lens base in the H-plane. This produces overlapping of consecutive beams near −1.5-dB level with respect to maximum. Beam overlap at −3 dB occurs for wider separation between consecutive feeds, compatible with the used feed aperture size (not shown). The figure shows excellent agreement between ILASH predictions and measurements and confirms the effectiveness of the proposed double-shell lens concept regarding beam scan linearity, internal reflections, beam shape, and Gaussissity preservation with extremely low scan loss up to 20° elevation. The achieved gain was of the order of 18–19 dBi and the scan loss better than 1 dB for 43–62 GHz. As a consequence of the GO-based lens design, the scan angle dependence on the feed off-axis position was practically independent of frequency, and the beam Gaussissity was better than 95 %.

The following example is intended to demonstrate the viability of the wideband XETS printed antenna with simple integral mixer described in section "Lens Feeds" as an integrated multi-beam lens feed and to show the effectiveness of the IF signal retrieval setup. Despite the tightly packed assembly of the XETS elements at the lens base, the proposed feed still offers good isolation between adjacent elements both at RF and IF, providing very stable radiation pattern and linear polarization over the bandwidth.

The above objectives can be demonstrated using a canonical single-material lens, the conclusions remaining valid for more sophisticated lens designs. A 68-mm diameter MACOR elliptical lens was used in this case (Costa and Fernandes 2007b) (Fig. 43a). The lens was designed to scan the angular interval from 0 to ±18° with beam overlap approximately at −3 dB for 43 GHz and at −4.5 dB for 62.5 GHz. The overlap level at the higher frequencies could have been improved with other lens designs, but this type of optimization exceeded the established objective for this example.

DOI 10.1007/978-981-4560-75-7_40-1

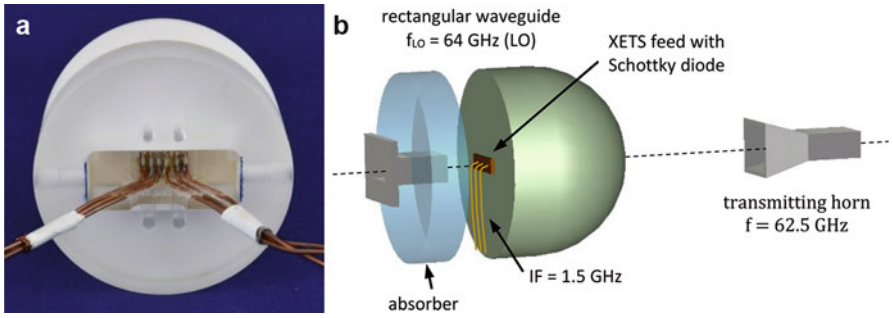


Fig. 43 (**a**) *Bottom* view of a MACOR elliptical lens fed by a multi-XETS assembly with an integrated mixer; (**b**) radiation pattern measurement setup at 62.5 GHz, using an integrated mixer at the base of the lens

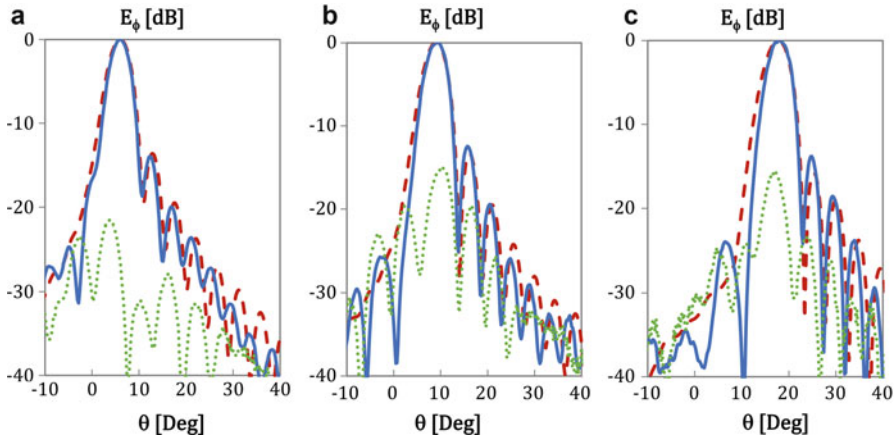


Fig. 44 Comparison between simulated and measurements radiation pattern at 62.5 GHz in the H-plane (*red dashed line*, simulations Co-pol; *blue line*, measurements Co-pol; and *green dotted line*, measurements X-pol) for a feed displacement of: (**a**) x = 2 mm; (**b**) x = 3 mm; (**c**) x = 6 mm

The radiation pattern measurement configuration is shown in Fig. 43b. The Schottky diode integrated into the XETS receives both the 62.5-GHz far-field Tx signal (through the lens) and the 64-GHz LO signal (through the air from the back), producing the desired 1.5-GHz IF signal. The LO assembly is fixed with respect to the lens, so that the LO signal amplitude illuminating the XETS does not change with lens rotation, whereas the IF signal amplitude changes only in response to lens radiation pattern at the TX frequency.

Examples of the H-plane measured and simulated radiation patterns are depicted in Fig. 44 at 62.5 GHz for three XETS elements (#2, #3, and #6). The curves are normalized to the respective maximum. The main lobe shape is quite stable, with the correct beam inclination despite the very close position of the XETS elements. Cross-polar level is below −15 dB, which is much better than what can be obtained, for instance, with the self-complementary log-periodic feed. This validates the expected feed performance.

Mechanical Beam-Steering Lenses

Mechanical beam-steering antennas tend to be superseded by electronic beam-steering antennas because the latter usually offer more compact and flexible solutions. But at millimeter waves, this trend is not so clear as losses and cost of phased arrays increase considerably. The main drawbacks of the mechanical beam-steering solutions are the mechanical complexity and the expensive millimeter-wave rotary joint.

DOI 10.1007/978-981-4560-75-7_40-1

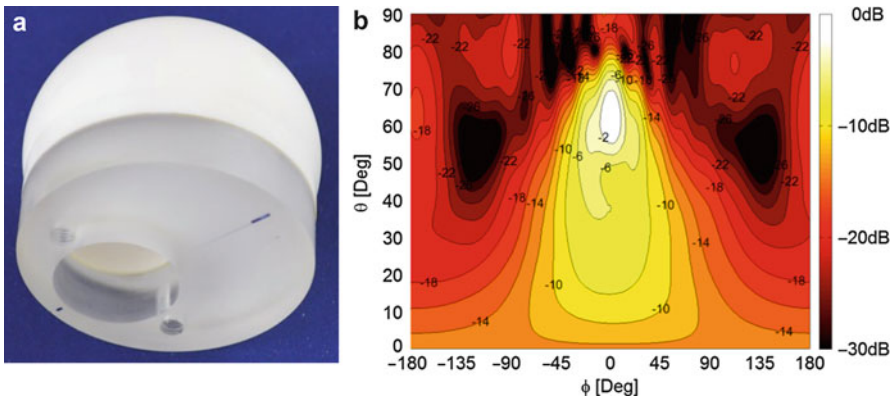


Fig. 45 (**a**) MACOR^{TM} axial-symmetric 50-mm diameter shaped lens; (**b**) radiation performance of the lens computed with the GO/PO method

A new approach for mechanical beam steering has been proposed in Costa et al. (2008b, 2009) where the feed remains stationary, while the beam tilt is obtained just by rotating or tilting an appropriately shaped dielectric lens that is held by a very simple mechanical structure in the near field of the fixed feed aperture.

One example of this configuration was intended for low Earth orbit satellite telemetry link at 26 GHz. The lens was required to produce a sector beam with shaping in elevation and to enable a simple mechanical azimuth scanning. The target gain versus elevation angle θ is given approximately by Sec(kθ). With an appropriate choice of the k shaping parameter, this elevation radiation pattern compensates path loss attenuation taking into account the Earth's curvature.

The designed shaped lens is axial-symmetric, but a spherical air cavity is excavated through its base at a calculated offset distance with respect to the lens symmetry axis (Fig. 45a). The feed aperture is at the center of this offset spherical air cavity. The asymmetry introduced by the feed position produces an adequate directive beam in the azimuth plane with Sec(kθ) shaping in elevation. This is an extremely simple solution where the feed is fixed and the lens rotates about the eccentric feed axis to provide the required beam scanning. This contactless feeding avoids the need for fault-prone rotary joints in long continuous usage. The assembly is very compact and lightweight (lens diameter is less than 5 wavelengths and height is less than 3 wavelengths) complying with mass and volume restrictions for satellites.

The same concept of a moving lens with fixed feed was extended for azimuth plus elevation mechanical scanning in a different application context (Costa et al. 2009). Wireless indoor transmission of uncompressed HD video signal between a video device like a camcorder and a HD TV display may soon become a requirement for mass-market home applications. Emerging standards are proposing to use the unlicensed spectrum from 57 to 66 GHz for such applications. For mass consumer applications, the radio link must use inexpensive low-power sources, and hence high-gain antennas (>20 dBi) are required to favor the link budget. The narrow beam needs to be steered to allow for the user mobility. Using the design formulation described in section "Beam-Steering Lens," a new configuration of a mechanical steerable beam antenna has been developed where a dielectric lens pivots in front of a single stationary moderate gain feed (Fig. 46a). Again, the feed is not in physical contact with the lens. As previously discussed, this eliminates the need for rotary joints and the associated drawbacks. The designed lens is made of polyethylene and is such that both the input and output lens surfaces are shaped to accommodate a beam collimation condition and a scan angle maximization condition. The lens output beam is aligned with the lens axis for all lens orientations within the operation cone. Thus, pivoting the lens around two of its main axis allows steering the beam in elevation and azimuth. The important difference with respect to conventional scanning lenses is that for all lens inclinations the focus is always exactly coincident with the

DOI 10.1007/978-981-4560-75-7_40-1

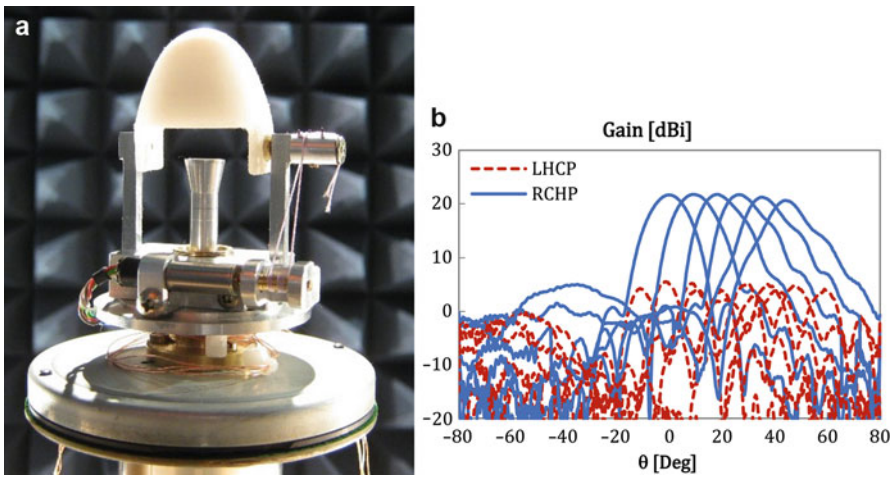


Fig. 46 (**a**) Photograph of the manufactured polyethylene lens plus horn feed; (**b**) measured radiation patterns of the lens antenna for several lens tilt angles

aperture phase center. In this way, aberrations are strongly reduced allowing wider scan angles than conventional solutions. The fabricated lens assembly demonstrated a −45° to +45° elevation scan capability over full azimuth, with 21-dB gain and better than +1.1-dB gain scan loss (Fig. 46b). The radiation efficiency is always above 95 %.

Conclusion

Rapid prototyping techniques and the trend to move into millimeter and submillimeter waves are bringing back the interest on dielectric lens antennas, namely, on integrated lens antenna configurations, taking advantage of the fast parallel advancements on low-cost fabrication of millimeter-wave integrated circuits. This integration prompts for mass-market applications at these bands.

It was shown in this chapter that dielectric lenses are very flexible to accommodate demanding specifications and lens design is reasonably simple. Although a fast progress is being made on low-profile planar antennas employing phase-shifting cells and metamaterials, dielectric lenses will remain a strong competitor at very small wavelengths owing to its effectiveness and simplicity for design and fabrication.

Cross-References

- ▶ Fresnel Zone Plate Lens Antenna
- ▶ Mm-Wave Sub-mm-Wave Antenna Measurement
- ▶ Reflectarray Antennas
- ▶ Reflector Antennas
- ▶ Theory of Transformation Optics in Antenna Design

DOI 10.1007/978-981-4560-75-7_40-1

References

Afsar MN, Li X, Chi H (1990) An automated 60 GHz open resonator system for precision dielectric measurements. IEEE Trans Microwave Theory Tech 38:1845–1853

Artemenko A, Mozharovskiy A, Maltsev A et al (2013a) Experimental characterization of E-band two-dimensional electronically beam-steerable integrated lens antennas. IEEE Antennas Wirel Propag Lett 12:1188–1191

Artemenko A, Maltsev A, Mozharovskiy A et al (2013b) Millimeter-wave electronically steerable integrated lens antennas for WLAN/WPAN applications. IEEE Trans Antennas Propag 61:1665–1671

Bares B, Sauleau R (2007) Design and optimisation of axisymmetric millimetre-wave shaped lens antennas with directive, secant-squared and conical beams. IET Microwaves Antennas Propag 1:433–439

Bisognin A, Titz D, Ferrero F et al (2014) 3D printed plastic 60 GHz lens: enabling innovative millimeter wave antenna solution and system. In: Microwave symposium (IMS), IEEE MTT-S international, Tampa Bay, United States, pp 1–4

Bor J, Lafond O, Merlet H et al (2014) Technological process to control the foam dielectric constant application to microwave components and antennas. IEEE Trans Compon Packag Manuf Technol 4:938–942

Boriskin AV, Vorobyov A, Sauleau R (2011) Two-shell radially symmetric dielectric lenses as low-cost analogs of the Luneburg lens. IEEE Trans Antennas Propag 59:3089–3093

Born M, Wolf E (1959) Principles of optics. Pergamon, New York

Chen LF, Ong CK, Neo CP, Varadan VV, Varadan VK (2004) Microwave Theory and Techniques for Materials Characterization, in Microwave Electronics: Measurement and Materials Characterization, Wiley, Chichester, UK. doi: 10.1002/0470020466.ch2

Cornbleet S (1994) Microwave and geometrical optics. Academic, London

Costa JR, Fernandes CA (2007a) Broadband slot feed for integrated lens antennas. IEEE Antennas Wirel Propag Lett 6:396–400

Costa JR, Fernandes CA (2007b) Integrated imaging lens antenna with broadband feeds. In: Antennas and propagation, EuCAP 2007. The second European conference, Edinburgh, UK, pp 1–6

Costa JR, Silveirinha MG, Fernandes CA (2008a) Evaluation of a double-shell integrated scanning lens antenna. IEEE Antennas Wirel Propag Lett 7:781–784

Costa JR, Fernandes CA, Godi G et al (2008b) Compact Ka-band lens antennas for LEO satellites. IEEE Trans Antennas Propag 56:1251–1258

Costa JR, Lima EB, Fernandes CA (2009) Compact beam-steerable lens antenna for 60-GHz wireless communications. IEEE Trans Antennas Propag 57:2926–2933

Costa JR, Lima EB, Fernandes CA (2010) Antenna phase center determination from amplitude measurements using a focusing lens. In: Antennas and propagation society international symposium, IEEE

Do-Hoon K, Werner DH (2010) Transformation electromagnetics: an overview of the theory and applications. IEEE Antennas Propag Mag 52:24–46

Edwards JM, O'brient R, Lee AT et al (2012) Dual-polarized sinuous antennas on extended hemispherical silicon lenses. IEEE Trans Antennas Propag 60:4082–4091

Fernandes L (1995) Developing a system concept and technologies for mobile broadband communications. IEEE Pers Commun Mag 2:54

Fernandes CA (1999) Shaped dielectric lenses for wireless millimeter-wave communications. IEEE Antennas Propag Mag 41:141–150

Fernandes CA (2002) Shaped-beam antennas. In: Godara L (ed) Handbook of antennas in wireless communications. CRC Press, New York, ch 15

Fernandes CA, Anunciada LM (2001) Constant flux illumination of square cells for millimeter-wave wireless communications. IEEE Trans Microwave Theory Tech 49:2137–2141

Fernandes CA, Costa JR (2009) Permittivity measurement and anisotropy evaluation of dielectric materials at millimeter-waves. In: XIX Imeko world congress: fundamental and applied metrology, proceedings. IMEKO, Budapest, pp 673–677

Fernandes CA, Lima EB, Costa JR (2010) Broadband integrated lens for illuminating reflector antenna with constant aperture efficiency. IEEE Trans Antennas Propag 58:3805–3813

Fernandes CA, Lima EB, Costa JR (2011) Tapered waveguide feed for integrated dielectric lens antenna performance tests. In: EUROCON – international conference on computer as a tool (EUROCON), IEEE, Lisbon, Portugal, pp 1–4

Filipovic DF, Gearhart SS, Rebeiz GM (1993) Double-slot antennas on extended hemispherical and elliptical silicon dielectric lenses. IEEE Trans Microwave Theory Tech 41:1738–1749

Filipovic DF, Gauthier GP, Raman S et al (1997) Off-axis properties of silicon and quartz dielectric lens antennas. IEEE Trans Antennas Propag 45:760–766

Fuchs B, Lafond O, Rondineau S et al (2006) Design and characterization of half Maxwell fish-eye lens antennas in millimeter waves. IEEE Trans Microwave Theory Tech 54:2292–2300

Fuchs B, Le Coq L, Lafond O et al (2007a) Design optimization of multishell Luneburg lenses. IEEE Trans Antennas Propag 55:283–289

Fuchs B, Lafond O, Rondineau S et al (2007b) Off-axis performances of half Maxwell fish-eye lens antennas at 77 GHz. IEEE Trans Antennas Propag 55:479–482

Fuchs B, Lafond O, Palud S et al (2008a) Comparative design and analysis of Luneburg and half Maxwell fish-eye lens antennas. IEEE Trans Antennas Propag 56:3058–3062

Fuchs B, Palud S, Le Coq L et al (2008b) Scattering of spherically and hemispherically stratified lenses fed by any real source. IEEE Trans Antennas Propag 56:450–460

Hailu DM, Ehtezazi IA, Safavi-Naeini S (2009) Fast analysis of terahertz integrated lens antennas employing the spectral domain ray tracing method. IEEE Antennas Wirel Propag Lett 8:37–39

Hailu DM, Ehtezazi IA, Neshat M et al (2011) Hybrid spectral-domain ray tracing method for fast analysis of millimeter-wave and terahertz-integrated antennas. IEEE Trans Terahertz Sci Technol 1:425–434

Hirvonen TM, Vainikainen P, Lozowski A et al (1996) Measurement of dielectrics at 100 GHz with an open resonator connected to a network analyzer. IEEE Trans Instrum Meas 45:780–786

Ka Fai C, Rui L, Cheng J et al (2014) 77-GHz automotive radar sensor system with antenna integrated package. IEEE Trans Compon Packag Manuf Technol 4:352–359

Kay K (1965) Electromagnetic theory and geometrical optics. Interscience, New York

Kelleher K (1961) Scanning antennas, chapter 15. In: Jasik H (ed) Antenna engineering handbook. McGraw-Hill, New York

Kim KW, Rahmat-Samii Y (1998) Spherical Luneburg lens antennas: engineering characterizations including air gap effects. In: Antennas and propagation society international symposium, vol 2064. IEEE, Atlanta, GA, USA, pp 2062–2065

Kolundzija B, Djordjevic A (2002) Electromagnetic modelling of composite metallic and dielectric structures. Artech House, Norwood

Komiyama B, Kiyokawa M, Matsui T (1991) Open resonator for precision dielectric measurements in the 100 GHz band. IEEE Trans Microwave Theory Tech 39:1792–1796

Komljenovic T, Sauleau R, Sipus Z et al (2010) Layered circular-cylindrical dielectric lens antennas – synthesis and height reduction technique. IEEE Trans Antennas Propag 58:1783–1788

Lima E, Costa JR, Silveirinha MG et al (2008) ILASH – software tool for the design of integrated lens antennas. In: Antennas and propagation society international symposium, AP-S 2008. IEEE, San Diego, USA, pp 1–4

Ling H, Chou R, Lee S (1989) Shooting and bouncing rays: calculating the RCS of an arbitrarily shaped cavity. IEEE Trans Antennas Propag 37:194–205

Llombart N, Lee C, Alonso-Delpino M et al (2013) Silicon micromachined lens antenna for THz integrated heterodyne arrays. IEEE Trans Terahertz Sci Technol 3:515–523

Lodge OJ, Howard JL (1888) On electric radiation and its concentration by lenses. Proc Phys Soc Lond 10:143

Luneburg RK (1943) US Patent 2,328,157

Maciel JJ, Felsen LB (1989) Systematic study of fields due to extended apertures by Gaussian beam discretization. IEEE Trans Antennas Propag 37:884–892

Mateo-Segura C, Dyke A, Dyke H et al (2014) Flat Luneburg lens via transformation optics for directive antenna applications. IEEE Trans Antennas Propag 62:1945–1953

Maxwell JC (1860) Scientific papers, I. Dover, New York

Min L, Wei-Ren N, Kihun C et al (2014) A 3-D Luneburg lens antenna fabricated by polymer jetting rapid prototyping. IEEE Trans Antennas Propag 62:1799–1807

Mosallaei H, Rahmat-Samii Y (2001) Nonuniform Luneburg and two-shell lens antennas: radiation characteristics and design optimization. IEEE Trans Antennas Propag 49:60–69

Naruse M, Sekimoto Y, Noguchi T et al (2013) Optical efficiencies of lens-antenna coupled kinetic inductance detectors at 220 GHz. IEEE Trans Terahertz Sci Technol 3:180–186

Neto A (2010) UWB, non dispersive radiation from the planarly fed leaky lens antenna – part 1: theory and design. IEEE Trans Antennas Propag 58:2238–2247

Neto A, Maci S, De Maagt PJI (1998) Reflections inside an elliptical dielectric lens antenna. IEE Proc Microwaves Antennas Propag 145:243–247

Neto A, Borselli L, Maci S et al (1999) Input impedance of integrated elliptical lens antennas. IEE Proc Microwaves Antennas Propag 146:181–186

Neto A, Monni S, Nennie F (2010) UWB, non dispersive radiation from the planarly fed leaky lens antenna – part II: demonstrators and measurements. IEEE Trans Antennas Propag 58:2248–2258

Ngoc Tinh N, Sauleau R, Perez CJM (2009) Very broadband extended hemispherical lenses: role of matching layers for bandwidth enlargement. IEEE Trans Antennas Propag 57:1907–1913

Ngoc Tinh N, Delhote N, Ettorre M et al (2010) Design and characterization of 60-GHz integrated lens antennas fabricated through ceramic stereolithography. IEEE Trans Antennas Propag 58:2757–2762

Ngoc Tinh N, Sauleau R, Le Coq L (2011) Reduced-size double-shell lens antenna with flat-top radiation pattern for indoor communications at millimeter waves. IEEE Trans Antennas Propag 59:2424–2429

Ngoc Tinh N, Boriskin AV, Rolland A et al (2013) Shaped lens-like dome for UWB antennas with a gaussian-like radiation pattern. IEEE Trans Antennas Propag 61:1658–1664

Nguyen NT, Sauleau R, Martinez Perez CJ et al (2010) Finite-difference time-domain simulations of the effects of air gaps in double-shell extended hemispherical lenses. IET Microwaves Antennas Propag 4:35–42

Nguyen NT, Sauleau R, Ettorre M et al (2011) Focal array fed dielectric lenses: an attractive solution for beam reconfiguration at millimeter waves. IEEE Trans Antennas Propag 59:2152–2159

Nikolic N, James GL, Hellicar A et al (2012) Quarter-sphere Luneburg lens scanning antenna. In: 15th international symposium on antenna technology and applied electromagnetics (ANTEM), pp 1–4

Olver A, Clarricoats P, Kishk A, Shafai L (1994) Microwave horns and feeds. IEEE Press, New York, Chap. 11

Pasqualini D, Maci S (2004) High-frequency analysis of integrated dielectric lens antennas. IEEE Trans Antennas Propag 52:840–847

Pavacic AP, Del Rio DL, Mosig JR et al (2006) Three-dimensional ray-tracing to model internal reflections in off-axis lens antennas. IEEE Trans Antennas Propag 54:604–612

Peterson AF, Ray SL, Mittra R (1998) Computational methods of electromagnetics. IEEE Press, New York

Petosa A, Ittipiboon A (2000) Shadow blockage effects on the aperture efficiency of dielectric Fresnel lenses. IEE Proc Microwaves Antennas Propag 147:451–454

Piksa P, Zvanovec S, Cerny P (2011) Elliptic and hyperbolic dielectric lens antennas in mm-waves. Radioengineering 20:271

Rebeiz GM (1992) Millimeter-wave and terahertz integrated circuit antennas. Proc IEEE 80:1748–1770

Rolland A, Sauleau R, Le Coq L (2011) Flat-shaped dielectric lens antenna for 60-GHz applications. IEEE Trans Antennas Propag 59:4041–4048

Rutledge D, Neikirk D, Kasilingam D (1983) Integrated circuit antennas. In: Button K (ed) Infrared and millimeter-waves, vol 10. Academic, New York, pp 1–90

Salema C, Fernandes C, Jha R (1998) Solid dielectric horns. Artech House, Boston, Chap. 7

Sanford JR (1994) Scattering by spherically stratified microwave lens antennas. IEEE Trans Antennas Propag 42:690–698

Sato K, Ujiie H (2002) A plate Luneburg lens with the permittivity distribution controlled by hole density. Electron Commun Jpn (Part I: Communications) 85:1–12

Sauleau R, Bares B (2006) A complete procedure for the design and optimization of arbitrarily shaped integrated lens antennas. IEEE Trans Antennas Propag 54:1122–1133

Semenov AD, Richter H, Hubers HW et al (2007) Terahertz performance of integrated lens antennas with a hot-electron bolometer. IEEE Trans Microwave Theory Tech 55:239–247

Silveirinha MGMV, Fernandes CA (2000) Shaped double-shell dielectric lenses for wireless millimeter wave communications. In: Antennas and propagation society international symposium, vol 1673. IEEE, Salt Lake City, UT, USA, pp 1674–1677

Silveirinha MG, Fernandes CA, Costa JR (2014) A graphical aid for the complex permittivity measurement at microwave and millimeter wavelengths. IEEE Microwave Wireless Compon Lett 24:421–423

Silver S (1984) Microwave antenna theory and design. Peter Pereginus, London

Trichopoulos GC, Mumcu G, Sertel K et al (2010) A novel approach for improving off-axis pixel performance of terahertz focal plane arrays. IEEE Trans Microwave Theory Tech 58:2014–2021

Trichopoulos GC, Mosbacker HL, Burdette D et al (2013) A broadband focal plane array camera for real-time THz imaging applications. IEEE Trans Antennas Propag 61:1733–1740

Van Der Vorst MJM, De Maagt PJL (2002) Efficient body of revolution finite-difference time-domain modeling of integrated lens antennas. IEEE Microwave Wireless Compon Lett 12:258–260

Van Der Vorst MJM, De Maagt PJL, Herben MHAJ (1999) Effect of internal reflections on the radiation properties and input admittance of integrated lens antennas. IEEE Trans Microwave Theory Tech 47:1696–1704

DOI 10.1007/978-981-4560-75-7_40-1

Van Der Vorst MJM, De Maagt PJI, Neto A et al (2001) Effect of internal reflections on the radiation properties and input impedance of integrated lens antennas-comparison between theory and measurements. IEEE Trans Microwave Theory Tech 49:1118–1125

Wu X, Eleftheriades G, Van Deventer-Perkins T (2001) Design and characterization of single- and multiple-beam MM-wave circularly polarized substrate lens antennas for wireless communications. IEEE Trans Microwave Theory Tech 49:431–441

Xue L, Fusco VF (2007) 24 GHz automotive radar planar Luneburg lens. IET Microwaves Antennas Propag 1:624–628

Yurduseven O, Cavallo D, Neto A (2014) Wideband dielectric lens antenna with stable radiation patterns fed by coherent array of connected leaky slots. IEEE Trans Antennas Propag 62:1895–1902

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_41-1

Self-Complimentary and Broadband Antennas

Kunio Sawaya*
Innovation Center for Creation of a Resilient Society, Tohoku University, Sendai, Miyagi Prefecture, Japan

Abstract

Frequency independent and broadband antennas are reviewed form the view points of the self-complementarity and self-similarity. The theory of self-complementary antenna, original and modified self-complementary antennas, and so-called log-periodic antennas are discussed. Then the frequency independent antennas based on the self-similarity and other broadband antennas are presented.

Keywords

Self-complementary antenna; Constant impedance; Transposed excitation; Log-periodic antenna; Self-similar antenna; Biconical antenna; Spiral antenna; Discone antenna; Bow-tie antenna; Tapered slot antenna

Introduction

Frequency independent antennas can be divided into two types of antennas: one is the self-complementary type and the other is the self-similar type antennas. Self-complementary antennas are based on the Mushiake relationship and have the constant input impedance with respect to frequency (Mushiake 1949). Since there is an infinite variety of self-complementary structures, many self-complementary and modified self-complementary antennas have been proposed. So called log-periodic antennas are considered to be derivatives of the self-complementary antennas. Some broadband antennas are based on both the self-complementarity and self-similarity such as the two-arm spiral antennas.

Self-similar object is defined by the fact that the structure is similar to a part of itself. Antennas having self-similar structures are called "self-similar antennas." If the linear dimensions of the self-similar antenna are reduced by a given factor, the electromagnetic properties remain the same, as long as the wavelength is reduced by the same ratio. Self-similar structures are divided into two types: one is the continuous structure composed of continuous shape, and the other is the log-periodic type or fractal structure composed of log-periodic elements. When the end effects of the self-similar antenna having continuous structure can be negligible, the electromagnetic properties are the same at all frequency, and the continuous self-similar antennas have the frequency-independent properties. On the other hand, the electromagnetic characteristics of the log-periodic or fractal self-similar antennas remain in the same, merely at the discrete points of the wavelength that have the same ratio with the self-similar ratio of the structure, and they cannot be the frequency-independent antennas.

In this chapter, the theory of self-complementary antenna, original and modified self-complementary antennas, and so-called log-periodic antennas are reviewed. Then frequency independent antennas based on the self-similarity and other broadband antennas are presented.

*Email: sawaya@ecei.tohoku.ac.jp

DOI 10.1007/978-981-4560-75-7_41-1

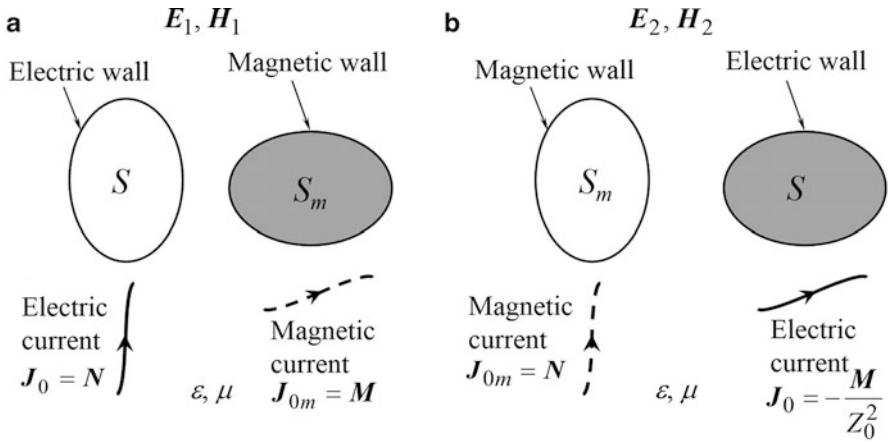


Fig. 1 Duality property of the electromagnetic field. (**a**) Structure #1. (**b**) Structure #2

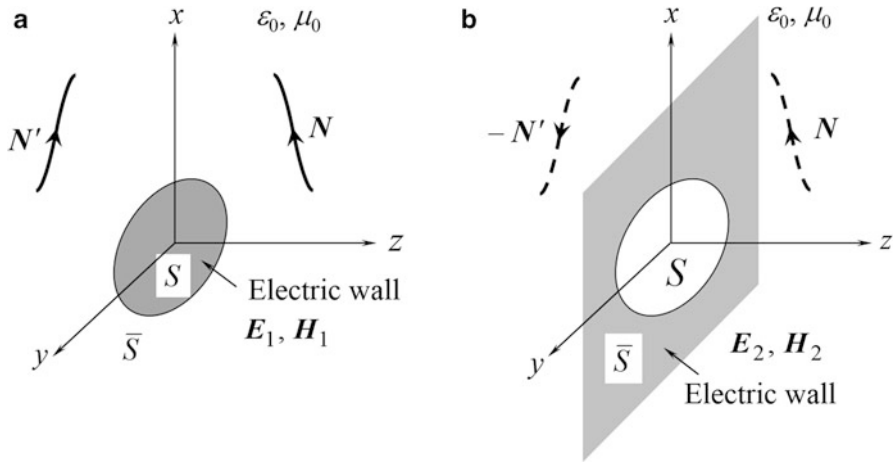


Fig. 2 Babinet's principle to electromagnetic fields. (**a**) Symmetric electric current sources. (**b**) Anti-symmetric magnetic current sources

Self-Complementary Antenna

Theory of Self-Complementary Antenna

There is a duality property between electromagnetic fields ($\boldsymbol{E}_1$, $\boldsymbol{H}_1$) in structure #1 and ($\boldsymbol{E}_2$, $\boldsymbol{H}_2$) in structure #2 shown in Fig. 1, which is expressed by

$$\boldsymbol{E}_1 = Z_0^2 \boldsymbol{H}_2, \boldsymbol{H}_1 = -\boldsymbol{E}_2, \tag{1}$$

if all electric walls S, magnetic walls S_m, electric currents $\boldsymbol{J}_0 = \boldsymbol{N}$, and magnetic currents $\boldsymbol{J}_{0m} = \boldsymbol{M}$ in structure #1 are respectively interchanged for magnetic walls S_m, electric walls S, magnetic currents $\boldsymbol{J}_{0m} = \boldsymbol{N}$, and electric currents $\boldsymbol{J}_0 = -\boldsymbol{N}/Z_0^2$ in structure #2, where $Z_0 = \sqrt{\mu/\varepsilon}$ is the intrinsic impedance of the medium.

Let us consider a structure #1 composed symmetric sources of electric currents with a conducting plane S placed in xy plane shown in Fig. 2a and another structure #2 composed anti-symmetric sources of magnetic currents with a conducting plane $\overline{S}$ in xy plane shown in Fig. 2b. The surface $\overline{S}$ in Fig. 2a and the

DOI 10.1007/978-981-4560-75-7_41-1

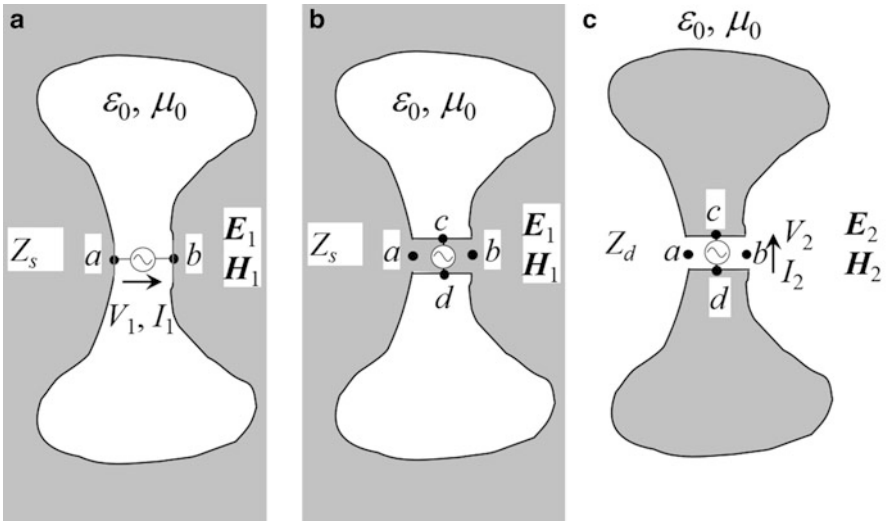


Fig. 3 Slot antenna and planar antenna of arbitrary shape (Mushiake 1949; Uda and Mushiake 1949). (**a**) Slot antenna fed by electric current. (**b**) Slot antenna fed by magnetic current. (**c**) Planar antenna fed by electric current

surface S in Fig. 2b are considered to be magnetic walls because of the symmetric electric currents and anti-symmetric magnetic currents, respectively. By applying the duality property, the relation between the electromagnetic fields in Fig. 2a, b is expressed by

$$\boldsymbol{E}_1 = \pm Z_0^2 \boldsymbol{H}_2,\ \boldsymbol{H}_1 = \mp \boldsymbol{E}_2,\ z \gtrless 0. \tag{2}$$

This relation is called "Babinet's principle" to electromagnetic fields, which had been derived in Japan during mid-1940s.

Mushiake applied Babinet's principle in electromagnetic fields to the analysis of a slot antenna having arbitrary shape shown in Fig. 3a. Let the slot antenna having voltage V_1 and current I_1 at the driving point generate electromagnetic field $\boldsymbol{E}_1$ and $\boldsymbol{H}_1$ as shown in Fig. 3a, which is equivalent to a slot antenna fed by magnetic current shown in Fig. 3b. The complementary planar antenna with driving voltage V_2 and current I_2 generate electromagnetic field $\boldsymbol{E}_2$ and $\boldsymbol{H}_2$ as shown in Fig. 3c. The driving point voltages V_1 and V_2 are expressed in terms of the driving currents I_1 and I_2 as

$$\begin{aligned} V_1 &= -\int_a^b \boldsymbol{E}_1 \cdot d\boldsymbol{s} = \int_a^b \boldsymbol{H}_2 \cdot d\mathbf{s} = \frac{I_2}{2} \\ V_2 &= -\int_d^c \boldsymbol{E}_2 \cdot d\boldsymbol{s} = -Z_0^2 \int_d^c \boldsymbol{H}_1 \cdot d\boldsymbol{s} = \frac{Z_0^2 I_1}{2}, \end{aligned} \tag{3}$$

where

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} \cong 120\pi \ \ [\Omega], \tag{4}$$

is the intrinsic impedance in vacuum space.

By using Eq. 3, the input impedance of the slot antenna Z_s shown in Fig. 3a can be obtained by the following equation (Mushiake 1949; Uda and Mushiake 1949)

DOI 10.1007/978-981-4560-75-7_41-1

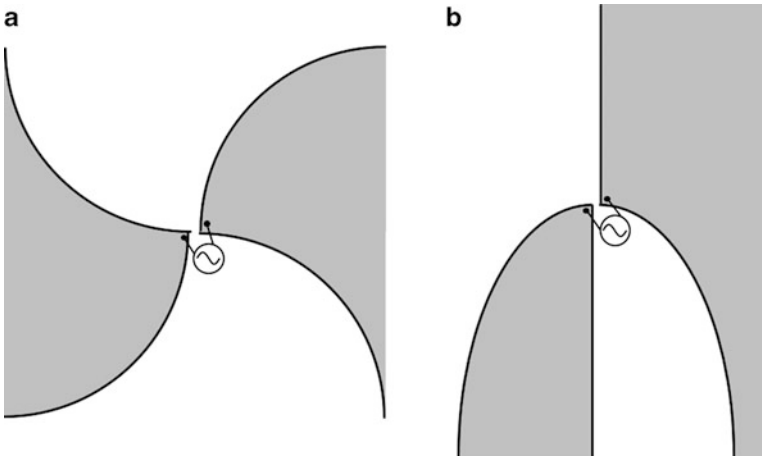


Fig. 4 Self-complementary antennas. (**a**) Balanced and rotationally symmetric type. (**b**) Unbalanced and axially symmetric type

$$Z_s = \frac{(Z_0/2)^2}{Z_d}, \tag{5}$$

where Z_d is the input impedance of complementary planar antenna shown in Fig. 3c.

Equation 5 had been already obtained in Japan to obtain the input impedance of slit antennas earlier than the paper of Booker (1946) as pointed out by Mushiake (2004), and the paper by Asami et al. (1947) was a more detailed report of the unpublished earlier paper. (The slot antenna was called the slit antennas in Japan at that time.) However, the structures reported in these papers had been restricted for the thin slot antennas and the thin wire antennas. On the other hand, Eq. 5 derived by Mushiake and Uda (Mushiake 1949; Uda and Mushiake 1949) does not have such limitation and can be applied for slot antennas and planar antennas with arbitrary shape.

Mushiake also originated the self-complementary structures and discovered that the input impedance of the antennas shown in Fig. 4 is constant independent of the frequency and given by

$$Z_{in} = \frac{Z_0}{2} \cong 60\pi \ \ [\Omega], \tag{6}$$

since arbitrarily shaped four lines for the boundary of two conducting sheets are exactly identical and the complementary structure in Fig. 4 is exactly the same to the original structure in Fig. 4 (Mushiake 1949; Uda and Mushiake 1949). Equation 6 is known as "Mushiake relationship" and this antenna is called "self-complementary antenna" (Rumsey 1957; 1966). The principle of the constant impedance of the self-complementary antennas is called the "principle of self-complementarity."

Modification of Self-Complementary Antennas

The self-complementary antennas are very interesting since there is an infinite variety of self-complementary structures as shown in Fig. 5. Although infinite structures are required for the constant impedance of the self-complementary antennas, the structures having teeth or notches and monopoles shown in Fig. 5a, b can realize the constant impedance with a finite conducting plane, because teeth or notches and monopoles radiate electromagnetic power and the truncation effect can be reduced for the finite structures. Archimedean spiral structure shown in Fig. 5c can also reduce the truncation effect because of the radiation from the spiral arms. Figure 5d is a self-complementary antenna called "alternate-

DOI 10.1007/978-981-4560-75-7_41-1

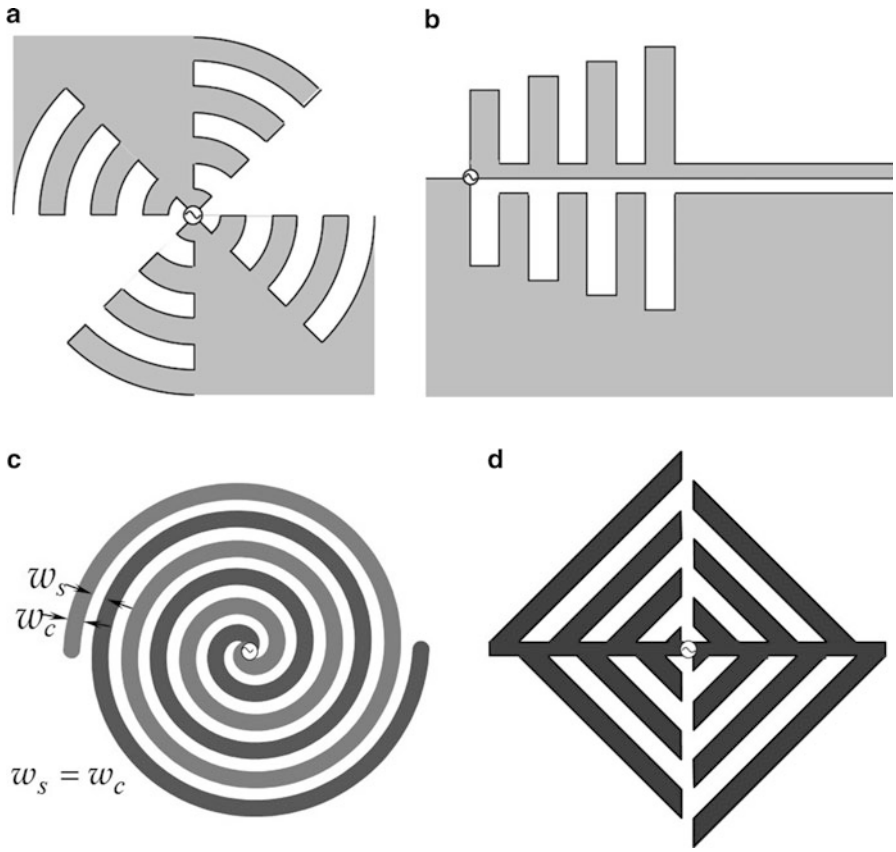


Fig. 5 Examples of self-complementary structures. (**a**) Balanced type with "teeth-type" structure. (**b**) Unbalanced type with notches and monopoles. (**c**) Archimedean-spiral structure. (**d**) Alternate-leaved structure

leaved structure" developed by Furuya et al. (1977). Since there is a lot of flexibility of the shape of the self-complementary antennas, it could be possible to obtain antennas having desired gain and/or desired radiation pattern maintaining the constant input impedance, e.g., an antenna having desired frequency dependence of gain.

Rumsey and DuHamel found the importance of the principle of self-complementarity (Rumsey 1957, 1966). DuHamel and Isbell (1957) proposed a self-complementary antenna with log-periodically spaced teeth as shown in Fig. 6 and showed that the input impedance is almost equal to 188 Ω under the condition that the operating frequency is higher than the frequency at which the length of the longest tooth is a quarter wavelength.

The most important structure for the broadband characteristics is the self-complementary structure. Nakano (2006) performed a numerical analysis of the toothed log-periodic antennas shown in Fig. 7 by using FDTD method. Figure 7a is a modified self-complementary antenna having trapezoidal log-periodic and point-symmetric structure. Figure 7b is another trapezoidal log-periodic and anti-complementary antenna having line-symmetric structure and is not a modified self-complementary antenna. Figure 8 shows the input-impedance of these antennas. In the case of modified self-complementary antenna shown in Fig. 7a, the input impedance is approximately $60\pi \approx 188\ \Omega$, although the structure is not an exact self-complementary structure. On the other hand, the input impedance of the anti-complementary antenna shown in Fig. 7b exhibits periodically oscillating behavior with respect to the logarithm of frequency and broadband characteristics are not obtained.

Deschamps (1959) proposed self-complementary multiterminal planar antennas having the input impedance depending on the structure. Mushiake and his colleagues also developed other types of self-

DOI 10.1007/978-981-4560-75-7_41-1

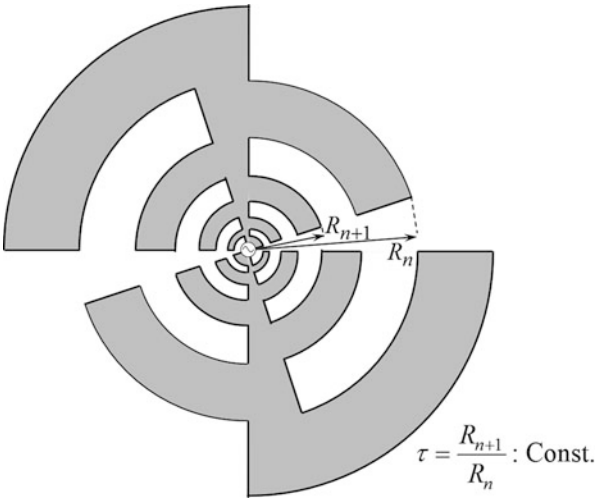


Fig. 6 Self-complementary antenna having log-periodic shape

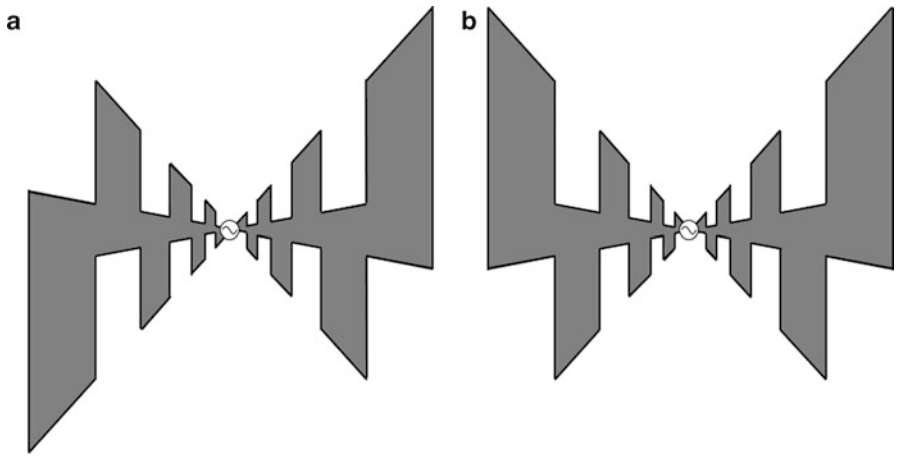


Fig. 7 Trapezoidal log-periodic antenna (Nakano 2006). (**a**) Self-complementary type. (**b**) Anti-complementary type

complementary antennas, such as rotationally symmetric four-terminal planar self-complementary antennas (Mushiake 1965), three-dimensional multi-planar self-complementary antennas, stacked self-complementary antennas, and monopole-slot antenna. Figure 5d is alternate-leaved self-complementary antenna (Furuya et al. 1977), which has not only the broadband input impedance but also frequency independent radiation pattern. Figure 9 shows the monopole-slot array antenna, whose input impedance was obtained theoretically as $Z_{in} = \sqrt{2}30\pi \cong 133\ \Omega$ by Ishizone et al. (1983). The details of these antennas are presented in the paper and book by Mushiake (1992, 1996). Also, the numerical investigation of the monopole-notch antennas shown in Fig. 5b was performed by Yamamoto et al. (1982) to obtain the relation between the structure parameters and the characteristics such as input impedance and the gain of the array.

DOI 10.1007/978-981-4560-75-7_41-1

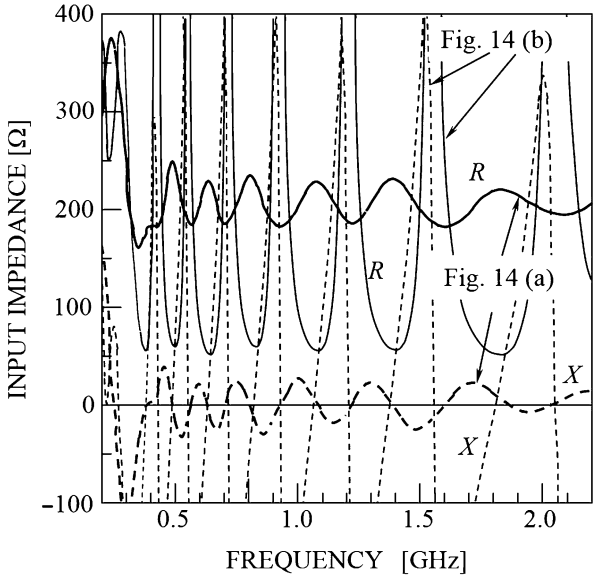


Fig. 8 Input impedance of trapezoidal log-periodic antenna (Nakano 2006)

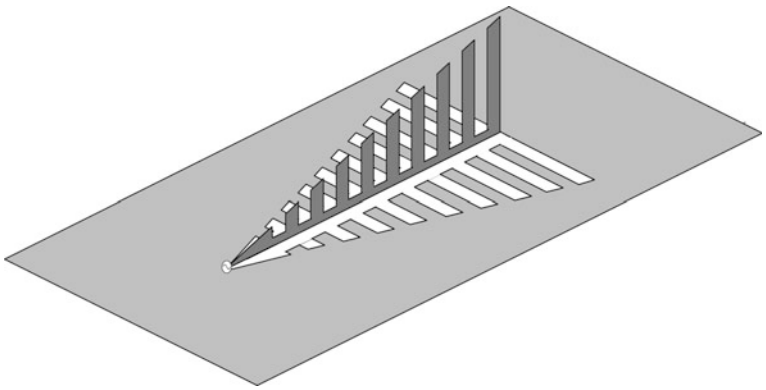


Fig. 9 Monopole-slot array antenna (Ishizone et al. 1983)

Modified Self-Complementary Antennas Having Log-Periodic Shape

The first structure of self-complementary antenna having log-periodic shape was the planar structure shown in Fig. 6. DuHamel and Ore (1958) deformed the self-complementary antenna and developed a planar antenna shown in Fig. 10, where the planar antenna is the original self-complementary antenna under the condition of $\alpha + \beta = \pi$. Since the planar structures radiate bidirectionally and unidirectional antenna was desired, they tried to deform the structure by folding the planar antenna and obtained a modified self-complementary type antenna called "log-periodic antenna" shown in Fig. 11. They also developed a log-periodic antenna composed of thin wires shown in Fig. 12, which is a wind-resistant structure.

The log-periodic antennas were further deformed to the log-periodic dipole array (LPDA) antenna by Isbell (1960) shown in Figs. 13 and 14, and design procedure of the LPDA was proposed by Carrel (1961), where theoretical investigation using sinusoidal current distribution on the dipole elements was performed. The parameters of the LPDA are the scale factor τ and space factor σ given by

DOI 10.1007/978-981-4560-75-7_41-1

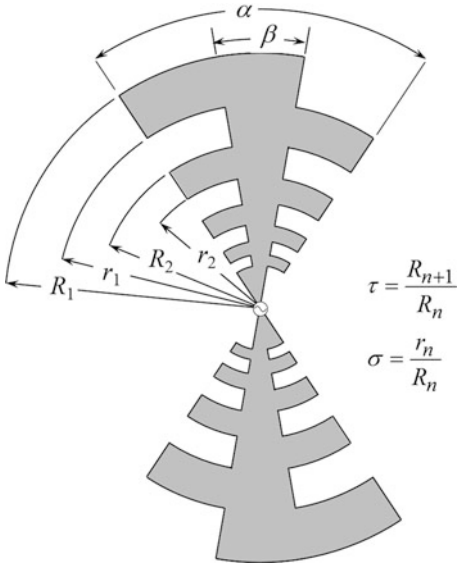


Fig. 10 Modified self-complementary antenna with log-periodic shape (DuHamel and Ore 1958)

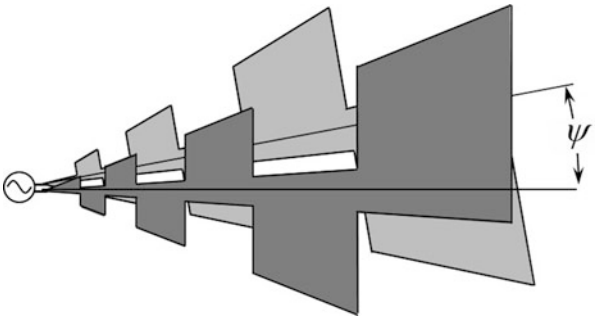


Fig. 11 Modified self-complementary antenna with log-periodic shape having directional radiation pattern (DuHamel and Ore 1958)

$$\tau = \frac{L_{n+1}}{L_n}$$
$$\sigma = \frac{d_n}{2L_n} = \frac{1-\tau}{4}\cot\alpha, \tag{7}$$

respectively. Carrel (1961) showed numerical results of the relation among the directivity, the scale factor τ and space factor σ as well as the consideration of the bandwidth and the input impedance. Peixeiro (1988) also showed the design method of the LPDA using numerical method and improved the Carrel's numerical results.

The LPDA has been widely used for extremely wide frequency operation in communications, TV reception and measurements, especially in the EMC measurements. Figure 15 shows the numerical and experimental input impedance of LPDA (Kim et al. 2001), where the numerical results were obtained by the method of moments using piece-wise sinusoidal basis and test functions (Richmond and Greary 1975; Tilston and Balmain 1990). As can be seen in Fig. 15, the input impedance of LPDA is approximately 50 Ω and the broadband characteristics of LPDA are demonstrated.

DOI 10.1007/978-981-4560-75-7_41-1

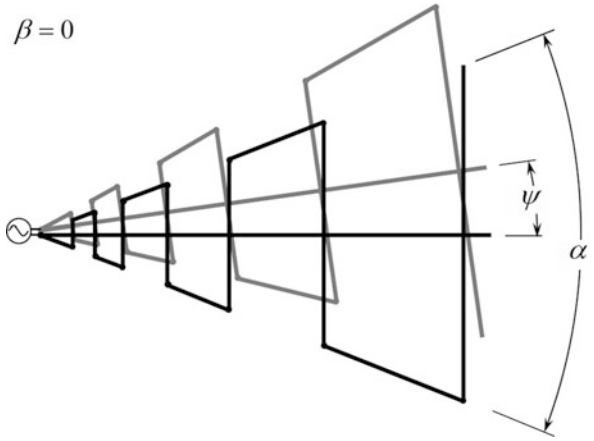


Fig. 12 Modified self-complementary antenna with log-periodic shape composed of thin wires (DuHamel and Ore 1958)

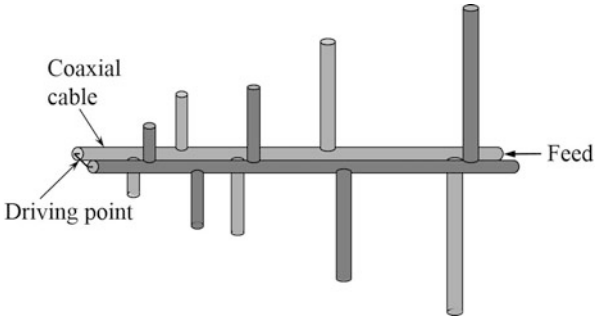


Fig. 13 Log-periodic dipole array (LPDA) (Isbell 1960)

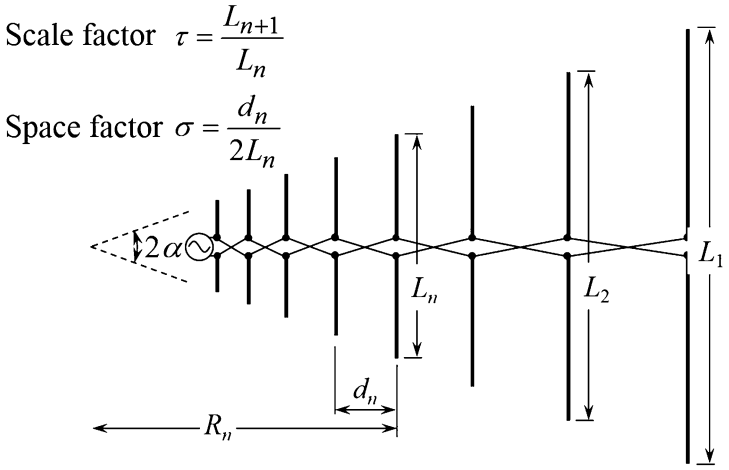


Fig. 14 Parameters of log-periodic dipole array (LPDA) (Isbell 1960)

As mentioned above, so-called log-periodic antennas were developed based on the principle of the self-complementarity. The broadband characteristics of the log-periodic antenna are obtained by the principle of self-complementarity rather than the log-periodic structure (Mushiake 1992, 1996). The most important structure in these broadband antennas is the transposed excitation shown in Fig. 14 (Mushiake 1999).

DOI 10.1007/978-981-4560-75-7_41-1

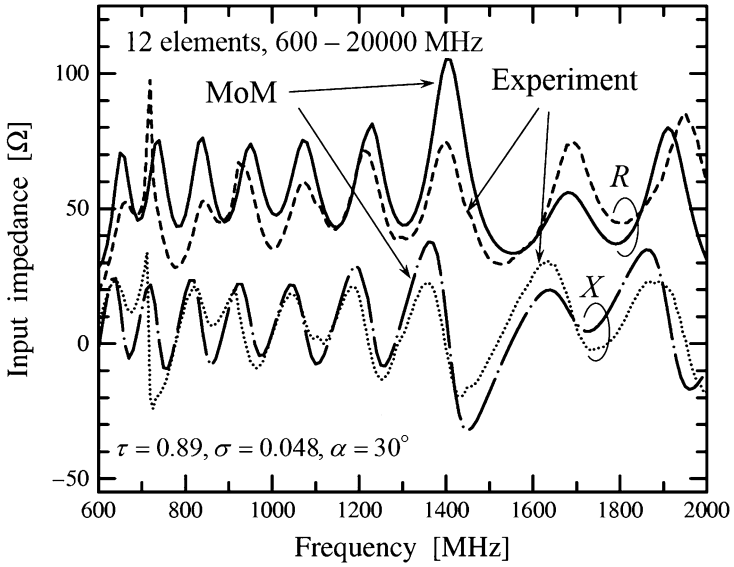


Fig. 15 Input impedance of log-periodic dipole array (LPDA). Comparison of numerical data by method of moments and experimental results (Kim et al. 2001)

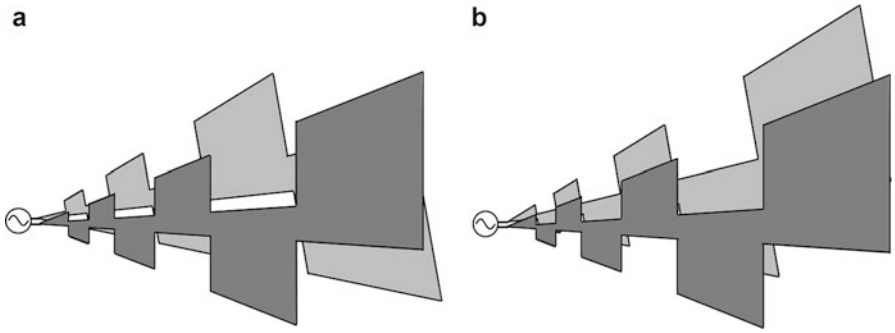


Fig. 16 Self-complementary type and anti-complementary log-periodic antennas (Folded trapezoidal shape) (Mushiake 1965)

Figure 17 shows a comparison of measured input resistance of self-complementary type log-periodic antenna shown in Fig. 16a and anti-complementary type log-periodic antenna shown in Fig. 16b (Mushiake 1965). In the case of anti-complementary type, frequency-independent characteristics are not obtained, whereas broadband property can be observed in the case of self-complementary type log-periodic antenna. Nakano (2006) also presented a numerical analysis of LPDA shown in Fig. 18, where Fig. 18a shows LPDA with transposed excitation and Fig. 18b is that with non-transposed excitation. The input impedance of the LPDA with transposed excitation shows almost constant impedance of about 100 Ω, but LPDA with non-transposed excitation shows strongly oscillatory curves as can be seen in Fig. 19. The broadband property of LPDA comes from transposed excitation of the folded up self-complementary antenna. The transposed excitation is inevitable outcome of the modification, and it is essential for the broadband LPDA. However, the log-periodic shape does not provide with broadband property to antennas (Log-periodic antenna "Any one of a class of antennas having a structural geometry such that its impedance and radiation characteristics repeat periodically as the logarithm of frequency." (IEEE Std 100-1992, Copyright © 1993, IEEE)). Therefore, LPDA is actually a modified self-complementary dipole array (MSCDA) with log-periodic shape (see Wikipedia "Self-complementary antenna" https://en.wikipedia.org/wiki/Self-complementary_antenna). Fractal antenna is another type of

DOI 10.1007/978-981-4560-75-7_41-1

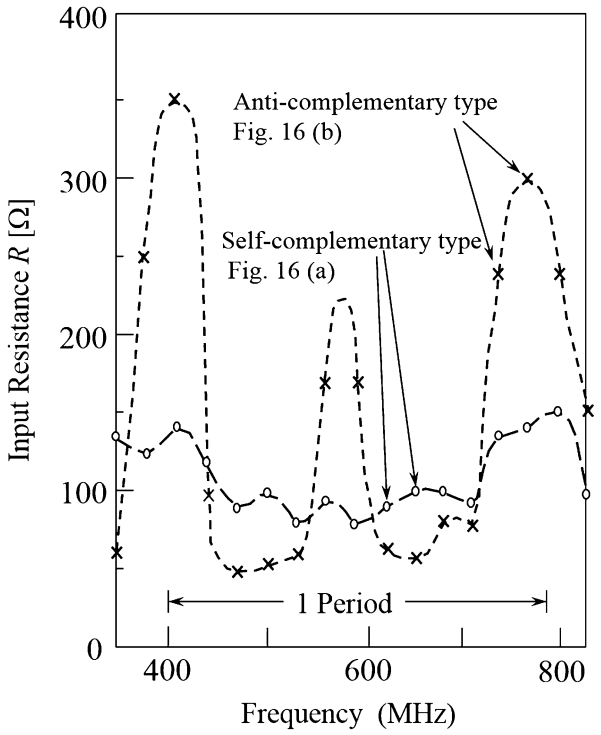


Fig. 17 Comparison of measured input resistance of log-periodic antennas (Mushiake 1965). (**a**) Transposed excitation. (**b**) Non-transposed excitation

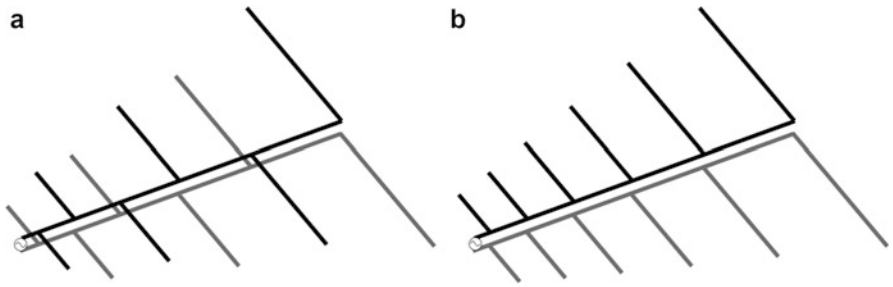


Fig. 18 Log-periodic dipole antennas. Comparison of transposed and non-transposed antennas (Nakano 2006)

log-periodic antenna but it has periodically oscillating characteristics with respect to the logarithm of frequency rather than the broadband properties (Puente et al. 1998).

Because of the broadband characteristics of the LPDA, several types of log-periodic antennas were proposed as described by Jordan et al. (1964). One of the log-periodic dipole array antennas is the log-periodic V array (LPVA) proposed by Mayes and Carrel (1960), where the V-shaped dipole array shown in Fig. 20 was employed and not only the λ/2-resonance of the V dipole elements but also the higher order resonance modes such as the 3λ/2-resonance and 5λ/2-resonance are utilized to maintain the unidirectional radiation pattern and increase the bandwidth. The characteristics of LPVA operating in the 3λ/2-resonance mode was analyzed numerically by Chan and Silvester (1975). LPVA is actually obtained by folding up of the square alternate-leaves type self-complementary antenna which has log-periodic shape in quite a large dimension with respect to the wavelength. Then, some tooth may be resonant at a frequency, by the basic property of the self-complementary antenna. The broadband property of LPVA

DOI 10.1007/978-981-4560-75-7_41-1

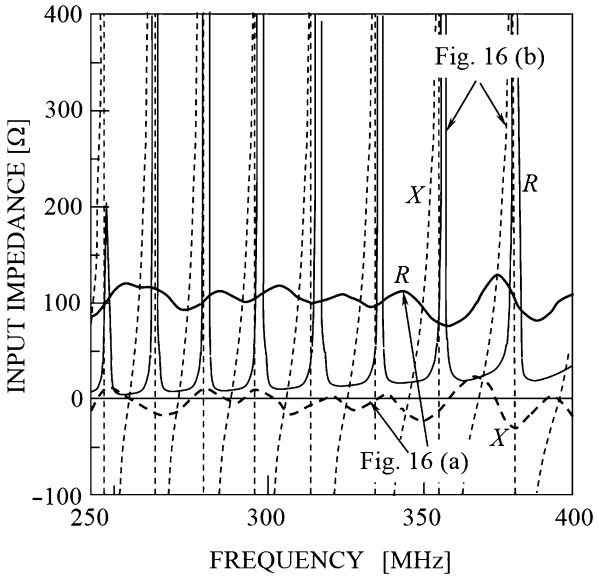


Fig. 19 Input impedance of log-periodic dipole array antenna. Comparison of transposed and non-transposed antennas (Nakano 2006)

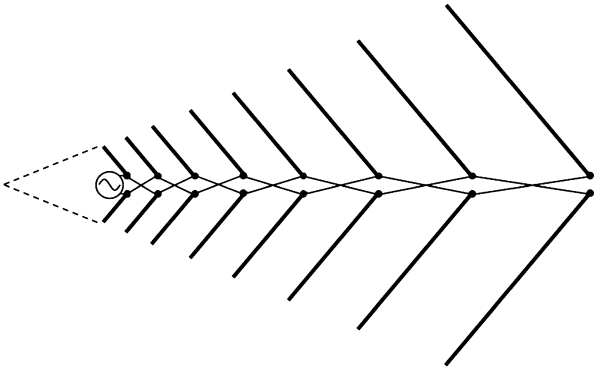


Fig. 20 Log-periodic V array (LPVA) antenna (Mayes and Carrel 1960)

actually comes from transposed excitation of the folded up self-complementary antenna. Further modification of LPVA leads to LPDA. Balmain and Mikhail (1969) proposed an inductively coupled log-periodic dipole array antenna, where the dipole elements are inductively coupled with transposed transmission line and do not have feeding gaps for unbroken dipoles. Since anomalous behavior of this antenna was observed, Oakes and Balmain (1973) optimized the structure in order to remove such behavior. Rojarayanont and Sekiguchi (1977) proposed a log-periodic loop array (LPLA) antennas shown in Fig. 21, where loop elements were employed and a bandwidth of about 2.5:1 is obtained. In these log-periodic antennas, transposed excitation is employed which yields the broadband characteristics.

DOI 10.1007/978-981-4560-75-7_41-1

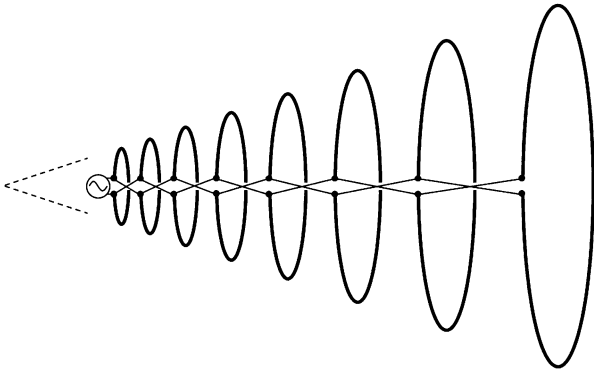


Fig. 21 Log-periodic loop array (LPLA) antenna (Rojarayanont and Sekiguchi 1977)

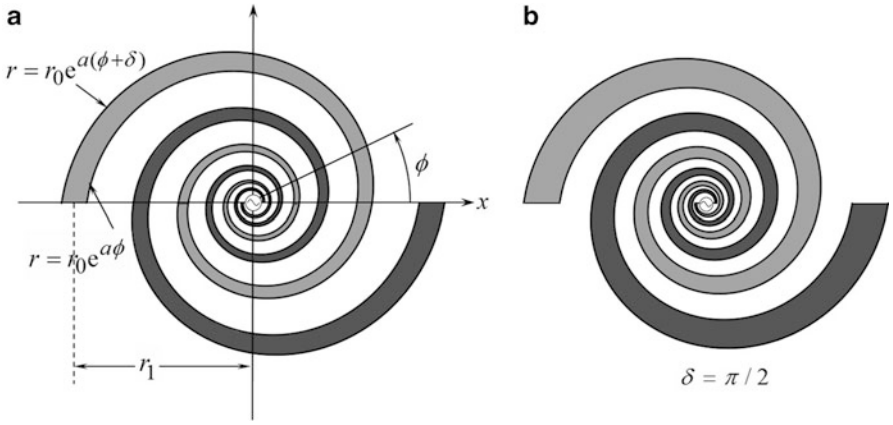


Fig. 22 Two-arm equiangular spiral antennas. (**a**) General equiangular spiral antenna. (**b**) Self-complementary and equiangular spiral antenna

Spiral Antennas

Planar Spiral Antennas

Planar log-periodic spiral antenna is one of the self-similar antennas. It radiates circularly polarized wave bidirectionally. Figure 22 illustrates the two-arm equiangular spiral antenna or log-spiral antenna (Dyson 1959a), where the outer and inner sides of the strip conductor are defined by

$$\begin{aligned} r &= r_0 e^{a(\phi+\delta)} \\ r &= r_0 e^{a\phi}, \end{aligned} \tag{10}$$

and δ is the parameter expressing the width of the strip conductor. The planar log-spiral antenna is the self-similar antennas and has the frequency independent properties. It is also noted that the structure becomes the original self-complementary antenna in the case of $\delta = \pi/2$ as shown in Fig. 22b.

An infinite structure is required for the complete self-similar antenna and truncation effect occurs in the cases of realistic finite structure. The lowest operating frequency of the log-spiral antennas may be given by

$$f_{\text{low}} = c/(2\pi r_1), \tag{11}$$

DOI 10.1007/978-981-4560-75-7_41-1

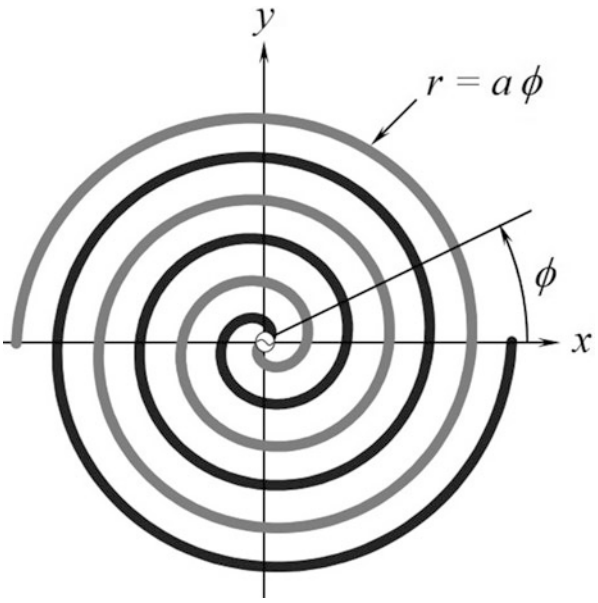


Fig. 23 General two-arm Archimedean-spiral antenna

where r_1 is the outer radius of the spiral and c corresponds the speed of light.

It could be possible to design four-arm or multi-arm spiral antennas, but multi-arm log-spiral antennas are not easy to fabricate because planar log-spiral antennas are composed of strip conductors with finite width.

Archimedean-Spiral Antennas

Archimedean-spiral antenna is another type of spiral antenna (Kaiser 1960). It also radiates circularly polarized wave bi-directionally. The shape of the conductors composing antenna is defined by

$$r = a\phi, \tag{12}$$

as shown in Fig. 23. When the width of the conducting strip w_c is equal to the width of gap between conducting strips w_s, the structure becomes the original self-complementary one as shown in Fig. 5c and constant impedance can be obtained. In the case of $w_c \neq w_s$, the Archimedean-spiral antennas are not self-complementary antenna or self-similar antenna, but broadband property is obtained, e.g., 10:1 bandwidths. This antenna radiates circularly polarized wave, but linearly polarized wave can be obtained by a pair of Archimedean-spiral antennas (Kaiser 1960). Four-arm Archimedean-spiral antenna is also available and was numerically analyzed by Nakano et al. (1983).

Conical Spiral Antennas

The planar spiral antennas are bidirectional antennas, and unidirectional antennas are desired. A method to obtain unidirectional antenna is a combination of the spiral antennas and reflecting plane or conducting cavity, but the presence of the reflector narrows the bandwidth significantly.

DOI 10.1007/978-981-4560-75-7_41-1

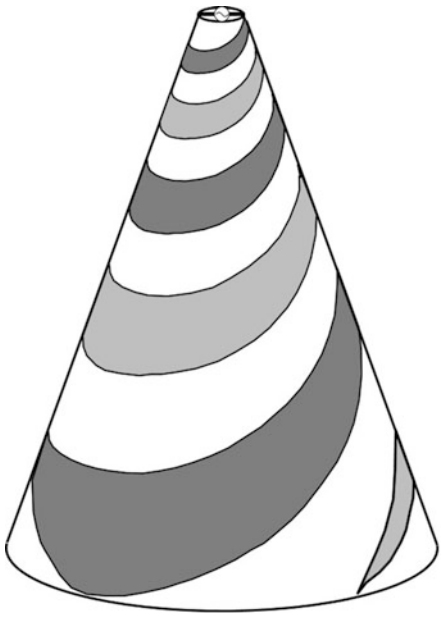


Fig. 24 Two-arm conical spiral antenna

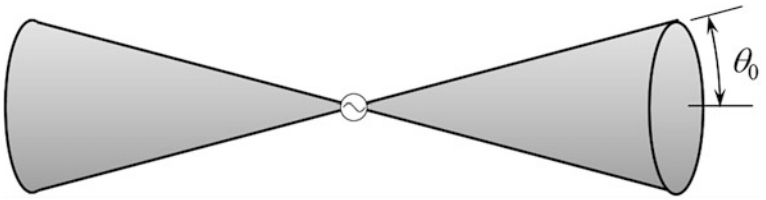


Fig. 25 Biconical antenna

Another method to realize the unidirectional radiation is the introduction of conical spiral antenna, where the log-spiral antenna is placed on a conical surface. Figure 24 shows the two-arm conical log-spiral antenna (Dyson 1959b). The conical spiral is one of the self-similar antennas and broadband characteristics can be obtained. The analysis of the conical log-spiral antenna was reported by Dyson (1965). Multi-arm conical log-spiral antennas were theoretically analyzed by Deschamps and Dyson (1971).

Other Broadband Antennas

Biconical and Bow-Tie Antennas

One of the continuous self-similar antennas is biconical antenna shown in Fig. 25. When the length of biconical antenna is infinite, it can be regarded as a TEM transmission line (Schelkunoff 1951; Tai 1948), and its characteristic impedance is given by

$$Z = \frac{1}{\pi}\sqrt{\frac{\mu_0}{\varepsilon_0}}\cot\frac{\theta}{2} \approx 120\cot\frac{\theta}{2}\,[\Omega], \tag{8}$$

DOI 10.1007/978-981-4560-75-7_41-1

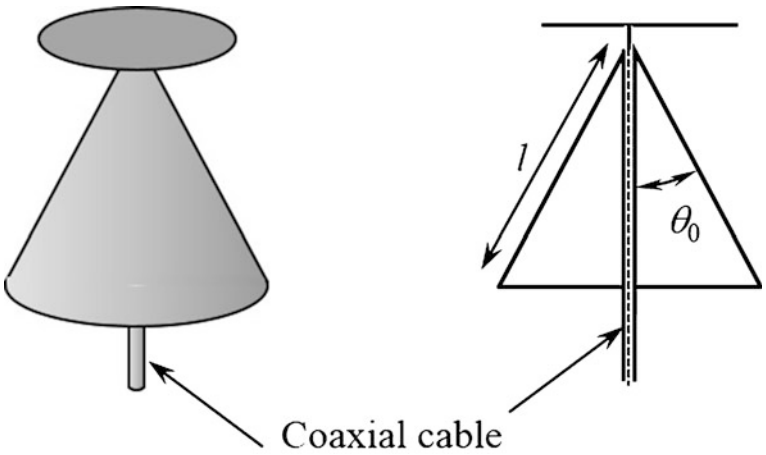


Fig. 26 Discone antenna (Kandoian 1946)

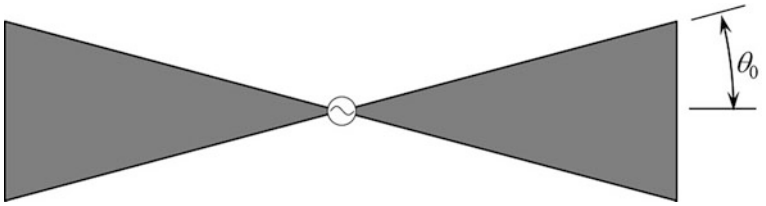


Fig. 27 Bow-tie antenna

In the case of finite length, the input impedance differs from the value of Eq. 8 but is almost constant as long as the total length is sufficiently large.

Figure 26 is the discone antenna composed of a disc and a cone proposed by Kandoian (1946). This antenna is considered to be a modification of biconical antenna and very convenient because of the small size and omni-directional pattern. Typical value of the cone angle is $\theta_0 = 30^\circ$ and the lowest operating frequency of the discone antenna is given by

$$f_{\text{low}} = c/(4l) \tag{9}$$

where l is the length of the cone.

Bow-tie antenna shown in Fig. 27 is another example of the continuous self-similar antennas. It is noted that this antenna becomes the planar self-complementary antenna when the angel is $\theta_0 = 45^\circ$. Bow-tie antennas are often used as pulse transmission and reception because the pulse distortion is small (Shlager et al. 1994). Resistively loaded bow-tie antenna are also used to improve the bandwidth, especially in the case of small bow-tie antenna (Shlager et al. 1994).

Tapered Slot Antennas

One of the other broadband antennas are tapered slot antenna (TSA). TSA is well known for the features of thin structure, low weight, easy to fabricate, broadband, and suitable for microwave integrated circuits. Figure 28 shows the linearly tapered slot antenna (LTSA) fabricated on a dielectric substrate (Yngvesson et al. 1985). LTSA is one of the self-similar antennas having broadband characteristics and unidirectional

DOI 10.1007/978-981-4560-75-7_41-1

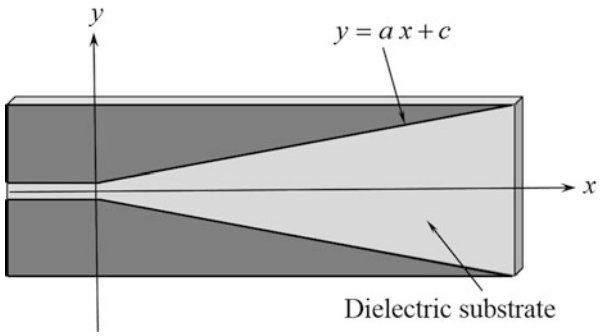


Fig. 28 Linearly tapered slot antenna (LTSA)

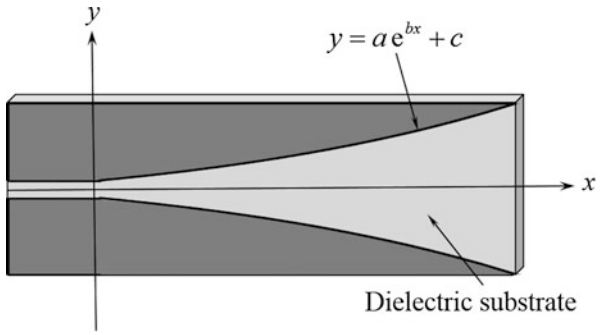


Fig. 29 Vivaldi antenna (Gibson 1979)

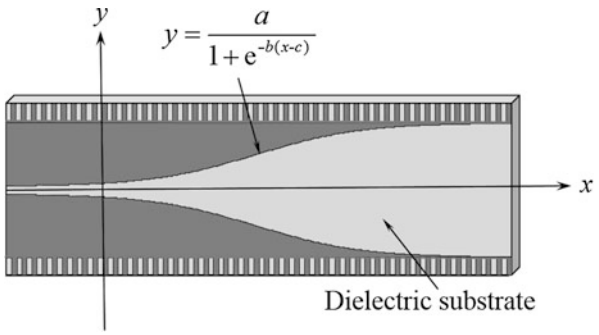


Fig. 30 Fermi antenna (Sugawara et al. 1997)

radiation pattern. In order to extend the bandwidth, Vivaldi antenna shown in Fig. 29 having exponential taper was proposed (Gibson 1979). TSAs were applied to the sensors of millimeter imaging and the elements of array antennas (Yngvesson et al. 1989).

Sugawara et al. (1997, 1998) proposed a TSA called "Fermi antennas" having a profile defined by the Fermi-Dirac function and the corrugation on the side of the substrate shown in Fig. 30. Fermi antenna have many parameters and it is not easy to design the Fermi antennas. Sato et al. (2005) analyzed the Fermi antenna using FDTD method and clarified the relation between the parameters and characteristics. They also showed that the Fermi antenna can yield almost the same radiation patterns in E-plane and H-plane, which are suitable for the millimeter-wave imaging using lens (Sato et al. 2003).

DOI 10.1007/978-981-4560-75-7_41-1

Summary

In this chapter, frequency independent and broadband antennas are reviewed form the view points of the self-complementarity and self-similarity. The theory of self-complementary antenna, original and modified self-complementary antennas, and so-called log-periodic antennas are discussed, where the transposed excitation based on the principle the self-complementarity is the most important to obtain the broadband antennas. Then the frequency independent antennas based on the self-similarity and other broadband antennas are presented.

The ultrawide band antennas described in this chapter are considered to be more and more important for the uses of the new wireless communications such as UWB (ultra wide band) communications and the cognitive radio communications.

Cross-References

- ▶ Broadband and Multiband Planar Antennas
- ▶ Spiral, Helical and Rod Antennas
- ▶ Tapered Slot Antennas
- ▶ Ultra-wideband Antennas

References

Asami Y, Matsumoto T, Matsuura S (1947) Study on the slit aerials. J IEE Japan 67(9):150–153 (in Japanese)

Balmain KG, Mikhail SW (1969) Loop coupling to a periodic dipole array. Electron Lett 5(11):228–229

Booker HG (1946) Slot aerials and their relation to complementary wire aerials. Proc IEE 90(4):620–629

Carrel R (1961) The design of log-periodic dipole antennas. IRE Int Conv Rec 1:61–75

Chan KK, Silvester P (1975) Analysis of the log-periodic V-dipole antenna. IEEE Trans Antennas Propagat 23(3):397–401

Deschamps GA (1959) Impedance properties of complementary multiterminal planar structures. IEEE Trans Antennas Propagat AP-7(5):371–379

Deschamps GA, Dyson JD (1971) The logarithmic spiral in a single-aperture multimode antenna system. IEEE Trans Antennas Propagat 19(1):90–96

DuHamel RH, Isbell DE (1957) Broadband logarithmically periodic antenna structures. IRE Natl Conv Rec 119–128

DuHamel RH, Ore FR (1958) Logarithmically periodic antenna design. IRE Natl Conv Rec 139–151

Dyson JD (1959a) The equiangular spiral antenna. IRE Trans Antennas Propagat 7(2):181–187

Dyson JD (1959b) The unidirectional equiangular spiral antenna. IRE Trans Antennas Propagat 7(4):329–334

Dyson JD (1965) The characteristics and design of the conical log-spiral antenna. IEEE Trans Antennas Propagat 13(4):488–499

Furuya T, Ishizone T, Mushiake Y (1977) Alternate-leaves type self-complememtary antenna and its application to high gain broad-band antennas. Technical report IECE AP77-43

Gibson PJ (1979) The Vivaldi aerial. In: Proceedings of the 9th European microwave conference, pp 101–105

Isbell DE (1960) Log periodic dipole arrays. IRE Trans Antennas Propagat 8(3):260–267

Ishizone T, Yokoyama Y, Nishimura S, Mushiake Y (1983) Unipole-slot array antennas. Trans IECE Japan J66-B(3):281–288 (in Japanese)
Jordan EC, Deschamps GA, Dyson JD, Mayes RE (1964) Developments in broadband antennas. IEEE Spectr 1(4):58–71
Kaiser J (1960) The Archimedean two-wire spiral antenna. IRE Trans Antennas Propagat 8(3):312–323
Kandoian AG (1946) Three new antenna types and their applications. Proc IRE Waves Electr 34(2):70w–75w
Kim D, Yamaguchi S, Chen Q, Sawaya K (2001) Bandwidth of phased array antennas using LPDA elements. Technical report IEICE AP2001-30
Mayes PE, Carrel RL (1960) Logarithmically periodic resonant-v arrays. University of Illinois, Urbana, Antenna Lab. Report 47
Mushiake Y (1949) The input impedances of slit antennas. J IEE Japan 69(3):87–88 (in Japanese)
Mushiake Y (1965) Constant impedance antennas. J IECE Japan 48(4):580–584 (in Japanese)
Mushiake Y (1992) Self-complementary antennas. IEEE Antennas Propagat Mag 34(6):23–39
Mushiake Y (1996) Self-complementary antennas –principle of self-complementarity for constant impedance. Springer, London
Mushiake Y (1999) Log-periodic structure provides no broad-band property for antennas. J IEICE 82(5):510–511 (in Japanese)
Mushiake Y (2004) A report on Japanese development of antennas: from the Yagi-Uda antenna to self-complementary antennas. IEEE Antennas Propagat Mag 46(4):47–60
Nakano H (2006) Recent progress in broadband antennas. In: Proceedings 2006 international symposium antennas and propagation (ISAP 2006) 2A1a-1, pp 1–4
Nakano H, Yamauchi J, Hashimoto S (1983) Numerical analysis of 4-arm Archimedean spiral antenna. Electron Lett 19(3):78–80
Oakes C, Balmain KG (1973) Optimization of the loop-coupled log-periodic antennas. IEEE Trans Antennas Propagat 21(2):148–153
Peixeiro C (1988) Design of log-periodic dipole antennas. IEE Proc 135(2):98–102
Puente C, Romeu J, Pous R, Cardama A (1998) On the behavior of the Sierpinski multiband antenna. IEEE Trans Antennas Propagat 46(4):517–524
Richmond JA, Greary NH (1975) Mutual impedance of nonplanar-skew sinusoidal dipoles. IEEE Trans Antennas Propagat AP-23(3):412–414
Rojarayanont B, Sekiguchi T (1977) A study on log-periodic loop antennas. Trans IECE Japan J60-B (8):583–589 (in Japanese)
Rumsey VH (1957) Frequency independent antennas. IRE Natl Conv Rec 114–118
Rumsey VH (1966) Frequency independent antennas. Academic, New York
Sato H, Arai N, Wagatsuma S, Sawaya K, Mizuno K (2003) Design of millimeter wave Fermi antenna with corrugation. IEICE Trans Commun J86-B(9):1851–1859 (in Japanese)
Sato H, Sawaya K, Wagatsuma S, Mizuno K (2005) Broadband FDTD analysis of Fermi antenna with corrugation. IEICE Trans Commun J88-B(9):1682–1692 (in Japanese)
Schelkunoff SA (1951) General theory of symmetric biconical antennas. J Appl Phys 22(11):1330–1332
Shlager KL, Smith GS, Maloney JG (1994) Optimization of bow-tie antennas for pulse radiation. IEEE Trans Antennas Propagat 42(7):975–982
Sugawara S, Maita Y, Adachi K, Mori K, Mizuno K (1997) A mm-wave tapered slot antenna with improved radiation pattern. IEEE MTT-S Int Microw Symp Dig 959–962

DOI 10.1007/978-981-4560-75-7_41-1

Sugawara S, Maita Y, Adachi K, Mori K, Mizuno K (1998) Characteristics of a mm-wave tapered slot antenna with corrugated edges. IEEE MTT-S Int Microw Symp Dig 533–536

Tai CT (1948) On the theory of biconical antennas. J Appl Phys 19(12):1155–1160

Tilston MA, Balmain KG (1990) On the suppression of asymmetric artifact arising in an implementation of the thin-wire method of moments. IEEE Trans Antennas Propagat 38(2):281–285

Uda S, Mushiake Y (1949) The input impedances of slit antennas. Tech Rep Tohoku Univ 14(1):46–59

Yamamoto Y, Sawaya K, Ishizone T, Mushiake Y (1982) Self-complementary monopole-notch array antennas. Trans IECE Japan J65-B(1):70–77 (in Japanese)

Yngvesson KS, Schaubert DH, Korzeniowski TL, Kollberg EL, Thungren T, Johansson JF (1985) Endfire tapered slot antennas on dielectric substrates. IEEE Trans Antennas Propagat 33(12):1392–1400

Yngvesson KS, Korzeniowski TL, Kim YS, Kollberg EL, Johansson JF (1989) The tapered slot antenna –a new integrated element for millimeter-wave applications. IEEE Trans Microwave Theory Tech 37(2):365–374

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_42-2

Fresnel Zone Plate Antenna

H. D. Hristov*
Department of Electronic Engineering, Federico Santa Maria University of Technology, Valparaiso, Chile

Abstract

The classic Fresnel zone plate has the advantage of being a flat, two-dimensional structure with a small thickness, light and easy to manufacture compared to the bulky refraction lens. In some cases, however, the zoned plate can be fabricated as a three-dimensional curvilinear assembly standing alone or conformal to some man-made or natural formation.

This chapter is devoted to aperture antennas based on flat or curvilinear Fresnel zone plate lenses or reflectors. It is written as a self-sufficient text, which brings together most of the standard knowledge and recent research on the Fresnel zone plate antennas.

Keywords

Diffraction; Focusing; Lens; Fresnel zones; Fresnel zone plate; Fresnel zone plate antenna

Introduction

The Fresnel zone plate antenna is a lens or reflector antenna consisting of two basic components: feed (dipole, horn, etc.) and Fresnel zone plate (FZP). The FZP can be transmissive (lens) or reflective (reflector) in the case of lens or reflector antenna, respectively. FZP lens or reflector is based on the Huygens-Fresnel diffraction principle and focuses an incident plane wave by diffraction, rather than by refraction. While trying to explain the free-space light propagation Fresnel proposed the concept of half-wave zones (known as Fresnel zones) outlined at a wave front plane. He disclosed that the light emitted from the nearby half-wave zones should be in a phase opposition (Fresnel 1866). A. J. Fresnel's portrait is pictured in Fig. 1.

The first optical half-wave FZP lens was designed and examined by J. L. Soret and was intended for the visible light focusing (Soret 1875). The original glass zone plate lens consisted of adjacent transparent and opaque circular Fresnel zones. Soret found that such a zone plate illuminated normally by a plane light wave produced infinite number of spherical waves focused at different focal spots (foci): primary ones with a focal distance (length) $F_1 = F = a_1^2/\lambda$ and secondary ones with focal distances $F/3, F/5$, etc., where a_1 is the radius of the first (central) zone circle and λ is the design light wavelength. The Soret-type lens is usually known as a binary-amplitude or simply binary FZP lens.

In 1871 Lord Rayleigh wrote in his laboratory notepad that the light intensity at the FZP focus could be increased four times if the opaque zones are made transparent and constructive. It was R. W. Wood, however, who put in practice Rayleigh's idea. He manufactured from glass and studied experimentally the so-called phase-reversing FZP lens, where all alternative zones were made transparent and productive, retarded or advanced in phase by 180°. Besides, Wood did the first ever practical application of two FZP

*Email: hristohristov@ymail.com
*Email: hristo.hristov@usm.cl

DOI 10.1007/978-981-4560-75-7_42-2

Fig. 1 Augustin Jean Fresnel's portrait framed by elliptical FZP lens

lenses serving as the objective and eyepiece of a light telescope. By means of this telescope Wood managed to see distinctively the lunar craters (Wood 1898, 1934).

Since the focusing principle of the FZP lens is valid for any frequency this lens can be operative not only in the optical range but also in the radio or quasioptical bands (here the term *radio* implies microwave, millimeter and terahertz (submillimeter) wave. The basic references on the theory, design, and production of the optical Fresnel zone plates and systems are collected in a book edited by Ojeda-Castañeda and Gómes-Reino (1996).

The radio wave FZP lenses originated directly from the optical ones and had similar characteristics and applications. From the very beginning the main efforts in studying radio-wave FZP lenses and antennas have been dedicated toward the growth of their focusing efficiency and resolution. The flat binary FZP lens is easy converted into a phase-reversing one if its opaque (metal or absorber) rings are replaced by transparent phase-shifting rings (dielectric, for example), having the same zone pattern but different profile thickness (depth). For making the flat zone plate competitive in focusing efficiency to the ordinary refraction lens different multiple phase-correcting techniques have been employed. Also, by placing a plane reflector behind the flat Fresnel zone lens, a FZP reflector (or folded FZP lens) is realized. And finally, at the expense of a bigger volume, the equal-in-aperture curvilinear FZP lenses or reflectors show the biggest possible focusing efficiency, much better imaging, and scanning properties. A futuristic view of a satellite communication FZP lens antenna conformal to a building cupola is illustrated in Fig. 2.

The early binary radio lens and reflector FZP antennas were patented about 80 years ago (Clavier and Darbord 1936; Bruce 1939, 1946). During the next three decades extensive studies and applications were carried out by numerous researchers. Both binary and dielectric phase-correcting FZP lenses were designed as focusing microwave antenna elements (Maddaus 1948; Van Buskirk and Hendrix 1961). Wiltse and Sobel headed several research projects resulting in the first effective millimeter and submillimeter (terahertz) phase-correcting FZP lenses and antennas (King et al. 1960; Sobel et al. 1961; Cotton et al. 1962; Wiltse 1985, 1998).

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_42-2

Fig. 2 Futuristic view of satellite FZP lens antenna conformal to building cupola

More about the evolution of radio wave FZP antennas until today, and about their present technological level and feasible applications are found in many journal and conference publications, and in several books (Goldsmith 1998; Hristov 2000; Guo and Barton 2002; Minin and Minin 2004, 2008a, Wiltse 2012, unpublished).

Flat Fresnel Zone Plates in Antenna Mode

Half-Wave (Binary and Phase-Reversing) FZP Lenses Depending on the zone pattern, profile shape, and plate structure, various radio FZP lens and reflector designs are possible. The flat half-wave FZP lenses, binary (Soret-type) and phase-reversal (Wood-type) have similar zone patterns but different cross section profiles and constructions. In general, each lens transforms a given curved wave front into another curved wave front, and therefore, it has two foci: transmitting and receiving, and two focal distances (lengths), F_1 and F_2, respectively. The lens antenna, however, is converting a plane wave front with $F_1 = \infty$ to a curved (spherical, for instance) wave front with F_2=F. It is understood that the lens with a single finite focal length is operating as a lens antenna in receive or transmit mode.

Zone Patterns of Binary FZP Lenses Figure 3 illustrates a binary FZP lens illuminated by an incident paraxial plane wave. This lens is acting like a receiving antenna (or in antenna receive mode). The binary FZP is named positive or negative if the odd zones are open or closed, respectively. The basic design parameters of the binary FZP lens are wavelength λ, total number of zones N, n^{th} zone outer radius b_n, lens diameter $D = 2b_N$, primary focal length F, and outmost zone width $\Delta b = \Delta b_N$. In the phase-correcting FZP lenses total lens thickness w is calculated.

Diffracted by the round FZP lens the plane wave is transformed into a spherical wave converging to a small spot or primary focal point P at $z = F$. More exactly, the focusing effect is a result of two wave phenomena: diffraction by the open zones and interference of diffracted waves in the focal region. Almost half of the electromagnetic energy illuminating the binary FZP lens is blocked by the obstructed Fresnel zones. Also, the phase in the open zones changes from 0° to 180°. Thus, compared to the ordinary refraction lens with the same aperture size, the binary FZP lens has much lower focusing efficiency.

DOI 10.1007/978-981-4560-75-7_42-2

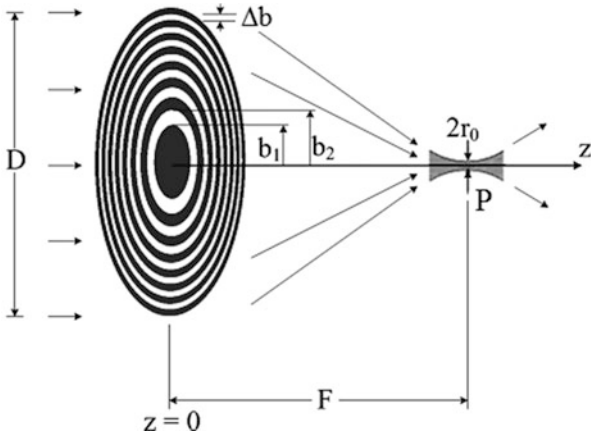


Fig. 3 Binary negative FZP lens operating in receive antenna mode

Zone patterns (front views) of circular, elliptic, and linear binary FZP lenses are shown in Fig. 4. For a paraxial plane wave the FZP patterns have symmetric shapes (circular, Fig. 4a, or linear, Fig. 4b). For an oblique plane wave incidence the zone patterns become asymmetric (elliptic, Fig. 4c, or linear, Fig. 4d). While the ring-shape lenses, circular or elliptical, focus the plane wave at a point (or make a plane-to-spherical wave conversion), the strip-shape lenses focus the plane wave along a focal line, or they convert the plane wave into a cylindrical wave. The opaque zones in the FZP lenses are usually metallized (covered by thin metal rings, strips, etc.) and are normally printed on a low-loss dielectric substrate.

Two identical linear zone plates crossed at right angle form the so-called linear FZP cross, Fig. 5a. More complex FZP planar patterns have also been investigated, as for example the hyperbolic FZP cross, Fig. 5b.

Profiles of Binary and Phase-Correcting FZP Lenses Figure 6 shows the transverse cross sections (profiles) of several flat half-wave FZP configurations. Fig. 6a displays the profile of a binary positive zone plate consisting of open (+) Fresnel zones and opaque (−) metal zone rings. Fig. 6b shows a binary negative zone plate. The binary FZP lens is typically manufactured by means of a microprinting technology, as illustrated in Fig. 6c, where a positive metal FZP rings are printed unilaterally on a thin, low-loss dielectric substrate. In Fig. 6d, e the phase reversing is made by means of Fresnel zone annular grooves cut in a flat dielectric plate with a permittivity ε. As a result, each groove with a depth w is alternated by a dielectric edge.

A microwave variety of Wood's zone plate is the invented by J. C. Wiltse zone plate of dielectric rings, Fig. 6f. This is a smooth, biplanar plate with increased focusing efficiency. The Wiltse lens is designed as a double-dielectric combination of Fresnel zone rings having equal thickness w. In Fig. 6f the permittivities of alternate dielectric rings are specified as ε_1 and ε_2. For $\varepsilon_1 = \varepsilon_0$ the double-dielectric phase-reversing lens becomes simpler and lighter, but some techniques for ring fixing are necessary. For example, in Fig. 6g, the rings are firmly fixed and encapsulated by two thin outer disks with a permittivity ε, which slightly increase the attenuation through the lens (Hristov 1996).

A bigger increase in the focusing efficiency is achieved by means of multiple phase corrections. For example, each full-wave zone, incorporating two Fresnel (half-wave) zones, can be subdivided into four subzones, and thus, a quarter-wave phase-correcting FZP lens can be realized. Four levels of stepwise phase correction per full-wave zone are completed by four stepped, annular grooves cut in a flat, single-permittivity dielectric plate, Fig. 7a, (Sobel et al. 1961; Garrett and Wiltse 1991).

A quarter-wave FZP lens with four distinct-permittivity dielectric rings is sketched in Fig. 7b, while Fig. 7c shows the profile of refraction plane-hyperbolic lens, which is obviously much thicker and

DOI 10.1007/978-981-4560-75-7_42-2

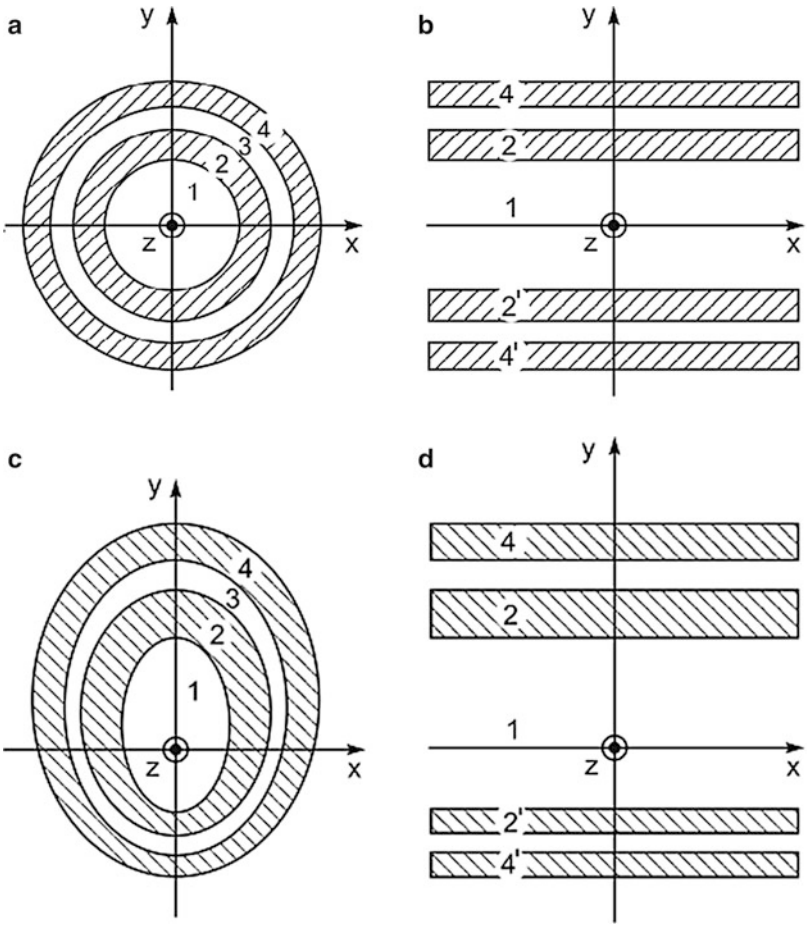


Fig. 4 Zone patterns of half-wave FZP lenses consisting of (**a**) concentric circular rings, (**b**) symmetric parallel strips, (**c**) concentric elliptical rings and (**d**) asymmetric parallel strips

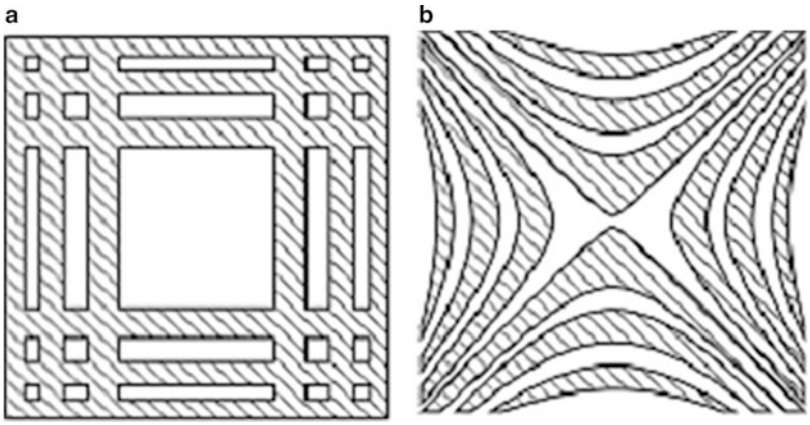


Fig. 5 Half-wave plane FZP lenses: (**a**) *front views* of linear FZP cross, (**b**) *front views* of hyperbolic FZP cross

weighty, and lossy as well. A traditional zoned refraction lens is shown in Fig. 7d. It comprises many small curved teeth-shaped edges, which at millimeter and terahertz frequencies become very problematic for production. Fig. 7e is a picture of a quarter-wave four-permittivity phase-correcting lens made by an

DOI 10.1007/978-981-4560-75-7_42-2

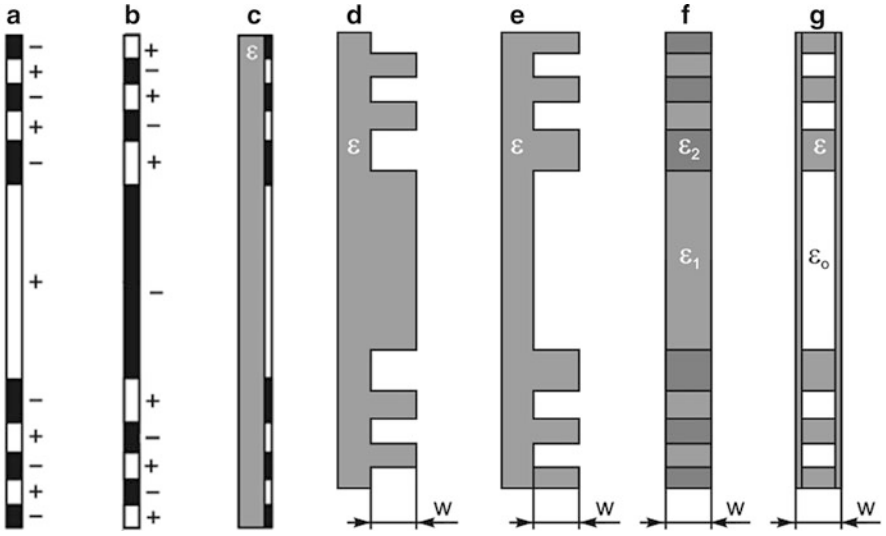


Fig. 6 Profiles of half-wave FZP lenses: (**a**) metal binary positive, (**b**) metal binary negative, (**c**) printed positive, (**d**) and (**e**) grooved phase-reversal, (**f**) bi-planar double-dielectric ring formation, and (**g**) air plus solid dielectric rings encapsulated by thin dielectric disks

appropriate subzone drilling of single-dielectric disk plate (Petosa and Ittipiboon 2003; Petosa et al. 2006).

Phase Correction in Flat FZP Lenses and by Use of Composite/Meta Media Additional thickness and weight reduction in the FZP lenses is possible by employing composite phase correction subzone rings, i.e., metal-dielectric structures instead of pure dielectric ones. Often such composite media are made as grid layers of metal obstacles or slots. The grid metal elements (dipoles, disks, rings, ordinary and Jerusalem crosses, etc.) are usually printed or embedded in a dielectric plate. Such a metal-dielectric plate is narrowband, frequency-selective surface (FSS) structure. Some technological examples are shown in Fig. 8, where (a) is a profile of double-metal grid embedded in a dielectric slab and (b) and (c) are front and profile views of metal double-square printed FSS array.

A thin phase-shifting surface (PSS) that consists of three metallic and two dielectric layers is described in Gagnon et al. (2010), Fig. 8d. The metallic layers consist of conducting elements, which are tuned to introduce the desired phase shifts. Such a frequency-selective surface structure has been applied for constructing phase-correcting zone plate lenses and lens antennas (Gagnon et al. 2012, 2013). The operations of the new FSS structures and FZP lenses/antennas based on them are well validated by measurements.

A low-profile Wood zone plate metalens based on the fishnet metamaterial is proposed in (Orazbayev et al. 2015). It is characterized by a refractive index less than unity (n = 0.51). The metalens is made of alternating dielectric and fishnet metamaterial concentric rings. The use of fishnet metamaterial allows reducing the reflections from the lens, while maintaining low profile, low cost and ease of manufacturing.

Profiles of FZP Reflectors By placing a plane reflector at a distance $d = \lambda/4$ behind the binary Fresnel zone plate the incident wave is brought back in antiphase, and the combination of FZP and reflector works as a FZP reflector (or folded binary FZP). Its cross-sectional view is sketched in Fig. 9a. An alternative FZP reflector is sketched in Fig. 9b, where a metal sheet is shaped in a zigzag manner for making the required phase-reversing steps. A similar FZP cross-sectional view of a metal quarter-wave stepped plate is shown in Fig. 9c. These two zone plate reflector structures can be encapsulated in a dielectric foam or honeycomb.

DOI 10.1007/978-981-4560-75-7_42-2

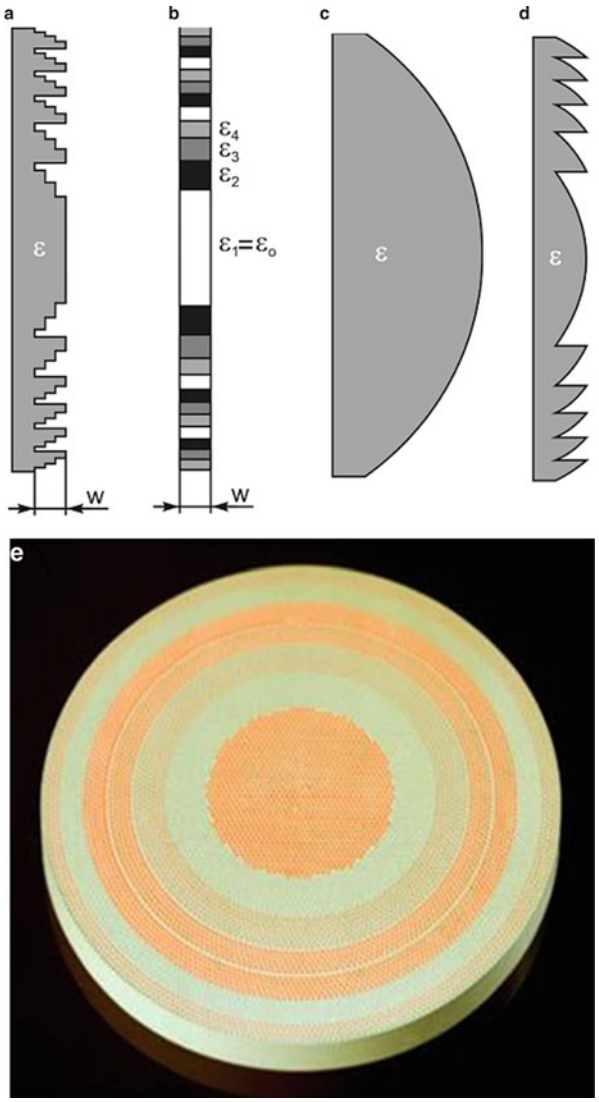


Fig. 7 More effective dielectric lenses: (**a**) grooved quarter-wave FZP lens, (**b**) quarter-wave four- dielectric ring FZP lens, (**c**) ordinary refraction lens, (**d**) zoned ordinary refraction lens and (**e**) drilledsingle-dielectric quarter-wave FZP lens

A printed millimeter-wave design of a FZP reflector is described in Huder and Menzel (1988). As shown in Fig. 9d it is fabricated by use of printed technology and consists of a ground plane disk reflector and FZP rings etched on a dielectric substrate, w_1 in width.

Based on the half-wave and quarter-wave grooved single-dielectric zone FZP lenses, FZP reflectors sketched in Fig. 9e and Fig. 9f, respectively, can be constructed. Similarly, the multi-dielectric quarter-wave FZP lens is transformed into FZP reflector (with a thickness $w_1 = w/2$, Fig. 9g. Obviously, the grooved and multi-dielectric FZP reflectors are roughly twice thinner compared to the respective FZP lenses.

FZP Design Considerations Cross-sectional design and building of phase-correcting dielectric zone plates involves the choice of the following structural features:

1. $\bar{f}$-focal number: FZP lens is characterized by its focal number $\tilde{f} = F/D$, where D is the lens diameter and F is the primary focal length. Typically, the optical FZP lenses have very big $\bar{f}$-number. The radio wave Fresnel zone plates have much smaller focal number and require more exact design equations and

DOI 10.1007/978-981-4560-75-7_42-2

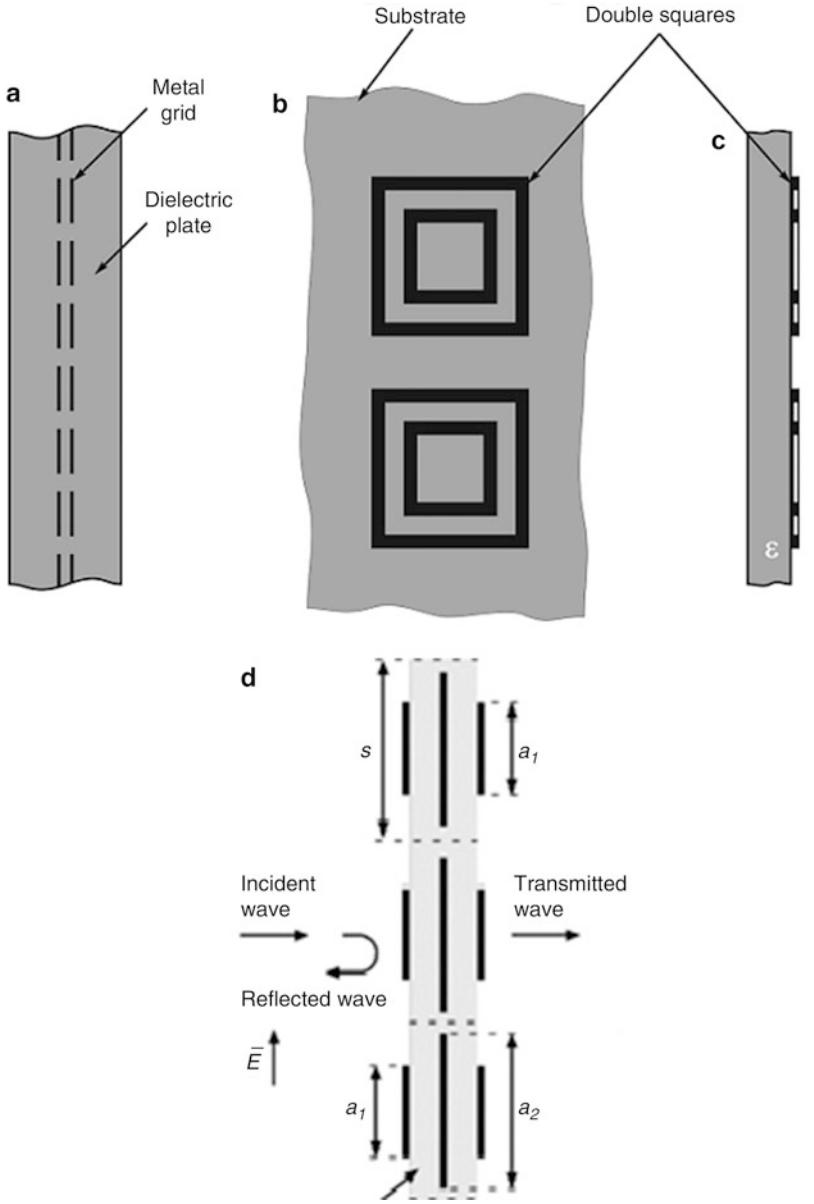


Fig. 8 Metal-dielectric phase-shifting media: (**a**) double-metal grid in dielectric slab, (**b**) front and (**c**) profile views of phase-correcting FSS of printed double-square metal rings, and (**d**) phase-correcting FSS of three metallic and two dielectric layers

methods for electromagnetic analysis. This is especially valid for the short-focal zone plates, with $\tilde{f} < 0.5$.

2. Zone pattern: number of FZP zones/subzones and their outlines (circular, elliptical, linear, cross-type, etc.).
3. FZP profile: straight-line or curvilinear (spherical, parabolic, conical, etc.).
4. Phase correction structure: selected according to the required focusing gain, focusing aperture efficiency, frequency bandwidth, lens wall thickness, and others.

FZP Fabrication Different precise technologies are employed for production of radio wave Fresnel zone plate lenses, especially of those for millimeter waves and terahertz frequencies:

DOI 10.1007/978-981-4560-75-7_42-2

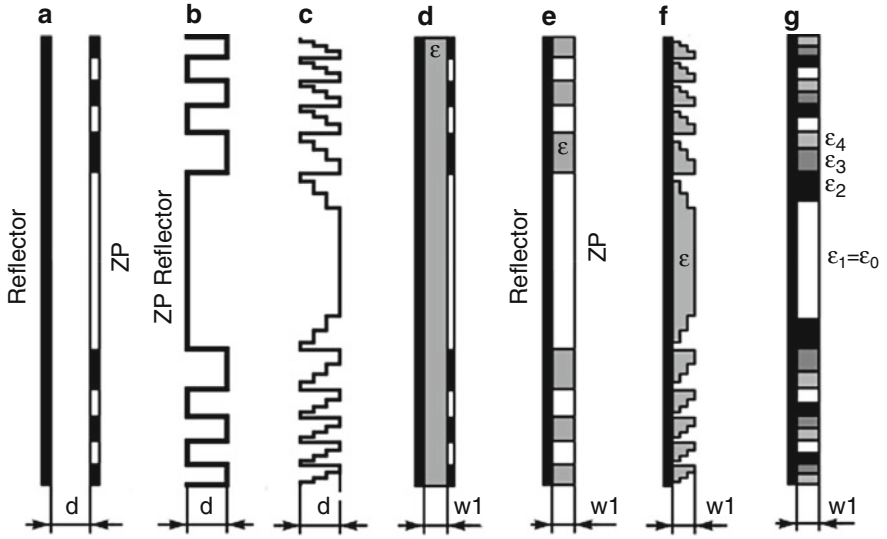


Fig. 9 Profiles of flat FZP reflectors: (**a**) FZP in front of reflector (folded binary FZP), (**b**) binary zigzag reflector, (**c**) quarter-wave zigzag reflector, (**d**) folded half-wave FZP printed on dielectric substrate, (**e**) half-wave dielectric FZP with reflector, (**f**) quarter-wave dielectric FZP with reflector and (**g**) multi-dielectric quarter-wave FZP with reflector

1. Metal (foil) zone elements printing on a thin dielectric substrate without a ground (FZP lenses) or grounded (FZP reflectors)
2. Cutting and milling for making phase-correcting steps or grooves, having in mind that at 1 THz and above an eighth-wavelength correction step can be less than 50 μm
3. Creation by drilling of multipermittivity zones in single-dielectric zone plates
4. Assembling of dielectric rings in multi-dielectric zoned plates
5. Making composite metal-dielectric or solid-state zone structures by use of microelectronic technologies (for instance, photolithography or chemical depositing)

The dielectric material for FZP fabrication has to be a low-loss one, with a proper design permittivity. The internal losses in the dielectric FZP lenses increase significantly with the frequency, especially over the millimeter and terahertz ranges. Typical dielectric materials include polymers, like polystyrene (Rexolite), Teflon, TPX, Tsurupica, etc., and dielectric crystals, like quartz, sapphire, and silicon (Goldsmith 1998).

Zone Radius of Half-Wave FZP Lens Geometry of plane-wave illuminated half-wave FZP is presented in Fig. 10a. Diffracted by the receiving FZP lens plane wave is transformed into infinite in number primary and secondary spherical waves converging to the respective primary and secondary foci. Because the energy at the primary focus $P(0, 0, F)$ is predominant, the secondary foci of the Fresnel zone plate are neglected. Thus, it is supposed that there is only one focus, the primary one. The reverse (transmit) wave transformation is taking place as follows: the spherical wave front radiated from a focal point source is transformed by the FZP lens into a plane wave. As was mentioned before, the phase in each Fresnel zone is not a constant but varies from 0 deg at the origin O to 180 deg at point Q_n. Thus, the half-wave (Fresnel zone) ray path difference is $\mathrm{RPD} = \overline{Q_nP} - \overline{OP} = \left(F^2 + b_n^2\right)^{1/2} = n\lambda/2$. If $b_n^2 << F^2$, the outer radius b_n of the n^{th} half-wave zone is easily found by

DOI 10.1007/978-981-4560-75-7_42-2

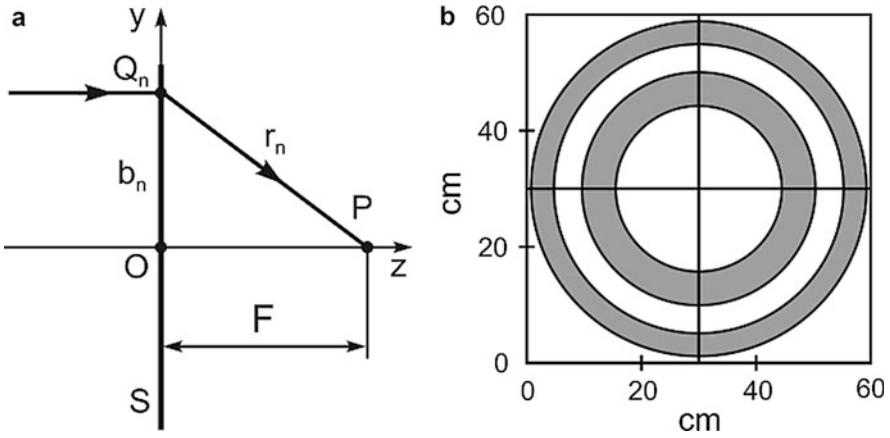


Fig. 10 Half-wave planar FZP axially illuminated by plane wave: (**a**) converging geometry and (**b**) Fresnel zone pattern

$$b_n = \sqrt{n\lambda F + \left(\frac{n\lambda}{2}\right)^2}, \quad n = 1,\ 2,\ 3,\ \ldots, N \tag{1}$$

where N is the total number of half-wave zones.

Equation 1 is usually applied to radio Fresnel zone plate lenses. For the optical waves, $(n\lambda/2)^2 << n\lambda F$, and Eq. 1 is further approximated to

$$b_n = \sqrt{n\lambda F} \tag{2}$$

The circular zone pattern drawn in Fig. 10b is calculated from Eq. 1 for λ =3.2 cm and $F = D = 60$ cm, or $F/D = 1$ and $D/\lambda = 18.75 >> 1$.

Zone Radius of Full-Wave FZP Lens Two adjacent half-wave (Fresnel) zones numbered by m and (m-1) form a full-wave zone, which corresponds to a ray path difference $RPD = m\lambda$, where $m = 1, 2, \ldots, M$, $m = n/2$, and $M = N/2$. Thus, the m^{th} full-wave zone has an outer radius b_m found by Eq. 3, similar to Eq. 1

$$b_m = \sqrt{2m\lambda F + (m\lambda)^2} \tag{3}$$

Zone Radius of Half-Wave FZP Lens with Reference Phase Recently, the classical concept for the zero initial phase at the FZP origin O is reconsidered (Webb 2003). An extra initial phase θ_{ref} named reference phase is introduced as a new variable, which for the positive FZP lens is changed from 0 to 180 deg, and for the negative FZP lens – from 180 to 360 deg. If the reference phase is expressed as $\theta_{ref} = 360\Delta_{ref}/\lambda$, the path ray difference RPD is increased by $\Delta_{ref} = \lambda\theta_{ref}/360$ or $RPD = \sqrt{F^2 + b_n^2} - F = n\lambda/2 + \Delta_{ref}$. In result, the zone radius equation, or Eq. 1, obtains a more general form

$$b_n(\theta_{ref}) = \sqrt{2F\left(\frac{n\lambda}{2} + \frac{\lambda\theta_{ref}}{360}\right) + \left(\frac{n\lambda}{2} + \frac{\lambda\theta_{ref}}{360}\right)^2} \tag{4}$$

DOI 10.1007/978-981-4560-75-7_42-2

For $n = 0$, an initial reference zone is defined with a radius $b_0 = b_{ref}$ given by

$$b_{ref} = \sqrt{2F\frac{\lambda\theta_{ref}}{360} + \left(\frac{\lambda\theta_{ref}}{360}\right)^2} \tag{5}$$

Obviously, a zero reference radius ($b_{ref} = 0$) corresponds to $\theta_{ref} = 0$, which matches the classical Fresnel-Soret half-wave zone plate. In fact, this extra free parameter, the variable radius $R_0 = b_{ref,}$ was introduced about 15 years earlier than the parameter reference phase (Minin and Minin 2004, 2008a). By changing the reference phase or radius, valuable improvement in the FZP lens/antenna performance is obtained. Also, there is a difference in frequency and phase response between the main lobe and sidelobes. This new finding can be utilized for multipath-fading and radar sidelobe clutter reductions and communication security enhancement (Webb et al. 2011).

Zone Radius of FZP Lens with Multiphase Correction In contrast to the traditional refractive lens the diffractive FZP lens does not make smooth phase-error compensation but a discrete or step-phase one. In other words, the FZP lens is acting as a stepwise zone/subzone phase converter. Each full-wave zone is divided into $Q = 2, 4, 8$, etc. (usually even in number) subzones, and the corresponding lenses are termed as half-wave, quarter-wave, eighth-wave, etc. The phase in each s^{th} subzone differs by $2\pi/Q$ relative to the phase in the adjacent subzone (Guo and Barton 1993b). More details for the phase-correcting mechanism in the FZP lens are found in Guo and Barton (2002). For $Q \to \infty$ the Fresnel zone plate is converted into the well-known Fresnel zone refractive lens, Fig. 7d.

In the classical subzone FZP lens, where $\theta_{ref} = 0$ and $RPD = \lambda s/Q$, the s^{th} subzone radius b_s is computed by

$$b_s = \sqrt{2F\left(\frac{s\lambda}{Q}\right) + \left(\frac{s\lambda}{Q}\right)^2} \tag{6}$$

where $s = 1, 2, 3, \ldots S$, $s = mQ,$, $S = MQ$, m and M are the current and total number of the full-wave zones, respectively.

If the extra reference phase is included (or if $\theta_{ref} \neq 0$) the subzone radius equation becomes

$$b_s = \sqrt{2F\left(\frac{s\lambda}{Q} + \frac{\lambda\theta_{ref}}{360}\right) + \left(\frac{s\lambda}{Q} + \frac{\lambda\theta_{ref}}{360}\right)^2} \tag{7}$$

For $\theta_{ref} = 0$ and $Q = 2$, $s = n$ and $S = N$, the multi-half-wave subzone FZP becomes a half-wave lens, and Eq. 7 converts to Eq. 1.

Thickness of Binary Metal FZP Lens The binary FZP is usually produced by a skinny metal sheet printed on a thin nongrounded dielectric substrate, Fig. 6c, while in the case of corresponding reflector the printed FZP lens is backed by the substrate ground, Fig. 9d.

Thickness of Grooved-Dielectric Multiphase-correcting FZP Lens The FZP phase correction structures are made most frequently of low-loss dielectric plates. The plates are grooved in a stepwise manner as shown in Fig. 7a, for lenses, and in Fig. 9f, for reflectors, where the phase is changed as a staircase of Q levels and $Q-1$ steps. Each step has a width w_1 given by

DOI 10.1007/978-981-4560-75-7_42-2

$$w_1 = \frac{\lambda}{Q(\sqrt{\varepsilon} - 1)} \tag{8}$$

where λ is the free-space design wavelength and ε is the relative dielectric permittivity.

In the case of phase reversal dielectric lens with one grooved step or two phase levels ($Q = 2$), the FZP lens thickness w is

$$w = \frac{\lambda}{2(\sqrt{\varepsilon} - 1)} \tag{9}$$

The total grooved-dielectric FZP lens of $Q-1$ steps per full wave has a thickness $w = (Q-1)\, w_1$, or

$$w = \frac{(Q-1)\lambda}{Q(\sqrt{\varepsilon} - 1)} \tag{10}$$

The supporting substrate slab of the FZP lens can be made with the same permittivity ε or distinct permittivity ε_s. As in the refraction lenses, the thickness w_s of the FZP supporting structure can be chosen to make a negligible influence on the FZP operation and better interface match between the lens and free space, and thus to improve the FZP efficiency.

According to Eq. 9, for $\varepsilon = 4$, the phase reversal FZP thickness is equal to the free-space half wavelength or $w = \lambda/2$. For a quarter-wave grooved-dielectric FZP lens ($Q = 4$) with a permittivity $\varepsilon = 4$, the Eq. 10 gives $w = 3\lambda/4(\sqrt{\varepsilon} - 1) = 0.75\lambda$.

For a better impedance match between the lens and free space, the permittivities of the lens ε and supporting layer ε_s are connected by the equation $\varepsilon_s = \sqrt{\varepsilon}$. In addition, the thickness of the supporting layer is given by the approximation $w_s \approx \lambda/4\sqrt{\varepsilon_s}$. Having depths w and w_s calculated, the total thickness w_t of the grooved-dielectric FZP lens is

$$w_t = w + w_s \tag{11}$$

Thickness of Multi-Dielectric Ring FZP Lens Consider initially the thickness (depth) equation for double-dielectric FZP lens illuminated by a paraxial plane wave (Black and Wiltse 1987)

$$w = \frac{\lambda}{2\left(\sqrt{\varepsilon_2} - \sqrt{\varepsilon_1}\right)} \tag{12}$$

where ε_2 and ε_1 are the relative permittivities of the two dielectrics (plastics).

It is evident from Eq. 12 that for a given design wavelength the minimum depth corresponds to the maximum ratio between ε_2 and ε_1. These relations are displayed graphically in Fig. 11a, (Wiltse 2012), where ε_r corresponds to ε_2. The FZP lens depth of cut (in cm) versus the frequency (in GHz) is plotted for three permittivity ratios: $\varepsilon_2/\varepsilon_1 = 1.6$, 2.1, and 2.54.

Particular depths of two plane-convex ordinary lenses, plane-spherical and plane-hyperbolic, Fig. 11b, are compared to the depth of same-diameter phase-reversing FZP lens, Fig. 11c. All lenses are designed for $f = 140$ GHz, $F = 200$ mm, and $D = 200$ mm. The hyperbolic and spherical refraction lenses are 8 and 11 times thicker, compared to the FZP lens, respectively. They are also much heavier and difficult to manufacture in comparison to the FZP lens. These advantages are at the expense of about two and half times smaller focusing efficiency.

DOI 10.1007/978-981-4560-75-7_42-2

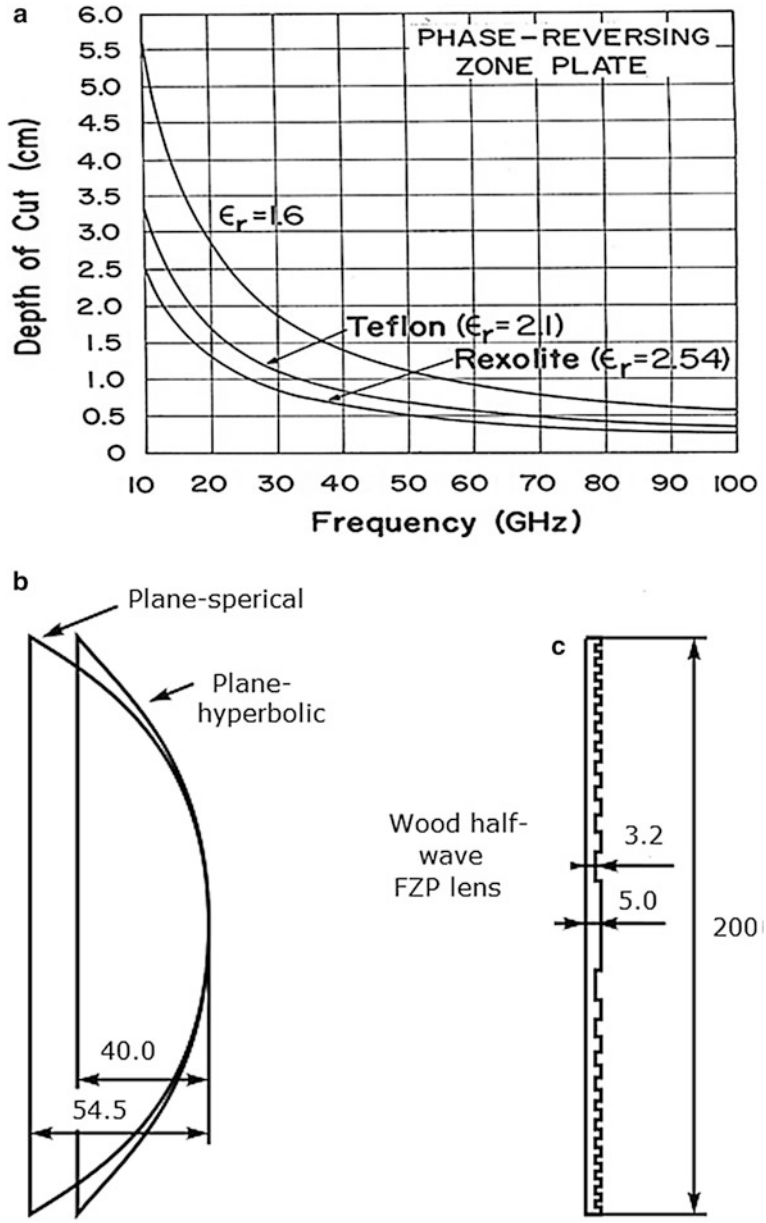


Fig. 11 Depths of refraction and diffraction lenses: (**a**) phase-reversing FZP depth of cut versus frequency, (**b**) ordinary refraction lenses and (**c**) grooved phase-reversing zone plate

Unlike the staircase grooved-dielectric FZP the thickness of flat multi-dielectric FZP lens (or Wiltse lens) is preserved constant. A general permittivity equation for the Wiltse lens was derived and published recently in Hristov and Rodriguez (2013). For a normal plane wave incidence (antenna receive mode) the wave goes through the first subzone thickness without reflection if the following standing-wave condition is met: $w_1 = \kappa\lambda/(2\sqrt{\varepsilon_1})$, where λ is the free-space design wavelength. The optimum ratio $\xi_{s,1}$ between the permittivities of s^{th} and 1st zone ring dielectrics is found by

$$\xi_{s,1} = \frac{\varepsilon_s}{\varepsilon_1} = \left[1 + \frac{2}{k}\left(1 - \frac{s-1}{Q}\right)\right]^2 \tag{13}$$

DOI 10.1007/978-981-4560-75-7_42-2

Here $s = 1, 2, 3, \ldots$ is a current subzone number in the first full-wave zone and $k = 1, 2, 3$, etc. is an integer number of half-waves in the s^{th}dielectric ring. Also, for a minimum FZP lens depth, k is set to 1 and the lens thickness w is fixed equal to that of the first zone/subzone, or $w = w_1$.

Equation 13 is valid for ideal, loss-free dielectrics. If the FZP lens is made of realistic microwave dielectrics, and the s^{th} subzone dielectric ring has a complex transmission coefficient $\tilde{T}_s = |\tilde{T}_s|\exp(j\Phi_t)$, where the attenuation A_s of the s^{th} dielectric ring is given by $A_s = 10\log\left(1/|\tilde{T}_s|^2\right)$ in decibels. This attenuation factor should include both the internal dielectric losses and losses due to the multiple reflections.

For a quarter-wave multi-dielectric FZP lens with $Q = 4$, $\varepsilon_1 = 1$, and $k = 1$ the ring permittivities are found by Eq. 13 as follows: $\varepsilon_2 = 6.25, \varepsilon_3 = 4$, and $\varepsilon_4 = 2.25$. Due to the radial phase periodicity, Eq. 13 is correct and recurrently applicable to all full-wave zones of the lens. Thus, for the second full-wave zone $\varepsilon_5 = 1$, $\varepsilon_6 = 6.25$, $\varepsilon_7 = 4$, and $\varepsilon_8 = 2.25$.

Essentials in Fresnel Zone Plate Operation

Multifocal Action of Binary FZP Lens Suppose a zone plate lens illuminated by a plane wave with a design wavelength λ_1 as shown in Fig. 3. According to the solution in Sussman (1960) the zone plate diffraction lens produces multiple foci at $z = \pi F/w$, primary and secondary. Here $w = \pm(2k + 1)\pi$, $k = 0, 1, 2, \ldots$ The multifocal action is illustrated in Fig. 12a, where the plus signs correspond to the real foci, $P_1(+F_1)$, $P_3(+F_3)$, $P_5(+F_5)$, etc., located at the points $z = F_1 = F$, $z = F_3 = F/3$, $z = F/5$, and so on. Along the + z axis the Fresnel zone plate behaves as a multifocal converging lens. On the other hand, the foci $P_1'(-F)$, $P_3'(-F/3)$, $P_5'(-F/5)$ are virtual images of the real foci. Thus, the binary or Soret zone plate acts simultaneously as two ray-converging multifocal lenses. The points P_1 and P_1' are the first-order or primary foci, and all others are high-order (secondary) foci. The Fresnel zone plate of metal rings has no virtual but only real foci, corresponding to the transmitted and reflected waves.

The normalized intensity versus the real focus number is graphed in Fig. 12b. The intensity maxima are found at the odd foci: $I_1(w = \pi) : I_3(w = 3\pi) : I_5(w = 5\pi), \ldots = 1: (1/3)^2: (1/5)^2$, etc. Thus, the high-order foci 3, 5, and so on have much smaller focal intensities compared to the primary focus intensity. On the other hand, the intensity at the primary focus is $1/\pi^2$ times weaker than the focal intensity of the equal-in-diameter refraction lens.

If the fields focused at the two primary foci $z = +F$ and $z = -F$ are received and brought together in phase, the focusing efficiency of the binary zone plate will be increased twice (or by 3 dB). In practice, such in-phase combination can be done by one of the following techniques: (i) feed-line power combining by use of two feeds, feed lines, phase shifter, and power combiner and (ii) free-space power combining by means of folded binary FZP reflector with diameter $D = 2b_N$ as shown in Fig. 9a.

Multifrequency Focusing of Binary FZP Lens This phenomenon corresponds to the axial multiple focusing of the Fresnel zone plate. Both phenomena are linked by Eq. 1 and can be easier explained by Eq. 2. From Eq. 2 follows that for a given FZP diameter the change of the focal distance F for $\lambda = \lambda_1 =$ const or the change of wavelength λ for $F =$ const leads to the multifocal or multifrequency effect, respectively.

Initially, the multifrequency phenomenon will be discussed taking as an example a binary FZP lens of four Fresnel zones designed for a focal length F and wavelength λ_1. Fig. 13a illustrates the Fresnel zone assembly with all zones open. The plane waves going through the positive and negative zones will cancel at P_1 and will not produce a focusing effect, i.e., the wave is propagating without any change. In Fig. 13b

DOI 10.1007/978-981-4560-75-7_42-2

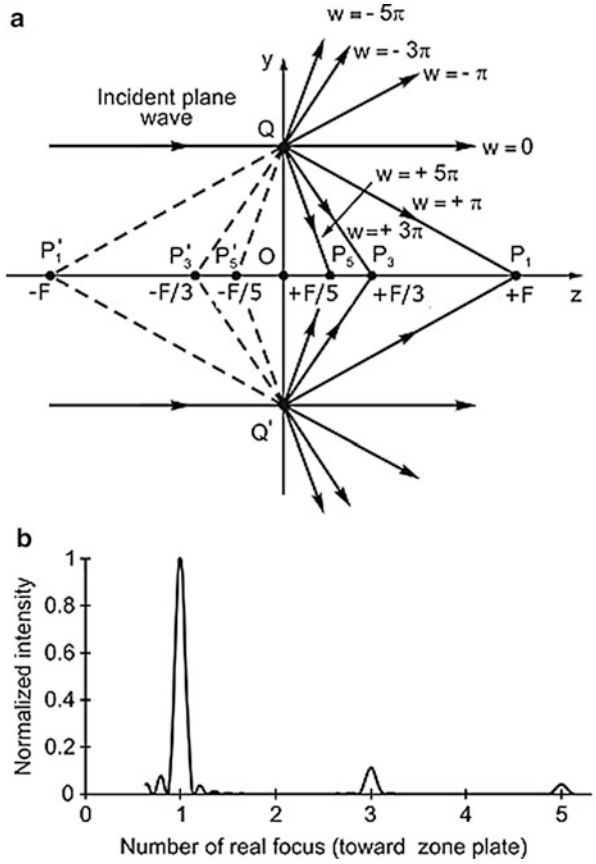


Fig. 12 Diffraction mechanism of binary zone plate illuminated by plane wave with design frequency f_1 wavelength λ_1: (**a**) forming primary and secondary foci, and (**b**) focusing intensity at the primary and secondary foci

the negative (even) zones are blocked by metal (or thin absorbing) rings (Soret lens). The focusing is produced by the open (white) zones only, which fields add in phase at the primary focus P_1.

Now consider the same FZP lens construction illuminated by a plane wave with a wavelength λ_3 equal to the third harmonic wavelength, or $\lambda_3 = \lambda_1/3$. According to Eq. 2, for the same lens diameter D and design focal length F, the threefold smaller wavelength λ_3 requires three times bigger number of subzones S, or $S = 3N$. This means that in each real half-wave zone corresponding to the design wavelength λ_1, three virtual λ_3 subzones are enclosed: two positive (in white) and one negative (in gray), Fig. 13c. Thus, for the s^{th} subzone radius Eq. 2 can be rewritten as follows:

$$b_s = \sqrt{s\lambda_3 F} = \sqrt{3n\lambda_3 F} \tag{14}$$

Thus, the subzone radii can be calculated by Eq. 14, where $s = 3n$ is the current subzone number. In Fig. 13c the fields produced by the 2nd (gray) and (white) subzones will cancel leaving only the 1st subzone to contribute at P_1. Similar physical explanations are valid for all other odd harmonic wavelengths: λ_5, λ_7, and so on.

Theoretically, for each odd wavelength harmonic the FZP lens produces one and the same field intensity (focusing gain) at the primary focus P_1. Nevertheless, the FZP aperture efficiency for the higher harmonics is dropping rapidly compared to the efficiency at the design wavelength λ_1. If the FZP focusing efficiency at λ_1 is set to $\eta_1 = 100\ \%$, the higher odd harmonics will be focused at P_1 with much lesser efficiencies: $\eta_3 = \eta_1(1/3)^2 \approx 11\ \%$, $\eta_5 = \eta_1(1/5)^2 \approx 4\ \%$, and so on Hristov (2011).

DOI 10.1007/978-981-4560-75-7_42-2

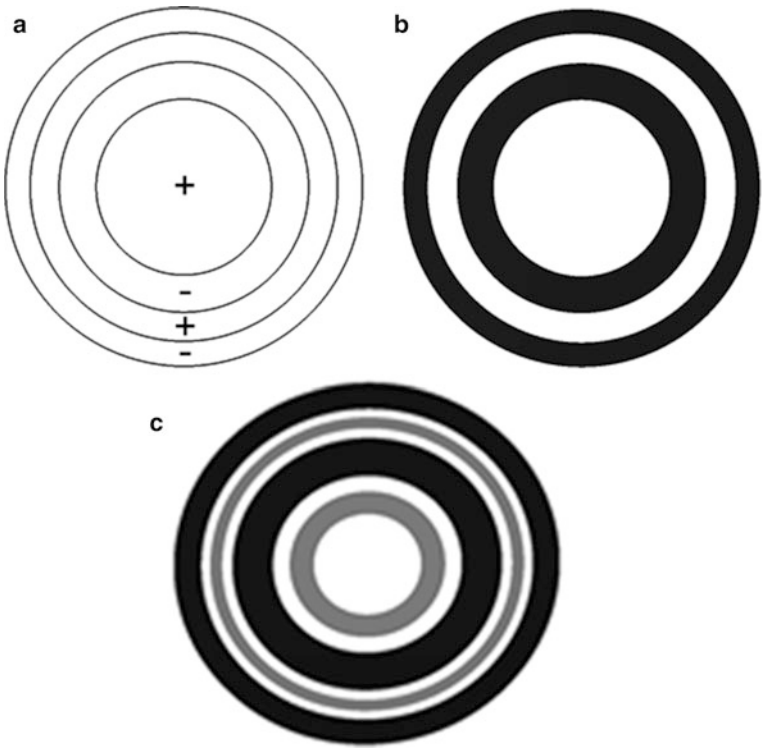


Fig. 13 Harmonic focusing action of binary FZP lens: (**a**) Fresnel zone assembly of all-open open Fresnel zones, (**b**) positive binary FZP lens with even zones blocked (focusing at design wavelength λ_1), (**c**) focusing at 3rd harmonic wavelength λ_3

Phase Correction Mechanism Detailed theory and interpretation of the phase correction mechanism is exposed in Guo and Barton (1993a, b, 1995, 2002). As was explained above, the zoned plate is a stepwise phase front converter, where each full-wave Fresnel zone is usually divided into half-wave, quarter-wave, eighth-wave, and so on subzones.

Consider first the phase correction procedure for the half-wave or Wood phase-reversing zone plate. Suppose that the zone plate is illuminated by a spherical wave front from a point source P at a focal distance F from the plate. The radial phase variation (or phase error) in the outer plane of the zone plate is described by a quadratic function Φ_1 that can be expressed as a function of Fresnel zone number as follows:

$$\Phi_1(n) = -\frac{2\pi}{\lambda}\left(\sqrt{F^2 + b_n^2} - F\right) \tag{15}$$

where b_n is the Fresnel zone radius calculated for a given F and λ by Eqs. 1 or 2.

The calculated Φ_1 is drawn as a dash-dotted line in Fig. 14a. The conversion of spherical phase front into a plane one makes zero phase error in the aperture. Ideally, this can be achieved by a refractive, plane-convex (or convex-plane) dielectric lens, which creates a phase-reversed quadratic function $\Phi_2(n) = -\Phi_1(n)$, drawn by the dashed line. Thus, the sum of the initial phase function $\Phi_1(n)$ and the lens phase correction function $\Phi_2(n)$ will be zero, and thus, the spherical phase error in the lens aperture will be entirely compensated. However, the Fresnel zone plate makes incomplete phase compensation. For example, the phase reversal zone plate produces staircase compensation (dotted line) with steps of 180 deg each. The sum of the spherical phase error function and staircase compensation function is a tooth-shape phase error curve (solid line), which varies within each Fresnel zone from 0 to −180 deg.

DOI 10.1007/978-981-4560-75-7_42-2

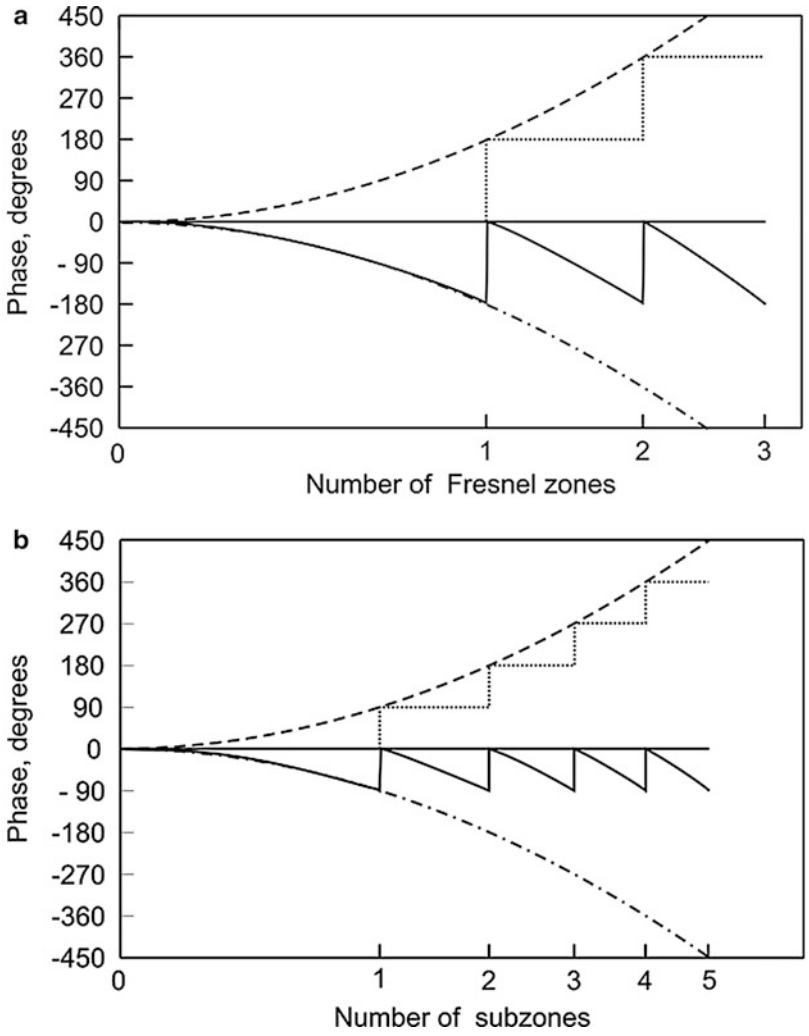


Fig. 14 Stepwise phase-front correction in (**a**) half-wave (phase-reversing) and (**b**) quarter-wave zone plates

Similar reasoning is valid for the quarter-wave zone plate, illustrated graphically in Fig. 14b. There, the resultant aperture phase error function does not exceed −90 deg within each subzone.

Diffraction Focusing Efficiency of FZP Lens Consider the focusing efficiency of ideal FZP lens, free of material and polarization loss. This efficiency depends only on the phase increment in the compensation staircase function and is expressed by the simplified equation (Garrett and Wiltse 1991)

$$\eta_f = \frac{\sin^2(\Delta\Phi/2)}{(\Delta\Phi/2)^2} \tag{16}$$

where $\Delta\Phi$ is a step increment in the staircase phase correction function. From Eq. 16 is easy calculated that the diffraction efficiency of the stepwise phase-corrected half-wave, quarter-wave, and eighth-wave zone plate lenses is equal to 40.5 %, 81 %, and 95 %. If $\Delta\Phi \to 0$ the zone plate diffraction efficiency tends to unity. Therefore, a good efficiency enhancement is obtained for $\Delta\Phi \leq 90$ deg. Actually, if the material, polarization, and reflection loss factors are taken into account the realistic FZP lens focusing efficiency becomes rather smaller.

DOI 10.1007/978-981-4560-75-7_42-2

Resolution of Binary FZP Lens A high resolution is needed when the lens is used for imaging and narrow-beam antenna design. By definition, the so-called Rayleigh resolution criterion is commonly expressed by the resolving (resolution) angle δ_0, in radians (Myers 1951)

$$\delta_0 = \chi \frac{\lambda}{D} \tag{17}$$

In case of the binary FZP $\chi = 1.22$ for number of zones N is bigger than about 200. For smaller N, say $N < 50$, the resolution constant χ varies with N. It is found that $\chi(N) < 1.22$ for the positive FZP lenses (odd zones open), and $\chi(N) > 1.22$ for the negative FZP lenses (even zones open).

According to the axial Airy diffraction pattern the resolving angle δ_0 is directly related to the first null radius r_0 and primary focal length F by the equation $\tan\delta_0 = r_0/F$. For a high resolution δ_0 is small enough and Eq. 17 is approximated to $\delta_0 \approx r_0/F$. Thus, taking into account Eq. 17 and the previous approximate equation for δ_0, the lens radius r_0 can be expressed as $r_0 = \chi(M)\lambda F/D$. The outmost zone width Δb is related to the resolving lens radius by the equation $r_0 = \chi(M)\Delta b$ (Sun and Cai 1991; Hristov 2011).

Curvilinear Fresnel Zone Plates in Antenna Mode

Axially Symmetric Zone Plate Structures For improving the focusing, resolving and frequency properties of the FZP lens it can be made like a thin bent plate (shell) with zone shapes similar to those in the plane FZP lens. In general, the curved FZP can be fabricated with an arbitrary shape provided the Fresnel zones are drawn in accordance with the Fresnel-zone/subzone ray path difference (RPD) condition (Hristov 2000).

From practical and theoretical points of view the most appropriate is a zone plate surface expressed mathematically as an axially rotational curve profile function. Examples of FZP lenses with simple rotational profiles are shown in Fig. 15a for spherical and in Fig. 15b for conical FZP lenses. The conical zone plate is especially suitable for building of collapsible umbrella-like FZP lens/reflector antennas.

Cylindrical Zone Plate Structures Figure 16 illustrates the zone patterns of (a) cylindrical rim ring-type and (b) strip-type FZP lenses. The cylindrical rim-type arrangement consists of circular Fresnel zone rims (rings) and converges the incident plane wave to a focal point P(z = F), while the strip-type FZP structure focuses the plane wave along the focal line P′P″.

Zone Pattern of Spherical Phase-Correcting FZP Lens Figure 17a is the geometry of a radio wave spherical zone plate (shell) with a radius R and sphere center at point C. Here the half-wave FZP lens is considered, with a zone pattern consisting of N half-wave (Fresnel) zones, specific for the binary (Soret-type) and phase reversal (Wood-type). The zone plate is convex-side illuminated by a plane wave and is operating in antenna mode. The approximate equation for the outer radius b_n of the n^{th} zone is derived on the basis of the Fresnel zone ray path difference condition

$$\mathrm{RPD} = \sqrt{\left[F - \left(R - \sqrt{R^2 - b_n^2}\right)\right]^2 + b_n^2} - F = \frac{n\lambda}{2} \tag{18}$$

where $n = 1, 2, 3, \ldots, N$.

From Eq. 18 the half-wave zone radius b_n is obtained (Dey and Khastgir 1973a, b, c)

DOI 10.1007/978-981-4560-75-7_42-2

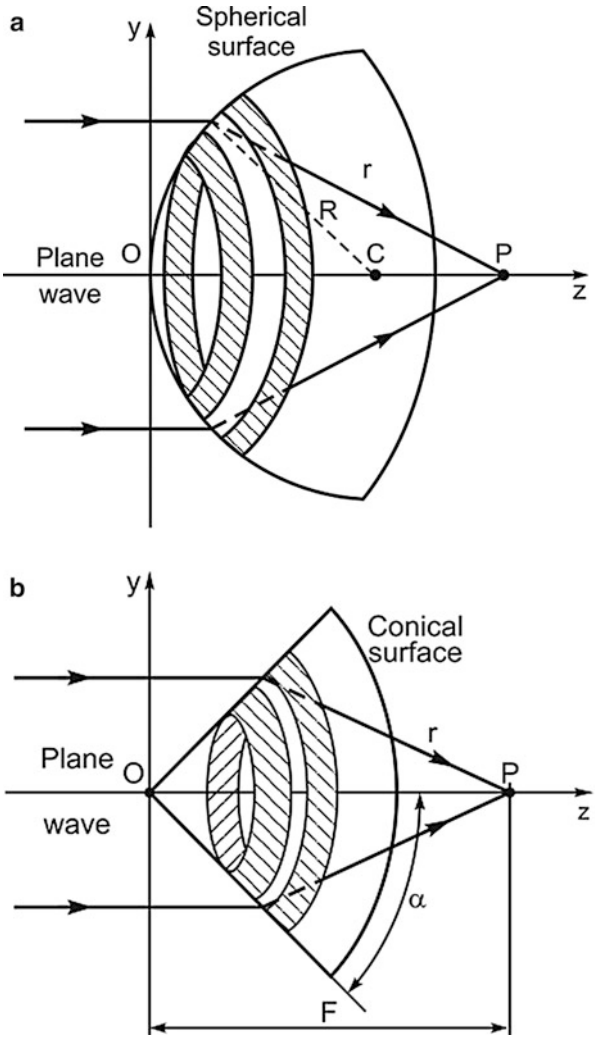


Fig. 15 Zone patterns of half-wave convex-side-illuminated FZP lenses with (**a**) spherical surface and (**b**) conical surface

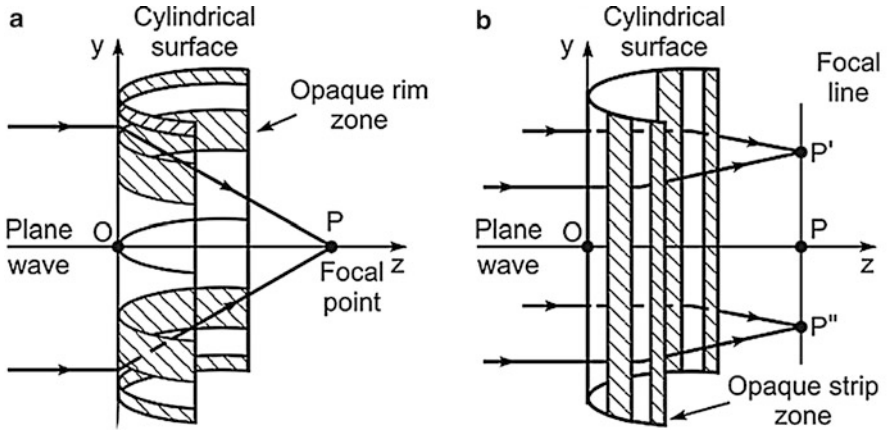


Fig. 16 Zone patterns of convex-side illuminated cylindrical FZP with (**a**) rim-shape and (**b**) strip-shape pattern structures

DOI 10.1007/978-981-4560-75-7_42-2

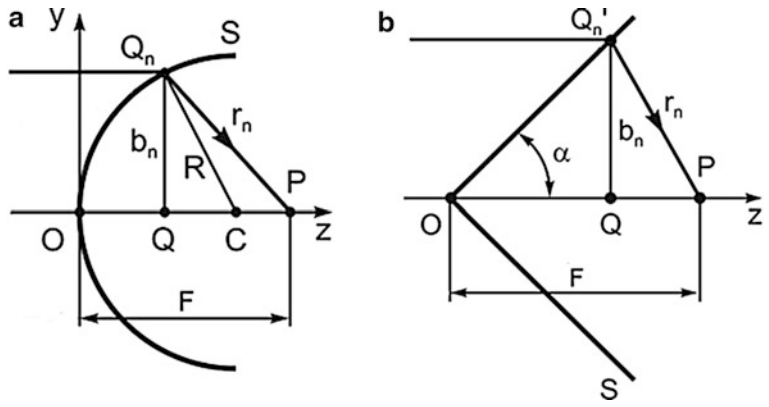


Fig. 17 Half-wave curvilinear FZP lens convex-side illuminated by plane wave: (**a**) spherical lens converging geometry and (**b**) conical lens converging geometry

$$b_n = \sqrt{R^2 - \left(\sqrt{R^2 + n\lambda(R - F)} - \frac{n\lambda}{2}\right)^2} \tag{19}$$

Equation 19 is simplified considerably for $F = R$. In this particular occasion the plane wave is focused at the sphere center C

$$b_n = \sqrt{n\lambda F - \left(\frac{n\lambda}{2}\right)^2} \tag{20}$$

The latter formula is very similar to the flat FZP zone radius Eq. 1, and therefore, both equations can be combined in one

$$b_n = \sqrt{n\lambda F \pm \left(\frac{n\lambda}{2}\right)^2} \tag{21}$$

where the plus and minus signs are valid respectively for the half-wave flat and spherical FZP lenses, with the same diameters and primary focal lengths. And finally, for a paraxial optical/quasioptical FZP $F >> \lambda$ and $n\lambda F >> (n\lambda/2)^2$, and Eq. 21 is converted to Eq. 2.

For the multiphase correcting FZP lens each full-wave zone is subdivided into s subzones, with $s = 1, 2, 3, \ldots S$, $s = mQ$, and $S = MQ$; Q is the number of subzones in a single full-wave zone; m and M are the current and total number of the full-wave zones, respectively. Thus, the FZP RPD condition becomes equal to $s\lambda/Q$, and the equation for the s^{th} radius is found as follows:

$$b_s = \sqrt{R^2 - \left(\sqrt{R^2 + 2(R - F)\left(\frac{s\lambda}{Q}\right)} - \frac{s\lambda}{Q}\right)^2} \tag{22}$$

Zone Pattern of Conical Phase-Correcting FZP Lens Figure 15b represents a rotational conical FZP lens convex-side illuminated by an axially incident plane wave. Here, for the half-wave phase-correcting conical FZP a specific design parameter is defined: cone opening semi-angle $\alpha = \text{atan}(s_n/OQ) = \text{const}$.

DOI 10.1007/978-981-4560-75-7_42-2

The n^{th} outer Fresnel zone radius b_n in the half-wave zone pattern projected in the plane $z = OQ$ is given by the following equation (Minin and Minin 2004):

$$b_n = \sqrt{2F\frac{n\lambda}{2} + \left(\frac{n\lambda}{2}\right)^2 + \left(\frac{n\lambda}{2\tan\alpha}\right)^2} - \frac{n\lambda}{2\tan\alpha} \tag{23}$$

For the general case of conical zone plate, the s^{th} subzone radius b_s is easy found from Eq. 23 after replacing b_n by b_s and $n\lambda/2$ by $s\lambda/Q$

$$b_s = \sqrt{2F\frac{s\lambda}{Q} + \left(\frac{s\lambda}{Q}\right)^2 + \left(\frac{s\lambda}{Q\tan\alpha}\right)^2} - \frac{s\lambda}{Q\tan\alpha} \tag{24}$$

where $s = 1, 2, 3, \ldots S$, $s = mQ$, and $S = MQ$ are parameters and relations similar to those for the spherical FZP lens.

The conical FZP lens with a half-opening angle $\alpha = 90$ deg becomes a flat FZP lens. If in addition $Q = 2$, and $s = n$, Eq. 24 reduces to Eq. 1.

The spherical and conical Fresnel zone plates can be turned also to the incident plane wave by their concave sides. Such FZP lenses are specified as concave lenses operating in antenna mode.

Fresnel Zone Plate Focusing Methods

Optical Diffraction Method for Focusing Analysis Classically, the FZP focusing and imaging or FZP antenna radiation analysis is completed by use of optical methods (GO, UTD, Kirchhoff's diffraction integral (KDI), etc.). Next is described an approximate diffraction focusing theory valid for any axially symmetric curvilinear zone plate with a specified profile function according to Hristov (2000). The zone plate geometry with a profile curve function $y(z)$ illuminated by a plane wave is illustrated in Fig. 18a. All open zone arcs are approximated by the corresponding chords. P is an axial running point illuminated by the elementary area $dS = y(z)d\varphi(dz/\cos\alpha_n)$.

According to the diffraction theory the field produced by the n^{th} zone at point P can be written as $dE_n = jE(Q)I(\vartheta_n, \vartheta_n')\exp(-j\beta r(z))dS/\lambda r(z)$, where the inclination factor is approximated as in the case of planar zone plate, or $I(\vartheta_n, \vartheta_n') \approx (1 + \cos\vartheta_n)/2, \beta = 2\pi/\lambda\, dz/dl = \cos\alpha_n = \sin\eta_n$. The n^{th}-zone field at point P is found as a surface integral, which after integration with respect to the rotational coordinate φ reduces to the next linear integral

$$E_n(P) = j\frac{\pi E_0}{\lambda}\int_{z_{n-1}}^{z_n} y(z)\frac{1+\cos\vartheta_n}{\cos\alpha_n}\frac{\exp[-j\beta(r(z)+z)]}{r(z)}dz \tag{25}$$

with $r(z) = \sqrt{(z_p - z)^2 + y(z)^2}$.

The integration limits z_n and z_{n-1} are calculated by the approximated RPD condition

$$z_{n,n-1} = z_p - b_{n,n-1} + \sqrt{(n, n-1)\lambda b_{n,n-1} - b_{n,n-1}^2} \tag{26}$$

where $b_{n,n-1}$ can be calculated in advance for each specific zone plate profile.

DOI 10.1007/978-981-4560-75-7_42-2

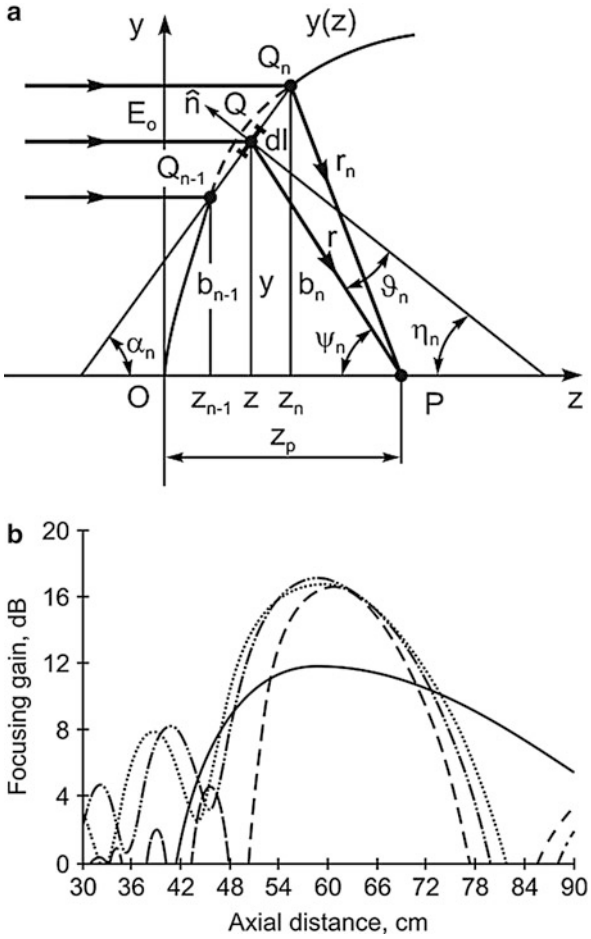


Fig. 18 Axial focusing of plane and curved FZP lenses: (**a**) geometry for derivation of general focusing theory, and (**b**) focusing gain versus axial distance from the lens for planar (*solid line*), spherical (*dotted line*), parabolic (*dashed-dotted line*) and conical (*dashed line*) zone plates

According to Fig. 18a, the total focused field at the focal point P is created by the odd open zones, or

$$E(P) = \sum_{n-1}^{N} E_n(P), \tag{27}$$

with $n = 1, 3, 5, \ldots, N$

The zone plate focusing gain G_f in decibels is found by the equation $G_f = 10\log|E_f(P)/E_0(P)|^2$, where $E_f(P)$ is the FZP focused electric field and $E_0(P)$ is the plane-wave electric field at P if the FZP lens is not present.

The profile functions of three studied curved zone plates are: $y(z) = \sqrt{R^2 - (R - z)^2}$ for a spherical zone plate, $y(z) = \sqrt{4f_P z}$ for a parabolic zone plate, where f_P is the parabolic surface focal length, and $y(z) = z\tan\alpha$ for a conical zone plate, Fig. 17b, with $\alpha = \alpha_n = \text{const}$. The flat zone plate is defined as a limiting case of the conical zone plate, or for $\alpha \to \pi/2$.

The above equations have been applied for numerical computations and comparison between the focusing properties of the planar, spherical, parabolic, and conical convex zone plates. All zone plates have the same focal length ($z_p = F = 60$ cm) and aperture diameter $D = 2b_N = 60$ cm. The spherical surface radius R is also set to 60 cm, the parabolic surface has a focal length $f_p = 15$ cm and an opening

DOI 10.1007/978-981-4560-75-7_42-2

conical half-angle is $\alpha = \pi/4$. The plane zone plate is defined for $\alpha = \pi/4 - 10^{-3}$, all lenses are designed for the wavelength $\lambda = 3.2$ cm. For the listed dimensions and design wavelength the binary flat and curved FZP lenses have two and three open zones, respectively. From Fig. 18b is concluded that for the same in size apertures and focal lengths, the focusing superiority of the curved zone plates over the plane zone plate is well expressed. On the other hand, the focusing gains of the curved FZP lenses do not differ considerably one from another, while the conical zone plate lens displays the best focusing resolution.

Full Electromagnetic Simulation Methods For the binary FZP with noncircular outline (square or hexagonal) or more complex phase-correcting dielectric/metal-dielectric zone plates and antennas, the more precise full-wave electromagnetic methods are preferable. These are mainly the integral equation method (IEM), method of moments (MoM), finite-element method (FEM), finite-difference time-domain method (FDTD) and its varieties. The full electromagnetic methods give very realistic simulation outcomes for any FZP lens focal number F/D, while the KDI theory is an appropriate tool only for the FZP lenses with F/D equal or bigger than about 0.5–1.0.

The body-of-revolution (BOR)-FDTD method was applied recently in Reid and Smith (2006, 2007) for a full electromagnetic analysis of binary or grooved-dielectric microwave and millimeter wave FZP lenses and antennas. In the cited papers precise sets of design and optimization graphs for the FZP lenses and antennas are given.

In Fig. 19a, b, two sets of design graphs for the binary and quarter-wave FZP lens antennas are illustrated. From these graphs is easy found that for $F/\lambda = 40$ and $D/\lambda = 60$, the focusing gain FG of the binary FZP lens is 26 dB. dB, while of the quarter-wave FZP lens is 32 dB.

Flat FZP Lens Antennas

Diffraction Radiation Theory In this section the radiation theory of the flat FZP antennas is based on the vector Kirchoff's diffraction integral (Jackson 1975; Black and Wiltse 1987; Leiten and Herben 1992; Baggen and Herben 1993; Van Houten and Herben 1994; Hristov and Herben 1995; Hristov 2000). The diffraction integral is applied here to the lens antennas based on the simplest doublet of microwave half-wave FZP lenses: binary FZP lens, Fig. 6a, and phase-reversing FZP lens, Fig. 6e. The binary FZP lens consists of coaxial thin metal rings, while the phase reversal FZP is an assembly of solid dielectric phase-shifting rings of permittivity $\varepsilon = 4$, loss tangent $\tan\delta = 0.001$, and ring thickness $w = \lambda/2$. Both FZP lens antennas, illustrated in Fig. 20a, b, respectively, are illuminated by equal round corrugated feed-horns with phase centers set on the z-axis at a focal point $z = -F$. Here, by F is labeled the FZP lens antenna primary focal length.

Each FZP antenna has rotational symmetry around the z-axis and radiation aperture centered in the xOy plane. The odd half-wave (Fresnel) zones are left open (positive lens) and even ones are covered by metal/dielectric rings. The zone radii $b_1, b_2, b_3, \ldots$, are calculated by Eq. 1. The horn radiation pattern is modeled by the cosine scalar equation $G_f(\psi) = 2(q+1)\cos^q\psi$ (Silver 1984). The feed radiation pattern is shaped by changing the parameter q depending on the lens aperture edge illumination (EIL) taper, which is typically from -8 to -12 dB down to the center field maximum.

From structural and theoretical points of view the phase reversal dielectric FZP antenna is a more complex configuration. The theory of the binary FZP antenna is regarded as a particular case of the phase reversal FZP antenna theory.

The incident free-space ray $\rho(\psi)$ associated with the feed-horn spherical wave propagates through the phase-shifting dielectric ring as a refraction ray. The ray tracing through this phase reversal FZP lens, containing solid-dielectric and air-transparent rings, is illustrated in Fig. 20b.

DOI 10.1007/978-981-4560-75-7_42-2

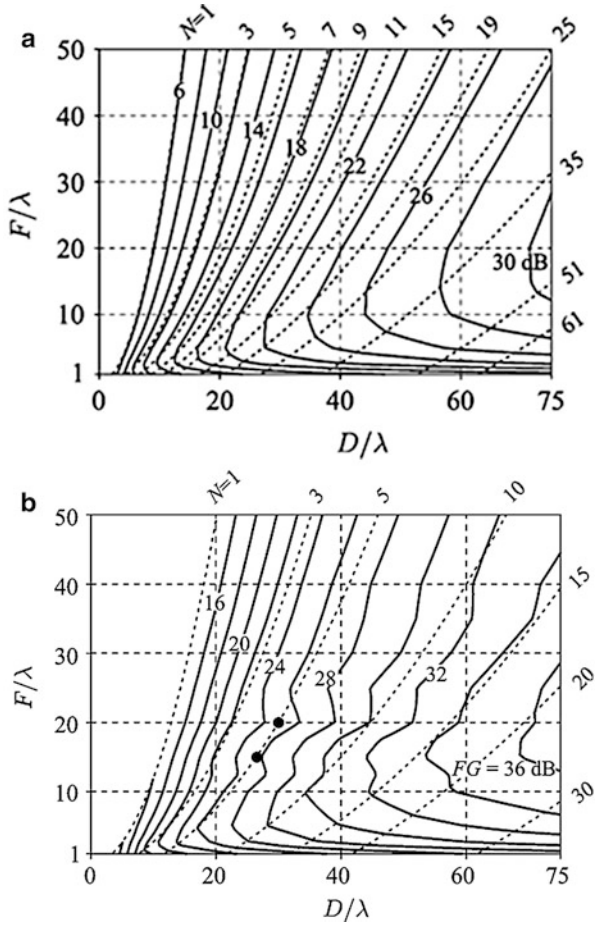


Fig. 19 Design graphs for (**a**) binary and (**b**) quarter-wave phase-correcting FZP lenses. *Solid lines* are contours of constant focusing gain FG, in decibels, *dashed lines* are for constant N

The refraction through the dielectric ring of a linearly polarized wave is characterized by multiple complex transmission coefficients: $T^{||} = \left|T^{||}\right|\exp\left(j\Phi_t^{||}\right)$ (for a parallel or magnetic polarization) and $T^{\perp} = |T^{\perp}|\exp\left(j\Phi_t^{\perp}\right)$ (for a perpendicular or electric polarization). In the lens output (aperture) plane $II - II'$ the refracted ray gives rise to an electric field $\mathbf{E}_d(\psi,\ \xi)$, which can be expressed as

$$\mathbf{E}_d(\psi,\ \xi) = C_f\sqrt{G_f(\psi)}\frac{\exp(-j\beta L(\psi))}{F + w}\cos\psi\ \mathbf{p}'_d(\psi,\ \xi) \tag{28}$$

where $\mathbf{p}'_d(\psi,\ \xi) = s_d\left(-T^{||}\cos\xi\hat{\mathbf{e}}_\psi + T^{\perp}\sin\xi\hat{\mathbf{e}}_\xi\right)$, $C_f = \sqrt{P_t Z_0/2\pi}$, and $L(\psi) = F/\cos\psi + \varepsilon w/\sqrt{\varepsilon - \sin^2\psi}$.

The multiple transmission coefficient $T^{||,\perp}$ of a dielectric ring is supposed equal to that of an infinite-in-extent dielectric plate with the same thickness w and is calculated according to Burnside and Burgener (1983).

The far electric field components $E_\theta^{(d)}$ and $E_\varphi^{(d)}$ are found as follows (Van Houten and Herben 1994):

DOI 10.1007/978-981-4560-75-7_42-2

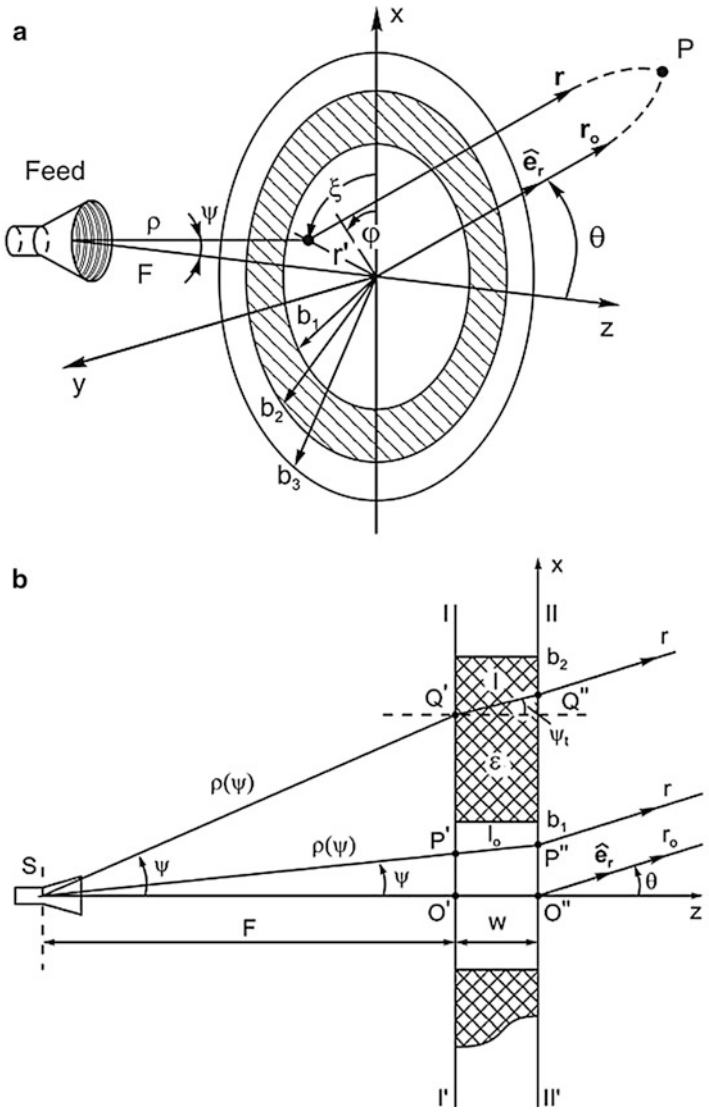


Fig. 20 Radiation geometry of: (**a**) binary and (**b**) phase-reversal FZP lens antennas fed by corrugated circular horns

$$E_{\theta}^{(d)}(\theta,\varphi) = -\pi C \cos\varphi \sum_{n} \int_{\psi_{n-1}}^{\psi_n} \exp(M_d(\psi)) O_d(\psi) I_{\theta}^{(d)}(\theta,\psi) d\psi \tag{29}$$

$$E_{\varphi}^{(d)}(\theta,\varphi) = -\pi C \sin\varphi \cos\theta \sum_{n} \int_{\psi_{n-1}}^{\psi_n} \exp(M_d(\psi)) O_d(\psi) I_{\varphi}^{(d)}(\theta,\psi) d\psi \tag{30}$$

where C, $\psi_{n,n-1}$ and the integral functions $O_d(\psi)$, $M_d(\psi)$, $N_d(\theta,\psi)$, $I_{\theta}^{(d)}(\theta,\psi)$, and $I_{\varphi}^{(d)}(\theta,\psi)$ can be also found in Hristov and Herben (1995), Hristov (2000).

DOI 10.1007/978-981-4560-75-7_42-2

The components $E_\theta^{(o)}(\theta, \varphi)$ and $E_\varphi^{(o)}(\theta, \varphi)$ of far field radiated by the open apertures are easily obtained from Eqs. 29 and 30 having in mind that the ray tracing is through the air rings with $\varepsilon = 1$ and $T^{||} = T^{\perp} = 1$. Thus, the above integral functions are considerably simplified.

The far-field E_θ and E_φ of the phase reversal FZP antenna are sums of the air-ring and dielectric-ring far fields

$$E_\theta(\theta, \varphi) = E_\theta^{(o)}(\theta, \varphi) + E_\theta^{(d)}(\theta, \varphi) \tag{31}$$

$$E_\varphi(\theta, \varphi) = E_\varphi^{(o)}(\theta, \varphi) + E_\varphi^{(d)}(\theta, \varphi) \tag{32}$$

The total far-field electric vector is a sum of the FZP antenna E_θ and E_φ vector components

$$\mathbf{E}(\theta, \varphi) = \hat{\mathbf{e}}_\theta E_\theta(\theta,\varphi) + \hat{\mathbf{e}}_\varphi E_\varphi(\theta, \varphi) \tag{33}$$

and the antenna directive gain pattern is calculated by

$$G(\theta, \varphi) = 10 \log\left(\frac{2\pi r^2}{Z_0 P_t}|\mathbf{E}(\theta, \varphi)|^2\right) \tag{34}$$

The quarter-wave FZP antenna comprising four dielectric rings in a full-wave zone is treated similarly. The numerical comparison between 30-GHz flat binary, phase-reversing, and quarter-wave FZP lens antennas with dimensions $F/\lambda = 15$, $D/\lambda = 18$, $EIL = -10$ dB gives roughly the following antenna radiation efficiencies: 12 %, 30 %, and 50 %. The efficiencies of the phase reversal and quarter-wave FZP lenses are for the following dielectric permittivities: $\varepsilon = [1,\ 4]$ and $\varepsilon = [1,\ 6.25,\ 4,\ 2.25\ ]$, respectively. All dielectric rings have delta tangent equal to 0.001.

Some Numerical Results for Flat FZP Lens Antennas Following the diffraction theory applied to the flat binary and dielectric phase reversal FZP lens antennas some numerical and experimental results for exemplary FZP antenna models are displayed in graph and table forms. The binary FZP antennas are made of positive or negative FZP lenses, which are supposed to have similar focusing characteristics. This opinion is checked numerically on 11.1-GHz flat FZP antennas with a lens aperture diameter $D = 100$ cm and a focal length $F = 200$ cm (or for $\tilde{f} = F/D = 2$). The lenses are illuminated by corrugated feed horns producing edge illumination level $EIL = -11$ dB.

In Fig. 21a are drawn the E-plane directive gain radiation patterns of both binary positive (solid line) and binary negative (dashed line) FZP lens antennas, in comparison with the phase reversal (dotted line) FZP lens antenna (Hristov 2000). It is evident that there is no notable difference between the pattern main lobes. The pattern distinction is bigger in the sidelobe areas. The positive FZP lens antenna with odd open zones has 0.6 dB higher peak gain, while the negative one with even open zones has a narrower beamwidth and lower near sidelobes. The positive zone plate has a central zone open that ensures a minimum reflection from the plate to the feed, better feed matching and somewhat larger frequency bandwidth. In Fig. 21b are contrasted the normalized radiation patterns of phase-reversal (dashed line) and quarter-wave (solid line) FZP lens antennas equal in diameter and focal length.

A combination of the Kirchhoff diffraction theory (for the main-lobe region) with the uniform theory of diffraction (UTD) for the sidelobe regions, leads to much closer simulated and measured radiation patterns (Baggen and Herben 1993; Sluijter et al. 1995). In Fig. 22a, the measured far-field pattern (solid line) is compared to the computed one (dotted line). Directive gain function versus frequency for the

DOI 10.1007/978-981-4560-75-7_42-2

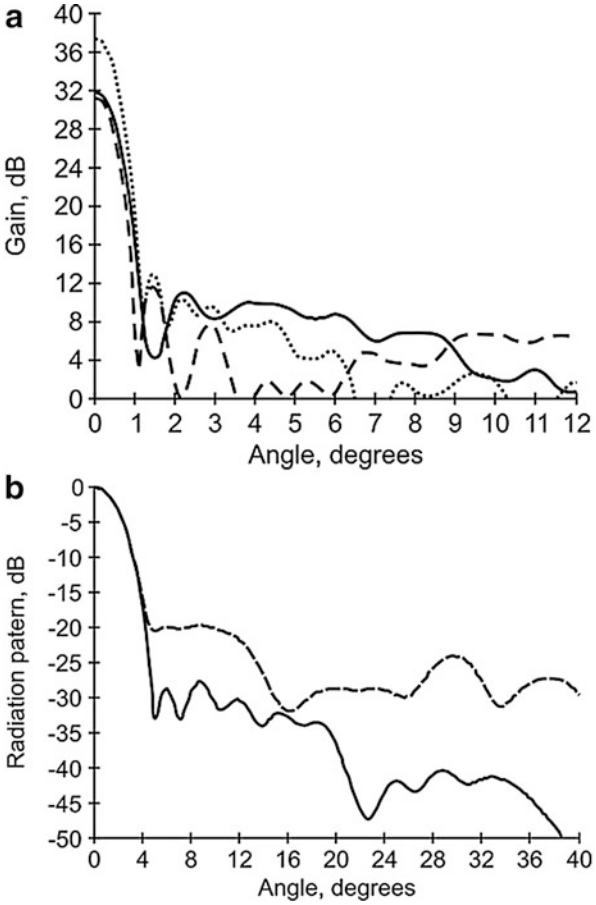


Fig. 21 Directive gain radiation patterns of: (**a**) binary FZP lens antennas, positive FZP (*solid line*), negative FZP (*dashed line*) and phase-reversal FZP lens antenna (*dotted line*), and (**b**) phase-reversal (*dashed line*) and quarter-wave (*solid line*)

same binary FZP lens antenna (D = 1.12 m, F = 0.71 m and design frequency f = 11.5 GHz) is illustrated in Fig. 22b. The theoretical graph is drawn with a solid line and measured data are shown as spread circles.

As is described above in the text, the binary FZP lens holds a multiharmonic focusing behavior. Similar actions hold also the binary, phase-reversing, and other phase-correcting FZP lens antennas. In Fig. 23 is demonstrated the harmonic gain and efficiency behavior of the phase-reversing (solid line) and quarter-wave (dashed line) FZP lens antennas, designed for a frequency f = 30 GHz, focal length F = 264 mm, aperture diameter D = 301.5 mm, and edge illumination level EIL = −10 dB (Hristov 2000). The gain curves are drawn in Fig. 23a and the aperture efficiency curves are shown in Fig. 23b, both in the range 10–180 GHz. The phase-reversing FZP lens antenna has almost constant directive gain of about 37.5 dB but very different efficiency maxima at 30 GHz, 90 GHz, 150 GHz, etc., or similarly to the binary FZP lens antenna. It is evident that the phase-reversing FZP lens antenna behaves as a frequency-selective antenna with a multiple-frequency ratio 1:3:9, ... The antenna aperture efficiency is 38.2 % at the design frequency of 30 GHz and quickly goes down for the harmonic frequencies of 90 GHz, 150 GHz, and so on.

Rather different is the frequency performance of quarter-wave FZP lens antenna with first three periodic gain maxima of about 41 dB taking place at 30 GHz, 60 GHz, and 90 GHz, or according to the multiple-frequency ratio 1:2:3. . . The antenna has a satisfactory efficiency of about 60 % at 30 GHz and 37.3 % at 60 GHz.

DOI 10.1007/978-981-4560-75-7_42-2

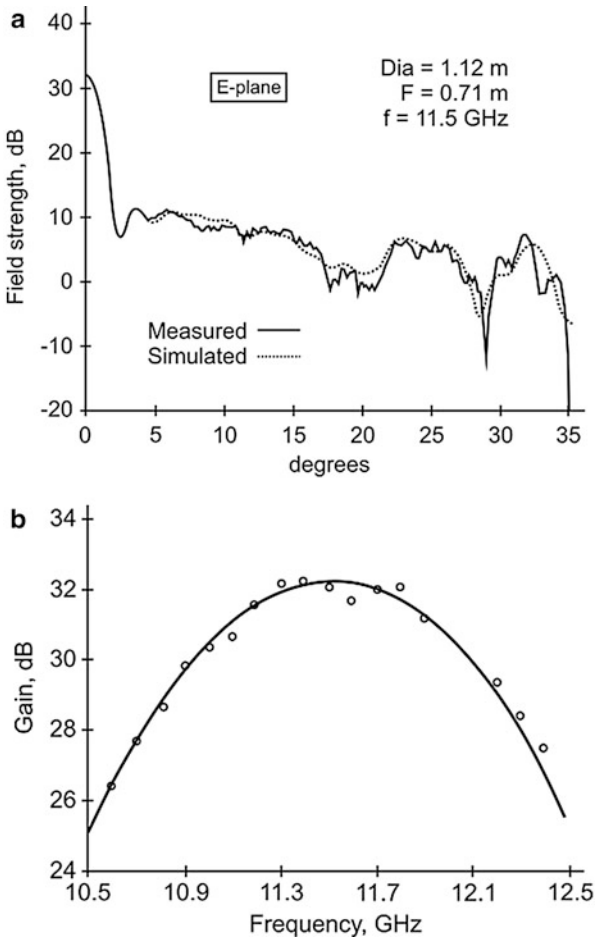


Fig. 22 Radiation field patterns of binary FZP lens antenna: (**a**) E-plane gain radiation pattern, theoretical (*dotted line*) and measured (*solid line*) and (**b**) Gain versus frequency, theoretical (*solid line*) and measured (*circles*)

Some Early FZP Lens Antennas and Applications One of the first binary microwave FZP antenna is published in Van Buskirk and Hendrix (1961). These authors built and measured X-band FZP antennas comprising binary FZP lenses with one, two, or three open half-wave zones. Two types of feeds, a half-wave dipole and turnstile, have been explored. Fig. 24 shows the full 360-degree experimental radiation pattern of the FZP antenna with a binary lens of three open zones. This radiation pattern is of great interest because clearly shows the forward (transmission peak) and backward (reflection peak) pattern areas. Both transmission (front) and reflection (back) beams have almost the same peak levels and 3-dB beamwidth of about 4 deg, which is nearly identical to the beamwidth of parabolic antenna with the same aperture diameter. As expected, the aperture radiation efficiency of the binary FZP antenna is very low, only about 10 % versus 55–70 % for the typical parabolic antenna.

A polyethylene phase reversal zone plate lens fed by a rectangular horn is integrated in a 33-GHz simple and low-cost homodyne Doppler radar, Fig. 25 (Lazarus et al. 1979). The zone plate had the following dimensions: $w = 9$ mm, $F = 180$ mm, and $D = 300$ mm ($\tilde{f} = F/D = 0.6$).

A 21-inch-diameter single-dielectric quarter-wave FZP lens antenna is utilized in a millimeter-wave radio transceiver, Fig. 26 (Norden Systems 1984; Thornton and Strozyk 1983). At 37.8 GHz the transceiver FZP antenna has a gain of 41.5 dB, beamwidth of 1.1 deg, and sidelobes less than −24 dB.

DOI 10.1007/978-981-4560-75-7_42-2

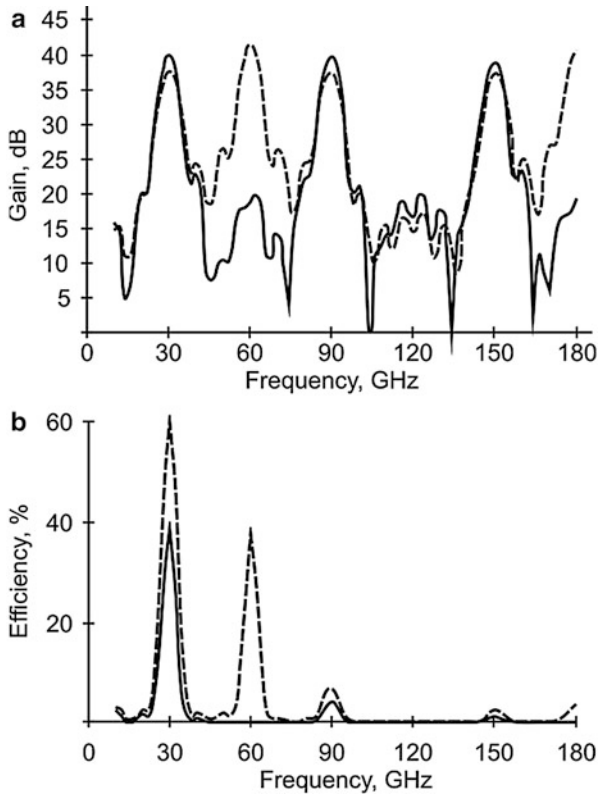


Fig. 23 Gain and efficiency of FZP lens antennas: (**a**) directive gain and (**b**) radiation efficiency of phase-reversal (*solid line*) and quarter-wave (*dashed line*) versus frequency

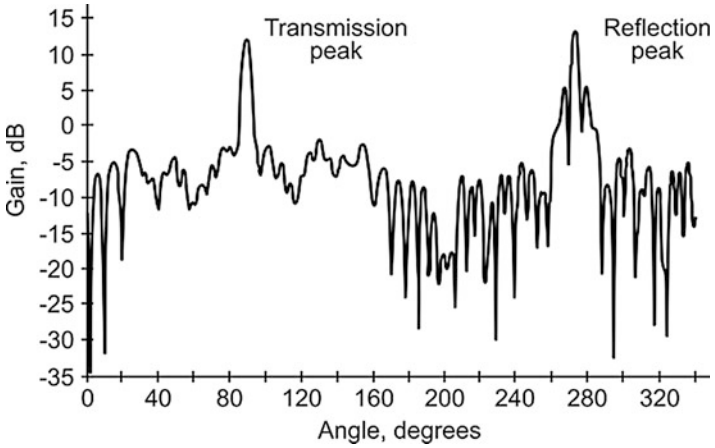


Fig. 24 Experimental radiation pattern of binary FZP antenna in angular sector 0–360 deg

Off-Axis Scanning in Flat Zone Plate Lens Antenna with Circular Fresnel Zones Scanning properties of the folded zone plate and antenna were initially discussed in Van Buskirk and Hendrix (1961), where is supposed that the maximum scan (off-axis) angle should not exceed ±20 deg. The first application of FZP antenna for scanned satellite TV reception is described in Shuter et al. (1984). The antenna main lobe was pointed to the satellite simply by scanning the feed-horn ±15 degrees off axis with a signal loss not more than 1 dB.

DOI 10.1007/978-981-4560-75-7_42-2

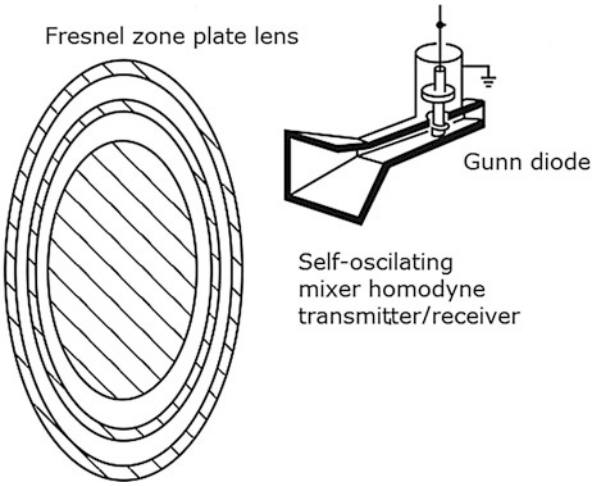


Fig. 25 Doppler radar antenna with phase-reversing FZP lens antenna

Fig. 26 Millimeter-wave radio transceiver with quarter-wave FZP lens antenna

A comprehensive theoretical study comparing the scan performances of a binary round-zone FZP antenna and a parabolic dish antenna is completed in Baggen et al. (1993). The main beam scanning is illustrated in Fig. 27. There, by displacing the feed instead of rotating the whole antenna, TV signals from different satellites are received. It is proved that the binary FZP lens with circular zones has nearly circular scan curve, while the scan curve of the parabolic reflector is the so-called Petzval line. From this follows that the main lobe scan angle of the FZP antenna is somewhat bigger than that of the parabolic antenna.

Flat Offset Antenna with Elliptical Zone Plate For realizing a more efficient off-axis scanning in a larger angular sector the circular zone plate should be redesigned to an elliptical zone plate.

A binary elliptical zone plate made of metal (foil) zone rings and attached to an office window or wall is illustrated in Fig. 28 (Mawzones Ltd Leaflet 1992). The FZP antenna consisting of attached zone plate and a feed-horn could be applied, for example, as a terminal device in a direct communication satellite link.

Beam-Scanning in Binary FZP Antenna Based on Photoconductivity A binary FZP antenna lens studied in Hajian et al. (2003) includes a semiconductor wafer (silicon or gallium arsenide) in which a

DOI 10.1007/978-981-4560-75-7_42-2

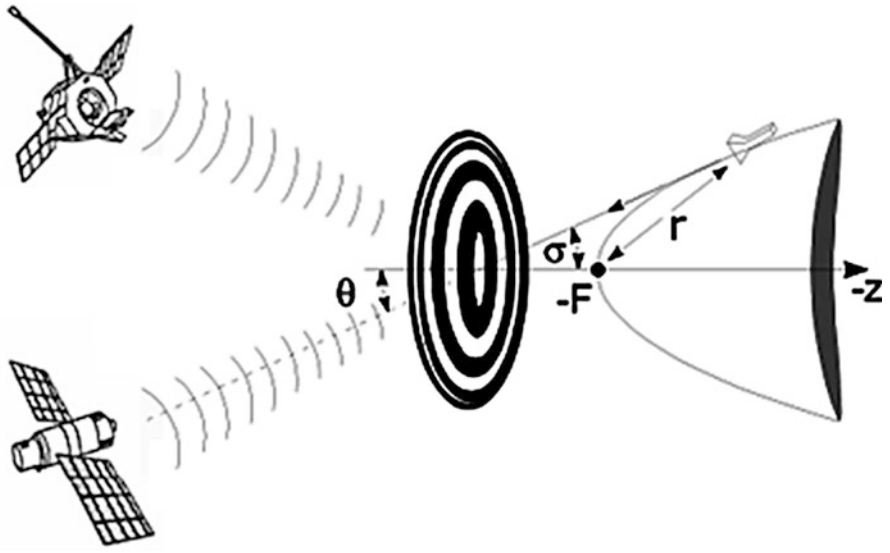


Fig. 27 Defocused Fresnel zone plate antenna for satellite scan purposes

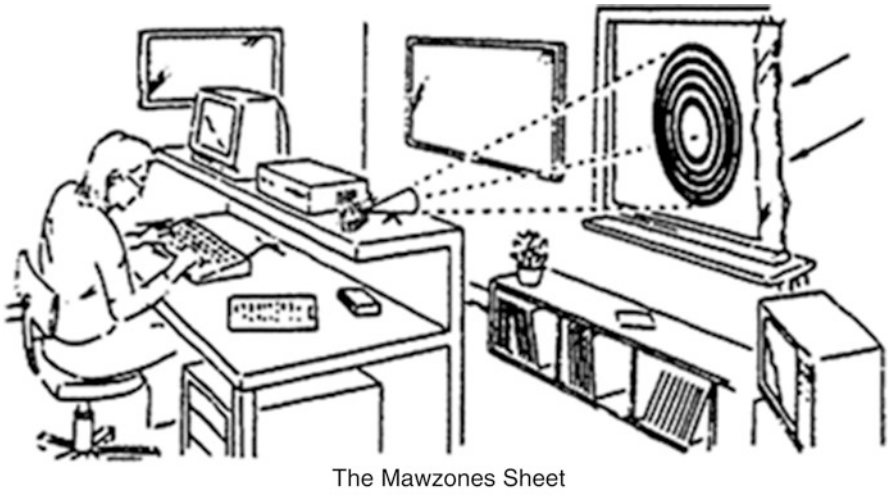


Fig. 28 Attached to window off-set binary Fresnel zone plate as an element of computer satellite link antenna

spatially variable density of the charge carriers is realized through a selective optical (laser) illumination. It is proved theoretically that the beam of such a circular-zone FZP antenna can be scanned by reconfiguring the wafer masking simply by changing the wafer conductivity in the zones by a variable laser illumination.

Low-Profile Fresnel Zone Plate Antenna Arrays

The microwave and millimeter wave FZP lenses are (i) high-profile (or long-focal) for a focal factor $\tilde{f} = F/D \geq 0.5 - 1.0$ or (ii) low-profile (short-focal) if $\tilde{f} = F/D \geq 0.25 - 0.5$. The low-profile lens antennas have a small overall volume of the combined lens and feed, and are suitable for effective IC (integrated circuit) arraying.

DOI 10.1007/978-981-4560-75-7_42-2

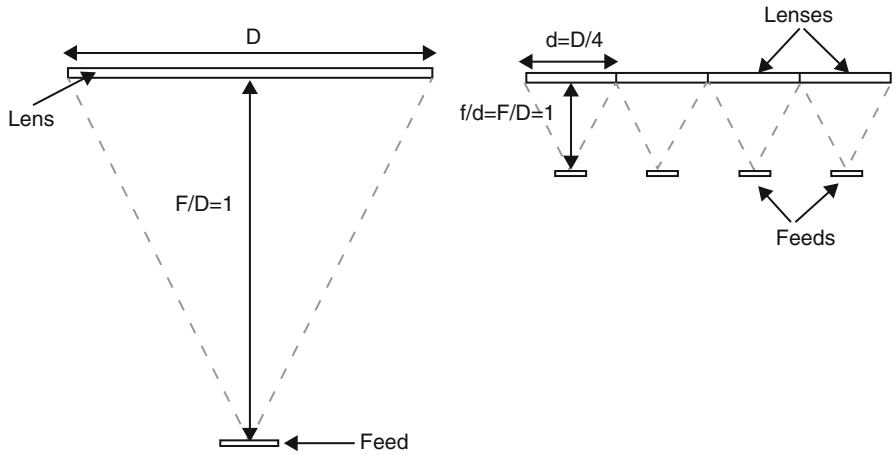


Fig. 29 Profile comparison between big single FZP lens antenna (*left*) and array of small FZP lens antennas (*right*)

A concept for the FZP lens antenna depth reduction involves replacing a single large-diameter lens with an array of small-diameter FZP lens antennas 2006 ili 2004??? Koi publkatcii sa tuk? This concept is illustrated in Fig. 29.

Let the single big FZP lens has a focal factor $\tilde{f} = F/D = 1$, and the same ratio $f/d = 1$ is fixed for the linear array of small FZP lenses. The array lens diameter is $d = D/x$, where $x = 4$ is the number of array lens elements. Thus, the antenna depth reduction of the lens antenna is made possible because each array element diameter is a portion of the big antenna diameter or d=D/4.

The research and applications of FZP antenna arrays are still in their initial stage. Almost all new original ideas and results related to this topic along with many others like the aperture shapes of FZP antenna array elements, FZP reference phase, FZP subwavelength resolution, etc. can be found in the recently published book (Minin and Minin 2008a).

Flat FZP Reflector Antennas

Folded (Reflector) FZP Antenna: Initial Arrangement The antenna based on the binary FZP lens of metal rings is a simple but not effective lens. It has two principal foci: one for transmission through the open zones and another for reflection by the metal zone rings. By placing a plane quarter-wavelength reflector behind the metal zone plate the incident plane wave transmitted through the open zones is turned back and interferes constructively at the reflection focus P_1, Fig. 30a (Van Buskirk and Hendrix 1961). This folded binary FZP lens acts like a phase reversal zone plate antenna. Fig. 30b is an artistic view of the Van Buskirk's and Hendrix's feasible construction of a simple and cheap radio telescope based on the FZP reflector.

By simple changes Eqs. 29 and 30 for the far-field components are transformed to similar equations for the folded zone plate antenna, as follows:

$$E_\theta(\theta,\varphi) = -\pi C\cos\varphi \sum_{n=1}^{N} (-1)^n \int_{\psi_{n-1}}^{\psi_n} O(\psi) e^{M(\psi)} I_\theta(\theta,\psi) d\psi \tag{35}$$

$$E_\varphi(\theta,\varphi) = -\pi C\cos\varphi\cos\theta \sum_{n=1}^{N} (-1)^n \int_{\psi_{n-1}}^{\psi_n} O(\psi) e^{M(\psi)} I_\varphi(\theta,\psi) d\psi \tag{36}$$

DOI 10.1007/978-981-4560-75-7_42-2

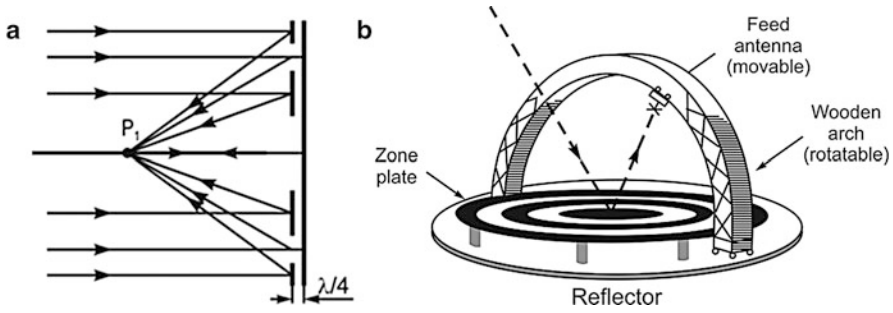


Fig. 30 Reflector (folded) zone plate antenna: (**a**) operation principle and (**b**) artistic view of FZP reflector radio telescope based on folded zone plate antenna

with n = 1, 2, 3,. . . N.

FZP Antennas with Machined and Printed Reflectors A typical FZP reflector antenna consisting of precisely machined grooved-dielectric zone plate with a disk reflector and a feed-horn is drawn in Fig. 31a.

Most likely, the first printed zone plate reflector antenna is described by Huder and Menzel (1988). This is a single-layer 94-GHz folded zone plate antenna manufactured by means of a standard microstrip technology, comprising a ground plane 125 mm in diameter and Fresnel-zone metal rings spaced from the ground plane by a dielectric substrate 0.508 mm in depth. The antenna is fed by an open-ended waveguide WR-10 with an aperture placed at a focal distance of 80 mm. The measured antenna parameters for an amplitude edge taper of −10 dB are: directive gain of 35 dB, 3-dB beamwidth of 1.6 deg, sidelobe levels less than −19 dB, cross-polar level less than −25 dB, and mismatch loss equal to 15dB.

A printed three-layer zone plate reflector is illustrated in Fig. 31b (Guo and Barton 1991, 1992, 2002). Shown above is the reflector front view, and below is drawn its cross-sectional view. More exactly, the plate consists of three layers metal rings separated by three dielectric substrates. On the basis of this quarter-wave zone plate a reflector antenna has been developed, fed by a polyrod feed that yields about −11 dB EIL level. The maximum antenna gain of 34.75 dB occurs at 11.8 GHz, and a 3-dB frequency bandwidth of about 13 % is estimated from the measured gain versus frequency curve.

Figure 32a illustrates a printed resonant-ring structure of quarter-wave FZP reflector (Guo 1993). The first three subzones are covered by arrays of different in diameter rings and the last one is left blank. A reflector FZP prototype has a diameter of 594 mm and focal length of 475.2 mm. The FZP reflector fed by a helical antenna forms an 11.4-GHz reflector antenna with a circular polarization, gain of 33.4 dB, radiation efficiency of 43 %, and frequency bandwidth of about 10 %.

Figure 32b shows a 9.375-GHz folded zone plate fed by a monofiliar backfire helical antenna (BHA) with the following design parameters: pitch angle of 26 deg, helical circumference C = 0.83 λ, number of turns N = 6, and small reflector diameter of 7.9 mm (Yamauchi et al. 1990). The backfire helix plus FZP reflector antenna has a focal length of 40 mm, diameter of 260 mm, and the next measured parameters: a peak gain of 21 dB at 9.375 GHz and a gain greater than 20 dB in the frequency range from 9 to 9.8 GHz. The axial polarization ratio is less than 1.5 and VSWR is less than 1.6 over the same frequency range.

A FZP-FSS compound lens, Fig. 33a, consisting of a binary Fresnel zone plate (FZP) and frequency-selective screen (FSS) was proposed and studied recently (Fan et al. 2010). The FZP has eight circular zones totally, four open and four closed, and the FSS is a square array of 40 × 40 four-leg loaded elements, Fig. 33b, cut in a thin metal sheet. The FZP-FSS lens is designed for a paraxial plane-wave illumination at the frequency of 12 GHz (λ = 2.5 cm), with a focal length F = 15 cm. The compound lens square side is 16λ = 40 cm. Both lenses, FZP and FZP-FSS, are simulated and analyzed numerically

DOI 10.1007/978-981-4560-75-7_42-2

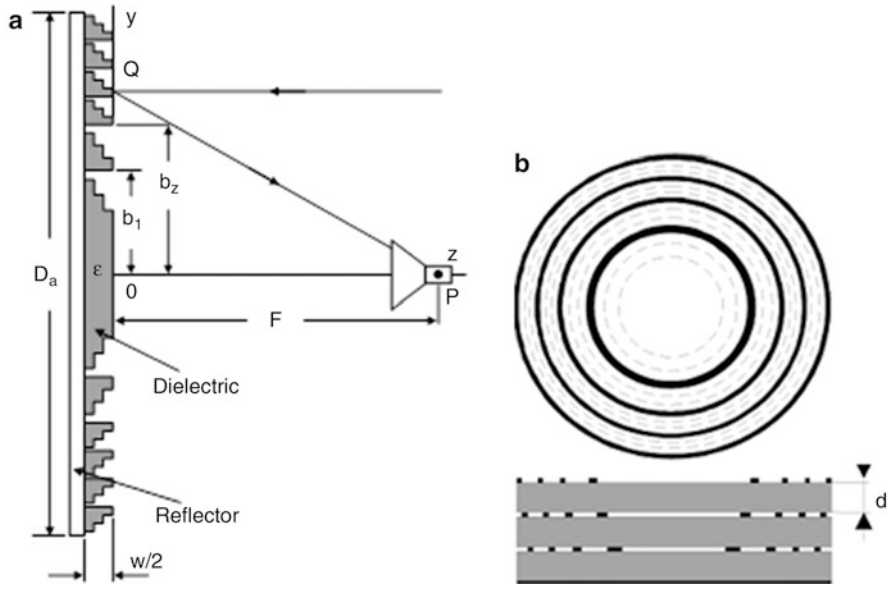


Fig. 31 Design drafts of: (**a**) quarter-wave grooved-dielectric FZP reflector antenna and (**b**) three-layer FZP reflector

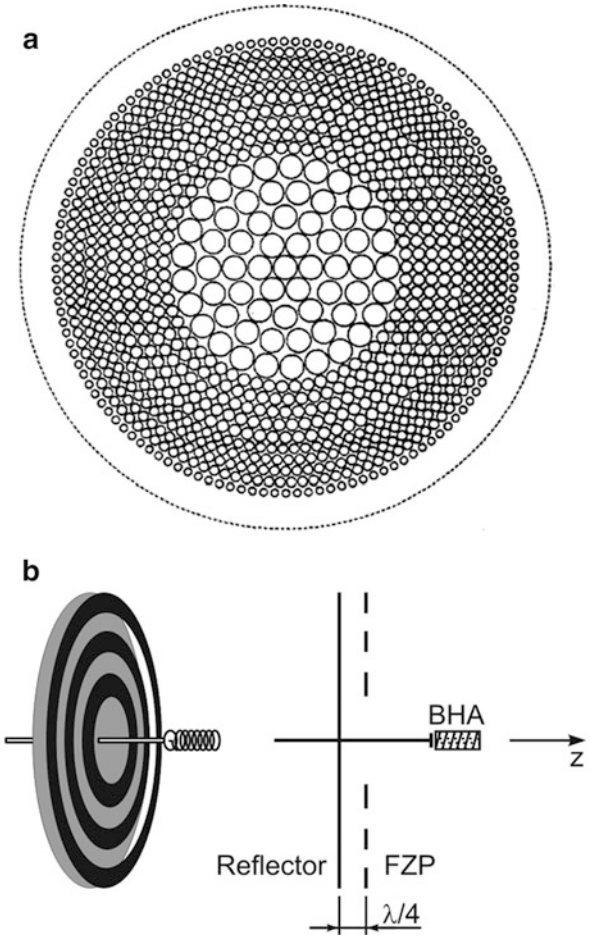


Fig. 32 FZP reflector antennas: (**a**) printed ring arrangement in quarter-wave FZP reflector and (**b**) folded zone plate fed by monofiliar backfire helical antenna (BHA)

DOI 10.1007/978-981-4560-75-7_42-2

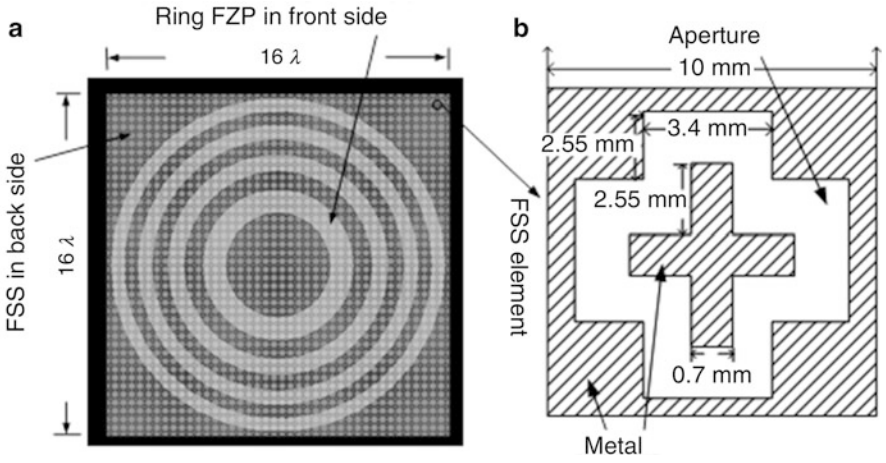


Fig. 33 FZP-FSS design illustrations: (**a**) FZP-FSS computer model, (**b**) geometry of the FSS four-leg cross element

by means of a specially developed hybrid PSTD-FDTD algorithm and software. The PSTD-FDTD results are contrasted with those obtained by lens prototype measurements, Fig. 34a. Some attractive focusing and spectral properties of the FZP-FSS lens compared to the same-size FZP lens have been found: (i) frequency filtering enhancement, (ii) about 2 dB increase in the peak focusing intensity, and (iii) more than 4 dB reduction in the first off-axis maximum.

The compound FZP-FSS lens has somewhat better 3dB focal resolution δ equal to 5.5 deg versus 5.8 deg for the single FZP lens. As seen in Fig. 34b the intensity of FZP-FSS lens at focal point ($Z = F$, $X = Y = 0$) is 19.1 dB.

Offset Reflector FZP Antenna Designs An effective X-band four-layer offset antenna is reported in Guo et al. (1994). Fig. 35a shows the front and sectional view of the antenna FZP reflector, which produces five phase shifts ($Q = 5$) in each full-wave zone: 0°, 72°, 144°, 216°, and 288°. The substrate layer depth d is calculated by $d = \lambda\sqrt{1 - \sin^2\theta_{off}/\varepsilon}/Q\sqrt{\varepsilon}$, where ε is the relative permittivity of each dielectric substrate and θ_{off} is the offset angle. Measured prototype of this offset FZP antenna is designed for 10.39 GHz and has a reflector 320 × 320 mm^2 in size and 20-degree offset angle. Its focal length is 190 mm. The antenna is fed by a rectangular horn with aperture dimensions 41 mm and 28 mm. The measured antenna radiation efficiency is 61 %, which is a very good value for such a complicated design.

A simple 1.5 × 1.5 m^2 offset antenna with a phase-reversing zone plate outside antenna is described in Sazonov (1999). The antenna receives 12.5-GHz DBS TV signals with a similar quality as a standard 1.2-m parabolic antenna.

Curvilinear FZP Antennas

Axially Symmetric FZP Antennas The binary 2D or 3D curvilinear (3D) Fresnel zone plate lens can be formed on an arbitrary-shaped surface, but the zone plate lens with a rotational silhouette is most simple and efficient. The spherical, cylindrical, and conical FZP lenses are next in simplicity to the flat Fresnel zone plate. The 3D or curvilinear FZP lens has more degrees of freedom for parametric optimization compared to the flat FZP lens. For example, in the axially symmetric conical lens a third dimension is the lens length (thickness) t, which is related to the cone semi-angle α.

DOI 10.1007/978-981-4560-75-7_42-2

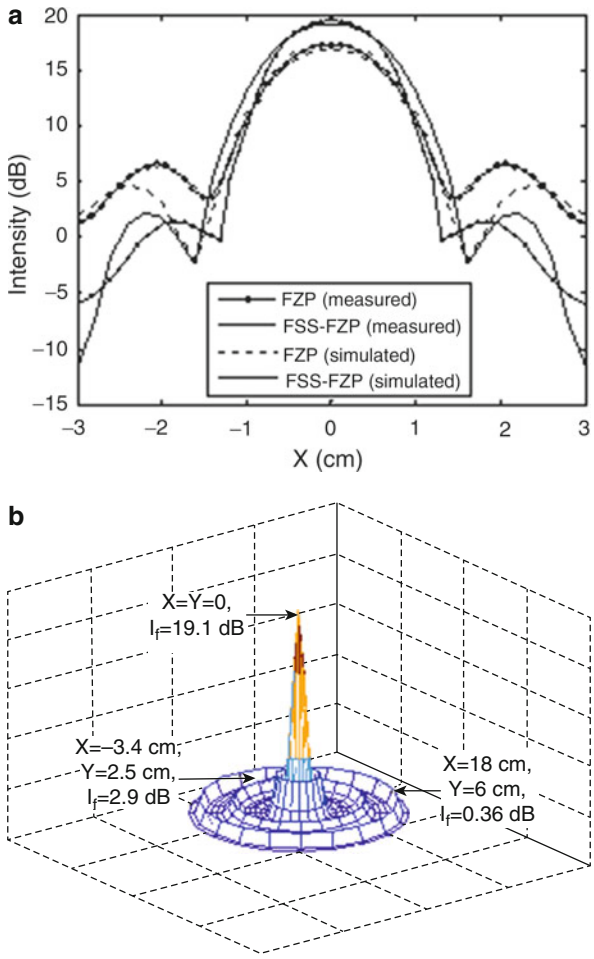


Fig. 34 Intensity distribution: (**a**) of FZP and FZP-FSS lenses along *X*-axis, (simulated and measured) and (**b**) 3D graph of intensity versus axes X and Y. Both intensity figures are valid for the transverse focal plane Z = F

Typically, the known curvilinear dielectric FZP configurations consist of phase-correcting elements (ribs/corrugations) that follow the FZ surface curvature and are difficult to produce especially at millimeter wave and terahertz frequencies.

First most comprehensive analytical studies of the curvilinear FZP lenses and antennas go back to the early 1970s, when K. Dey and P. Khastgir developed the diffraction theory of microwave curvilinear zone plate antennas. Rigorous theoretical investigations of the spherical and parabolic zone plate antennas has been made on the basis of the Kirchhoff diffraction integral (Dey and Khastgir 1973a, b, c; Khastgir et al. 1973). It has been shown that for a given antenna aperture the microwave spherical and paraboloidal zone plate lenses/antennas are superior to the flat zone plate lenses/antennas in two aspects: (i) greater number of zones can be constructed, and in result, (ii) bigger focusing gain and resolution can be attained along the zone plate axis. The problem of determining the off-axis defocus of curvilinear zone plate in the specific case of paraboloid of revolution has been analyzed later on in Khastgir and Bhomwmick (1978).

In order to obtain exact solutions Dey and coauthors have completed their theoretical studies for each zone plate curvature separately. An approximated universal solution of the Kirchhoff diffraction integral (KDI) for any lens profile shape is presented in Kamburov et al. (2005). This KDI theory is valid for a half-wave axially symmetric FZP antenna and is based on the conical-segment lens profile approximation. Vector equation for the electric far field is derived from which follow expressions for the antenna co- and

DOI 10.1007/978-981-4560-75-7_42-2

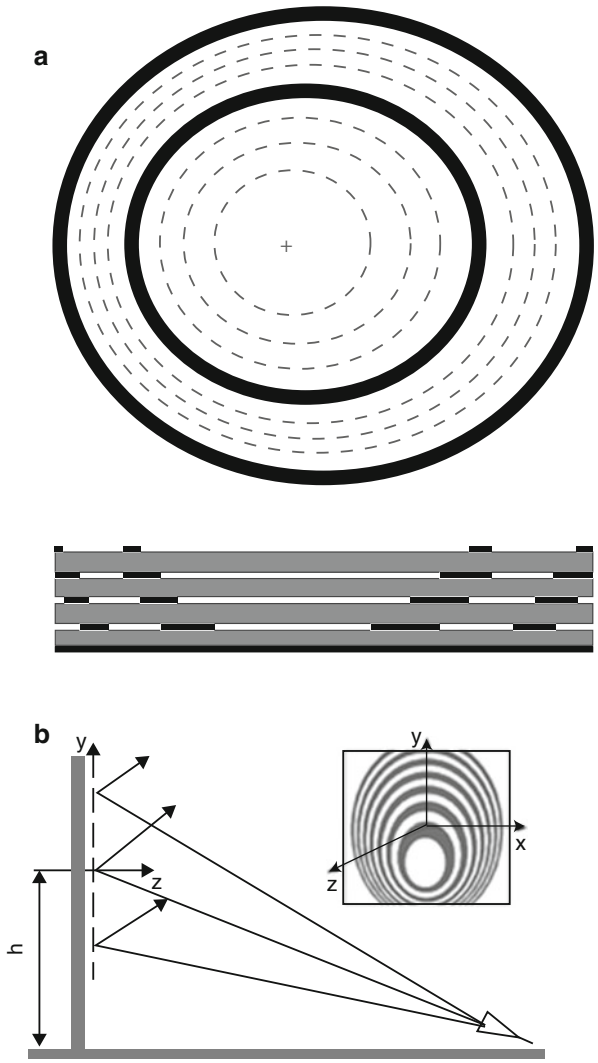


Fig. 35 Offset zone plate reflector antennas: (**a**) printed four-layer offset reflector and (**b**) offset FZP reflector antenna with elliptic zones

cross-polar radiation patterns, directive gain, and radiation efficiency. The proposed theory is utilized for the numerical analysis and comparison of 140-GHz curvilinear half-open FZP lens antennas with flat, conical, parabolic, and spherical apertures with the same diameter.

Figure 36a shows the geometry of the curvilinear half-open FZP lens antenna with a feed located at the lens primary focus P_1 (Hristov et al. 2005). The curvilinear FZP lens is assumed with an axially symmetric surface. The focal length of the lens is $F = F_a + t$. It is concave-side lens illuminated by a spherical wave from the feed at P_1. The m^{th} and $(m-1)^{th}$ zone radii b_m and b_{m-1} and are calculated by Eq. 24. Figure 36b shows the numerical copolar radiation patterns of two FZP lens antennas designed for a frequency of 2 GHz (Hristov and Feick 2001). Each antenna is fed by a scalar horn. The first antenna lens is a big flat zone plate, 30 m in diameter. The second FZP lens is a semispherical plate (shell) with the same aperture diameter. The flat and spherical FZP lenses also have equal radii and focal lengths or $R = F = 15$ m.

Table 1 displays the main parameters of both antennas. Due to the surface curvature the semispherical FZP lens has 1.65 times bigger number of zones than the flat FZP lens. For achievement of one and the same field edge illumination of −10 dB, the feed gains of both antennas should be different. From Table 1

DOI 10.1007/978-981-4560-75-7_42-2

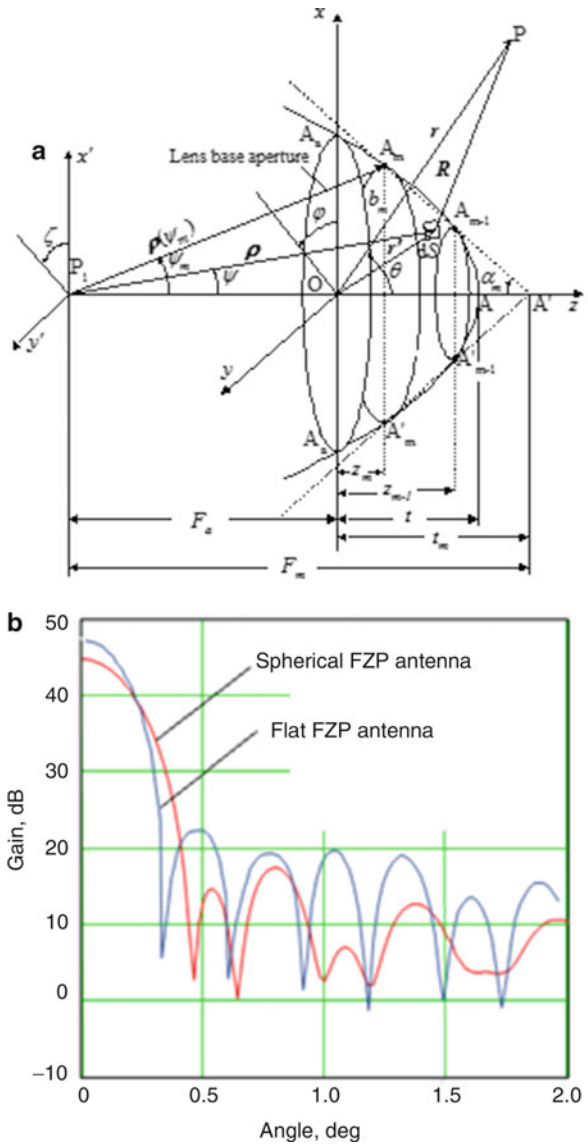


Fig. 36 Far-field of FZP lens antenna: geometry of curvilinear of flat antenna surface and (**b**) gain radiation patterns of flat (*blue line*) and spherical (*red line*) binary FZP lens antennas

Table 1 Comparative table for flat and semi-spherical FZP lens antennas

Parameters/FZP antennas	Number of zones	Feed gain (dB)	Antenna gain (dB)	Beamwidth (deg)	Antenna efficiency (%)
Flat FZPlens antenna	82	10.6	44.8	0.33	7.7
Semi-spherical FZP lens antenna	135	4.3	47.7	0.28	15.2

it is concluded that for the same-size aperture the curvilinear FZP lens antenna has much bigger gain and efficiency, and slightly narrower main beam compared to the corresponding flat FZP lens antenna.

A parabolic FZP antenna comprising metal zone-plate rings in front of a solid parabolic reflector was proposed, fabricated, and reported in Delmas et al. (1993). More exactly, the parabolic zone plate reflector antenna is designed and built to have two symmetric diffraction focal points P_1 and P_2 with an angular

DOI 10.1007/978-981-4560-75-7_42-2

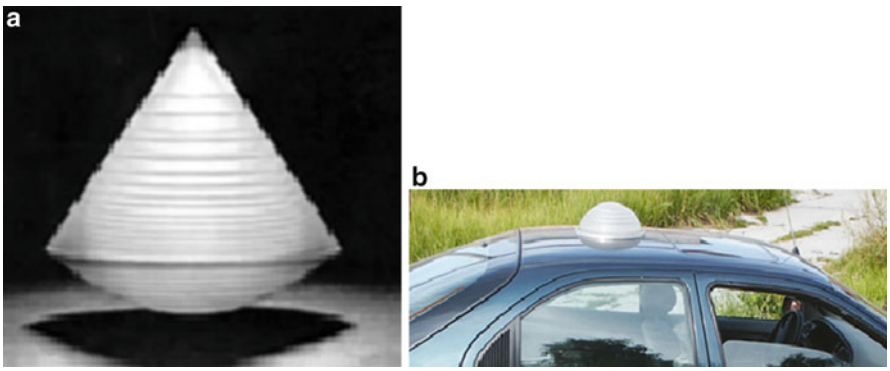


Fig. 37 Curvilinear FZP prototypes: (**a**) composite curvilinear FZP lens antenna and (**b**) car-roof top semi-spherical FZP lens antenna for satellite TV signal reception

Table 2 Characteristics of parabolic dish and FZP antennas

Characteristics	**Parabolic (dish) antenna**	**FZP (diffractive) antenna**
Optic schematics	Reflector antenna	Lens or reflector antenna
Feed blockage	Yes	Yes or no
Surface shape	Fixed (parabolic)	Arbitrary
Building material	Metal or carbon	Lowloss dielectric (plastic)
Precision in machining	$\pm\lambda/32$	in the range $\pm\lambda/5 - \pm\lambda/10$
Satellite pointing	By whole antenna rotating	By moving receiver only
Multi-beam mode	Constrained	$\pm$ 15–20 deg
Frequency band	Wide	Variable narrow bands

spacing of about 40° between them. The antenna prototype is utilized for simultaneous reception of TV programs from two different geostationary satellites. At a design frequency of 12 GHz, the basic antenna dimensions and parameters of the single-layer zone plate backed by a parabolic reflector of 900 mm in diameter are: parabolic surface focal length $F_p = 290$ mm, lateral diffraction focal length $= 620$ mm, total number of zones $N = 10$, zone plate height above the parabolic reflector $= 6.25$ mm, peak gain $= 35$ dB, 3-dB beamwidth $= 2°$, sidelobe level $= -20$ dB, and cross-polar level $= -22$ dB.

Important theoretical and practical progress in the area of curvilinear FZP lenses and antennas is done by I. Minin and O. Minin. These researchers have developed and examined experimentally many spherical, parabolic, and conical dielectric FZP lenses and antennas in the frequency range 40–95 GHz (Minin and Minin 1988, 1989, 2004, 1990, 2005a, 2008a). One of the studied FZP antenna models has a half-wave phase reversal spherical zone plate with $D/\lambda = 36$ and $D/F = 1.48$. The measured antenna gain is 34 dB, and the H- and E-plane sidelobe levels are less than -28 dB and -18 dB, respectively.

Composite conical-parabolic lens antenna consisting of two portions, conical and parabolic, is pictured in Fig. 37a. A semispherical FZP lens antenna mounted on a car roof, Fig. 37b, serves for satellite TV signal reception. The shown FZP antenna prototypes are fabricated from grooved-dielectric shells, which serve simultaneously as FZP lenses and antenna radomes.

A very useful table comparison between the parabolic (reflective) antenna and FZP (diffractive) antennas is found in Minin and Minin (2008a). Slightly altered, the table is presented here as Table 2. The characteristics included are mainly related to the antennas for satellite TV reception and communications.

DOI 10.1007/978-981-4560-75-7_42-2

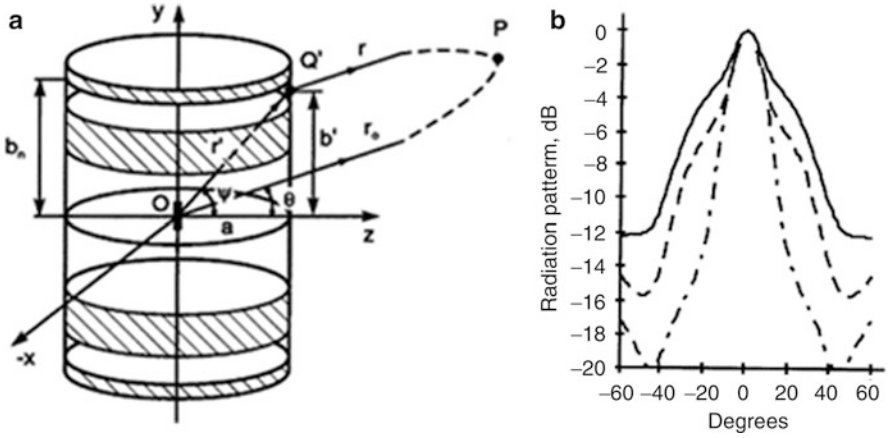


Fig. 38 Cylindrical FZP antennas: (**a**) geometry of original cylindrical FZP lens antenna and (**b**) yOz-plane radiation patterns of binary (*solid line*), phase-reversing (*dashed line*) and quarter-wave (*dash-dotted line*) FZP antennas

Cylindrical FZP Lens Antennas Cylindrical FZP lens antenna with an uniform horizontal radiation pattern is proposed and examined in Ji and Fujita (1996). The rim height b_n, Fig. 38a, is calculated by the corresponding to the flat FZP zone radius equation, Eq. 2. I. Minin and O. Minin suggested more exact empiric equation for the rim height, which leads to antenna gain increase.

By virtue, the binary cylindrical FZP antenna is a low-efficiency radiation structure. The opaque Fresnel zones in the examined antenna are thin circular metal rings and the feed is a half-wave dipole located at the mid-point O, Fig. 38a.

Based on the original binary FZP configuration several modifications with enhanced radiation efficiency are examined numerically (Hristov 1999). As in the flat FZP lens the efficiency is grown by replacing the binary zone plate structure with a phase-reversing or quarter-wave phase-correcting one. A 9-GHz cylindrical Fresnel zone antenna 80 mm in diameter is studied numerically. The vertical yOz-plane radiation patterns with main-beam maxima in the horizontal plane xOz are illustrated in Fig. 38b. The antenna directive gain corresponding to each antenna pattern is 5.1 dB, 6.4 dB and 7.7 dB for the binary, phase-reversing and quarter-wave phase-corrected lens antenna.

If the upper half of the cylindrical Fresnel zone antenna is set on a ground plate and fed by a quarter-wavelength monopole, an asymmetric zone-lens antenna is realized (Hristov 1999). As expected, the asymmetric antenna has a pattern beam slightly tilted above the horizon.

Both symmetric and asymmetric cylindrical Fresnel zone antennas are quite suitable as base station or car mobile antennas in wireless communication and broadcasting radio systems.

FZP Antennas for Terahertz Frequencies

Introduction Text The terahertz frequencies range from 100 to 10,000 GHz or from 0.1 to 10 THz (Lee 2009). Often, the frequencies between 100 and 1000 GHz are called low-terahertz or sub-terahertz frequencies to which correspond free-space wavelengths between 3 and 0.3 mm. Two significant problems arise at terahertz wavelengths: one is keeping losses low, which leads to the use of low-loss solid dielectrics or low-permittivity, foamed dielectric materials (Wiltse 2012). Depending on the number of phase correction steps the groove dielectric FZP lens might be small fractions of the wavelength.

Example: for a frequency of 500 GHz (or $\lambda = 600$ μm) and dielectric permittivity $\varepsilon = 2.1$ (Teflon), the height w_1 of each quarter-wave groove calculated by Eq. 8 is equal to 333 μm. This groove size is still likely for machining. At higher frequencies and finer phase correction a plastic with a permittivity around

DOI 10.1007/978-981-4560-75-7_42-2

1.03–1.05 (a dense foam, for instance) might be more appropriate. For the same frequency and a permittivity of 1.05 the groove depth is 6 mm, or the FZP lens becomes easy for machining but rather thick. For higher terahertz frequencies special microelectronic technologies like a chemical deposition might be necessary. Examples of terahertz FZP lenses produced by use of lithography and reactive ion etching on high-resistivity silicon are demonstrated in Walsby et al. (2002), Wang et al. (2002).

The design and practical basics of millimeter-wave and low-terahertz phase-correcting zone plates for constructing wireless transmission antennas were reported half a century ago (Sobel et al. 1961; Cohn et al. 1962; Cotton et al. 1962). More recently, important theoretical and design methods have been developed and described in Goldsmith and Moore (1984), Wiltse (1994, 2004), Goldsmith (1998).

Low-Terahertz FZP Antenna for Integrated-Circuit (IC) System A folded FZP lens antenna integrated with a bolometer device is proposed, fabricated, and studied in detail in Gouker and Smith (1992). The binary zone plate lens is printed on a grounded fused silica substrate, where the ground metal plane is serving as a FZP reflector. On the opposite side a strip dipole is printed at the focal point $z = F$. The energy collected by the resonant feed-dipole antenna is detected by a bismuth bolometer at the dipole terminals. All components are made by use of simple integrated-circuit (IC) fabrication technique. Another feature of the FZP antenna design is the small focal length–to–aperture diameter ratio (F/D ranges from 0.1 to 0.5). The antenna models described are designed for a frequency of 230 GHz ($\lambda \cong 1.3$ mm). Based on the diffraction Kirchhoff integral the authors developed a far-field theory for the prediction of antenna radiation pattern. Measurements of FZP antenna prototype with $F = 10$ mm and $D = 21.4$ mm have shown a good agreement with the diffraction theory. The average experimental gain is 23.3 dB, and the E-plane 3-dB beamwidth is 4.2 deg.

Low-Terahertz Cavity-Backed FZP Lens Antenna A compact 270-GHz cavity-backed binary Fresnel zone plate lens antenna integrated in a low-temperature cofired ceramic (LTCC) is proposed, fabricated, and studied in Xu et al. (2013), Fig. 39, cross-section and face views. The gain of this low-terahertz FZP lens antenna is enhanced by using a back cavity formed via sidewall and ground plane. A compact feed transition is realized for a broader-band input match.

The measured results show that the cavity-backed FZP antenna achieves a peak gain of 20.8 dB at 270 GHz and a 3-dB gain bandwidth of 9.1 GHz (or about 4 %). These results reasonably validate the design concept, fabrication, and measurement. The proposed compact FZP antenna is a candidate for the future terahertz substrate-integrated planar antenna systems.

Microwave FZP Antenna Operation at Terahertz Frequencies The frequency and space domain antenna harmonic operation of a binary 90-GHz FZP lens antenna is studied by use of vectorial Kirchhoff-Fresnel diffraction theory in the frequency band 50–1550 GHz (Hristov 2011). The lens is made of ten thin metal rings printed on a thin low-loss substrate, whose material loss is taken into account in the process of calculation. The lens antenna has a diameter $D = 100$ mm and focal length $F = 66.7$ mm. The outmost (10th) Fresnel zone is obstructed by a narrow ring of width $\Delta b = 2.83$ mm. The feed horn is supposed to preserve its directive gain function $G_f(\psi) = 2(q + 1)\cos^q\psi$ one and the same in the whole frequency band. The binary FZP lens of the numerically studied antenna produces at primary (design) frequency $f_1 = 90$ GHz and all odd secondary terahertz harmonics ($3f_1 = 270$ GHz,$5f_1 = 450$ GHz f_0, etc.) one and the same directive gain of about 29 dB, Fig. 40a. The comb-shaped frequency response promises attractive filtering capacity of FZP lens antenna, because for the present FZP design all harmonic gain peaks are identical in shape, bandwidth (16.5 GHz), and top values. As expected, the antenna aperture efficiency rapidly goes down with the frequency, following the square of the reverse odd harmonic ratio 1, $(1/3)^2$, $(1/5)^2$and so on. It is clear that at the terahertz harmonic frequencies the antenna

DOI 10.1007/978-981-4560-75-7_42-2

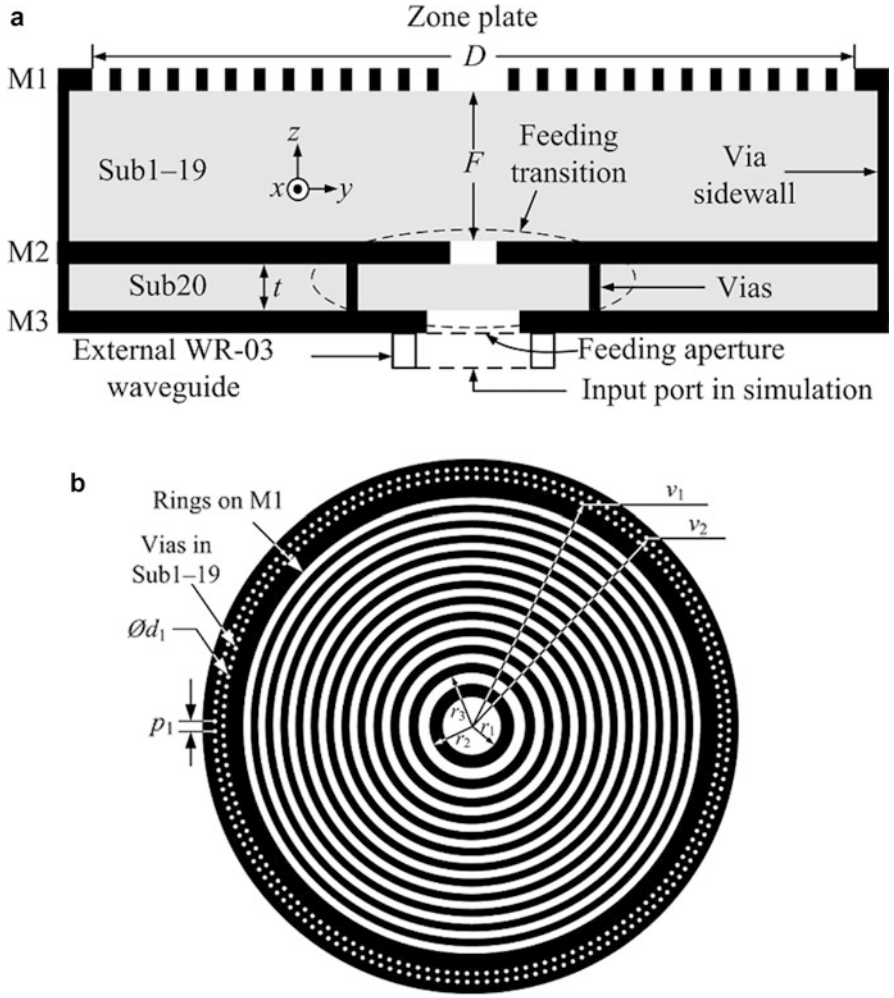


Fig. 39 Cavity-backed FZP lens antenna for 270-GHz frequency band: (**a**) cross-sectional view and (**b**) face view

radiation aperture is not utilized efficiently. Such a low-efficiency performance, however, is typical for all ultrawide band or frequency-independent antennas. For instance, the Archimedean antenna designed for the microwave band 1–20 GHz at the frequency of 10 GHz has about 0.5 % aperture efficiency only.

It is also proved that the FZP antenna pattern beamwidth closely follows the Rayleigh resolution criterion for all terahertz harmonics. It is clarified that if the antenna resolution is of a primary

importance and the low efficiency is not a vital limitation, or may be increased by additional amplification, the microwave FZP lens antenna can be appropriate for terahertz applications too. The FZP resolution increases for the higher harmonics, Fig. 40b, and is 2.4 deg, 0.26 deg, and 0.15 deg for the frequencies 90, 810, and 1350 GHz, respectively. Also, the studied FZP lens antenna has very high axial polarization ratio (AR >55 dB at 90 GHz and >75 dB at 1350 GHz).

Unique and expensive microtechnology is not required for producing the simple and regular-size FZP microwave structure. Applications where the spatial and frequency resolutions are key factors already have been set up in the terahertz metrology, imaging tomography for medical and security use, spectral analysis of organic and nonorganic materials, and radio astronomy among many others.

Microwave and Terahertz FZP Lens Antennas Contrasted to Ordinary Lens Antenna Compared to the bulky and heavy ordinary dielectric lens the grooved-dielectric FZP lens has the very important

DOI 10.1007/978-981-4560-75-7_42-2

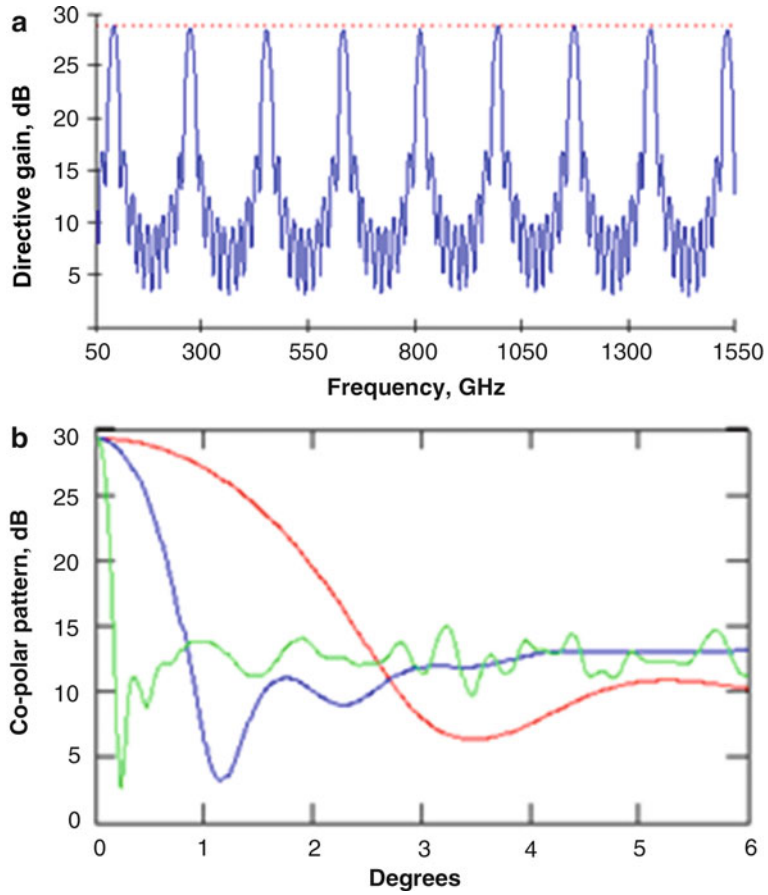


Fig. 40 Frequency- and space-domain action of binary FZP lens antenna: (**a**) directive gain versus frequency and (**b**) radiation co-polar patterns for three harmonic frequencies, f1 = 90 GHz (*red line*), f9 = 810 GHz (*blue line*) and f15 = 1350 GHz (*green line*)

advantage of being thin, light, and easy for production, Fig. 41a. Computer-simulated model of quarter-wave FZP antenna is shown in Fig. 41b. The terahertz design data for the ordinary and FZP lens antennas subject of comparison are $f = 1.5$ THz or $\lambda = 0.2$ mm, $F = 4.55$mm and $D = 4.83$ mm (Rodriguez et al. 2011; Hristov and Rodriguez 2013). The terahertz lenses are modeled for TPX polymer with $\varepsilon = 2.09$ and $\tan\delta = 0.0132$. Similar microwave lenses are designed for $f = 38$ GHz and have $F = 180$ mm and $D = 190.7$ mm. For the design microwave frequency the same TPX polymer has $\varepsilon = 2.13$ and $\tan\delta = 0.0043$. Both antennas are fed by round 16-dB each corrugated horns. A number of terahertz and microwave FZP lens antennas are numerically studied and contrasted to an ordinary plane-hyperbolic lens by use of CST Microwave Studio software. It is found by Fig. 42 that for four or more FZP phase correction steps both antenna types, FZP and ordinary, have close realized gains, similar beamwidths, and comparable cross-polar and mismatch qualities. For example, at the microwave frequency of 38 GHz the eight-step microwave FZP antenna gives way to the ordinary lens antenna by 1.7 dB, Fig. 42b, while for the corresponding 1.5-GHz antenna the gain difference is very small as seen by Fig. 42a. The FZP lens antennas (microwave or terahertz) are narrowband compared to the ordinary ones. Nevertheless, at 1.5 THz the absolute frequency band is more than 500 GHz, which in fact is a very big frequency range for most wireless applications.

DOI 10.1007/978-981-4560-75-7_42-2

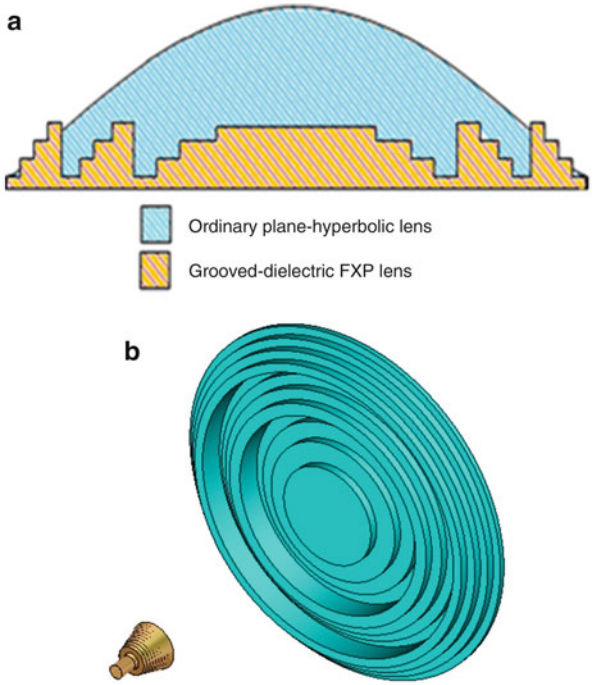


Fig. 41 Dielectric lenses and antenna: (**a**) comparative profiles of ordinary and FZP lenses and (**b**) computer-simulated model of 1.5-THz grooved dielectric quarter-wave FZP lens antenna with corrugated feed-horn

Flat-Ring Conical Fresnel Zone Antenna: Low-Terahertz Antenna Project

 # 1. Introduction A lens antenna based on a cone-like Fresnel zone (CFZ) lens consisting of flat dielectric rings is subject to a recently completed project (Rodriguez et al. 2011; Hristov et al. 2012). The 3D lens originates from Wood's flat phase reversal FZP. A low-terahertz CFZ flat-ring lens assembly conformal to a truncated conical surface is designed and studied by CST computer simulation. Each CFZ lens is characterized by the cone semi angle α in degrees and number of rings ν. For a shortness, each cone-shape FZ lens is abbreviated as CFZ.ν.α. The CFZ lens antenna is compared to flat FZP and ordinary lens antennas same in aperture size and focal length.

 # 2. Lens Geometry and Antenna Model The geometry for analysis of conical Fresnel zone lens with two flat dielectric rings is shown in Fig. 43a. The lens convex side is illuminated by an axial plane wave. CFZ.6.75 lens antenna fed by a rectangular open waveguide (OWG) is shown in Fig. 43b.

 # 3. Design Equations for CFZ Lens For a given design wavelength λ_0, cone semi-angle α, and focal length F, the n^{th} zone radius R_n and axial zone location Z_n are calculated by the next system of equations

$$R_n = \sqrt{2\Lambda(n)F + \Lambda^2(n) + \left(\frac{\Lambda(n)}{\tan\alpha}\right)^2} - \frac{\Lambda(n)}{\tan\alpha} \tag{37}$$

$$Z_n = \frac{R_n}{\tan\alpha} \tag{38}$$

DOI 10.1007/978-981-4560-75-7_42-2

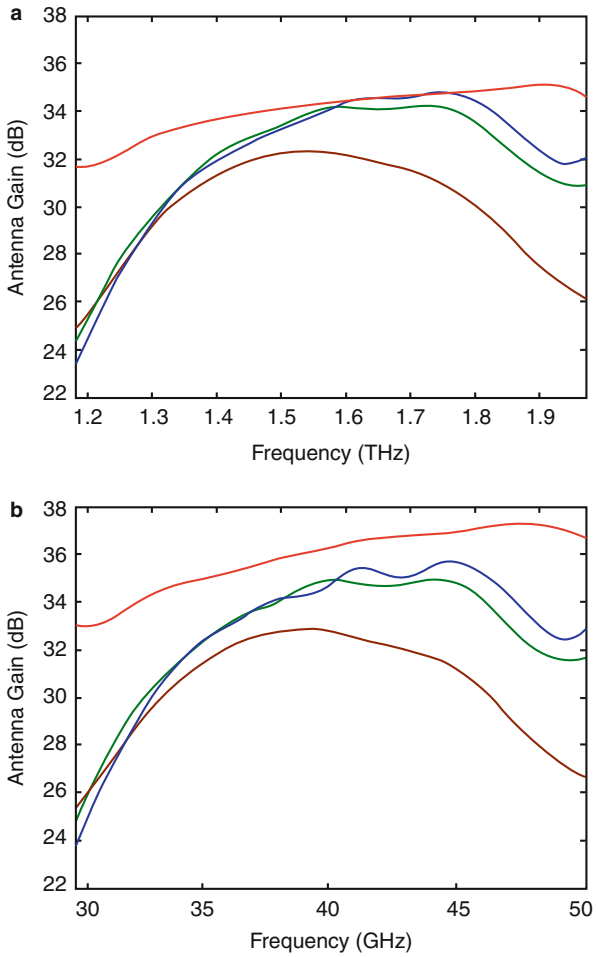


Fig. 42 Realized antenna gain graphs versus frequency for : 2-step FZP (*brown line*), 4-step FZP (*green line*), 8-step FZP (*blue line*) and plane-hyperbolic (*red line*) lenses, for (**a**) terahertz and (**b**) microwave bands, respectively

where $n = 1, 2, \ldots, N$ is an integer sequence, $N = 2\nu$ is the number of all half-wave (Fresnel) zones, and $\Lambda(n) = n\lambda/2$ is a half-wave zone function.

The odd zones in the CFZ lenses are open apertures with an air permittivity $\varepsilon_1 = 1$. The phase-reversing rings on the even zone apertures are made of low-loss dielectric with a preferable permittivity $\varepsilon_2 \geq 2$. In the conical lenses made of flat dielectric rings in air the ring thickness d is calculated by $d = \lambda/2\left(\sqrt{\varepsilon_2 - 1}\right)$. Except in the air, the dielectric ring assembly can be embedded in a low-permittivity solid or liquid media, with $1.03 < \varepsilon_1 < 1.07$.

 # 4. Design Data for Projected CFZ Antennas Design frequency $f = 229$ GHz, focal length $F = 30$ mm, ≈ 42.6 mm ($\tilde{f} = F/D = 0.704$), $\alpha = 75$ deg, $\varepsilon = 2.25$, and $\tan\delta = 0.0004$. For the listed design frequency and dimensions the CFZ lens has six dielectric rings of depth $d = 1.1$ mm each. According to the above abbreviation the lens is marked as CFZ.6.75.

 # 5. Lens Focusing Field Intensity Plots The yOz-plane computer-simulated focusing plots of the plane-wave illuminated lens CFZ.6.75 and classical plane-hyperbolic (PH) lens are displayed in Fig. 44a, b, respectively. It is evident that the lens CFZ.6.75 has a slightly smaller resolving angle δ_0 (transverse

DOI 10.1007/978-981-4560-75-7_42-2

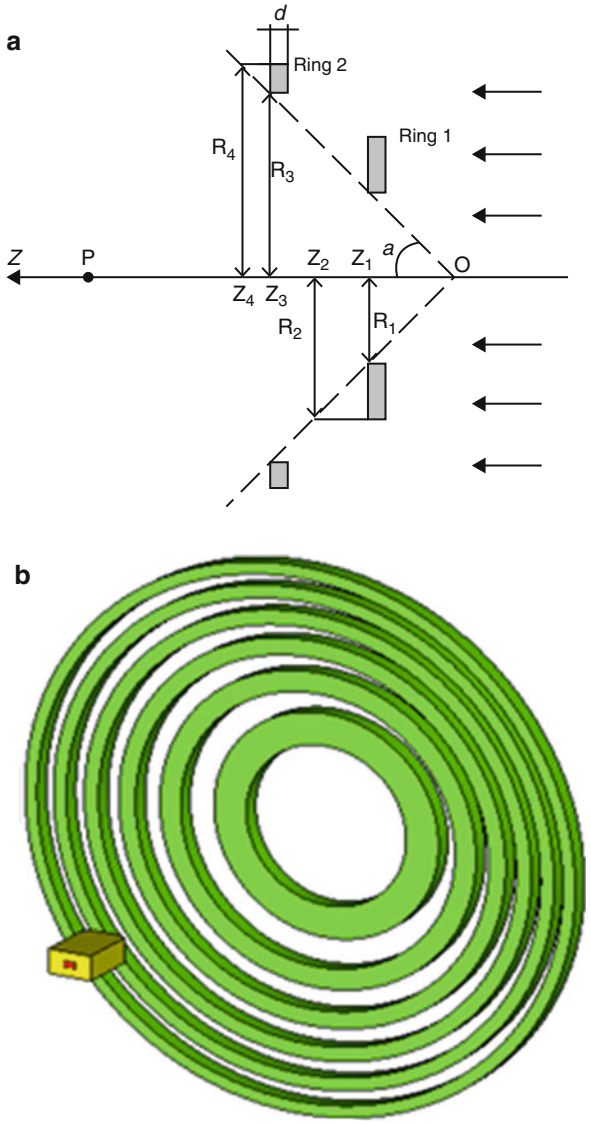


Fig. 43 Conical flat-ring FZ lens antennas: (**a**) lens geometry and (**b**) 3D CFZ.6.75 antenna with OWG feed

resolution) and much better axial resolution. Such a big axial focusing enhancement leads to a smaller spherical aberration and 3D focusing spot compared to the classical PH lens.

 # 6. Lens Antenna with CFZ.6.75 Lens The E-plane (red line), H-plane (blue line), and 45-degree-plane (green line) radiation patterns of the CFZ.6.75 and corresponding plane-hyperbolic lens antennas are contrasted in Fig. 45. The E- and H-plane radiation patterns of each antenna are well overlapping.

In Table 3 are listed the main radiation parameters of the three antennas with: equal in aperture and in focal length (i) flat FZP lens, (ii) cone-shape CFZ.6.75, and (iii) classical PH lens. The abbreviations for the main antenna parameters are: G is the normalized gain, HPBW-half-power beamwidth, SL-maximum sidelobe level, and XP-maximum cross-polarization level.

DOI 10.1007/978-981-4560-75-7_42-2

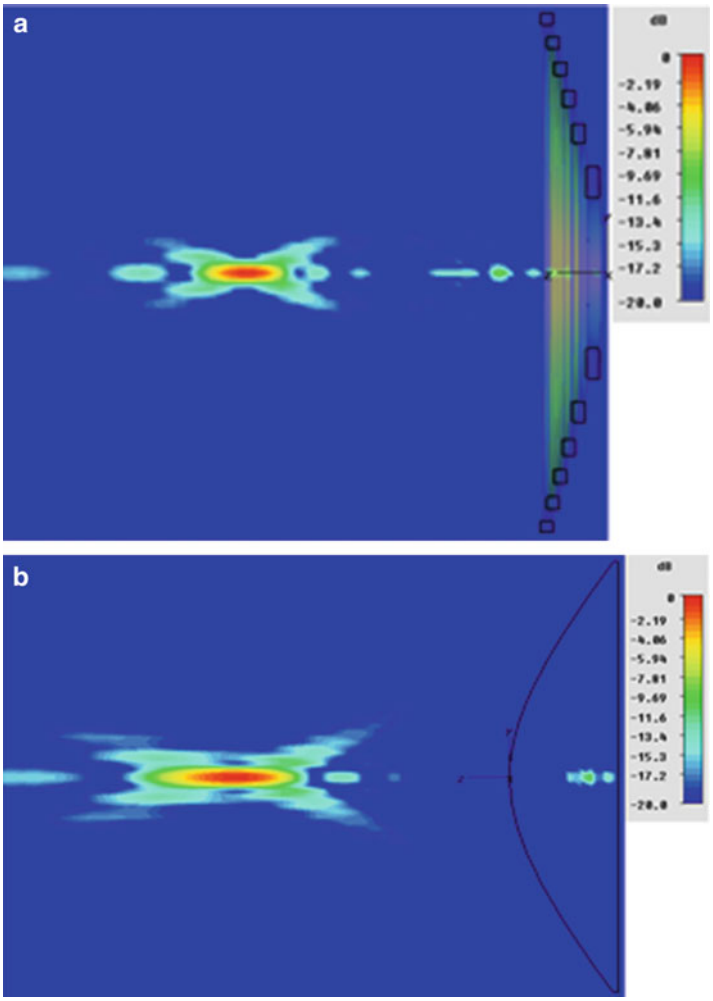


Fig. 44 Plots of yOz-focusing gain of plane-wave illuminated (**a**) curvilinear lens CFZ.6.75 and (**b**) plane-hyperbolic (PH) lens

 # 7. Conclusions For the same design wavelength of 1.31 mm, focal length of 30 mm, and aperture diameter of about 42.6 mm the studied flat FZP lens antenna is inferior in gain compared to both phase-reversing CFZ.6.75 and PH lens antennas. The latter two antennas feature similar radiation characteristics, but the CFZ.6.75 antenna is much lighter and has important structural and technological advantages. Certainly, the CFZ lens antenna with finer lens phase correction (quarter-wave, eight-wave, and so on) will become superior in radiation compared to the corresponding ordinary refraction lens. The main disadvantage of the conical lens antenna is its bigger length and occupied volume.

DOI 10.1007/978-981-4560-75-7_42-2

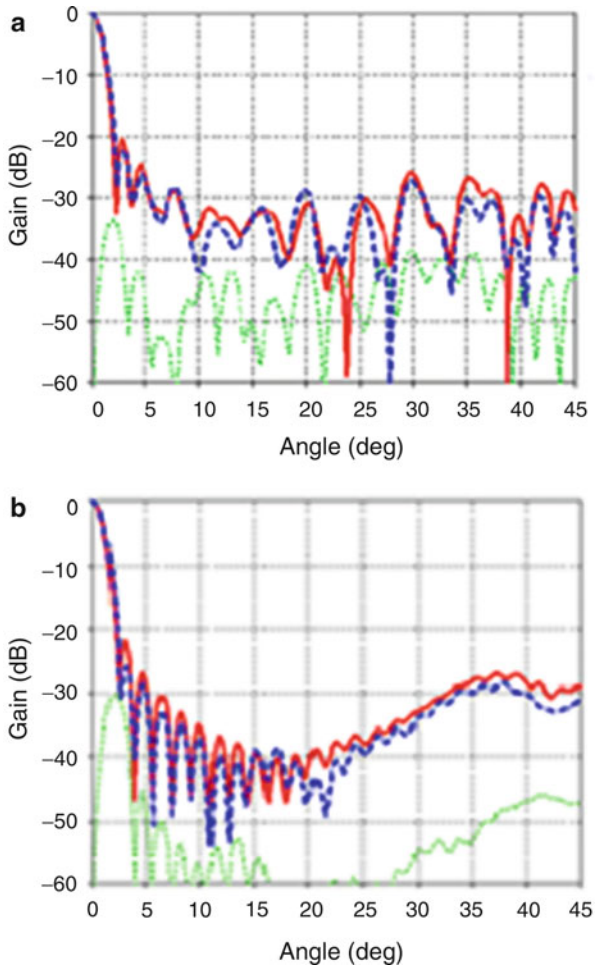


Fig. 45 Co-and cross-polarization radiation patterns of: (**a**) CFZ.6.75 lens antenna and (**b**) PH lens antenna

Table 3 Radiation parameters of lens antennas

Antennas	G (dB)	HPBW (deg)	SL (dB)	XP (dB)
(i) Flat FZ lens antenna	30.7	1.85	−20.5	−37.0
(ii) CFZ.6.75 lens antenna	32.0	1.85	−19.8	−33.1
(iii) Plane-hyperbolic (PH) lens antenna	32.6	1.95	−22.6	−30.3

References

Baggen LC, Herben M (1993) Design procedure for a Fresnel-zone plate antenna. Int J Infrared Millim Waves 6:1341–1352

Baggen LC, Jeronimus CJ, Herben M (1993) The scan performance of the Fresnel zone plate antenna: a comparison with the parabolic reflector antenna. Microw Opt Technol Lett 13:769–774 (Correction in vol 14, p 138)

Black D, Wiltse J (1987) Millimeter-wave characteristics of phase-correcting Fresnel zone plates. IEEE Trans Microw Theory Technol 12:1122–1128

Bruce E (1939) Directive radio system. US Patent 2.169.553

Bruce E (1946) Directive radio system. US Patent 2.412.202

DOI 10.1007/978-981-4560-75-7_42-2

Burnside WD, Burgener KW (1983) High frequency scattering by a thin lossless dielectric slab. IEEE Trans Antennas Propag 1:104–110

Clavier A, Darbord R (1936) Directional radio transmission system. US Patent 2.043.347

Cohn M, Wentworth F, Sobel F, Wiltse J (1962) Radiometer instrumentation for the 1 to 2 millimeter wavelength region. Proc IRE 1962 Nat Aerospace Electronics Conf: 537–541

Cotton J, Sobel F, Cohn M, Wiltse J (1962) Millimeter wave research. Electronics Communications Inc., Dayton: 512–518

Delmas J J, Toutain S, Landrac G, Cousin P (1993) TDF antenna for multi-satellite reception using 3D Fresnel principle and multilayer structure. IEEE Int Antenna Propagat Symp Digest, vol 2, Ann Arbor: 1647–1650

Dey K, Khastgir P (1973a) Comparative focusing properties of spherical and plane microwave zone plate antennas. Int J Electron 35:497–506

Dey K, Khastgir P (1973b) A study of the characteristics of a microwave spherical zone plate antenna. Int J Electron 35:97–103

Dey K, Khastgir P (1973c) A theoretical study of the axial field amplitude of microwave paraboloidal, spherical and plane zone plate antennas. J Inst Electron Telecommun Eng: 697–700

Fresnel AJ (1866) Calcul de l'intensité de la lumiére au centre de l'ombre d'un écran. Euvres Complét 1:365–372

Fan Y, Ooi B-L, Hristov HD, Leong M-S (2010) Compound diffractive lens consisting of Fresnel zone plate and frequency selective screen. IEEE Trans Antennas Propag 6:1842–1847

Gagnon N, Petosa A, McNamara DA (2010) Thin microwave quasi-transparent phase-shifting surface (PSS). IEEE Trans Antennas Propag 4:1193–1201

Gagnon N, Petosa A, McNamara DA (2012) Printed hybrid lens antennas. IEEE Trans Antennas Propag 5:2514–2518

Gagnon N, Petosa A, McNamara DA (2013) Research and development on phase-shifting surfaces (PSSs). IEEE Antennas Propag Mag 2:29–48

Garrett J, Wiltse JC (1991) Fresnel zone plate antennas at millimeter wavelengths. Int J Infrared Millim Waves 12:195–220

Goldsmith PF, Moore EL (1984) Gaussian optics lens antennas. Microw J 7:153

Goldsmith PF (1998) Quasioptical Systems: Gaussian Beam Quasioptical Propagation. IEEE Press, Piscataway

Guo YJ, Barton S, Wright T (1991) Design of high-efficiency Fresnel zone plate antennas. IEEE Antennas Propag Symp, London, Ontario: 182–185

Guo YJ, Barton S (1992) A high-efficiency quarter-wave zone plate reflector. IEEE Microw Guided Wave Lett 12:470–471

Guo YJ, Barton S (1993a) Fresnel zone plate reflector incorporating rings. IEEE Microw Guided Wave Lett 3:417–419

Guo YJ, Barton S (1993b) On the subzone phase correction of Fresnel zone plate antennas. Microw Opt Technol Lett 6:840–843

Guo YJ, Sassi IH, Barton S (1994) Multilayer offset Fresnel zone plate reflector. IEEE Microw Guided Wave Lett 6:196–198

Guo YJ, Barton S (1995) Phase correcting zonal reflector incorporating rings. IEEE Trans Antennas Propag 43:350–355

Guo YJ, Barton S (2002) Fresnel zone antennas. Kluwer, Norwell

Gouker MA, Smith GS (1992) A millimeter-wave integrated-circuit antenna based on the Fresnel zone plate. IEEE Trans Microw Theory Tech 5:968–977

DOI 10.1007/978-981-4560-75-7_42-2

Hajian M, de Vree GA, Ligthart LP (2003) Electromagnetic analysis of beam-scanning antenna at millimeter-wave band based on photoconductivity using Fresnel–zone-plate technique. IEEE Antennas Propag Mag 5:13–25

Hristov HD, Herben M (1995) Millimeter-wave Fresnel zone plate lens and antenna. IEEE Trans Microw Theory Tech 2770–2785

Hristov HD (1996) The multi-dielectric Fresnel zone plate antenna-a new candidate for DBS reception. IEEE Int Antennas Propagat Symp, Baltimore 1:746–749

Hristov HD (1999) Variety of cylindrical Fresnel zone plate antennas. IEEE Int Antennas Propagat Symp Digest 2:750–753

Hristov HD (2000) Fresnel Zones in Wireless Links, Zone Plate Lenses and Antennas. Artech House, Boston-London

Hristov HD, Feick R (2001) The dome-like Fresnel-zone antennas (or how to convert a dome into antenna). IEEE USNC/URSI Radio Science Meeting Digest, Boston: 46

Hristov HD, Kamburov LP, Urumov JR, Feick R (2005) Focusing characteristics of curvilinear half-open Fresnel zone plate lenses: plane wave illumination. IEEE Trans Antennas Propagat 6:1912–1919

Hristov HD (2011) Terahertz harmonic operation of microwave Fresnel zone plate and antenna: frequency filtering and space resolution properties. Int J Antennas Propagat 12:1. doi:10.1155/2011/541734, 8 pages

Hristov HD, Rodriguez JM, Grote W (2012) The grooved-dielectric Fresnel zone plate: effective terahertz lens and antenna. Microw Opt Technol Lett 6:1943–1948

Hristov HD, Rodriguez JM (2013) Design equation for Fresnel zone plate lens. IEEE Microw Wirel Compon Lett 11:574–576

Huder B, Menzel W (1988) Flat printed reflector antenna for mm-wave applications. Electron Lett 24:318–319

Jackson JD (1975) Classical electrodynamics. Wiley, New York

Ji Y, Fujita MA (1996) A cylindrical Fresnel zone antenna. IEEE Trans Antennas Propagat: 1301–1303

Kamburov LP, Hristov HD, Urumov JR, Feick R (2005) Curvilinear Fresnel-zone lens antenna: vector radiation theory. Int J Infrared MilimWaves 11:1593–1611

Kamburov LP, Rodriguez JM, Urumov JR, Hristov HD (2014) Millimeter-wave conical Fresnel zone lens of flat dielectric rings. IEEE Trans Antennas Propagat 4: 2140–2148.

King M, Rodgers J, Sobel F, Wentworth F, Wiltse J (1960) Quasi-optical components and surface waveguides for 100- to 300-Gc frequency range. Electronic Communications, Inc. Report No. 2 on Contract AF19 (604)-5475

Khastgir P, Bhomwmick KN (1978) Analysis of the off-axis defocus of microwave zone plate. Indian J Pure Appl Phys 16:96–101

Khastgir P, Chakravorty JN, Dey K (1973) Microwave paraboloidal, spherical and plane zone plate antennas: a comparative study. Indian J Radio Space Phys 1:47–50

Lazarus M, Silvertown A, Novak S (1979) Fresnel–zone plate aids low-cost Doppler design. Microwaves 11:78–80

Lee Y-S (2009) Principles of terahertz science and technology. Springer Science, LLC, New York

Leiten L, Herben M (1992) Vectorial far field of the Fresnel-zone plate antenna: a comparison with the parabolic reflector antenna. Microw Opt Technol Lett 5:49–56

Maddaus AI (1948) Fresnel zone plate antenna. Naval Research Lab, Washington, DC, Report R-3293

Mawzones Ltd (1992) Leaflet. Herts

Minin I, Minin O (1988) Diffraction lenses on parabolic surfaces. Comput Opt 3:21–29

Minin I, Minin O (1989) Invariant properties of elements of diffraction quasioptics. Comput Opt 6:89–97

Minin I, Minin O (1990) Paraboloidal zone plates: an experimental study. Comput Opt 1:5–9

DOI 10.1007/978-981-4560-75-7_42-2

Minin I, Minin O (2004) Diffraction Optics of Millimeter Waves. IOP Institute of Physics, London
Minin I, Minin O (2005a) Three dimensional Fresnel antennas. In: Tazon A (Ed) Advances on Antennas, Reflectors, and Beam Control. Research Singpost, India: 115–148
Minin I, Minin O, Webb G (2005b) Flat and conformal zone plate antennas with new capibilities. Applied Electromagnetics and Communications, Int Conf, Dubrovnik, Croatia: 405–408.
Minin I, Minin O (2008a) Basic principles of Fresnel antenna arrays. Springer Academic, Berlin
Minin I, Minin O (2008b) Development and application of 3D diffractive antennas. TELE-Satell Broadband 5:14–16. www.TELE-satellite.com
Mottier P, Valette S (1981) Integrated Fresnel lens on thermally oxidized silicon substrate. Appl Optics 20:1630–1634
Myers OE (1951) Studies of transmission zone plates. Am J Phys 19:359–365
Norden Systems, Inc (1984) Millimeter-wave radio series 380. Data sheet
Ojeda-Castañeda J, Gómes-Reino C (Eds) (1996) Selected papers on zone plates, vol 128, SPIE milestone series. Bellingham, Washington, DC
Orazbaev B, Beruete M, Pacheco-Pena V, Crespo G, Teniente J Navarro-Cia M (2015) Wood zone plate fishnet metalens. E Sciences, EPJ Appl Metamat 2: 6 pages
Petosa A, Ittipiboon A (2003) Design and performance of a perforated dielectric Fresnel lens. IEE Proc Microw Antennas Propagat 10:309–314
Petosa A, Ittipiboon A, Thirakone S (2006) Investigation on arrays of perforated dielectric Fresnel lenses. IEE Proc Microw Antennas Propagat 3:270–276
Reid DR, Smith G (2006) A full electromagnetic analysis for the Soret and folded zone plate antennas. IEEE Trans Antennas Propagat 12:3638–3646
Reid DR, Smith G (2007) Full electromagnetic analysis of grooved-dielectric zone plate antennas for microwave and millimeter-wave applications. IEEE Trans Antennas Propagat: 2138–2146
Rodriguez JM, Hristov HD, Grote W (2011) Fresnel zone plate and ordinary lens antennas: comparative study at microwave and terahertz frequencies. 41st EU Microwave Conf (EuMC-20011): 894–987
Sanyal GS, Singh M (1968) Fresnel zone plate antenna. J Inst Telecommun Eng, India: 265–281
Sazonov D (1999) Computer-aided design of holographic antennas. IEEE Int Antennas and Propagat Symp, Orlando, Fl, Symp Digest, 2:738–741
Silver S (1984) Microwave antenna theory and design. Peter Peregrines, London
Shuter WL, Chan CP, Li EW, Yeung AK (1984) A metal plate Fresnel zone lens for 4 GHz satellite TV reception. IEEE Trans Antennas Propagat 3:306–307
Sluijter J, Herben MHA, Vullers OJG (1995) Experimental validation of PO/UTD applied to Fresnel zone plate antenna. Microw Opt Technol Lett 2:111–113
Sobel F, Wentworth FL, Wiltse JC (1961) Quasi-optical surface waveguides and other components for the 100- to 300- Gc Region. IRE Trans Microw Theory Tech MTT-9:512–518
Soret J (1875) Ueber die durch Kreisgitter erzeugten Diffractionsphunomene. Annalen der Physik und Chemie 156:99–113
Suhara T, Kobayashi K, Nishihara H, Koyama J (1982) Graded-index Fresnel lenses for integrated optics. Appl Opt:1966–1971
Sun J-A, Cai A (1991) Archaic focusing of Fresnel zone plates. J Opt Soc Am 1: 33–35.
Sussman M (1960) Elementary diffraction theory of zone plates. Am J Physics 11: 394–398
Thornton, Strozyk (1983) MCPR-An LPI wideband cable replacement radio. Proc IEEE Southcon-83, Atlanta GA: 21/2-1–21/2-14

DOI 10.1007/978-981-4560-75-7_42-2

Van Buskirk LF, Hendrix CE (1961) The zone plate as a radio frequency focusing element. IRE Trans Antennas Propagat 9:319–320

Van Houten JM, Herben M (1994) Analysis of phase correcting Fresnel-zone plate antenna with dielectric/transparent zones. J EM Waves Applications 8:847–858

Walsby ED, Wang S, Xu J, Yuan T, Blaikie R, Durbin SMT, Cumming DRS (2002) Multilevel silicon diffractive optics for terahertz waves. J Vac Sci Technol B 6:2780–2783

Wang S, Yuan T, Walsby ED, Blaikie RJ, Durbin SM, Cumming DRS, Xu J, Chang X-C (2002) Characterization of T-ray binary lenses. Opt Lett 13:1183–1185

Webb G (2003) New variable for Fresnel zone plate antennas. In: Proceedings of the 2003 Antenna Applications Symp, Allerton Park, Monticello

Webb G Minin IV, Minin OV (2011) Variable reference phase in diffractive antennas: review, applications and new results. IEEE Antennas Propag Mag 2:78–94

Wiltse J (1985) The Fresnel zone-plate lens. Proc SPIE Symp: 41–47

Wiltse JC (1994) Millimeter wave Fresnel zone plate antennas. Chapter 11, Handbook of Millimeter wave and MW Engineering for Communications and Radar (book), SPIE CR54. Bellingham: 272–293

Wiltse JC (1998) High-efficiency, high-gain Fresnel zone plate antennas. Proc SPIE 3375:286–290

Wiltse JC (2004) Diffraction optics for terahertz waves. Proc SPIE 5411:127–135

Wiltse JC (2012) Fresnel zone plates antennas at terahertz, millimeter-waves and microwave frequencies. Unpublished book

Wood RW (1898) Phase-reversal zone plates and diffraction telescopes. Philos Mag 45:511–523

Wood RW (1934) Physical optics, 3rd edn. The MacMillan Co, New York

Xu J, Chen ZN, Qing X (2013) 270-GHz LTCC-integrated high-gain cavity-backed Fresnel zone plate lens antenna. IEEE Trans Antennas Propagat 4:1679–1687

Yamauchi S, Honma S, Honma T, Nacano H (1990) Focusing properties of Fresnel zone-plate and its applications to a helix radiating a circularly polarized. Electronics Communications Jpn 9:107–113

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_43-1

Optical Nanoantennas

Robert D. Nevels* and Hasan Tahir Abbas
Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA

Abstract

An overview of the field of optical plasmonic antennas is presented in this chapter. After a brief introduction and historical review, the theory of surface plasmon polaritons which leads to a set of overall observations as to the requirements and restrictions placed on the operation of plasmonic waveguides and antennas is presented. Both a single metal-dielectric interface and two interfaces between a metal sheet with dielectrics on either side are considered. In the second section the physical principles of operation and mathematical design criteria are presented for several common optical antennas including on-surface metallic structures and free standing particles. The third section covers the basic theory of aperture radiators along with a more detailed description of some popular designs. Current applications of optical nanoantennas are presented along with a discussion on some future directions in optical nanoantenna research.

Keywords

Optical antenna; Nanoantenna; Surface plasmon polaritons; Plasmonics; Negative dielectric constant; Near infrared; Electromagnetic; Propagation

Introduction

Optical nanoantennas are nano-size objects that transmit or receive electromagnetic fields through their intrinsic plasmonic behavior in the optical or near infrared range. Optical antennas have become the object of great interest to engineers due to advancements in nano device construction technologies. In particular, commercial computer aided design (CAD) programs that allow the characterization of materials with a negative real part permittivity and the lower cost and advancements in electron beam (e-beam) etching machines permitting the creation of nanocircuits with line widths on the order of 10 nm or smaller have made it possible for universities and research labs to not only numerically simulate, but also build models and measure the properties of optical antennas. The fact that practical research and development is now available to the academic and industrial community has led to several new applications and technologies including dramatically improved spectroscopy (Ouyang et al. 1992; Nie and Emory 1997; Kneipp et al. 1997), disease and toxin sensors (Arduini et al. 2010; Nevels et al. 2012), wireless communication with nano-circuitry (Adato et al. 2011), and the creation of nanocircuits using subwavelength lithography (Sotomayor Torres et al. 2003; Ishihara et al. 2006). The opportunities for antenna engineers in this vast new landscape are great, but much work remains to be done on analytical and numerical characterization as well as in developing an understanding as to how electromagnetic and quantum processes can be incorporated into a coherent mathematical theory that will allow ready design of such devices.

*Email: nevels@ece.tamu.edu

DOI 10.1007/978-981-4560-75-7_43-1

Dimensions of optical antennas are in the nanometer (nm) range of visible and near infrared light. Typical antenna element shapes include the dipole, bowtie, aperture, and sphere, as shown in Fig. 1. Figure 1d illustrates the oscillating dipole like behavior produced by plasmonic action in a nanometer scale gold sphere illuminated at 800 nm where gold has a dielectric constant $\varepsilon_r = -20.277 - j2.07$. Optical antennas are always fabricated out of noble metals, most commonly gold and silver and less commonly with aluminum, chromium, and copper. Unfortunately there are some negative aspects of these materials that make them difficult to work with in engineering applications. In an ambient environment silver can form a silver sulfide layer that inhibits the propagation of plasmon waves (Nevels and Michalski 2014) and as such is not always a good candidate for an optical antenna. Aluminum has a higher dielectric function imaginary part than does gold, increasing from a small difference at 550 nm to a significantly larger value at 830 nm where it becomes exceptionally lossy. During the construction of a nanoantenna, ion beam bombardment can cause gold to melt making it hard to form a smooth metallic structure that is true to its intended design (Farahani 2006). Nevertheless, noble metals are necessary for the applications described in this chapter because their plasma frequency, a crucial element of the design, lies in the wavelength range of visible light.

The majority of current research papers center on free space designs, but for most practical applications nanoantennas will be etched on a substrate. The type of substrate depends on the particular application, to some extent determined by the mechanism used to excite the antenna, but for many industrial applications cost and availability considerations have driven the move to silicon dioxide (SiO_2), often described in technical papers simply as "glass." Typical signal generation is via off-board laser light or by an on-board transmission line. Unlike classical antenna signals which can be generated at virtually any microwave carrier frequency, there are a select few frequencies where low cost optical sources are available, although specialized lasers can operate at frequencies throughout the optical spectrum. In optical terminology the signal has traditionally been specified in wavelengths rather than frequency, however here both terms will be used. Some common wavelengths where research is being carried out are in the neighborhood of 550, 630, and 820 nm. A more complete list of various laser types along with their operating frequencies can be found (Weber 2001).

In the optical realm, in addition to referring to "wavelength" rather than "frequency," other commonly encountered terms can at first seem odd or even incorrect to someone accustomed to working in the classical microwave domain. The term "antenna" itself is generally associated with the radiator or receiver component of a circuit, but in the nanometer regime it is also often used when describing a resonator, a device that is not intended for transmission or reception of signals. Also, the term "field intensity" is usually associated with the electric or magnetic field intensity in a microwave system, but in optics it is

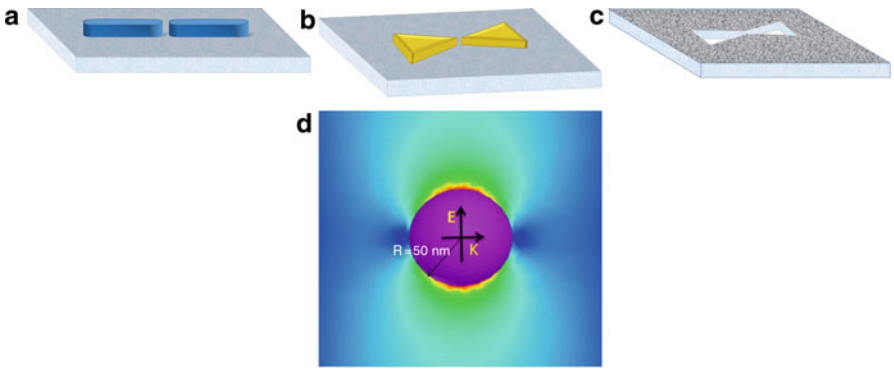


Fig. 1 Common nanoantenna elements (**a**) dipole (**b**) bowtie (**c**) aperture and (**d**) sphere. The dipole, bowtie, and aperture antennas are typically etched on a low loss substrate such as silicon dioxide while spheres may be mounted at the tip of a cone shaped substrate or arranged in some fashion on a flat substrate

DOI 10.1007/978-981-4560-75-7_43-1

used to describe what an antenna engineer would refer to as the "radiation intensity," which is power density multiplied by r^2.

A counter-intuitive oddity of noble metals at optical frequencies is that they are no longer perfect conductors, but rather they have some of the properties of a dielectric with, as mentioned above, a permittivity whose real part is negative. This turns out to be a remarkable advantage over dielectrics with a positive real part, due to the fact that the wavelength of the current in these metals can be much less than the wavelength of the free space incident field. A solid nanorod still resonates around one-half wavelength of its surface current but, given an incident field frequency, one must use numerical or approximate analytical methods to determine the shorter than free space wavelength of the antenna at resonance. While optical antennas make excellent resonators, due in part to a tightly bound plasmon current, the wavelength mismatch between air and metal, which does not exist at microwave frequencies, is a significant factor hampering their radiation and receiving properties. To understand how the radiation efficiency of an optical antenna can be improved it is necessary to first study the unusual properties of waves on metals at optical frequencies. The essentials of optical wave propagation and radiation are presented in the following section.

Electromagnetic Theory of Surface Plasmon Polaritons

The physical mechanism for wave propagation at optical frequencies is very different from that of propagating waves at microwave frequencies although the mathematics for determining the wave behavior is virtually the same and can be laid out in a classical Sommerfeld integral analysis. At optical frequencies surface plasmon polaritons (SPPs), electromagnetic surface waves created by coherent charge oscillations in an electron gas in a metal, propagate at a metal-dielectric interface (Ritchie 1957; Otto 1976; Raether 1988). Resonant plasmonic oscillations can also occur in the confined space of nanoparticles (Nie and Emory 1997). Fortunately, in most cases where optical antennas are concerned the quantum mechanisms which create SPPs or local confined plasmons can be expressed in terms of the material properties of metals and nanoparticles, such as shape, size and permittivity (Kelly et al. 2003). Because the study of nanoantennas relies on the behavior of plasmonic waves, it resides in the research area described as "plasmonics," a subfield of nanophotonics (Maier and Atwater 2005; Park 2009). In this section some of the basic theory of SPPs will be presented in order to aid the reader in understanding the underlying principles that determine the behavior of nanoantennas. First, the fundamental problem of electromagnetic propagation on a planar dielectric-metal interface at optical frequencies is considered, followed by a discussion on the role played by quantum effects in determining the allowable frequency range and loss mechanisms in nano-device design. This presentation does not require an understanding of quantum mechanics, but occasionally the language of quantum electronics is used to describe the effects that determine the properties of the constitutive parameters at optical frequencies.

Single Boundary Structure

Assuming a metal can have a complex permittivity at optical wavelengths, the propagation constants k_{xi} and k_{zi} at the planar interface between dielectric and metal half-spaces, $i = 1,2$ respectively, can be found starting with expressions for transverse magnetic (TM) plane waves in these two regions as shown in Fig. 2. The electromagnetic field expressions in region 1 where

DOI 10.1007/978-981-4560-75-7_43-1

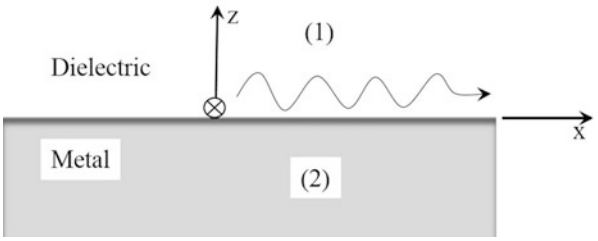


Fig. 2 Dielectric and metal half-spaces with a planar boundary

$z \geq 0$ are,

$$\mathbf{E}_1 = (E_{x1}\hat{\mathbf{x}} + E_{z1}\hat{\mathbf{z}})e^{-j(\mathrm{k}_x x + k_{z1} z)} \tag{1a}$$

$$\mathbf{H}_1 = H_{y1}e^{-j(k_x x + k_{z1}\mathrm{z})}\hat{\mathbf{y}} \tag{1b}$$

and in region 2 where $z \leq 0$,

$$\mathbf{E}_2 = (E_{x2}\hat{\mathbf{x}} + E_{z2}\hat{\mathbf{z}})e^{-j(\mathrm{k}_x x - k_{z2} z)} \tag{2a}$$

$$\mathbf{H}_2 = H_{y2}e^{-j(k_x x - k_{z2}\mathrm{z})}\hat{\mathbf{y}} \tag{2b}$$

A positive sign is chosen for k_{z2} to ensure propagation in the negative z direction. The Im(k_z) must therefore be negative in order to be bounded at infinity.

Ampere's law, $\nabla \times \mathbf{H} = j\omega\varepsilon\mathbf{E}$, applied to Eqs. 1 and 2 yields the boundary conditions

$$k_{z1}H_{y1} = \omega\varepsilon_1 E_{x1} \tag{3a}$$

$$k_{z2}H_{y2} = -\omega\varepsilon_2 E_{x2} \tag{3b}$$

where $\varepsilon_{1,2}$ are the permittivities of the dielectric and metal regions respectively. Since metal at optical frequencies has the behavior of a lossy dielectric, continuity of tangential electric and magnetic fields, $E_{x1} = E_{x2}$ and $H_{y1} = H_{y2}$, can be applied at the boundary along with Eq. 3. This gives the dispersion relation for waves on the metal-air boundary

$$\frac{k_{z1}}{\varepsilon_1} + \frac{k_{z2}}{\varepsilon_2} = 0 \tag{4}$$

Helmholtz equation, $\nabla^2\mathbf{E} + k_i^2\mathbf{E} = 0$, where $i = 1,2$, and assuming the permeabilities of all regions are that of air, lead to the dispersion equations for the two regions,

$$k_x^2 + k_{zi}^2 = \varepsilon_i\left(\frac{\omega}{c}\right)^2 \doteq \varepsilon_i k_o^2 \tag{5}$$

where c is the speed of light in air. Equations 4 and 5 are combined to produce the SPP propagation constant

DOI 10.1007/978-981-4560-75-7_43-1

$$k_x = k_o \left(\frac{\varepsilon_{r1}\varepsilon_{r2}}{\varepsilon_{r1} + \varepsilon_{r2}} \right)^{1/2} \tag{6}$$

Where k_o is the free space wavenumber and the subscript r is used to indicate a dielectric constant quantity. If the permittivity of the dielectric is real and that of the metal is complex, $\varepsilon_2 = \varepsilon_2' - j\varepsilon_2''$, and $|\varepsilon_2'| \gg \varepsilon_2''$ the complex propagation constant in Eq. 6 can be expressed as (Raether 1988).

$$k_x = k_o \left(\frac{\varepsilon_1 \varepsilon_2'}{\varepsilon_1 + \varepsilon_2'} \right)^{1/2} - jk_o \frac{\varepsilon_2''}{2(\varepsilon_2')^2} \left(\frac{\varepsilon_1 \varepsilon_2'}{\varepsilon_1 + \varepsilon_2'} \right)^{3/2} = k_x' - jk_x'' \tag{7}$$

Similarly, from Eqs. 5 and 6, k_{zi} becomes approximately,

$$k_{z1} = k_o \left(\frac{\varepsilon_1{}^2}{\varepsilon_1 + \varepsilon_2'} \right)^{1/2} + jk_o \frac{\varepsilon_1 \varepsilon_2''}{2(\varepsilon_1 + \varepsilon_2')^{3/2}} = k_{z1}' + jk_{z1}'' \tag{8}$$

$$k_{z2} = k_o \left(\frac{\varepsilon_2'{}^2}{\varepsilon_1 + \varepsilon_2'} \right)^{1/2} - jk_o \frac{\varepsilon_2''(2\varepsilon_1 + \varepsilon_2')}{2(\varepsilon_1 + \varepsilon_2')^{3/2}} = k_{z2}' - jk_{z2}'' \tag{9}$$

From Eqs. 4, 5, 6, 7, 8, and 9 some very important conclusions can be drawn concerning the nature of the two half-space materials and the properties of nano surface waves on their intersecting boundary:

(a) First, if loss is neglected ($\varepsilon'' \simeq 0$), a valid assumption for noble metals over significant segments of the optical range, so that the permittivities of both regions are real, the equality in Eq. 4 will not hold if these permittivities are both positive or both negative. But, if the permittivity in the dielectric region (1) is positive, $\varepsilon_1 > 0$, and the real part of the permittivity in the metal region (2) is negative, $\varepsilon'_2 < 0$, then it is possible to satisfy Eq. 4.

(b) Nevertheless, a wave still cannot exist if the term under the square root in Eq. 7 is negative because this would lead to a totally complex k_x, which would mean the fields in Eqs. 1 and 2 will exponentially grow or decay rather than propagate along the boundary. However, if

$$\varepsilon_1 > 0,\ \varepsilon_2' < 0 \ \text{ and } \ |\varepsilon_2'| > \varepsilon_1 \tag{10}$$

then the term under the square root in Eq. 7 is positive and propagation of a wave at the boundary is assured. This condition (10) causes the signs of the terms under the square roots in Eqs. 8 and 9 to be negative, but Eq. 4 will still hold as long as same sign for $\sqrt{-1}$ is chosen for both square roots.

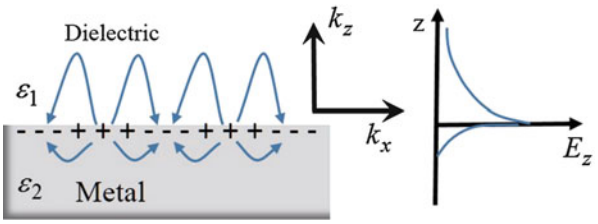


Fig. 3 Surface plasmon propagation along a dielectric-metal boundary and exponential decay perpendicular to the boundary

DOI 10.1007/978-981-4560-75-7_43-1

(c) If the conditions in Eq. 10 are met then from Eq. 6 $k_{x>} k_o$ and therefore k_z in Eq. 5 must be complex. According to Eqs. 1 and 2 this results in exponential decay of the field away from the boundary in both the positive and negative z-directions as long as $\sqrt{-1} = -j$ is chosen for the square root terms in Eqs. 8 and 9. Exponential decay of the surface plasmon wave into the dielectric and metal is illustrated in Fig. 3.

(d) According to Snell's law the component of the wavenumber k_x tangential to the boundary is the same in both mediums. If medium 1 is air, the x-component of a plane wave incident upon the boundary at an angle θ from normal is $k_{x1} = k_o \sin\theta$, which is less than k_o. However, according to Eqs. 6 and 10 surface plasmons cannot exist unless $k_x > k_o$. Therefore, surface plasmons cannot be excited by shining light on a flat metal surface.

(e) From Eq. 7 the velocity of a surface plasmon wave on a planar boundary is

$$v_{sp} = \omega/k'_x = c\left(\frac{\varepsilon_1 + \varepsilon'_2}{\varepsilon_1\varepsilon'_2}\right)^{1/2} \tag{11}$$

and its wavelength is

$$\lambda_{sp} = \frac{2\pi}{k'_x} = \lambda_o\left(\frac{\varepsilon_1 + \varepsilon'_2}{\varepsilon_1\varepsilon'_2}\right)^{1/2} \tag{12}$$

(f) The exponential decay of a plasmon along its propagation direction is determined by the second term in Eq. 7. The "propagation length" of a plasmon wave is the distance $x = L$ at which the wave decays to $1/e$ of its initial value. The plasmon propagation length on a planar surface is therefore

$$L = 1/k''_x = \frac{2\left(\varepsilon'_2\right)^2}{k_o\varepsilon''_2}\left(\frac{\varepsilon_1 + \varepsilon'_2}{\varepsilon_1\varepsilon'_2}\right)^{3/2} \tag{13}$$

(g) If the conditions in Eq. 10 hold, the second term on the right hand side in Eqs. 8 and 9 are small so, neglecting these terms, perpendicular to the boundary the 1/e decay distance is (Raether 1988)

$$z_1 = \left|\frac{1}{k'_{z1}}\right| = \frac{1}{k_o}\left|\left(\frac{\varepsilon_1 + \varepsilon'_2}{{\varepsilon_1}^2}\right)^{1/2}\right| \tag{14}$$

into the dielectric and

$$z_2 = \left|\frac{1}{k'_{z2}}\right| = \frac{1}{k_o}\left|\left(\frac{\varepsilon_1 + \varepsilon'_2}{{\varepsilon'_2}^2}\right)^{1/2}\right| \tag{15}$$

into the metal.

(h) Power and energy relationships can be found by assuming a value for the magnetic field amplitude and by using Amperes law to determine the electric field.

The conditions described in Eq. 10 are met by noble metals in air for the optical and near-infrared ranges in the case of a TM polarized wave. A similar analysis will show that a transverse electric (TE) polarized wave cannot exist at a dielectric-metal boundary.

DOI 10.1007/978-981-4560-75-7_43-1

In practical applications one should take care in the selection of the metal to be used in a nanoantenna design. Some metals, such as silver, tend to develop an oxidation layer, which can severely decrease propagation length and increase the antenna impedance. For a more detailed discussion concerning the effect of a silver sulfide oxidation layer on a silver substrate plasmon wave, the reader is referred to (Nevels and Michalski 2014).

Figure 4 shows the propagation length as a function of wavelength for three metals, gold, silver and aluminum (Homola 2006). The practitioner is cautioned that although the propagation length of a surface plasmon does increase as frequency decreases, the analysis leading to this figure is based on an idealization of the physical problem. Here, a pure wave propagation condition was assumed in the formulation of equations 1 and 2. In the practical case a source, such as the feed point of an antenna, must be included in the problem formulation. If the source is included, one sees that as the frequency decreases less energy is transferred into the surface plasmon wave and more into the space wave – the field that is radiated directly into space from the source point. In the far infrared range the surface plasmon essentially ceases to exist. Plasmon wave behavior is discussed in more detail in the section of this chapter concerning aperture antennas.

Quantum Effects

The above analysis considers only the classical electromagnetic aspects of propagation on a dielectric metal boundary. However, quantum mechanical effects inherent in metals and dielectrics enter the picture through the permittivity, thereby determining its value as well as the frequency range over which surface plasmon polaritons can propagate. Essentially quantum processes reveal the frequency range over which the metal has a permittivity with a negative real part and where losses begin to have a serious effect on wave propagation within that frequency range. In the following the influence of plasma resonance and atomic collisions on SPP propagation are briefly explored. Although the properties described below are best understood using quantum mechanics, the mathematics below is classical, primarily developed by concatenating harmonic oscillator models. However, the differences in the results of the two theories is small, allowing classical electromagnetic calculations without the necessity of delving into quantum theory.

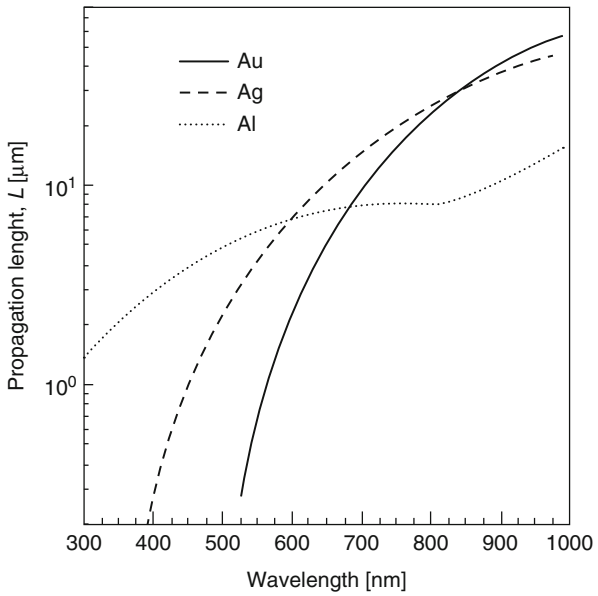


Fig. 4 Propagation length of a surface plasmon propagating along the interface between a dielectric (refractive index 1.32) and a metal as a function of wavelength calculated for gold (Au), silver (Ag), and aluminum (Al)

DOI 10.1007/978-981-4560-75-7_43-1

Free electrons can exist near the top of the energy band distribution in a metal. The interaction of these electrons with photons and the long-range Coulomb force of atoms create an electron oscillation known as a plasmon. Taking these effects into account, the dielectric function of metal can be expressed in terms of the Drude model (Born and Wolf 1970),

$$\varepsilon(\omega) = \left(1 - \frac{\omega_p^2}{\omega(\omega - j\nu)}\right) = \left(1 - \frac{\omega_p^2}{\omega^2 + \nu^2}\right) - j\frac{\nu\omega_p^2}{\omega(\omega^2 + \nu^2)} = \varepsilon_r - j\varepsilon_i \tag{16}$$

where $\nu = 1/\tau$ is the collision frequency, τ is the time between collisions, and ω_p is the plasma frequency

$$\omega_p = \sqrt{\frac{Ne^2}{\varepsilon_o m_e}} \tag{17}$$

where e and m_e are the electron charge and mass respectively and N is the concentration of free electrons. Plasma frequencies for metals are in the visible and ultraviolet ranges. Typical values of plasma and collision frequencies are $\omega_p = 0.2321 \times 10^{16}$Hz and $\nu = 5.513 \times 10^{12}$Hz for gold and $\omega_p = 0.2068 \times 10^{16}$Hz and $\nu = 4.449 \times 10^{12}$Hz for silver, however impurities and other factors can affect these numbers so the literature does not have a unique set of values for these parameters. A comprehensive list of plasma and collision frequencies published by a variety of researchers for gold, silver and other noble metals can be found at (Moroz 2009).

Collisions are primarily between electrons and relatively large lattice vibrations (phonons), so for most metals at room temperature $\nu \ll \omega_p$ (Bohren and Huffman 2004), which reduces Eq. 16 to approximately

$$\varepsilon(\omega) \simeq 1 - \frac{\omega_p^2}{\omega^2} \tag{18}$$

This important expression shows that when $\omega < \omega_p$ then $\varepsilon(\omega) < 0$ which in part satisfies the conditions in Eq. 10. Therefore surface plasmon polaritons can only exist below the plasma frequency. Equation 18 substituted into Eq. 6 gives a simplified Drude model for the plasmon wave vector:

$$k_x \simeq k_1 \left(\frac{\left(\omega^2 - \omega_p^2\right)}{\omega^2\varepsilon_{r1} + \left(\omega^2 - \omega_p^2\right)}\right)^{1/2} \tag{19}$$

The Drude model in the form given in Eq. 16 is widely used in nanoantenna analytical and numerical calculations for determining the permittivity, and therefore the propagation constant, of the plasmon polariton of noble metals. However, this free electron model fails at frequencies higher than about 850 THz (Archambault et al. 2009), due to bound charge effects, and must be modified by incorporating additional terms (Bohren and Huffman 2004). This becomes a tedious process but reasonable accuracy can be achieved. Another option, preferred by the authors, is to use measured data (Lynch and Hunter 1998) with a partial fraction fit (Michalski 2013). Figure 5 shows the dielectric function for gold and silver over the frequency range from 0.5 eV (120 THz) to 6.5 eV (1.57 PHz) obtained from measured data via a partial fraction fit. The solid lines are the partial fraction fit for the real part, ε', and the dashed lines are for the imaginary part, ε'', of the dielectric function.

DOI 10.1007/978-981-4560-75-7_43-1

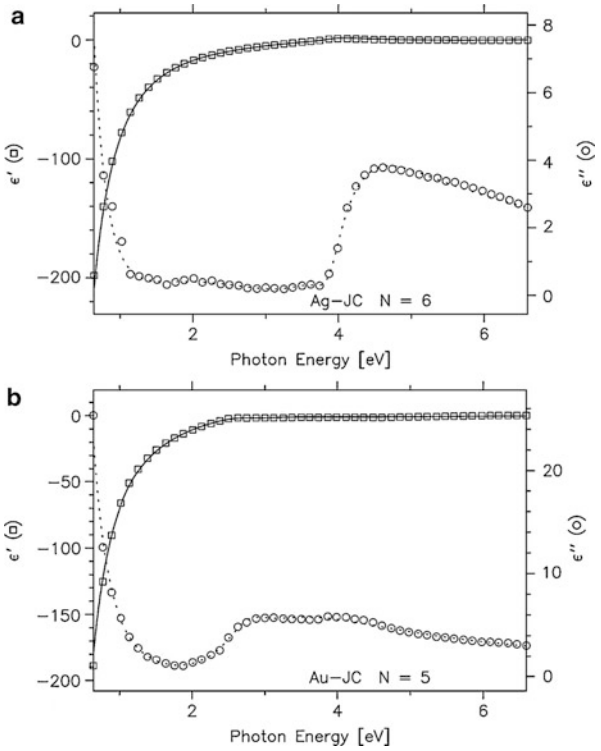


Fig. 5 Dielectric function for (**a**) Silver and (**b**) Gold as a function of photon energy $E = hf$ where f is frequency and $h = 6.626068 \times 10^{-34} \mathrm{m}^2\,\mathrm{kg/s}$ is Planck's constant. The solid lines are the partial fraction fit for ε' and the dashed line is for ε''. The circles and squares are from measured data originally supplied by (Johnson and Christy 1972)

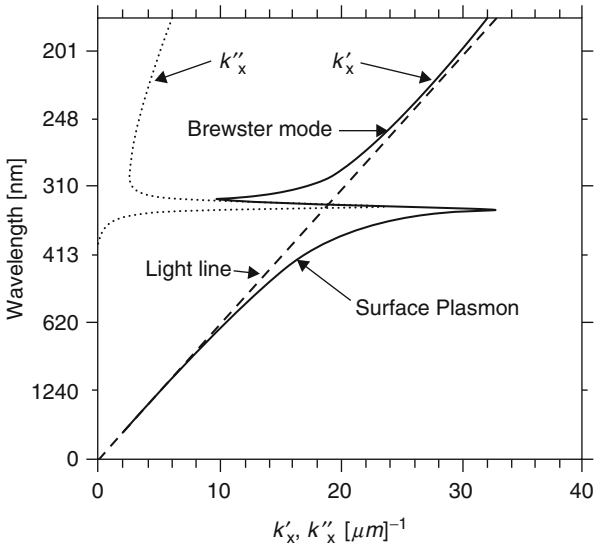


Fig. 6 SPP dispersion curve for ω versus k'_x and k''_x for an air-silver interface. The lower part of the curve is the non-radiative surface plasmon region which is capped by a high loss backbending region, and a Brewster mode radiative region above

The significance of the data in Fig. 5 becomes clearer in Fig. 6 where the surface plasmon dispersion curve for frequency versus the propagation, k'_x, and attenuation, k''_x, constants at an air-silver boundary are plotted. The lower part of the curve is the non-radiative surface plasmon region followed above by an anomalous dispersion region and above that a Brewster mode radiative region. The slanted dashed line is

DOI 10.1007/978-981-4560-75-7_43-1

the free space light line $k = \omega/c$. The Brewster mode portion of the dispersion curve is in the fast wave region to the left of the light line. It is the locus of the values of k_x of the plane wave incident on the boundary at the Brewster angle. These waves do not set up a surface wave, but rather carry their power directly into the metal without reflection. Below the Brewster mode section is a region of anomalous dispersion, often described in the literature as "backbending," which has a group velocity that is virtually zero, and is accompanied by high loss.

The surface plasmon portion of the dispersion curve is of primary importance to nanoantenna design. The plasmon region lies to the right of the light line and therefore it is a slow wave since the phase velocity here is less than the speed of light. Because surface plasmons have a shorter wavelength than light they are prevented from radiating from a flat planar surface. Of particular significance is the decreasing slope of the dispersion curve, tending toward horizontal near where it joins the anomalous dispersion segment. In the plasmon region the surface plasmon propagation constant increases dramatically with a small increase in frequency. The horizontal dashed line at 380 nm intersects the light line at approximately $k_o = 17$ rad/μm while at the same wavelength $k_x' = 21$ rad/μm, which means the surface plasmon wavelength can be much less than the free space wavelength. Therefore, a resonant half-wavelength nano-dipole antenna can be much smaller than a half-wavelength measured in air at the same frequency. As the plasmon wavelength decreases near the anomalous dispersion region the propagation constant approaches a limiting value. The Drude model provides the estimate for this limiting frequency,

$$\omega_{sp} \simeq \omega_p / \sqrt{1 + \varepsilon_{r1}} \tag{20}$$

obtained by setting the denominator in Eq. 19 equal to zero, which is equivalent to assuming the propagation constant in Eq. 19 approaches infinity. This is not what happens to the propagation constant, as can be seen in Fig. 6 where real data is used, but it does provide a reasonable estimate of the limiting frequency. Nanoantennas must be designed to operate below this frequency, thereby avoiding a group velocity approaching zero and dramatically increased losses.

Although several key aspects of wave propagation on nano-structures have been presented here, all analysis has been performed assuming propagation is on a surface between two half spaces. However, most nanoantennas have air above and are etched in a noble metal, such as gold, on a dielectric substrate, such as silicon dioxide. The metal antenna therefore lies between two different materials. Nanoantennas are typically excited at a corner or edge where both upper and lower surfaces are exposed to the incoming electromagnetic field. Because the plasmon propagation constant (7) is a function of the permittivity in the regions above and below an interface, it will be helpful to understand how different phase velocities (Eq. 11) and wavelengths (Eq. 12) on the two sides of a nano-antenna will affect its performance. This issue is briefly addressed in the following paragraphs.

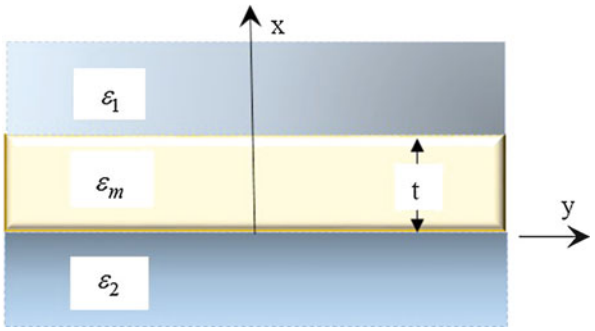


Fig. 7 Thin metallic film between two dielectric regions with different permittivities

DOI 10.1007/978-981-4560-75-7_43-1

Two Boundary Structures

Surface plasmon polaritons on a metal film have properties that do not exist on an interface between metal-dielectric half spaces. To analyze the mode configuration on a metal strip of thickness t consider the configuration shown in Fig. 7 where the dielectrics are assumed to extend from the surface of the metal to infinity and possess different permittivities, ε_1 and ε_2. The permittivity of the metal is ε_m and all three regions have the permeability of free space. The waves in each region are assumed to be TM polarized with a magnetic field component in the y-direction in each region given by

$$\mathbf{H}_1 = H_{y1} e^{-k_{z1} z - j k_x x} \hat{\mathbf{y}} \text{ dielectric region 1 } (z \geq t) \tag{21a}$$

$$\mathbf{H}_m = \left(H_{ya} e^{k_{zm} z} + H_{yb} e^{-k_{zm} z}\right) e^{-j k_x x} \hat{\mathbf{y}} \text{ metal region } m \text{ (t} \geq z \geq 0) \tag{21b}$$

$$\mathbf{H}_2 = H_{y2} e^{k_{z2} z - j k_x x} \hat{\mathbf{y}} \text{ dielectric region 2 } (z \leq 0) \tag{21c}$$

The dispersion relations for the three regions are,

$$k_{zi}^2 = k_x^2 - k_o^2 \varepsilon_i \text{ and } i \doteq 1, 2, m \tag{22}$$

and the boundary conditions satisfied by $\mathbf{H}_m$ at each interface are

$$H_{ym} = H_{yi} \tag{23a}$$

$$\frac{1}{\varepsilon_m} \frac{\partial}{\partial z} H_{ym} = \frac{1}{\varepsilon_i} \frac{\partial}{\partial z} H_{yi} \tag{23b}$$

Where $i = 1,2$. Each of these conditions are applied to Eq. 21 at each boundary yielding a set of four homogeneous linear equations which can be combined to produce

$$\frac{H_{yb}}{H_{ya}} = \frac{\left(\frac{\varepsilon_2 k_{zm}}{\varepsilon_m k_{z2}} - 1\right)}{\left(\frac{\varepsilon_2 k_{zm}}{\varepsilon_m k_{z2}} + 1\right)}, \frac{H_{yb} e^{-k_{zm} t}}{H_{ya} e^{k_{zm} t}} = \frac{\left(\frac{\varepsilon_1 k_{zm}}{\varepsilon_m k_{z1}} + 1\right)}{\left(\frac{\varepsilon_1 k_{zm}}{\varepsilon_m k_{z1}} - 1\right)} \tag{24}$$

The equations in 24 are equated to produce the dispersion relation

$$\left(\frac{\varepsilon_m k_{z1}}{\varepsilon_1 k_{zm}} + 1\right)\left(\frac{\varepsilon_m k_{z2}}{\varepsilon_2 k_{zm}} + 1\right) = \left(\frac{\varepsilon_m k_{z1}}{\varepsilon_1 k_{zm}} - 1\right)\left(\frac{\varepsilon_m k_{z2}}{\varepsilon_2 k_{zm}} - 1\right) e^{-2 k_{zm} t} \tag{25}$$

for a metal sandwiched between two different dielectrics. Equation 25 can be verified by allowing the metal thickness to become large, $k_{zm} t \gg 0$, in which case the right hand side goes to zero. This yields the pair of equations

$$\frac{\varepsilon_m k_{z1}}{\varepsilon_1 k_{zm}} + 1 = 0, \cdots \frac{\varepsilon_m k_{z2}}{\varepsilon_2 k_{zm}} + 1 = 0 \tag{26}$$

recognized from Eq. 4 to be the dispersion equations for surface plasmon polaritons at the boundaries between two individual dielectric-metal half-spaces.

DOI 10.1007/978-981-4560-75-7_43-1

If ε_1 and ε_2 are positive and $\varepsilon_m < 0$, the right hand side of Eq. 25 is positive, which means the left hand side must also be positive. This is possible if $\left(\frac{\varepsilon_m k_{z1}}{\varepsilon_1 k_{zm}} + 1\right)$ and $\left(\frac{\varepsilon_m k_{z2}}{\varepsilon_2 k_{zm}} + 1\right)$ are both positive or both negative. Under the assumption that both are negative, that $\varepsilon_1 > \varepsilon_2$, and $|\varepsilon_m| > \varepsilon_1$ the characteristic Eq. 25 becomes (Durach et al. 2004),

$$k_{zm}t = \coth^{-1}\left(\frac{|\varepsilon_m| k_{z1}}{\varepsilon_1 k_{zm}}\right) + \coth^{-1}\left(\frac{|\varepsilon_m| k_{z2}}{\varepsilon_2 k_{zm}}\right) \tag{27}$$

Similarly, assuming that both expressions in brackets on the left hand side of Eq. 25 are positive and $|\varepsilon| < \frac{\varepsilon_1 \varepsilon_2}{\varepsilon_1 - \varepsilon_2}$ the characteristic equation becomes (Durach et al. 2004),

$$k_{zm}t = \tanh^{-1}\left(\frac{|\varepsilon_m| k_{z1}}{\varepsilon_1 k_{zm}}\right) + \tanh^{-1}\left(\frac{|\varepsilon_m| k_{z2}}{\varepsilon_2 k_{zm}}\right) \tag{28}$$

Equations 22 together with 27 or 28 can be used to solve for the wavenumbers in the even and odd mode cases, referring respectively to the symmetric or asymmetric form of the tangential (z-directed) component of the electric field about the center of the metal layer (Burke et al. 1986).

When the metal plate is thick the waves on the two sides of the metal slab do not interact, as indicated in Eq. 26. However, as the slab becomes thinner the electromagnetic fields on the two sides do interact and the frequency splits into a low frequency even mode (Eq. 27) and a higher frequency odd mode (Eq. 28). Although the tangential electric field is not entirely symmetric in the metal region when $\varepsilon_1 \neq \varepsilon_2$, more of the field lies inside the metal in the even mode case (Eq. 27) than it does in the basically asymmetric odd mode case (Eq. 28). As the metal layer becomes thinner the damping of the odd mode decreases, approximately as the square of the thickness, allowing this mode to travel a greater distance than a plasmon polariton on the interface between two half spaces. However, the basically symmetric mode damping decreases with decreasing thickness, allowing propagation distances one or two orders of magnitude greater than those of thick metal construction (Sarid 1981). The physical reason attributed to this phenomena is that in the asymmetric case, where the null is in the center of the metal, more of the field resides outside as the thickness of the metal decreases lessening the presence of the tangential electric field in the metal thereby reducing Joule heat which in turn decreases the electron collision rate resulting in less damping of the wave. This analysis suggests that the efficiency of an optical antenna can be improved by decreasing the thickness of the metallic material with which it is constructed and operating in the higher odd mode frequency range.

When the metal width is finite and the environment is symmetric ($\varepsilon_1 = \varepsilon_2$), there are four fundamental modes and a number of higher order modes. These modes can be classified according to the cross-sectional width and thickness, similar to a rectangular waveguide and will become successively cut off as the thickness decreases leaving one Gaussian like mode. Efficient optical fiber coupling to this plasmonic mode in a metal guide has been demonstrated (Berini 2000), raising the possibility of similar coupling between optical antennas and metal strip plasmonic waveguides designed with the appropriate width and thickness so as to operate in this particular mode. Closer to the nanoantenna case, in an asymmetric environment ($\varepsilon_1 \neq \varepsilon_2$) there is no pure TM mode. It has been shown that modes in this case can change their symmetry properties with changes in both transverse directions, width and thickness, and that all modes show a cutoff height that increases both with decreasing width and increasing dielectric mismatch between the sub- and superstrate (Berini 2000, 2001). Unfortunately, this includes air-metal-glass configurations commonly used in the design of optical antennas and waveguides. Thus the complicated

DOI 10.1007/978-981-4560-75-7_43-1

spatial profiles shown in modes in an asymmetric environment prohibit efficient excitation techniques (Maier and Atwater 2005; Yang et al. 1991).

In this section it has been shown that in the optical frequency range electromagnetic wave propagation, which is via a surface plasmon polariton, can take place on an air-dielectric boundary if one of the two materials has a dielectric constant with a negative real part obeying the conditions in Eq. 10. The dispersion equation for this wave is given in Eq. 7 and presented graphically in Fig. 6, which shows that a surface plasmon is a slow wave, i.e., slower than the speed of light in the dielectric medium. The Drude model, which is relatively accurate below, and not too close to, the plasma frequency ω_p and useful for analytical calculations, was given in Eq. 16. This model, or its approximation in Eq. 17, is commonly used to obtain approximate velocity (Eq. 11), wavelength (Eq. 12), propagation distance (Eq. 13), and exponential decay distance away from the boundary (Eqs. 14 and 15) of the plasmon wave. It also provides an estimate of maximum frequency (Eq. 20) at which a plasmon wave can propagate. Although Fig. 6 shows that the surface plasmon wavelength can be much less than the free space wavelength, close to the maximum frequency the plasmon wave will experience severe attenuation.

Because optical antennas are operated in a frequency range which is at the overlapping boundary between electromagnetic and quantum theories, in the paragraphs above some of the quantum justification for the unusual behavior of materials in the near infrared and optical domains has been presented. This discussion was then followed by an asymmetric two boundary case development of the dispersion equations modeling a typical optical antenna structure in cross-section, showing that the interaction between the plasmon waves on the two boundaries breaks into symmetric and asymmetric modes, (Eqs. 27 and 28). The symmetric mode is preferred when coupling to another device and the asymmetric mode is preferred for long distance propagation in a plasmon waveguide. This section then concluded with a discussion of the finite width metal guide that has been shown to have a complicated mode structure, hampering efforts to obtain a mathematical expression for impedance matching. Current attempts at obtaining a mathematical model and along with design methods that improve impedance matching and enhance optical antenna radiation will be covered in the individual antenna discussions that follow.

Design of Optical Antennas

The plasmonic nature of nanoantenna operation prevents the designer from describing antenna performance with formulas that have been developed over many decades at microwave and lower frequencies. Dielectric properties of metals are only one factor preventing a direct correspondence between established antenna theory and actual optical antenna operation. A traditional antenna concept that the wave vector for the current is the same as the free-space wave vector must be abandoned, dispersion effects must be accounted for and losses not generally expected in a microwave frequency "perfect conductor" become dependent on the size and shape of the antenna.

The optical antenna, although not a traditional perfect conductor, cannot be analyzed as though it is a pure dielectric rod as would be the case for an optical fiber. But even if an optical antenna could be modeled as a pure dielectric, the tips of a rod represent the famous unsolved dielectric wedge problem. As described above, the fast wave current on a perfect conductor must be replaced by a slow wave plasmon current tightly coupled to the antenna surface. Radiation on an optical dipole still occurs at the tips of the rods and can still be viewed as current induced. The primary difficulty that must be faced when attempting an analytical analysis is that for practical reasons the design requires placing the antenna on an interface between a dielectric below, such as a silicon dioxide substrate, and air above. Although numerical techniques can be used to accurately analyze the characteristics of the structure, similar to microstrip

DOI 10.1007/978-981-4560-75-7_43-1

patch antennas, no exact analytical methods are available. Analytical approximations and numerical attempts to obtain the basic characteristics of an optical antenna, the resonant frequency, input impedance, gain, directivity and efficiency, are presented in this section.

Dipole and Patch Antennas

Size and shape can affect the design frequency of an optical antenna in ways not typically expected in microwave frequency antenna engineering. For example a rectangular microstrip patch is expected to resonate when the incident field is polarized in the direction of the $\lambda/2$ dimension of the patch, while the transverse dimension is flexibly left to something greater or less than $\lambda/2$. However, as shown in Fig. 8 both the width and length of a metal plate must be considered in designing an optical antenna. A narrow nanodipole resonating at a particular frequency, as shown in Fig. 8a, does not maintain resonance when widened. However, the first resonance can be regained by changing both the length and width of the dipole as shown in Fig. 8b. The mechanism causing this phenomena has to do with the manner in which surface plasmons are excited. The incident wave with a free space wavenumber transfers energy into the plasmon wave having a much larger wavenumber at locations where the interaction between the incident wave and the antenna produce the largest concentration of higher order modes, which will couple energy into the plasmon mode. This interaction is primarily seen on the corners of the dipole in Fig. 8b. Similar to plane waves in a waveguide, plasmon waves fan out from the corners taking a circuitous route bouncing from the edges of the plate before phase matching produces a resonance condition. Further complicating the picture is the chemical potential and the "lighting rod" effect whereby opposite charges form a high concentration of the field across the antenna center gap. However, plasmons in the narrower dipole, Fig. 8a, follow a shorter more direct route to resonance than does the plasmon wave on the wider dipole Fig. 8b thereby experiencing less attenuation and ultimately creating a region of higher field enhancement in the gap than does the wide dipole. Figure 8c shows that as the dipole widens the resonant frequency shifts and the enhancement decreases. Enhancement is defined as the ratio of the total to the incident

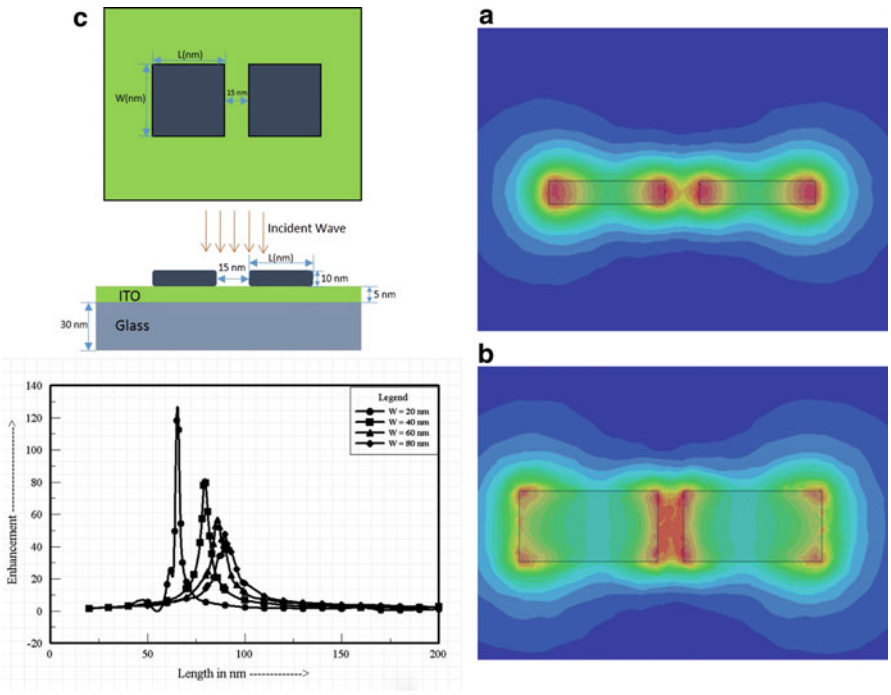


Fig. 8 Maximum enhancement of a gold nanodipole excited at 550 nm (**a**) narrow and (**b**) wide profile. A narrow structure does not maintain resonance when widened without also changing its length

DOI 10.1007/978-981-4560-75-7_43-1

electric field in the gap region, although in some publications this ratio is squared and therefore described as intensity enhancement.

The behavior of the narrower dipole shown in Fig. 8a is more closely aligned with what one would expect at microwave frequencies whereby increasing the length of each nanorod in plasmon half-wavelength intervals will produce successive resonance and antiresonance in the gap input impedance. Notice that the total length of the dipole is a full plasmon wavelength λ_p plus the gap separation distance, as opposed to the typical $\lambda_o/2$ dipole first resonance of a perfectly conducting microwave antenna with a negligible gap width. In terms of plasmon wavelengths, a first resonance does occur in an isolated $\lambda_p/2$ nanorod or a vertical $\lambda_p/4$ nanorod on a metal substrate (Taminiau et al. 2007).

Although an exact determination of the plasmonic wavelength in a dielectric nanorod is not mathematically feasible, (Novotny 2008) has derived an effective wavelength λ_{eff} based on the assumption that the antenna is a thin dielectric cylinder operating in the TM_0 cylindrical waveguide mode. The antenna is immersed in a material with dielectric constant ε_s, a Drude dielectric function is used to describe the metal rod material and the apparent increase in the antenna length at the rod ends is accounted for by an additional reactance term. This scaling law is

$$\lambda_{eff} = n_1 + n_2\left(\frac{\lambda}{\lambda_p}\right) \tag{29}$$

Where λ is the wavelength of the external region, λ_p is the metal plasma wavelength, and n_1, n_2, coefficients with dimensions of length that depend on antenna geometry and static dielectric properties, are

$$n_1 = 2\pi R[13.74 - 0.12(\varepsilon_\infty + \varepsilon_s 141.04)/\varepsilon_s - 2\pi]; n_2 = 0.24\pi \mathrm{R}\sqrt{\varepsilon_\infty + \varepsilon_s 141.04}/\varepsilon_s \tag{30}$$

In Eq. 30 R is the radius of the cylinder, and ε_∞ is the dielectric constant correction in the Drude formula for $\omega \gg \omega_p$. For gold $\varepsilon_\infty \simeq 11$, $\lambda_p = 138\,\mathrm{nm}$ and for silver $\varepsilon_\infty \simeq 3.5$, $\lambda_p = 135\,\mathrm{nm}$ (Novotny 2008).

In the case of an optical patch or dipole antenna it is necessary to account for a phase shift acquired by the surface plasmon polariton upon reflection at rod, strip, or gap terminations. The resonance wavelength λ of metal strip resonators of thickness t and width w are determined approximately by (Søndergaard and Bozhevolnyi 2007)

$$w\frac{2\pi}{\lambda}n_{slow} = m\pi - \phi \tag{31}$$

where n_{slow} is the real part of the mode index of surface plasmon polaritons propagating in a metal film with the same thickness t as the strip, $m = 1,2,3,\ldots$ is the order of the resonance, and ϕ is the phase of the reflection coefficient at strip terminations. For a symmetric structure where $\varepsilon_d = \varepsilon_1 = \varepsilon_2$, once t, and ε_m are selected and the phase shift ϕ determined, n_{slow} can be obtained from Eqs. 22 and 26 or 27 or for thin strips one can use the approximation (Søndergaard et al. 2008)

$$n_{slow} \simeq \sqrt{n_d^2 + \frac{4n_d^4}{k_0^2 t^2 n_m^4}} \tag{32}$$

where n_d and n_m are the indices of refraction of the dielectric and metal and $k_o = 2\pi/\lambda$. The wavelength of the slow surface plasmon polariton is then given by $\lambda_{slow} = \lambda/n_{slow}$.

DOI 10.1007/978-981-4560-75-7_43-1

Analytical models are preferable because clear trends in antenna performance can be seen by adjusting mathematical parameters that represent physical quantities such as the antenna width or length and the excitation frequency. On the other hand numerical techniques can provide both the accuracy and flexibility needed to model complex antenna structures. However, purely mathematical solutions can become clouded with long complex expressions and special functions while a purely numerical analysis provided by the powerful commercial codes that have become available these last few years, while furnishing useful visual data in terms of graphs and charts, do not always provide a way to understand the multifaceted interaction between the variables of an antenna system. Recently a compromise approach for optical antennas has been advanced, advocated primarily by Engheta and Alù (Engheta et al. 2005; Alù et al. 2007; Zhao et al. 2011; Agio and Alù 2013), whereby the antenna and field excitation are replaced by an equivalent lumped circuit model based on numerical analysis data. The basic model, shown in Fig. 9, is centered on a Thévenin equivalent circuit for the antenna with a capacitance accounting for a non-negligible gap impedance Z_g along with the intrinsic impedance Z_a of the dipole. The parallel combination of these two impedances forms the input impedance of the circuit. The input impedance is calculated, by driving the antenna at the gap with an arbitrary source voltage V_g and calculating, with full-wave simulations, the displacement current I_g flowing across the arms at the terminals in the region of the gap $Z_{in} = V_g/I_g$.

The intrinsic antenna impedance Z_a is extracted from the input impedance by first taking the parallel combination of the gap and load impedances according to

$$Z_{in} = \frac{1}{1/Z_a + 1/Z_g};\ Z_g = 1/j\omega C;\ C = \varepsilon_0 S/g \tag{33}$$

where S and g are respectively the gap cross-sectional area and gap height. Setting $Z_a = R_a + jX_a$ and $Z_{in} = R_o + jX_o$, Eq. 33 is rearranged to give:

$$R_a = \frac{R_o}{1 + \omega C\left(2X_o + \omega C\left(R_o^2 + X_o^2\right)\right)} \tag{34}$$

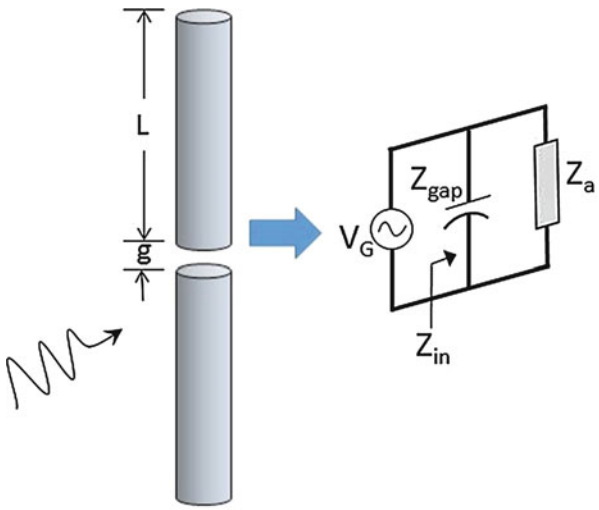


Fig. 9 Optical antenna with equivalent circuit model. The driving voltage could be an incident electromagnetic field, a transmission line connected to the antenna, or a quantum emitter

DOI 10.1007/978-981-4560-75-7_43-1

$$X_a = \frac{X_o + \omega C\left(R_o^2 + X_o^2\right)}{1 + \omega C\left(2X_o + \omega C\left(R_o^2 + X_o^2\right)\right)} \tag{35}$$

Since full wave simulations have provided the Z_{in} components R_o and X_o the dipole intrinsic impedance is determined mathematically from Eqs. 34 to 35. The intrinsic impedance Z_a, which is unaffected by antenna loading either by a transmission line or by altering the gap capacitance, is now completely determined.

The antenna is resonant when the input reactance $X_o = 0$ which from Eq. 35 is

$$\omega_0 = \frac{X_a}{C\left(R_a^2 + X_a^2\right)} \tag{36}$$

The "open circuit" or first resonance has been shown to have a very low value, whereas the "short circuit" or antiresonance has a high impedance, on the order of kilo-ohms. The feed line or other loading conditions will determine which operating frequency is selected. In the event of loading the capacitance becomes $C = \varepsilon_L S/g$, the new resistance and reactance terms in the input impedance $Z_{in} = R_{in} + jX_{in}$ are again found numerically and the equivalent dielectric constant ε_L can be calculated from Eq. 33 with ε_0 replaced by ε_L. Extra circuit elements can be added to account for the transmission feed line or a quantum radiator such as is the case in molecular spectroscopy, or a different circuit can be created for single quantum radiators without a gap (Agio and Alù 2013).

Efficiency is an important measure used to characterize antenna performance. The overall antenna efficiency can be broken down into the product of several measures of efficiency, the most important of which are the radiation efficiency and the reflection efficiency, $\eta = \eta_{rad}\,\eta_{ref}$, where the maximum value of each efficiency is one. The reflection efficiency determines how well the transmission feed line characteristic impedance is matched to the antenna input impedance. Its value, given in terms of the transmission line reflection coefficient Γ, is $\eta_{ref} = 1 - |\Gamma|^2$ where $|\Gamma| \leq 1$. The radiation efficiency is ratio of the total power radiated P_{rad} to the power received P_{in} by the antenna. The input power is assumed to be the sum of the radiated power and the power losses on the antenna. Provided the antenna is fed at the antenna maximum current point, which is most often the case with optical antennas operating at or below the lowest resonant frequency. At the first antiresonance, the radiation efficiency can be expressed in terms the radiation resistance R_{rad} and the ohmic loss resistance R_L experienced by the current flow in the antenna elements. The radiation efficiency is therefore

$$\eta_{rad} = \frac{R_{rad}}{R_{rad} + R_{in}\sin^2\left(\pi L_{eff}/\lambda_{eff}\right)}; R_{rad} = 2P_{rad}/I_{\max}^2 \tag{37}$$

The ohmic loss resistance R_{in} is the real part of the input impedance. The optical dipole radiation resistance can be determined analytically by first hypothesizing the current at resonance. Taking into account the effective length and wavelength of an optical antenna, L_{eff} and λ_{eff} respectively which can be calculated using Eq. 29, the standard microwave expression for a dipole current is modified to become (Alù et al. 2007),

$$I(z) = \frac{I_o \sin\left[\pi\left(L_{eff} - 2|z|\right)/\lambda_{eff}\right]}{\sin\left(\pi L_{eff}/\lambda_{eff}\right)} \tag{38}$$

DOI 10.1007/978-981-4560-75-7_43-1

Where I_o is the numerical displacement current at the feed point $z = 0$. This current Eq. 38 is used to find an analytical expression for the power radiated by the optical dipole P_{rad} which is then substituted into Eq. 37 to find the efficiency.

Antenna directivity defined as $D_o =$ (time average power density maximized with angle)/(average total power radiated through a sphere), can again be analytically calculated with the formulas above as can gain which is the product of the efficiency times the directivity, $G_o = \eta_{rad} D_o$. Keep in mind that this exercise has been a combination of numerical and analytical calculations, forming a hybrid approach that seems best suited for understanding complex interactions as well as for overall engineering optical antenna design. Although the example of a dipole has been used here, many other structures have been successfully investigated using this equivalent technique (Zhao et al. 2011; Agio and Alù 2013). For design data for the dipole based on measurements the reader is referred to references (Schuck et al. 2005; Fischer and Martin 2008; Muskens et al. 2007).

Bowtie Antenna

Compared to a dipole antenna, the bowtie antenna has the advantage of broadband operation yet it is a simple design. The feed point of a bowtie antenna is a gap at the vertex of two tip to tip apposal triangular electrically conducting arms. If extended to infinity while the angles of the triangle vertices remain unchanged this antenna would be frequency independent. When the incident field polarization is perpendicular to the gap, the antenna can be driven into resonance. Figure 10 shows the intensity distribution for a 520 nm resonant bowtie design made out of gold on a SiO_2 substrate and excited by a normally incident plane wave. The numerical simulation in Fig. 10a shows that a high field intensity distribution occurs not only at corners and in the gap region but also along each of the sides, likely due to the sharp edge design. In Fig. 10b a plot of the electric field vectors immediately above the surface of antenna indicate the largest field components are just above the gap region but tend to decrease in amplitude as they spread out and terminate at the wider ends of the antenna arms where a much lower intensity is seen in Fig. 10a. The intensity of the field in the bowtie gap has been shown to be strongly dependent upon the gap width (Schuck et al. 2005). For the design presented in (Schuck et al. 2005), starting on the order of 50 nm the resonant intensity climbs exponentially as the gap separation narrows, however as the gap separation widens the intensity tends to level out at around 100 nm. Although very little work has been carried out concerning mathematical design criteria for the construction of on-surface optical antennas, optimal design dimensions can be obtained through published data gathered in experiments. For optical frequency design data based on measurements for the bowtie antenna the reader is referred to reference (Fischer and Martin 2008). Studies have also been carried out to determine optimum

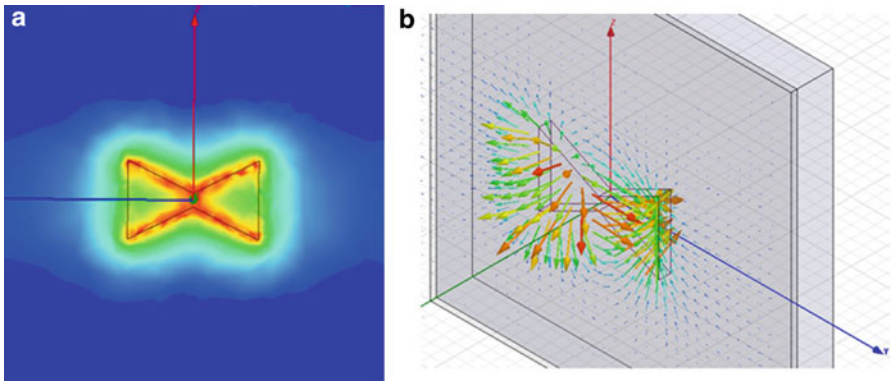


Fig. 10 Bowtie antenna with (**a**) highest field intensity shown in red around the edges and corners of the triangular arms and (**b**) electric field vectors where the largest amplitude field is displayed in red occurs in the gap between the two arms

DOI 10.1007/978-981-4560-75-7_43-1

bow angle, the outside angle formed by the edges of the two apposal arms of a bowtie antenna. It has been reported that the strongest enhancement can be obtained with a bow angle of 90° (Fischer and Martin 2008).

Unlike microstrip antennas printed on a thin substrate with a metal backing focusing radiation into the air region above, optical antennas are a single metal radiator lying on a dielectric-air interface. There can be no ground plane below because there are no perfect conductors at optical frequencies. Although radiation is in both directions, most of the energy radiated from an optical antenna tends to couple into the higher density substrate. In the microwave case a TM_0 mode surface wave can be launched into the substrate because this mode has no cutoff wavelength, but the fundamental mode in a typical microwave patch antenna is not conducive to launching TM modes. As long as the substrate is thin a microwave patch antenna suffers little substrate induced loss. As the substrate thickens energy from the microwave patch antenna is coupled into surface wave modes, seriously degrading the efficiency of the antenna. However, an optical antenna radiates in both directions regardless of the substrate thickness, making it in any case less efficient than its microwave counterpart. One measurement study carried out in the infrared range (Fischer and Martin 2008) concluded that the dipole and bowtie efficiencies are respectively 20 % and 30 % whereas a typical microstrip patch antenna has an efficiency of 60 %. A similar conclusion was reached in a more recent study (Agio and Alù 2013) that included the optical dipole, bowtie, and dimer constructions. A dimer, which has two pancake shaped arms, has an efficiency level close to the bowtie design. Numerical investigations of the optical Yagi-Uda design show a strong emission into the substrate and a much smaller emission lobe into the air side of the surface (Hofmann et al. 2007; Kosako et al. 2010). Although most of the electromagnetic radiation is emitted into the critical angle below, the above surface emission still shows a strong directivity.

Yagi-Uda Antenna

The dipole and bowtie nanoantennas described above are most often used in applications where a wide beamwidth or broadband transmission is the primary objective. However as nano-science progresses, achieving highly directional beams in the optical regime will be necessary in order to, for instance, transmit signals from nanocircuits to a receiver above the surface with a minimum power load on the electronic system. Another significant application is the possibility of enhancing and directing the radiation from a nano-emitter, which will be discussed in the Applications section below. At the microwave level the Yagi-Uda antenna shown in Fig. 11 has proven to be highly directional for radiation of radio waves yet simplistic in construction. One would expect that basically similar design principles can be used to obtain directionality with an optical Yagi-Uda design. Over the last few years this has been shown to be the case primarily due to the fact that the elements of the Yagi antenna are nanorods, the

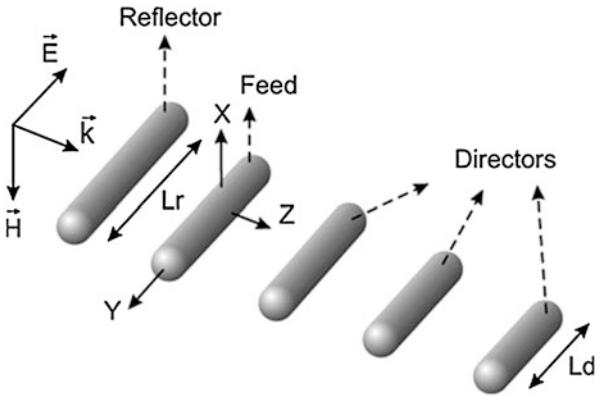


Fig. 11 Yagi-Uda optical antenna constructed with nano rods

DOI 10.1007/978-981-4560-75-7_43-1

properties of which have been the subject of extensive study by the optics community over the last 30 years.

The essential Yagi-Uda design consists of a resonant driven element flanked on one side by a longer passive reflector element and on the other by a series of shorter director elements, all of which lie in the same plane. Researchers have studied a variety of element types for the Yagi, including spheres and prolate spheroids, but low efficiency and difficulties in achieving smooth surfaces and precise positioning on the nano level has focused attention on the classical cylindrical rod design. In addition to surface smoothness and the precise rod length and positioning needed for optimum array performance, other difficulties encountered in the optical regime are the loss of energy coupled into the substrate, the difference in wavelength between free-space and the metallic rods, and field interaction with the air-substrate boundary causing a phase difference between surface plasmons on the two sides of the antenna. Naturally these effects can be modeled by pure numerical calculations but preventing unwanted consequences such as the substrate energy loss are still left open for future research. However detailed design rules have been developed (Hofmann et al. 2007) for the optical Yagi. As a rough estimate high directivity can be obtained with separation distances of about 0.25λ between the feed and the reflector and about 0.3λ between the feed and the director and between each of the director elements (Kosako et al. 2010). The length of the feed element is the effective resonant wavelength λ_{eff} of the antenna rods which is related to the incident free space wavelength λ by a simple relation $\lambda_{eff} = n_1 + n_2(\lambda/\lambda_p)$ given in Eq. 29 above with an accompanying discussion.

A scanning near-field optical microscope (SNOM) has been used to simultaneously image the amplitude and phase of the normal E-field component on a working Yagi-Uda nanoantenna (Dorfmuller et al. 2011). This data was processed to produce a time evolution animation of the fields of the antenna during reception, providing an unusual opportunity to check time domain numerical simulations of the received fields. These measurements showed that illumination of the antenna from the forward direction resulted in constructive interference of scattered light by the antenna elements which leads to a strong field enhancement in the feed element. Measurement data taken when illumination is from the rear of the antenna reveal that the destructive interference by the feed element suppresses strong fields.

Log-Periodic Antenna

To date most theoretical and experimental studies on directional optical antennas have focused on the Yagi-Uda design, shown in Fig. 11, which provides good directivity for a specific design frequency with a limited bandwidth. A possibility of improving both bandwidth and directivity may be met with a log-periodic antenna design. The name is derived from a progressive spacing between the elements of the antenna based on a logarithmic scale. A number of log-periodic design schemes have been employed at microwave frequencies. At optical frequencies several graded size and spacing designs have been proposed with elements varying in shape from spheres and ellipses to the classical elongated rod. Due to difficulties in creating precise size and shapes at the nano level most implementations are only approximations of the true log-periodic requirement.

In a log-periodic antenna arrangement (Balanis 2005) the arm lengths (l_n), spacing between arms (R_n) and rod diameters (d_n), increase logarithmically as defined by the inverse of the geometric ratio τ as given by

$$\frac{1}{\tau} = \frac{l_{n+1}}{l_n} = \frac{R_{n+1}}{R_n} = \frac{d_n}{d_{n+1}} \tag{39}$$

DOI 10.1007/978-981-4560-75-7_43-1

Another parameter associated with the array, but independent of the geometric ratio, is the scaling factor. The scaling factor is often expressed normalized by the element length and as such it is given by $\sigma = d_n/2l_n$. The half angle at the antenna apex is determined by the parameters τ and σ according to $\alpha = \tan^{-1}[(1-\tau)/4\sigma]$.

In one study (Pavlov et al. 2012) carried out measurements on an optical log-periodic antenna with gold elements on a glass substrate strictly following the design criteria above. However unlike other optical array designs, a gap was placed in the middle of each arm in order to enhance the field produced by plasmonic action. It was observed that the directionality of the antenna reaches a maximum with $N = 10$ elements but the optimum field enhancement value of $|E|^2_{FB}$ for the forward beam electric field is obtained between $N = 6$ and $N = 10$. A decrease of the intensity enhancement values, defined as $|E_{total}|^2/|E_{inc}|^2$ with increasing numbers of elements is associated with increasing plasmonic losses incurred in the addition of elements. Adding a gap did not affect the antenna patterns, however a significant increase in beam intensity was observed. As with other on-surface plasmonic antennas a significant portion of the radiated power is coupled into the substrate.

One of the drawbacks of the log-periodic design is that the currents in the optical version have the same phase in each element. In addition the elements are closely spaced producing a phase progression of the current in the direction of the longer elements. This produces an endfire beam in the direction of the longer elements and interference effects in the pattern. The standard design for a log-periodic antenna calls for crisscrossing the feed between adjacent elements thereby adding a 180^o degree phase shift to the terminal of each element. Since the shorter elements are spaced closer together and opposite in phase, very little energy is radiated from them and interference effects are minimal. The radiation pattern in this crisscross case tends to be toward the shorter elements (Balanis 2005). In the previous paragraph the design calling for dipoles rather than solid rods appears to be a step in the right direction. The plasmonics of optical frequency materials and construction methods on the nano level will determine the efficacy of adapting the standard microwave design where alternate dipoles are connected.

Subwavelength Particles

The spheroid is the basic subwavelength plasmonic particle employed in optical antennas. It is usually described in terms a major axis with diameter $2a$ and two minor axes with diameters $2b$ and $2c$. The electrical behavior of metallic particles can deviate remarkably from their bulk form, particularly in terms of their resonant behavior when illuminated by an electromagnetic field. Typically the imaginary part of a bulk form dielectric function will go through a maximum at resonance. However, the imaginary part of the dielectric function of a metal particle will have no maximum, but rather decreases monotonically in magnitude as the frequency increases. Nevertheless, when $\varepsilon_m = -2\varepsilon_s$, where ε_m is the real part of the particle dielectric function and ε_s is the dielectric constant of the surrounding medium, a small spherical metal particle will have a peak in its absorption cross-section at ω_F, which corresponds to the particle resonant or Fröhlich frequency (Bohren and Huffman 2004). If the damping coefficient is small the real part of the Drude model dielectric function in Eq. 16 gives a resonant frequency

$$\omega_{\mathrm{F}} = \frac{\omega_p}{\sqrt{1+2\varepsilon_s}} \tag{40}$$

for a small spherical particle. For metallic ellipsoids the absorption cross-section is a maximum when

$$\omega_{\mathrm{F}} = \omega_p \sqrt{\frac{L}{\varepsilon_s - L(\varepsilon_s - 1)}} \tag{41}$$

DOI 10.1007/978-981-4560-75-7_43-1

where L is a geometric function that takes into account the shape of the ellipsoid. Most nanoantennas are prolate spheroids, which have the property $b = c$ and a geometric function

$$L = \frac{1-e^2}{e^2}\left(-1 + \frac{1}{2e} ln \frac{1+e}{1-e}\right); e = 1 - \frac{a^2}{b^2} \tag{42}$$

The shape of a prolate spheroid ranges from a needle ($e = 1$) to a sphere ($e = 0$). For a sphere $L = 1/3$, which reduces Eqs. 40 to 39. For a more detailed discussion and formulas for other shapes, such as disk and associated variations, see (Bohren and Huffman 2004).

Various particle shapes have been discussed in the literature with the idea of creating an enhanced field concentration. Typically single solid particles, for example constructed with gold or silver, have the low field enhancement (see Fig. 1d). It was later found that dielectric spheres covered with a gold or silver shell exhibited stronger enhancement of the incident field. However no single particle has shown the promise of exceptionally high field concentration seen in the dimer configuration pictured in Fig. 12a.

Although enhancement schemes with multiple nanoparticles produce strong fields between particles many applications require the concentration of strong fields at locations where the antenna structure is removed from the high intensity area of the enhanced field. To this end there has been a good deal of research on self-similar particles. The Yagi-Uda and log-periodic antennas are two examples. Another example is shown in Fig. 12b, c, likely motivated by the Yagi-Uda design. When two self-similar gold spheres containing different radii are connected together, an enhanced field is produced at the end of the smaller sphere as shown in Fig. 12b, c is a continuation of this idea with additional spheres containing successively smaller radii. Although successful in design due to a considerably improved enhancement exterior to the tip, this antenna, like many free standing designs, does not offer enhancements exceeding that of the gold coated cone tip commonly used in SNOM applications described below.

Purcell Factor

Light emission from an optical quantum emitter such as a quantum dot or a nano-crystal can be enhanced when in close proximity to an intense electromagnetic source. This effect was first discovered at radio frequency by E. M. Purcell in 1946, when he observed an increase in the probability of spontaneous emission of an emitter placed in a resonator (Purcell 1946). However, controlling the degree of enhancement of a quantum emitter remains a major challenge. Optical microcavities or microresonators such as optical nanoantennas, generally used for emission enhancement can be mathematically quantified by what has become known as the Purcell factor F (Vahala 2003), defined as:

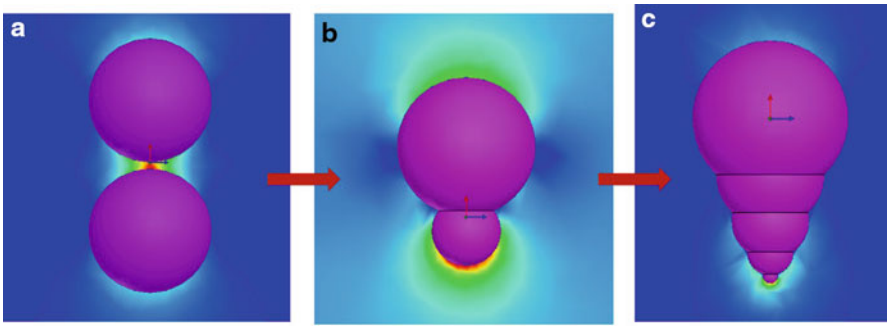


Fig. 12 Spherical particle nanoantennas arranged in (**a**) a two-sphere dimer configuration and as (**b**) two and (**c**) multiple attached self-similar spheres. An enhanced field region is seen to exist external to the self-similar configuration at the end of the smallest sphere

DOI 10.1007/978-981-4560-75-7_43-1

$$F = \frac{3}{4\pi^2}\left(\frac{\lambda}{n}\right)\left(\frac{Q}{V}\right) \tag{43}$$

where λ is the wavelength of the emitted photon, Q and V are the quality factor and mode volume respectively, and n is the refractive index of the surrounding medium. The Q corresponds to the temporal confinement while the mode volume V represents the spatial confinement. Although cavity based structures can provide very high Q (~10^4) (Song et al. 2005), they are consequently narrow bandwidth devices. This spectral limitation is a problem for contemporary solid-state quantum emitters which, despite many efforts in research, still have a broad spectral emission due to stability issues (Gaebel et al. 2004).

Optical nanoantennas, on the other hand have a much lower Q (~100) (Curto 2013) and hence are suitable for applications involving broadband sources. However, due to surface plasmon polaritons, they have the capability to highly confine the light source volume well beyond the diffraction limit (Maier 2006; Barthes et al. 2011). Coupling optical nanoantennas with quantum structures like quantum dots can result in highly directive quantum emitter intensity pattern, but with a mode volume as low as 0.002 $(\lambda/n)^3$ (Curto 2013) and therefore, a very high Purcell factor.

Aperture Antennas

So far in this chapter on-surface optical antennas have been emphasized primarily because there is considerable current scientific and industrial interest in the design of antennas for optical and near infrared communications. However, there is also significant industrial interest in antennas that are effective in focusing light to sub-wavelength dimensions. A primary application lies in the possibility of extending current photolithographic techniques to produce circuit chips on a sub-wavelength scale using relatively inexpensive optical equipment as compared to ion beam etching machines. Nano-aperture antennas have been the key component in sub-wavelength lithographic research and also been a key element in the development of several other technologies that will be discussed in the Applications section. Below a brief treatment of the theory of optical aperture antennas is presented followed by an overview of selected designs that have received popular attention in the literature.

Optical Aperture Antenna Theory

Although important nano level aperture optics was carried out in the study of fluorescence molecules in the 1980s (Fischer 1986), the report by (Ebbesen et al. 1998) of dramatically enhanced transmission through sub-wavelength holes in metal films, was an event that spurred widespread interest in optical aperture antennas. Because this phenomena, now known as extraordinary optical transmission (EOT), seemed to fly in the face of the widely accepted classical electromagnetic analysis by (Bethe 1944) and (Bouwkamp 1950), there has been considerable scientific interest in the subject reaching to the core of surface plasmon theory. According to the classical theory, the power transmitted through an aperture in an infinitely thin metallic screen scales as the inverse fourth power of the aperture size in terms of wavelengths. If the screen is made to be of finite thickness, then an additional reduction in the transmitted field strength will result, due to the below-cutoff dimensions of the slot. However, it was discovered, first experimentally and then confirmed by numerical analysis, as shown in Fig. 13, that for metals at optical frequencies there are situations where remarkable amounts of energy pass through sub-wavelength slits and holes. Figure 13 illustrates a cross-sectional view of a metallic sheet made of silver with two slots passing through the sheet. This nanostructure is excited by a plane wave from below. The magnitude of the field intensity shows a complicated structure with significant field at the slot corners, an interaction resulting in strong standing waves in the horizontal direction below, some weak field transmission through the slots, and a standing wave pattern on top of the center plate.

DOI 10.1007/978-981-4560-75-7_43-1

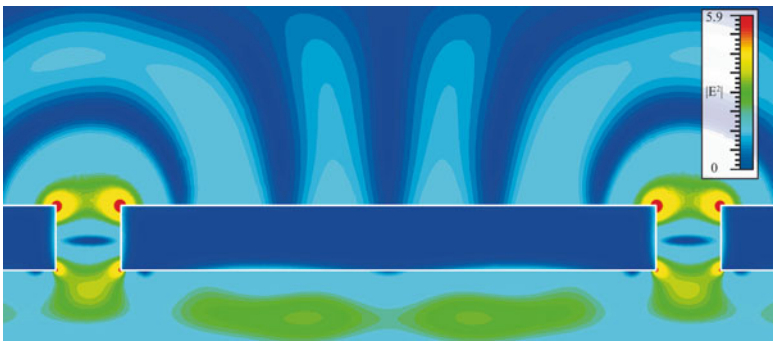


Fig. 13 Numerical calculation (CST©) showing a surface plasmon standing wave E^2 pattern on a thin silver plate with dielectric constant $\varepsilon_r = -18.242 + j1.195$ between two slots illuminated by a 660 nm plane wave from below. The center plate has a length $5/2\lambda_{spp}$ (1,650 nm) and each aperture has a width and thickness of $\lambda_{spp}/3$ (200 nm)

Although (Ebbesen et al. 1998) explained EOT in terms of surface plasmon polaritons, others (Lezec and Thio 2004; Gay et al. 2006) questioned the existence of surface plasmons and asserted that the observed phenomena could be explained by the existence of evanescent waves arising in a classical electromagnetics analysis not used by Bethe and Bouwkamp. However, a more detailed study that evolved through a series of technical papers confirmed the original EOT hypothesis (Lalanne and Hugonin 2006; Nevels and Michalski 2014). In the following a purely mathematical analysis is presented revealing the nature of transmission through sub-wavelength holes including the surface plasmon, a space wave in the vicinity of the aperture, and a lateral wave behavior on the metal surface at distances over 100 surface plasmon wavelengths removed from the aperture. The importance of this analysis lies in the fact that surface plasmon radiation plays a more significant role in optical aperture radiation than does diffraction. Also because many applications require optical apertures with narrowly focused beams it is important to know that significant energy is carried in surface plasmon polaritons on the surface of the metal on the side opposite the aperture excitation.

A thin slot aperture is modeled by placing a 2-D magnetic line current on the air-metal interface along the y-axis perpendicular to the page, as shown in Fig. 1. This arrangement gives rise to a transverse magnetic (TM) field, which has three non-zero electromagnetic field components H_y, E_x and E_z. A straightforward analysis yields the magnetic field along the air-metal boundary:

$$H_y(x) = -\frac{k_o}{2\pi\eta_o}\int_{-\infty}^{\infty}\tilde{G}(k_x)e^{-jk_x x}dk_x \tag{43}$$

$$\tilde{G}(k_x) = \frac{1}{\tilde{D}(k_x)}, \tilde{D}(k_x) = \frac{k_{z2}}{\varepsilon_2} + \frac{k_{z1}}{\varepsilon_1}, k_{z1,2} = \sqrt{k_{1,2}^2 - k_x^2} \tag{44}$$

where $\tilde{G}(k_x)$ is the Green function in terms of the spectral variable k_x, ε_1 is the dielectric constant of the upper half-space and ε_2, containing a negative real part, is the dielectric constant of the metal lower half-space. Assuming the permeabilities of the two regions are the same as that of free-space, the location of the surface wave pole k_{xp}, which can be found analytically by setting $\tilde{D}(k_x)$ to zero, is $k_{xp} = k_o\sqrt{\varepsilon_1\varepsilon_2/(\varepsilon_1+\varepsilon_2)}$ as predicted by Eq. 6. The corresponding residue is

DOI 10.1007/978-981-4560-75-7_43-1

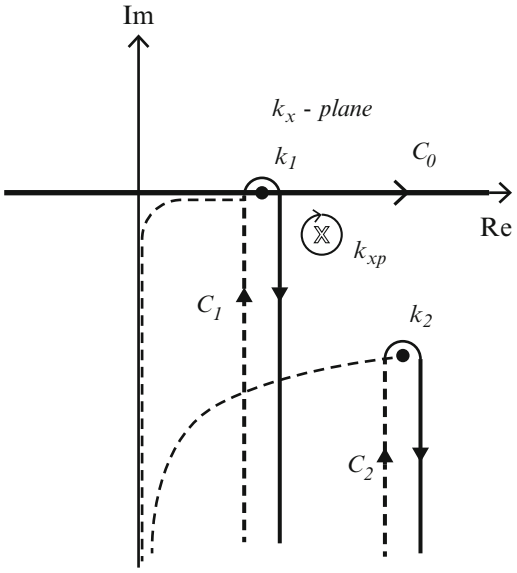


Fig. 14 Complex plane integration path around branch cuts connecting the branch points k_1 and k_2, the wavenumber for air and metal respectively, and the SPP pole k_{xp}

$$R_p = \frac{1}{\tilde{D}'(k_{xp})}, \tilde{D}'(k_x) = -k_x\left(\frac{1}{\varepsilon_2 k_{z2}} + \frac{1}{\varepsilon_1 k_{z1}}\right) \tag{45}$$

The spectral domain Green function $\tilde{G}(k_x)$ has branch points at k_1 and k_2, respectively the wavenumbers of air and the metal at optical frequencies. Air is lossless, so k_1 is real and therefore lies on the real axis, as shown in Fig. 14, whereas metal has significant loss, so k_2 is far removed from and below the real axis. The field can be obtained by integrating in the k_x-plane along the real axis, path C_0, from minus to plus infinity by means of the definitions $\mathrm{Im}(k_{z1}) < 0$ and $\mathrm{Im}(k_{z2}) < 0$ on the top sheet of the complex plane for both branch cuts. Integration on the real axis can be improved upon by deforming the integration path vertically so that C_0 is replaced by C_1 and C_2, paths of steepest descent in the k_x-plane. An important consequence of the optical properties of a noble metal is that the position of the pole k_{xp} is such that it is captured by the vertical path deformation. Had the real part of the dielectric constant of metal been positive, the pole would reside to the left of the integration path below k_1 and therefore it would not be captured by the integration path. It has been shown that poles captured by the steepest descent path are physical waves, in this case a surface plasmon wave, whereas poles not captured are not actual waves although they do influence the behavior of the field through their proximity to the integration path (Collin 2004). The contribution from the integral C_2 around the branch point at k_2 is negligible because the branch point lies well below the real axis so exponential decay due to the imaginary part of k_x is significant. The spatial domain Green function can now be expressed in terms of its pole and branch point components as,

$$G(x) = \underbrace{\int_{C_1} \widetilde{G}(k_x)\, e^{-jk_x x} dk_x}_{\text{Composite Wave}} + \underbrace{\int_{C_2} \widetilde{G}(k_x)\, e^{-jk_x x} dk_x}_{\text{Negligible}} - \underbrace{2\pi j R_p e^{-jk_{xp}x}}_{\text{SPP Wave}} \tag{46}$$

A careful analytical study of the composite wave shows that subtracting the surface plasmon polariton pole from the C_1 integrand yields a negligible result. However, when the pole is added back the resulting analytical expression yields the following limiting case forms (Nevels and Michalski 2014),

DOI 10.1007/978-981-4560-75-7_43-1

$$I_p(x) \sim \begin{cases} x^{-1/2} \text{ for small and moderate distance } x \\ x^{-3/2} \text{ for sufficiently large distance } x \end{cases} \tag{47}$$

At small or moderate distances from the aperture (magnetic line source) the magnetic field has a $x^{-1/2}$ behavior, the same as the small argument behavior of a Hankel function for a line source in free space and is often referred to as the "space wave". At large distances the $x^{-3/2}$ decay is typical of lateral wave behavior.

Numerical results displaying the composite wave (solid line) and surface plasmon polariton wave (dashed line) where the metal is silver and the line source is radiating at 633 nm and 2,500 nm are shown in Fig. 15 below. As predicted the dominant behavior of the magnetic field near the aperture is the $x^{-1/2}$ space

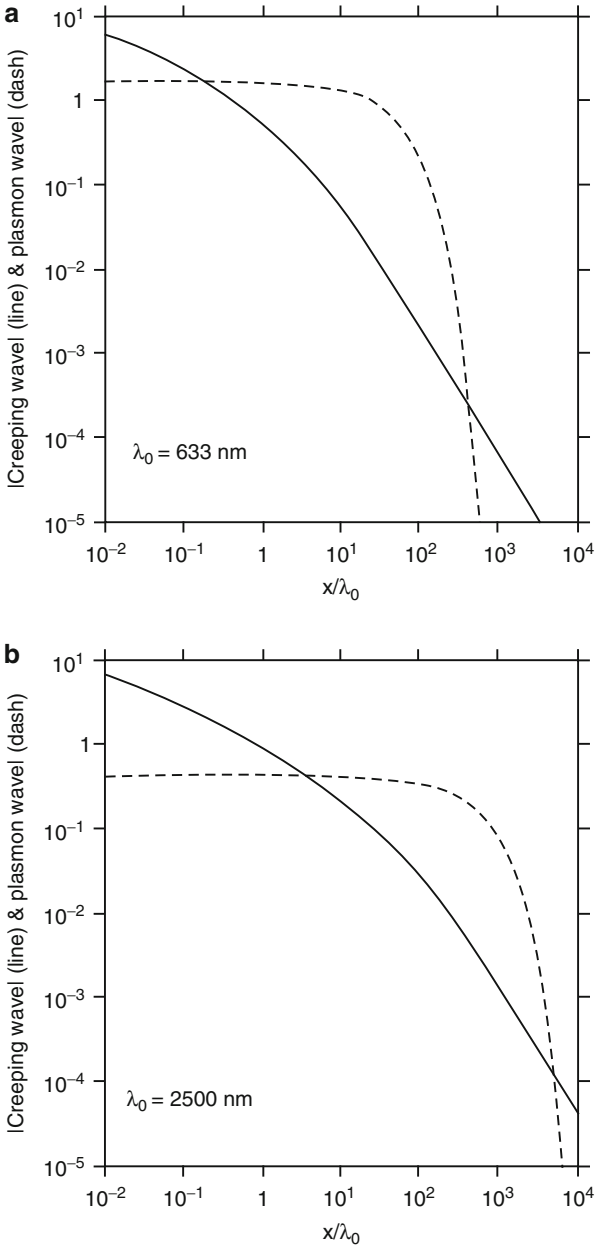


Fig. 15 Logarithmic scale comparison of composite wave and surface plasmon polariton wave normalized amplitude for an air-silver interface as a function of distance from the source at (**a**) 633 nm (**b**) 2,500 nm

DOI 10.1007/978-981-4560-75-7_43-1

wave. However at greater distances the plasmon wave becomes dominant, and in Fig. 15a close to 500 wavelengths where the plasmon wave tends to rapidly decay the lateral wave behavior $x^{-3/2}$ dominates. Figure 15b shows that at 2,500 nm the surface wave amplitude is reduced due to losses that enter the numerical calculation through the dielectric function given in Fig. 5. Eventually, at frequencies below the near infrared range, the metal begins to behave more like a perfect conductor and the space wave completely dominates the behavior of the field.

This exercise confirms the existence of the plasmon wave and displays its relationship to other wave phenomena produced in the aperture of a sub-wavelength slot and on the air-metal interface. More important from an antenna engineering standpoint is that energy carried by the plasmon wave through the slot does radiate into the space wave above the slot. The question, "Can a plasmon wave pass through a slot?" is answered by the analysis at the beginning of this chapter and in an excellent review covering analytical methods used to determine the properties of subwavelength apertures (Garcia-Vidal et al. 2010). Because a surface plasmon wave is tightly bound to the metal surface, decaying to $1/e$ at distances far less than a wavelength above the metal walls, its expression in a waveguide is not governed by the typical waveguide boundary conditions. Plasmon waves in the two walls of the slot, although not entirely independent of one another, are not subjected to cutoff conditions and can therefore carry much more energy than a waveguide evanescent mode or even an infinite sum of evanescent modes. The conclusion that can be drawn from this analysis is that at optical wavelengths a significant amount of plasmon energy can be transported through a metallic slot with a sub-wavelength width. Because the slot is sub-wavelength in width it is therefore possible to form an array of plasmonic radiators to focus an intense beam to a sub-wavelength line or spot.

Bowtie Aperture Antenna

Many aperture antennas are a simple Babinet equivalent of their metal counterpart. The aperture bowtie, also known as the diabolo antenna, is one such design. The aperture bowtie shown in Fig. 2c differs from the well-known metal bowtie nanoantenna in that the opposing pair of aperture triangles are connected through an aperture extension of the facing triangular tips. Thus instead of a large charge density accumulating between the triangles across the air gap, a high optical current density develops within the gap but between the narrowly spaced metal sides of the gap. A high intensity enhanced field component is achieved in the gap connecting the apposal triangular arms by polarizing the incident light parallel to the narrow gap connecting the arms. Polarizing the incident field across the narrow dimension of the gap will produce little or no field enhancement. One of the interesting properties of the aperture bowtie is that the magnetic field rather than the electric field is enhanced in the gap. By analogy, the metal bowtie behaves more like an electric dipole and the aperture bowtie more like a magnetic dipole. Numerical simulations have shown a 2,900-fold enhancement of the magnetic field at a wavelength of 2,540 nm, confined to a 40-by-40 nm region near the center of the nanoantenna.

Although in general aperture antenna enhancement is not as high as can be achieved with a similar metallic structure, apertures are easier to make and have the advantage of providing a visual objective. The aperture bowtie has been shown to have sufficient enhancement and confinement to be effective in nanolithography applications (Wang et al. 2006). The effectiveness of the aperture bowtie nanoantenna in nanolithographic applications is attributed to the fact that the magnetic field created in the aperture gap above the photoresist mask enters the mask in the perpendicular direction, easily penetrating into the metal, thereby leading to noticeable dissipation effects (Grosjean et al. 2011).

Optical tweezers are instruments that use light to move microscopic particles. Most often this requires a high powered focused laser beam to provide the necessary force to trap and move the particle. Optical tweezers are often needed in biological applications and are expected to have important applications in the construction of nano-machines. A variant on the bowtie design has been shown to create an optical vortex

DOI 10.1007/978-981-4560-75-7_43-1

in the electromagnetic field (Kang et al. 2011). The unusual field pattern is such that a small particle moving away from the center of the vortex experiences increasingly higher forces pointing back towards the center. The aperture bowtie provides the user the ability to see the particle while manipulations are carried out.

Concentric Rings

Perhaps the most useful aperture antenna enhancement design is a concentric ring structure. In cross-section the ring has the appearance of a shallow trough etched into the metallic surface. Concentric rings have been incorporated around a variety of aperture designs with the goal of concentrating the transmitted signal into a narrow beam or to gain maximized reception. A single radiating aperture must be large in terms of wavelengths and contain a flat planar phase-front in order to concentrate the signal. A common example at microwave frequencies is a parabolic dish antenna which in practice will have an aperture size of at least ten wavelengths. However, a typical single sub-wavelength aperture antenna will have a large beamwidth. In order to provide directionality, an optical aperture antenna can be modified by adding the concentric ring structure which in essence provides a Bragg reflection condition for plasmon current waves traveling out from center. Another way of looking at the concentric ring design is that it provides an external impedance match condition for the source driven aperture.

Again, because metal at optical frequencies has a behavior similar to that of a dielectric, at least in terms of boundary conditions, numerical solutions provide the only route to design criteria for concentric rings. The number of grooves, groove width, groove depth, periodicity of the grating, hole diameter, and the distance between the first groove and the driven aperture all play a role and are strongly interlinked in the design. However, a basic optimization criteria for bull's eye structures has been formulated based on extensive numerical studies. What follows is repeated from an excellent discussion on the subject, which can be found in more detail in a paper by (Mahboub et al. 2010): The first item to define is the desired resonance wavelength which in turn determines the period of the structure. The choice of the hole diameter will then be determined by whether one would like optimal efficiency as normalized to hole area or highest absolute transmission. For the former, the diameter should be about half the period, but for the latter the aperture size can be increased. It should be noted that as the hole size increases relative to the period, the spectrum will eventually broaden which is a trade-off. The groove width should also be around half the period with a depth to width ratio at 0.4. The number of grooves should be just enough to reach saturation, typically around six to ten grooves depending on the geometrical parameters.

Arrays

Many other specialized aperture designs have been reported in the literature, although design information is scarce. Usually the designer can refer to the current microwave antenna literature as a starting point for a numerical parameter investigation to determine that proper dimensions for an optical antenna design. External to the antenna the array factor and standard microwave array methods will typically be applicable to optical antenna array design, assuming the aperture field information is known. That is, a computer code must be used to determine the near-field and a suitable array factor can be applied to the aperture near-field information to find the far-field. The array elements can be square or circular or any one of the antennas discussed above. Rectangular shaped apertures are of particular interest since they are the Babinet equivalent of a dipole antenna. The half-wavelength resonance condition applies to the slot antenna of high conductivity and a half-wave slot leads to a resonantly enhanced and aperture confined electric field (Park 2009).

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_43-1

Applications of Optical Antenna

SNOM

Scanning near-field optical microscopy (SNOM), an important technique for visualizing biological systems, involves obtaining high resolution topographic and optical images with a cone shaped probe that scales down to the nano level at the tip. The need for high spatial, temporal and resolving power with subwavelength resolution was the driving factor in a move to the fiber aperture probe, essentially a tapered optical fiber with a nanoscale tip. Evanescent or non-propagating fields that exist only near the surface of the object carry the high frequency spatial information about the object and have intensities that drop off exponentially with distance from the object. Because of this, the detector must be placed very close to the sample in the near-field zone, typically a few nanometers. However, the aperture probe suffers from diffraction of light at the tip, which reduces its imaging resolution. Present research is centered on finding an antenna design that reduces the fiber beamwidth, thereby improving resolution. Although fabrication of tips true to the design criteria and surface smoothness are significant challenges to be overcome, a number of antenna structures such as the bowtie, Yagi and several plasmonic nanosphere probes have been tried out showing successful improvement in many cases.

Photon Emitters and Optical Fluorescence

Single photon emission from a quantum emitter generates a stream of photons. Single photons can be used in a number of technologies such as Computed Tomography, which is an imaging technique able to provide 3D information. A single quantum emitter positioned inside the subwavelength size feed gap of an optical antenna couples to the antenna, radiating in such a way as to be considered to be an impedance matching device. The density of the emitted photons thereby increases, improving the tomographic image.

Solid state light-emitting devices are expected to eventually replace fluorescent tubes as illumination sources. Quantum dot nanocrystals are very promising for light-emitting sources. However, their light-emission efficiencies are still substantially lower than those of fluorescent tubes. It has been observed that quantum emitters in close proximity to a plasmon producing metal surface causes plasmonic fluctuations of the free electron gas. The associated currents radiate, in many cases with a significant increase in intensity. Antenna structures such as the bowtie have been designed, optimized in size, shape, and material properties to increase the radiation efficiency of a nanoscopic optical source. However, for visual appeal, sources with greater optical bandwidth are needed (Farahani 2006; Curto et al. 2010).

Raman Spectroscopy

Molecules vibrate, rotate and translate in a number of ways when exposed to an electromagnetic field. Raman spectroscopy is an optical technique used to detect the vibrational modes. It relies upon detection of a very low intensity portion of the scattered wave spectrum called inelastic or Raman scattering. Because molecules of a particular type, such as anthrax or mold aflatoxins, have a certain Raman spectra they can be identified by this method. However, because the Raman spectra is very small, very intense laser light is required in order to move the Raman signal above the noise level. Plasmonic nanoantennas create highly enhanced local fields when pumped resonantly, leading to increased Raman scattering (Felidj et al. 2003).

Communication with Nanocircuitry

As circuit chips become smaller and processer speeds increase, wire leads to the circuitry become less practical and less efficient. However, nanoantennas can overcome these drawbacks, with more efficient and on-chip created optical antennas communicating over wireless links to external circuitry. Cross-talk

DOI 10.1007/978-981-4560-75-7_43-1

and the many problems with computer interconnects can be eliminated. One of the properties of on-surface optical antennas is that most of the radiated energy travels downward into the chip substrate. This can become an advantage where the construction of a via through the nanochip is impractical. A proper nanoantenna design would allow the circuitry to rapidly communicate back and forth directly through the substrate.

Conclusion

In this chapter an overview of the field of optical antennas has been presented along with some of the basic theory of plasmonics, the underpinning concept governing electromagnetic wave behavior at optical frequencies. It was shown that surface plasmon polaritons lead to a set of requirements and restrictions placed on the design of antennas constructed out of noble metals at optical and near infrared frequencies. An unusual aspect of plasmonic waves in metals is their subwavelength behavior, an advantage when detecting small particles but it can cause added difficulty when impedance matching an optical antenna to free space. It was shown that the surface plasmons explain the unusual amount of radiation coupled through a subwavelength aperture and that as a consequence aperture optical antennas are remarkably efficient, behaving in a manner similar to their Babinet equivalent counterpart. Spectroscopy, disease and toxin sensors, wireless communication with nano-circuitry and the creation of nano-circuits using subwavelength lithography are some of the early beneficiaries of this new technology. The field of nanoplasmonics offers great opportunities for optical frequency circuit and antenna engineers with the future easily as bright as it is for the modern day wireless engineer.

Cross-References

- ▶ Antennas in Medical Diagnosis and Treatment Systems
- ▶ Commercial Antenna Design Tools
- ▶ Numerical Modeling in Antenna Engineering
- ▶ Photomixer Antennas at Terahertz
- ▶ Radiation Efficiency Measurements of Small Antennas
- ▶ Terahetz Antennas and Measurement

References

Adato R, Yanik AA, Altug H (2011) On chip plasmonic monopole nano-antennas and circuits. Nano Lett 11:5219–5226

Agio M, Alù A (2013) Optical antennas. Cambridge University Press, Cambridge

Alù A, Salandrino A, Engheta N (2007) Coupling of optical lumped nanocircuit elements and effects of substrates. Opt Express 15(13):865–876

Archambault A, Teperik TV, Marquier F, Greffet JJ (2009) Surface plasmon Fourier optics. Phys Rev B 79:195414

Arduini F, Amine A, Moscone D, Palleschi G (2010) Biosensors based on cholinesterase inhibition for insecticides, nerve agents and aflatoxin B1 detection (review). Microchim Acta 170:193

Balanis CA (2005) Antenna theory analysis and design. Harper and Row, New York

DOI 10.1007/978-981-4560-75-7_43-1

Barthes J, Des Francs GC, Bouhelier A, Weeber JC, Dereux A (2011) Purcell factor for a point-like dipolar emitter coupled to a two-dimensional plasmonic waveguide. Phys Rev B 84(7):073403

Berini P (2000) Plasmon-polariton waves guided by thin lossy metal films of finite width: bound modes of symmetric structures. Phys Rev B 61:10484–10503

Berini P (2001) Plasmon-polariton waves guided by thin lossy metal films of finite width: bound modes of asymmetric structures. Phys Rev B 63:125417–125432

Bethe HA (1944) Theory of diffraction by small holes. Phys Rev 66:163–182

Bohren CF, Huffman DR (2004) Absorption and scattering of light by small particles. Wiley-VHC, Weinheim

Born M, Wolf E (1970) Principles of optics. Pergamon Press, Oxford

Bouwkamp CJ (1950) On Bethe's theory of diffraction by small holes. Philips Res Rep 5:321–332

Burke JJ, Stegeman GI, Tamir T (1986) Surface-polariton-like waves guided by thin, lossy metal films. Phys Rev B 33:5186–5201

Collin RE (2004) Hertzian dipole radiating over a lossy earth or sea: some early and late 20th-century controversies. IEEE Antennas Propagat Mag 46:64–79

Curto A (2013) Optical antennas control light emission. PhD thesis, ICFO – The Institute of Photonic Sciences

Curto AG, Volpe G, Taminiau TH, Kreuzer MP, Quidant R, van Hulst NF (2010) Unidirectional emission of a quantum dot coupled to a nanoantenna. Science 329:930–932

Dorfmuller J, Dregely D, Esslinger M, Khunsin W, Vogelgesang R, Kern K, Giessen H (2011) Near-field dynamics of optical Yagi-Uda nanoantennas. Nano Lett 11:2819–2824

Durach M, Rusina A, Ipatova IP (2004) Surface polaritons in layered semiconductor structures, section nanostructured materials – electronics, optics and devices, the 2nd Joint German-Russian Advanced Student School (JASS), St.-Petersburg

Ebbesen TW, Lezec HJ, Ghaemi HF, Thio T, Wolff PA (1998) Extraordinary optical transmission through sub-wavelength hole arrays. Nature 391:667–669

Engheta N, Salandrino A, Alù A (2005) Circuit elements at optical frequencies: nanoinductors, nanocapacitors, and nanoresistors. Phys Rev Lett 95:095504

Farahani JF (2006) Single emitters coupled to bow-tie nano-antennas. PhD thesis, University of Basel, Germany

Felidj N, Aubard J, Levi G, Krenn JR, Hohenau A, Schider G, Leitner A, Aussenegg FR (2003) Optimized surface-enhanced Raman scattering on gold nanoparticle arrays. Appl Phys Lett 82:3095–3097

Fischer UC (1986) Submicrometer aperture in a thin metal film as a probe of its microenvironment through enhanced light scattering and fluorescence. J Opt Soc Am B 3:1239–1244

Fischer H, Martin OJF (2008) Engineering the optical response of plasmonic nanoantennas. Opt Express 16:9144–9154

Gaebel T, Popa I, Gruber A, Domhan M, Jelezko F, Wrachtrup J (2004) Stable single-photon source in the near infrared. New J Phys 6(1):98

Garcia-Vidal FJ, Martin-Moreno L, Ebbesen TW, Kuipers L (2010) Light passing through subwavelength apertures. Rev Mod Phys 82:729–787

Gay G, Alloschery O, Viaris de Lesegno B, O'Dwyer C, Weiner J, Lezec HJ (2006) The optical response of nanostructured surfaces and the composite diffracted evanescent wave model. Nat Phys 2:262–267

Grosjean T, Mivelle M, Baida FI, Burr GW, Fischer UC (2011) Diabolo nanoantenna for enhancing and confining the magnetic optical field. Nano Lett 11:1009–1013

Hofmann HF, Kosako T, Kadoya Y (2007) Design parameters for a nano-optical Yagi-Uda antenna. New J Phys 9:217

Homola J (2006) Chapter: electromagnetic theory of surface plasmons, in surface plasmon resonance based sensors. Springer, Berlin

Ishihara K, Ohashi K, Ikari T, Minamide H, Yokoyama H, Shikata J, Ito H (2006) Terahertz-wave near-field imaging with subwavelength resolution using surface-wave-assistanted bow-tie aperture. Appl Phys Lett 89:201120

Johnson PB, Christy RW (1972) Optical constants of noble metals. Phys Rev B 6:4370–4379

Kang J-H, Kim K, Ee H-S, Lee Y-H, Yoon T-Y, Seo M-K, Park H-G (2011) Low-power nano-optical vortex trapping via plasmonic diabolo nanoantennas. Nat Comm 1592:582

Kelly KL, Coronado E, Zhao LL, Schatz GC (2003) The optical properties of metal nanoparticles: the influence of size, shape, and dielectric environment. J Phys Chem B 107:668–677

Kneipp K, Wang Y, Kneipp H, Perelman LT, Itzkan I, Dasari RR, Feld MS (1997) Single molecule detection using surface-enhanced Raman scattering (SERS). Phys Rev Lett 78:1667

Kosako T, Kadoya Y, Hofmann HF (2010) Directional control of light by a nano-optical Yagi-Uda antenna. Nat Photonics 4:312–315

Lalanne P, Hugonin JP (2006) Interaction between optical nano-objects at metallo-dielectric interfaces. Nat Phys 2:551–556

Lezec HJ, Thio T (2004) Diffracted evanescent wave model for enhanced and suppressed optical transmission through sub-wavelength hole arrays. Opt Express 12:3629–3651

Lynch DW, Hunter WR (1998) Comments on the optical constants of metals and an introduction to the data for several metals. In: Palik ED (ed) Handbook of optical constants of solids. Academic, San Diego

Mahboub O, Carretero Palacios S, Genet C, Garcia-Vidal FJ, Rodrigo SG, Martin-Moreno L, Ebbesen TW (2010) Optimization of bull's eye structures for transmission enhancement. Opt Express 18:124329

Maier SA (2006) Plasmonic field enhancement and SERS in the effective mode volume picture. Opt Express 14(5):1957–1964

Maier SA, Atwater HA (2005) Plasmonics: localization and guiding of electromagnetic energy in metal/dielectric structures. J Appl Phys 98:011101

Michalski KA (2013) On the low order partial fraction fitting of dielectric functions at optical wavelengths. IEEE Trans Antennas Propagat 61:6128–6135

Moroz A (2009) Wave scattering. www.wave-scattering.com/drudefit.html. Accessed Oct 2014

Muskens OL, Giannini V, Sánchez-Gil JA, Rivas JG (2007) Optical scattering resonances of single and coupled dimer plasmonic nanoantennas. Opt Express 15:17736–17746

Nevels RD, Michalski KA (2014) On the behavior of surface plasmons at a Metallo-Dielectric interface. J Lightwave Techno 32:3299–3305

Nevels R, Welch GR, Cremer PS, Hemmer P, Phillips T, Scully S, Sokolov AV, Svidzinsky AA, Xia H, Zheltikov A, Scully MO (2012) Configuration and detection of single molecules. Mol Phys 110:1993–2000

Nie S, Emory SR (1997) Probing single molecules and single nanoparticles by surface-enhanced Raman scattering. Science 275:1102–1106

Novotny L (2008) Effective wavelength scaling for optical antennas. Phys Rev Lett 98:266802

Otto A (1976) Spectroscopy of surface polaritons by attenuated total reflection, Chapter 13. In: Seraphin BO (ed) Optical properties of solids. North Holland, Amsterdam, pp 679–729

Ouyang F, Batson PE, Isaacson M (1992) Quantum size effects in the surface-plasmon excitation of small metallic particles by electron-energy-loss spectroscopy. Phys Rev B 46:15421–15425

Park Q-H (2009) Optical antennas and plasmonics. Contemp Phys 50:407–423

Pavlov RS, Curto AG, van Hulst NF (2012) Log-periodic optical antennas with broadband directivity. J Opt Comm 285:3334–3340

Purcell EM (1946) Spontaneous emission probabilities at radio frequencies. Phys Rev 69:681
Raether H (1988) Surface plasmons on smooth and rough surfaces and on gratings. Springer, Berlin
Ritchie RH (1957) Plasma losses by fast electrons in thin films. Phys Rev 106:874–881
Sarid D (1981) Long-range surface plasmons on very thin metal films. Phys Rev Lett 47:1927–1930
Schuck PJ, Fromm DP, Sundaramurthy A, Kino GS, Moemer WE (2005) Phys Rev Lett 94:17402
Søndergaard T, Bozhevolnyi SI (2007) Slow-plasmon resonant nanostructures: scattering and field enhancements. Phys Rev B 75:073402
Søndergaard T, Beermann J, Boltasseva A, Bozhevolnyi SI (2008) Slow-plasmon resonant-nanostrip antennas: analysis and demonstration. Phys Rev B 77:115420
Song BS, Noda S, Asano T, Akahane Y (2005) Ultra-high-q photonic double heterostructure nanocavity. Nat Mater 4(3):207–210
Sotomayor Torres CM, Zankovycha S, Seekampa J, Kama AP, Clavijo CC, Hoffmanna T, Ahopeltob J, Reutherc F, Pfeifferc K, Bleidiesselc G, Gruetznerc G, Maximovd MV, Heidarie B (2003) Nanoimprint lithography: an alternative nanofabrication approach. Mater Sci Eng C 23:23–31
Taminiau TH, Moerland RJ, Segerink FB, Kuipers L, van Hulst NF (2007) λ/4 resonance of an optical monopole antenna probed by single molecule fluorescence. Nano Lett 7:28–33
Vahala KJ (2003) Optical microcavities. Nature 424(6950):839–846
Wang L, Uppuluri SM, Jin EX, Xu X (2006) Nanolithography using high transmission nanoscale bowtie apertures. Nano Lett 6:361–364
Weber MJ (2001) Handbook of lasers. CRC Press, Boca Raton
Yang F, Sambles JR, Bradberry GW (1991) Long-range surface modes supported by thin films. Phys Rev B 44:5855–5872
Zhao Y, Engheta N, Alù A (2011) Effects of shape and loading of optical nanoantennas on their sensitivity and radiation properties. J Opt Soc Am B 28:1266–1274

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_44-1

Grid Antenna Arrays

Mei Sun[a]* and Yue Ping Zhang[b]
[a]Institute for Infocomm Research, A*STAR, Singapore, Singapore
[b]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore

Abstract

Grid antenna array (GAA) is a kind of planar array antenna with multiple rectangular loops. It can flexibly function either as a traveling-wave or non-traveling-wave antenna. This chapter lists clearly the different variations of the GAA and reviews briefly its theory development as well as applications. With a focus of the resonant GAA for 60-GHz millimeter-wave applications, the basic theory and operation are explored for the basic single-feed, dual-feed, and sub-array structures. More importantly, the further integration of the GAA as an antenna-in-package (AiP) module is presented with the design, fabrication, and test details at 60 GHz. The chapter finally reviews the state-of-the-art GAA millimeter-wave applications showing that the GAA really has wide applications from low microwave frequencies even to 120-GHz millimeter-wave frequencies.

Keywords

Grid antenna arrays; Array antenna; Microstrip grid array antenna; Millimeter-wave grid array antenna; Cross-mesh array antenna

Introduction

The grid antenna array (GAA) was first proposed by Kraus in 1964 (Kraus 1964a). It was a traveling-wave antenna of scanning beam in the backward direction against frequency. In 1966, Kraus filed a patent on his invention (Kraus 1966). Figure 1 shows the basic grid arrangement and its variations most in microstrip technology. The basic structures shown in Fig. 1a, b consist of rectangular meshes of microstrip lines on a dielectric substrate or metal lines in the air ($\varepsilon_r \geq 0$) and backed by a metallic ground plane. It can be fed at the mesh cell center as in Fig. 1a or at the mesh cell edge as in Fig. 1b with the other end terminated. The antenna can be fed by a metal via through an aperture on the ground plane such as by coaxial cable. Depending on the electrical length of the sides of the meshes, the grid antenna array may be resonant or nonresonant. Taking grid antenna array on a substrate as an example, for a resonant grid antenna array like in Fig. 1a, the sides of the meshes should be one wavelength by a half-wavelength in the dielectric and the instantaneous currents would be out of phase on the long sides of the meshes and in phase on the short sides of the meshes, respectively. As a result, the long sides of the meshes behave essentially as microstrip line elements and the short sides act as both radiating and microstrip line elements producing the main lobe of radiation in the boresight direction, while for a nonresonant traveling-wave grid antenna array like in Fig. 1b, the length of the short side of the meshes can be slightly more than one-third wavelength and the length of the long side of the meshes should be two times longer but three times shorter than the length of the short side of the meshes in the dielectric. Assuming that it is fed from one end, the currents in the

*Email: msun@i2r.a-star.edu.sg

DOI 10.1007/978-981-4560-75-7_44-1

short sides of the meshes follow a phase progression producing the maximum radiation in a backward angle-fire direction.

The work on GAA had been silent for years since its invention until M. Tiuri proposed a "chain antenna" in 1974 (Tiuri et al. 1974a) as in Fig. 1d. It was also a traveling-wave antenna with a scanning beam. Different from Kraus's design, Tiuri applied absorbing load at one end of the radiating array (Tiuri et al. 1974b). The world saw the first non-traveling-wave GAA in 1981 from R. Conti et al. (1981; Toth et al. 1983). He adopted four GAAs as in Fig. 1o with fixed broadside beams for the monopulse radar. Figure 1c shows their methods of amplitude control through control of microstrip line impedances to lower the first side lobe.

The grid antenna array has caught considerable attention of Nakano and his associates. Since the middle of 1990s, they have reported the design and analysis of various grid array antennas. Figure 1f shows their proposed center-fed dual-linearly polarized GAA by crossing the meshes in 1995 (Nakano et al. 1995); Fig. 1e shows a derivative of the GAA with transformed hexagonal loops, which was named honeycomb wire antenna, appeared from Nakano's group (Nakano et al. 1996, 1997); Fig. 1g shows their miniaturized grid antenna array by meandering the long sides of the meshes (Nakano and Kawano 1997; Nakano et al. 1998a); Fig. 1h shows the first circularly polarized GAA from Nakano's group realized by placing C-figure element above the radiating array in 1998 (Kawano and Nakano 1998); Fig. 1i also shows their circularly polarized grid antenna array by modifying the short sides of the meshes in 2007 (Nakano et al. 2007; Iitsuka et al. 2012); Fig. 1n shows their cross-mesh array, and with C-figure elements on top, this GAA can have multiple polarizations (namely, vertically linear polarization, horizontally linear polarization, and circularly polarization) (Kawano and Nakano 1999, 2000; Nakano et al. 2001a, b);

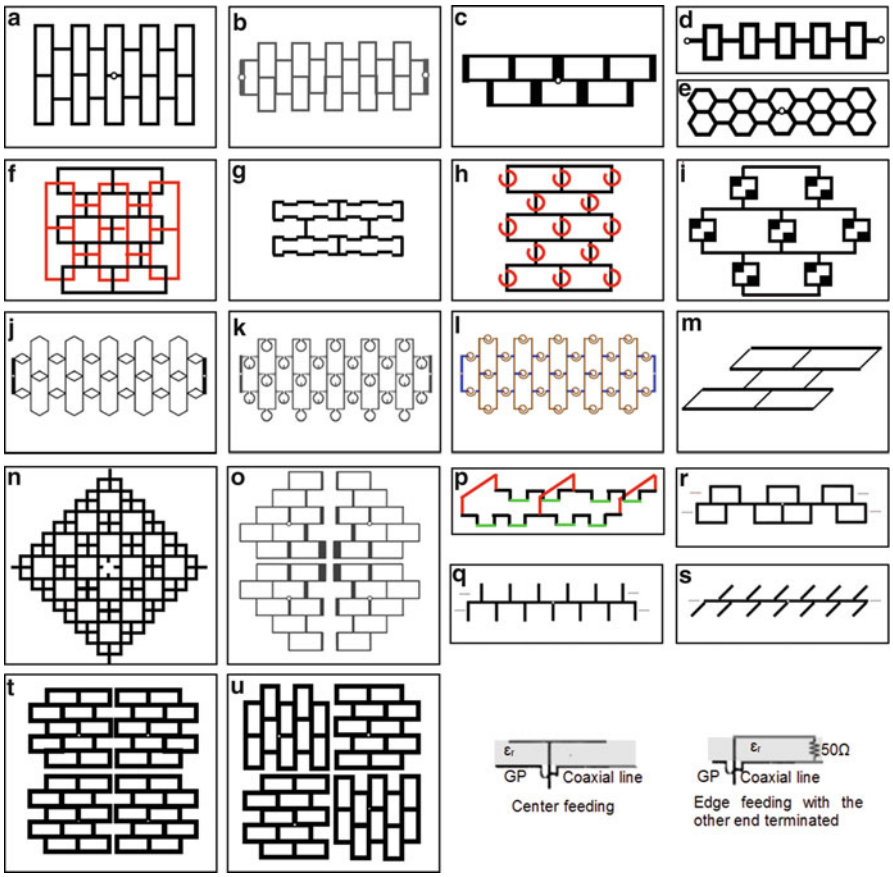


Fig. 1 The basic grid antenna array (**a**, **b**) and its variations (**c**–**u**)

DOI 10.1007/978-981-4560-75-7_44-1

Fig. 1j shows their proposed rhombic grid antenna array (RGAA) that has numerous bent radiation elements contributing to a more stable gain (Iitsuka et al. 2009a; Nakano et al. 2013a); Fig. 1k shows their loop-based GAA that can achieve a seamless CP frequency-scanning beam with perturbation elements added to the loops (Nakano et al. 2013b; Iitsuka et al. 2010); and Fig. 1l shows their tilted frequency-scanning CP GAA realized by adopting spiral elements as the radiation elements in 2010 (Iitsuka et al. 2010).

The above studies were all conducted at lower microwave frequencies. Zhang's group, for the first time, brought the GAA onto the stage of millimeter-wave application. They managed to implement a basic GAA as in Fig. 1a at 60 GHz on low temperature co-fired ceramic (LTCC) substrate to yield a standard surface-mounted chip package in 2008 (Sun and Zhang 2008; Zhang and Sun 2008). This antenna was the first millimeter-wave GAA and also was considered as an excellent antenna candidate for "antenna-in-package" technology (Sun and Zhang 2008). The idea of using GAA as a 60-GHz packaged antenna was further elaborated in Sun et al. (2009a, b). Targeted at millimeter-wave and integration purpose, they advocated GAA with a novel balanced feeding that eliminates the necessity of an extra lossy balun when the antenna is to be used in a differential system in 2011 (Zhang et al. 2011c). They also propose a new 45° linearly polarized grid antenna array by adjusting the angle between the long and short sides of the meshes as in Fig. 1m and a miniaturized grid antenna array by meandering the long sides and bending the short sides of the meshes in a multilayer metal structure as in Fig. 1p (Zhang and Sun 2008; Zhang et al. 2011c). The bending makes the large part of the short sides of the meshes further away from the ground plane, which may improve the radiation. Besides the millimeter-wave application, they also found the application of the GAA for automotive radar sensor and derived a comb array antenna for this purpose (Zhang et al. 2011a, b, 2012) as in Fig. 1r, q, s. In 2012, they also successfully adopt the GAA as a sub-element to further form a bigger array as an example in Fig. 1t (Zhang and Zhang 2012a, b, 2013). By dividing the large radiating array into small sub-grid arrays with multiple feedings, the accumulated phase error was minimized which resulted in more desirable radiation. The packaging scheme of such antennas was introduced in Zhang and Zhang (2012a) and Zhang et al. (2013a), and it was further integrated with a dual-band monopole as a mobile terminal antenna (Zhang et al. 2013d). More successfully, they realize a novel circularly polarized GAA by sequentially rotating the sub-GAA arrays as shown in Fig. 1u (Zhang and Zhang 2010, 2012a) and made into a package of such antennas in 2013 (Zhang et al. 2013b).

The analysis and synthesis of the GAA antenna has a long record. Wave velocities of the antenna were first analyzed in Kraus (1964b). In 1992, L. Hildebrand conducted his master thesis on the analysis of GAAs by the method of moments (MoM), which was the first attempt to analyze the characteristics of GAA (Hildebrand 1992). H. Nakano first analyzed the GAA by MoM in the GHz range in 1994 (Nakano et al. 1994). Later in mid-2000s, he also developed a fast MoM calculation technique for its analysis (Nakano et al. 1998b, 2001c, 2005). K. Palmer made the first attempt of GAA synthesis in 1997 (Palmer and Cloete 1997). The devoid of a systematic analysis and synthesis method of GAA had been lasting for 50 years until 2011, when the GAA's performance was characterized and a synthesis method was proposed by Zhang's group (Zhang and Zhang 2011). Radiation mechanism of the GAA was verified for the first time by pattern synthesis (Zhang et al. 2013c).

The literature review shows that GAA has wide applications in microwave frequencies. Zhang's group has made it an excellent antenna candidate for millimeter-wave "antenna-in-package" technology by providing various elegant package solutions as mentioned above. In addition, they successfully developed GAA for automotive radar sensor applications at 24 GHz (Zhang et al. 2011a, b, 2012). Inspired by their work, W. Menzel adopted the GAA for 79-GHz automotive radar (Bauer and Menzel 2011; Frei et al. 2011; Bauer et al. 2013), and T. Zwick implemented a D-band 120-GHz GAA on liquid crystal polymer (LCP) substrate as a lid of a surface-mounted chip package (Beer et al. 2013). All these demonstrate GAA's wide applications from microwave to millimeter-wave frequencies.

DOI 10.1007/978-981-4560-75-7_44-1

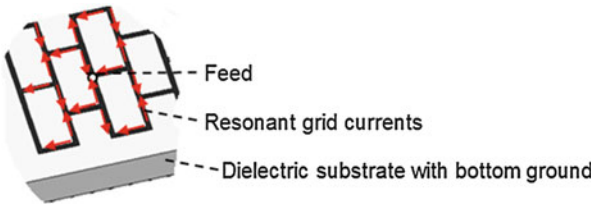


Fig. 2 Basic theory of the resonant grid array antenna

In this book chapter, although both resonant and nonresonant grid array antennas are useful for many applications, it will only focus on the more advanced technologies of the resonant GAA on the substrate at millimeter-wave frequencies. The basic theory and operation are presented in section "Basic Theory and Operation." The designs especially for integration purpose are presented in section "Design for Integration." The applications are presented in section "Millimeter Wave Applications." The chapter is summarized in section "Summary."

Basic Theory and Operation

Basic Theory

The GAA features a spatial scanning beam against input signal's frequency when designed as a traveling-wave antenna. Currents on the short side of each loop are in phase, which are responsible for the modulated equal phase front and co-polarization. Currents on the long side are out of phase so that they can cancel each other mostly in the far-field region and bring about cross-polarization. As a result, the short side of each loop is referred to as radiator or radiating element. As a traveling-wave antenna, it is usually excited at one end of the radiating array, while the other end of the radiating array is terminated either without (as in Kraus's design (Kraus 1964a)) or with an absorbing load (as in Tiuri's design (Tiuri et al. 1974a)). Phase shifting is generated when the input signal propagates from one end of radiating array to the other, which brings about a tilted equal phase front, hence a tilted beam. Because of the fixed dimensions of the loop, input signals at different frequencies can generate respective phase shifting, resulting in a modulated equal phase front against frequencies.

Though scanning beam can be achieved by a traveling-wave GAA without the requirement for an extra phase shifting or beam-forming network, it neither provides precise beam control in which a beam can appear at a specific locus nor a scanning beam at a fixed frequency. Hence, the GAA is not very popular as a traveling-wave antenna but finds wide application as a non-traveling-wave antenna.

As a non-traveling-wave antenna, the GAA is usually fed near its geometrical center. Because of the uniform phase degradation of EM waves that propagate from the feeding point to the periphery of the radiating array, the non-traveling-wave GAA features a fixed beam in the broadside direction. In this case, the short sides of each loop also carry in-phase current for effective radiation, while the currents on the long side are out of phase as shown in Fig. 2.

To elaborate the operation of the non-traveling-wave resonant GAA, the design of the single-feed, dual-feed, sub-array structures will be presented respectively in the following parts. Based on the A6 LTCC from Ferro, the grid array antennas are designed to operate in the 60-GHz band. The A6 ceramic tape has the dielectric constant of 5.9 and loss tangent of 0.0015 after firing, and the metallic paste is either silver or gold with good conductivity. The design determines the dielectric substrate dimensions, the number of meshes, the microstrip line impedances, and the excitation location with the associated diameters of the metal via and the aperture through the analysis of the HFSS simulations. It is demonstrated that the grid

DOI 10.1007/978-981-4560-75-7_44-1

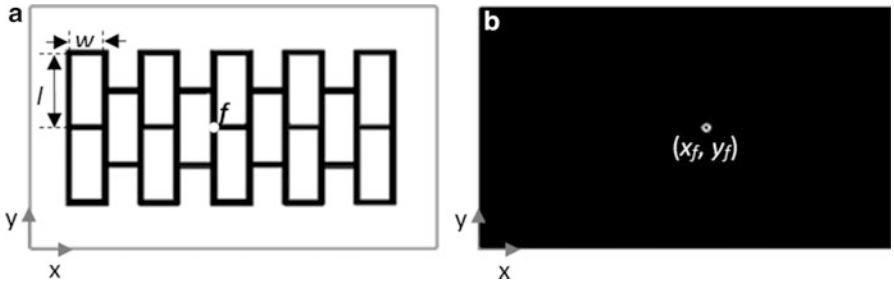


Fig. 3 The 60-GHz grid array antenna: (**a**) top and (**b**) bottom views

antenna array is easy to design, and more importantly, the grid antenna array is particularly suitable for fabrication in LTCC as its meshed structure avoids microfractures or warpage.

Basic Single-Feed Structure

The 60-GHz microstrip grid antenna array as shown in Fig. 3 consists of rectangular meshes of microstrip lines on a dielectric substrate backed by a metallic ground plane. It is fed at one of the nodes by a coaxial pin through an aperture on the ground plane. Based on the Ferro A6-S LTCC, the microstrip grid antenna array is designed to operate at 61.5 GHz with the maximum gain of ≥13.5 dBi, the impedance and radiation bandwidth of 7 GHz, and the efficiency of ≥80 %.

Given the specified gain value of 13.5 dBi and considering the various losses, one can find the required number of meshes to be at least 14, which leads to an estimation of the length and width of the substrate as 11.5 mm by 5 mm, respectively. The thickness of the substrate should be chosen to avoid the excitation of the TE_1 mode surface wave. Hence, a body size of $13.5 \times 8 \times 0.375$ mm^3 is finally determined by also taking the LTCC layout rule into account. For the sake of low cost and easy fabrication, the width and thickness of the microstrip lines are kept uniform as 0.15 and 0.01 mm, respectively. The mesh dimensions and the location of excitation f as well as the associated diameters of the metal via d_v and the aperture a_v on the ground plane are obtained from the HFSS simulations. They are $l = 2.5$ mm $\approx \lambda_g$, $w = 1.365$ mm $\approx \lambda_g/2$, $d_v = 0.1$ mm, $a_v = 0.3$ mm, $x_f = 4$ mm, and $y_f = 6.075$ mm, where λ_g is the guided wavelength.

Figure 4a, b shows the simulated antenna $|S_{11}|$ and peak realized gain as a function of frequency, respectively. The 10-dB impedance bandwidth is 13 GHz from 51 to 64 GHz or 26.7 % at 60 GHz. The maximum peak realized gain is 15 dBi at 61 GHz. The efficiency is also found to be better than 85 % over the required 7-GHz bandwidth. Figure 4c, d shows the simulated antenna radiation patterns in the xz- and yz-planes at 60 GHz, respectively. As expected, the antenna radiation patterns have the main beam of radiation in the boresight direction with side lobes lower than 20 dB and weak cross-polarization radiation components.

Basic Dual-Feed Structure

A basic dual-feed grid antenna array is shown in Fig. 5. The differential feeding scheme here is different from those in Kraus (1964a) and Nakano et al. (2005) where a gap is made on the short side of a mesh to connect to the differential source. The targeted specifications are the maximum gain of ≥15 dBi, the impedance and radiation bandwidth of 7 GHz, and the efficiency of ≥80 %. With the design steps similar to the single-feed one, the number of meshes of 14, substrate size of$13.5 \times 8 \times 0.385$ mm^3, the width and thickness of the microstrip lines are initially determined. Through HFSS simulation, the optimized mesh dimensions and the location of dual feeds as well as the associated diameters of the metal vias d_v and the apertures a_v on the ground plane are determined as $l = 2.5$ mm $\approx \lambda_g$, $w = 1.365$ mm $\approx \lambda_g/2$, $d_v = 0.1$ mm, $a_v = 0.3$ mm, $x_{f1} = 7.3$ mm, $y_{f1} = 3.98$ mm, $x_{f2} = 8.08$ mm, and $y_{f2} = 3.98$ mm.

DOI 10.1007/978-981-4560-75-7_44-1

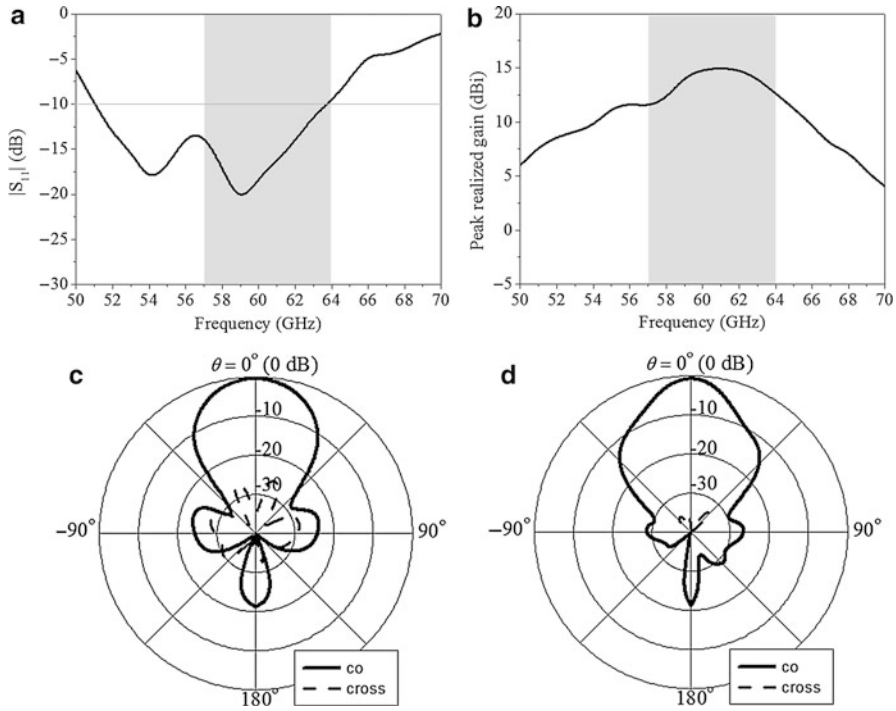


Fig. 4 Simulated grid antenna array performance: (**a**) $|S_{11}|$, (**b**) peak realized gain, (**c**) the xz-plane, and (**d**) the yz-plane radiation patterns at 60 GHz

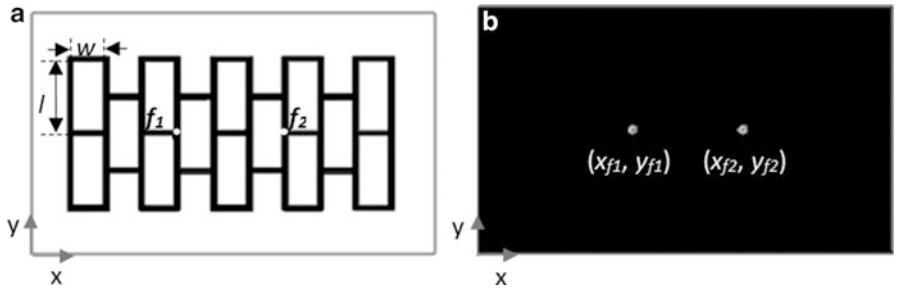


Fig. 5 The dual-feed grid array antenna: (**a**) top and (**b**) bottom views

Figure 6a compares the simulated current distributions on the grid at 61.5 GHz for the differential feed case with its single-feed counterpart. Note that the instantaneous currents do not distribute as shown in Kraus (1964a), Conti et al. (1981), and Palmer and Cloete (1997), that is, they are out of phase on the long sides of the meshes and in phase on the short sides of the meshes, respectively. Rather, they are only truly out of phase on the long sides of the meshes and in phase on the short sides of the meshes near the feeding points. This is because at such a high frequency, a slight mesh dimension change will cause a big change in signal phase over transmission, for example, at 60 GHz, a 70-μm variation in w will cause ~10° signal phase difference, thus making the control over phase synchronization of the far meshes from the source more difficult. The grid antenna array excited for differential operation has two source points, so it has more meshes of desirable current distributions than that excited for single-ended operation.

Figure 6b compares the simulated $|S_{11}|$. The impedance bandwidth is 10 GHz (16.7 %) from 53 to 63 GHz for the single end and 8 GHz (13 %) from 57 to 65 GHz for the differential operations, respectively. The wider impedance bandwidth for the single-ended antenna is the result of less number

DOI 10.1007/978-981-4560-75-7_44-1

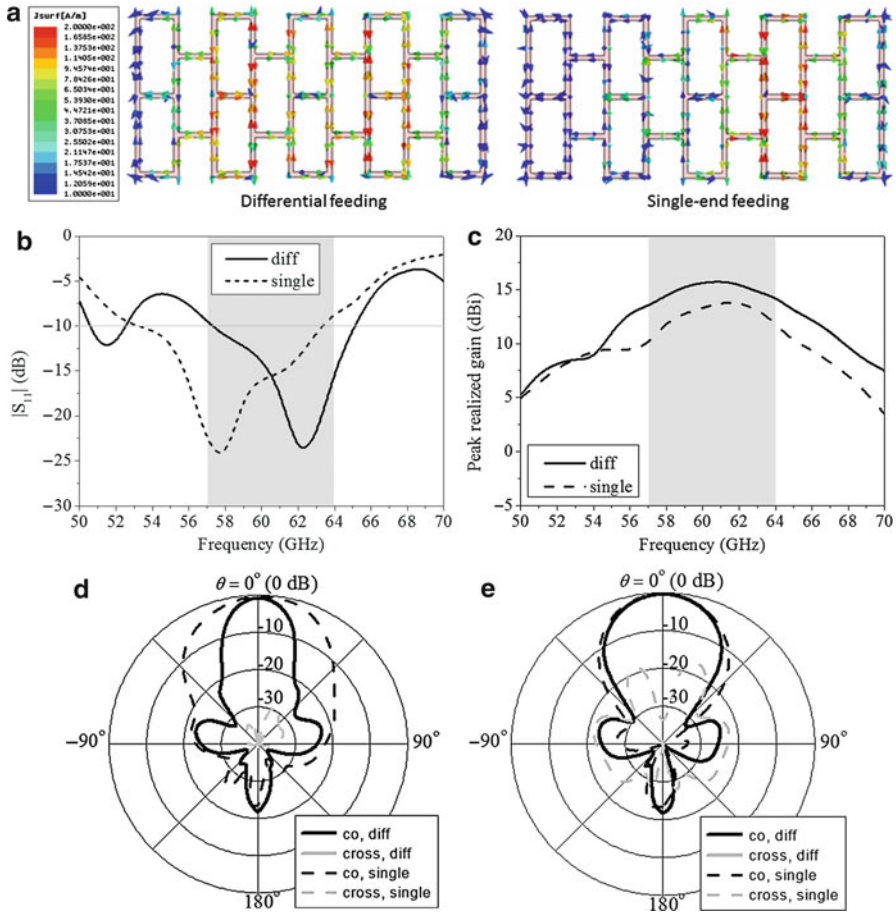


Fig. 6 The simulated results of the dual-feed grid antenna array compared with the single-feed one: (**a**) current distributions at 61.5 GHz, (**b**) $|S_{11}|$, (**c**) peak realized gain, (**d**) xz- and (**e**) yz-plane patterns at 61.5 GHz

of radiating elements that have in-phase currents. Figure 6c shows the simulated antenna peak gain. They are 16 and 13.5 dBi at 61.5 GHz for differential and single-ended operations, respectively. The 3-dB gain bandwidths are enough. The simulated efficiency is found better than 90 % for both cases. Figure 6d, e compares the simulated antenna radiation patterns in the xz- and yz-planes at 61.5 GHz. As expected, the main lobe and deep null of radiation appear in the boresight direction for the co- and cross-polarization fields, respectively. The differential excitation yields a sharper main beam in the xz-plane and a similar main beam in the yz-plane and much weaker cross-polarization field in both planes as compared with the single-ended operation, due to the current distributions shown in Fig. 6a. The front-to-back ratio is 21 dB.

Sub-array Structures

This section will introduce the GAA sub-array operation. Figure 7 shows the different structures formed by GAA sub-array element of (a). It is designed at 61.5 GHz as detailed in Zhang and Zhang (2014). It is interesting to note that on the sub-grid array, the currents on each short side are indeed in phase, while the currents on each long side are out of phase. As expected, the main beams of radiation are in the broadside direction as plotted in Zhang and Zhang (2014).

The formation of a large array with four sub-grid arrays and multiple feeds not only improves the unsynchronized current but also introduces more design flexibilities. For example, a linearly polarized (LP) GAA can be realized when sub-grid arrays in the upper and lower quadrants are fed out of phase simultaneously. Figure 7b shows the simulated current distribution at 61.5 GHz of the LP GAA of

DOI 10.1007/978-981-4560-75-7_44-1

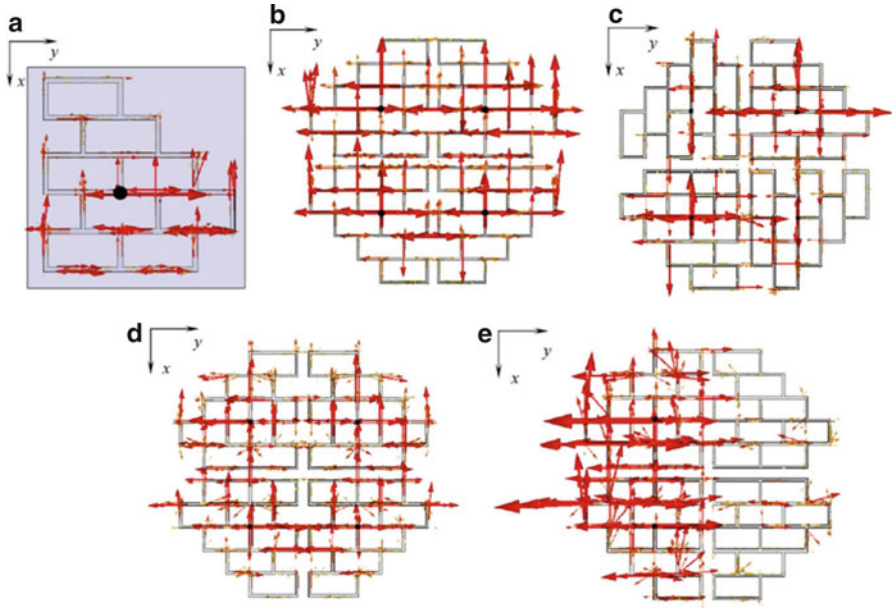


Fig. 7 The resonant GAA sub-array structures with denoted current distribution: (**a**) a GAA sub-array element, (**b**) a linearly polarized GAA, (**c**) a circularly polarized GAA, (**d**) a single balanced GAA, and (**e**) a single unbalanced GAA with four sub-arrays

32 loops formed by 4 sub-grid arrays and 4 feeds. It is clear from the figure that the current distribution is more satisfactory than that shown in Fig. 7a. The improvement is due to the fact that one feed excites 8 loops rather than 32 loops, leading to a smaller accumulated phase error. Although there are still incorrect current probably caused by the coupling between the sub-grid arrays and multiple feeds, the radiation patterns are improved with maximum radiation still in the broadside direction as detailed in in (Zhang and Zhang 2012b, 2013, 2014; Zhang et al. 2013a).

By sequential rotation, either a right-handed (RH) or left-handed (LH), the circularly polarized (CP) GAA can be realized. Figure 7c shows the current distribution at 61.5 GHz of the CP GAA of 32 grids formed by 4 sub-grid arrays and 4 feeds as detailed in (Zhang and Zhang 2010, 2012a; Zhang et al. 2013b), where good CP performance is achieved with satisfied axial ratio bandwidth and maximum radiation in the broadside direction.

Moreover, when the sub-grid arrays in the upper quadrants are driven with 0° phase and those in lower quadrants with 180° phase simultaneously in Fig. 7d, a single balanced antenna is realized which is compatible with a system of balanced output. While if the sub-grid arrays in the left quadrants and in the right quadrants are respectively driven differentially (that is to say, the sub-grid arrays in the left quadrants are driven by a differential pair and those in the right quadrants are driven by the other differential pair), two balanced antennas with one for transmission and the other for reception are realized. Figure 7d shows the simulated current distribution of the single balanced GAA at 61.5 GHz. Figure 7e shows the simulated current distribution when one of the dual balanced GAAs is excited at 61.5 GHz. Both radiation patterns have maximum radiation in the broadside direction (Zhang and Zhang 2014).

The peak realized gain results of the above GAAs are presented in Zhang and Zhang (2014). It is found that the GAA in Fig. 7b has the narrowest 3-dB gain bandwidth. The 3-dB gain bandwidth of the single unbalanced LP GAA in Fig. 7e is much wider, so is the single balanced LP GAA. Both the maximum gain values are enhanced by about 2 dB as compared with the GAA in Fig. 7b because of improved current distribution. As expected, the maximum gain values of the CP and dual balanced LP GAAs are about 3 dB lower.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_44-1

Design for Integration

GAA Integrated in a Grid Array Package

Figure 8 illustrates the ball grid array package with a microstrip grid antenna array connected to a 60-GHz receiver die. Figure 8a shows the simplified block diagram of the 60-GHz receiver die, including its critical input/output pads. The receiver uses a dual-conversion superheterodyne architecture with a 57–64 GHz RF range, 8.1–9.1 GHz IF range, and 16.3–18.3 GHz VCO range. The bandwidth of the in-phase/quadrature (I/Q) baseband signal is currently limited by the IF bandwidth of ±1 GHz. The critical input/output pads are in ground-signal-ground (GSG) coplanar configurations. The input pads require the package interconnection with 50-Ω characteristic impedance from 57 to 64 GHz, while the output pads also require 50-Ω characteristic impedance but at much lower frequencies up to 2 GHz as determined by the maximum bandwidth of the I/Q baseband signals. The receiver die size is 3.4 × 1.7 × 0.46 mm^3 (Pfeiffer et al. 2006).

Note that the package features standard wire bonding and there are four co-fired laminated ceramic layers for the package. The first ceramic layer is 0.385 mm thick, the second ceramic layer is 0.285 mm thick with an opening 3.8 × 2 mm^2, the third layer is 0.21 mm thick with an opening 5 × 3.2 mm^2, and the fourth layer is 0.385 mm thick with an opening 5 × 3.8 mm^2. These openings form the three-tier cavity that can house the 60-GHz receiver die. There are also five metallic layers for the package. The top layer provides the metallization for the microstrip grid array antenna, the first buried layer the metallization for the solid ground plane, the second buried layer the metallization for the antenna feeding traces and signal traces, the third buried layer the metallization for the signal traces, and the bottom exposed layer the metallization for the package ground plane and solder ball pads. The package has 58 input/outputs with a JEDEC standard solder ball pitch of 0.65 mm. Two dummy solder balls are attached to the two corners of

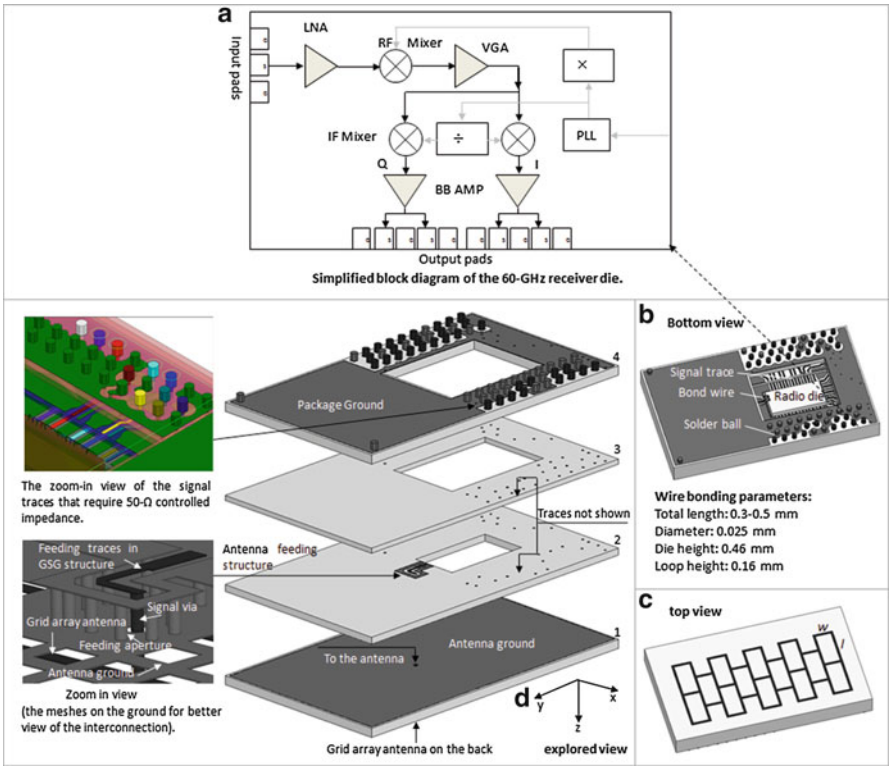


Fig. 8 The views of the ball grid array package with a grid antenna array and a 60-GHz die: (**a**) the simplified block diagram of the 60-GHz receiver die, (**a**) top view, (**b**) bottom view, and **c** explored view of the whole package

DOI 10.1007/978-981-4560-75-7_44-1

the package, respectively, for an enhanced attachment on the system PCB. The size of the whole package is $13.5 \times 8 \times 1.265$ mm^3.

Bond wires are used for chip-package interconnection. The input pads of the chip carrying RF signals from 57 to 64 GHz are connected to the antenna in package. The output pads of the chip carrying baseband signals up to 2 GHz are connected to signal traces in package. Here, bond wires between the receiver die and the antenna are the most critical because the discontinuity introduced by bond wires can significantly affect the performance of the entire 60-GHz radio. Nonetheless, the wire-bonding technique, well established in consumer electronics, remains a very attractive solution since it is robust and inexpensive. In addition, it has the advantage of being tolerant on die and package thermal expansion, an important requirement for many applications. There are methods to compensate the discontinuity on either or both sides of the die and package for millimeter-wave applications (Budka 2001; Sun et al. 2006). Figure 8 also shows wire-bonding parameters. The wire profile is based on the JEDEC standard. Note that the shortest bond wires are 300 μm long, which is almost a doubled length of the shortest bond wire supported by the current technology and would thus greatly improve the yield of assembly of the chip with the package.

A microstrip patch antenna array designed for 60-GHz applications has been reported (Lamminen et al. 2008). A complex feeding network, sophisticated process techniques, and embedded cavity are required to achieve enhanced bandwidth and efficiency. To avoid the problems of the microstrip patch array antenna, the microstrip grid antenna array is adopted. It operates in a resonant mode that requires the sides of the meshes one wavelength by one-half wavelength in the dielectric. Thus, the instantaneous current distribution would be out of phase on the long sides of the meshes and in phase on the short sides of the meshes, respectively. The resulting radiation would be broadside to the grid and linearly polarized parallel to the short sides of the meshes. For the sake of low cost and easy fabrication, the width and thickness of the microstrip lines are kept uniform as 0.15 and 0.01 mm, respectively. The mesh dimensions are $l = 2.5$ mm $\approx \lambda_g$, $w = 1.365$ mm $\approx \lambda_g/2$, where λ_g is the guided wavelength at 60 GHz.

Figure 8 also shows the zoom-in view of the antenna feeding network. It consists of such packaging elements as a coplanar waveguide (CPW) line cascaded first with a 90°-angled stripline and then vias in a GSG arrangement. It is known that the GSG arrangement not only minimizes potential electromagnetic interference but also improves the feeding performance. The signal via is finally fed to the antenna through an aperture in the antenna ground. The CPW feed line is designed to be 50 Ω with a pitch of 0.25 mm and a line width of 0.15 mm. The signal via, the aperture, and ground via have diameters of 0.1 mm, 0.3 mm, and 0.1 mm, respectively.

Figure 8 also shows the zoom-in view of the signal traces that require 50-Ω controlled impedance up to 2 GHz. These signal traces involve bond wires, CPW lines, vias, and solder balls. The signal traces in purple, red, aqua, and blue are denoted as signal traces 1, 2, 3, and 4, respectively. Signal traces 1 and 2 in a GSGSG configuration connect the on-chip Q signal output pads to the system board, so do signal traces 3 and 4 for the on-chip I signal output pads.

The package was fabricated in FERRO A6 LTCC ($\varepsilon_r = 5.9$ and $\tan\delta = 0.002$) with nine green tapes and gold metal in a panel size of 100×100 mm^2 by LTCC Boutique Foundry in Singapore Institute of Manufacturing Technology. Figure 9 shows the photo of a sample. The antenna function of the package was tested with a probe-based measurement setup (Zwick et al. 2004). Figure 9 also illustrates the test fixture to hold the package for testing. The package is supported by an FR4 fixture, which provides the good support for the probe to touch the antenna CPW feeding lines for measurement. This will also avoid the direct contact of the package on the metal probe station. A photo of a probe touching example is also shown in Fig. 9. The probe pitch is 250 μm to agree with the pitch of tested antenna CPW lines.

Figure 10 shows the simulated results with the measured results of the microstrip grid antenna array in the ball grid array package. No de-embedding was made to the measured results. It is evident from Fig. 10a that the measured 10-dB impedance bandwidth is 8.7 GHz from 56.3 to 65 GHz, which covers

DOI 10.1007/978-981-4560-75-7_44-1

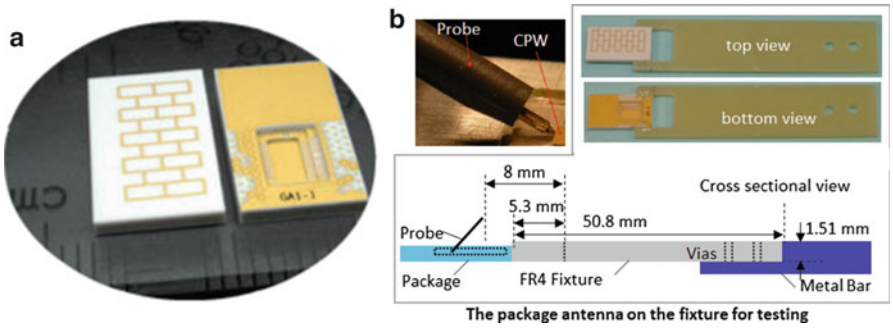


Fig. 9 Photo of the ball grid array package with the microstrip grid antenna array and illustration of the package antenna on the fixture for testing

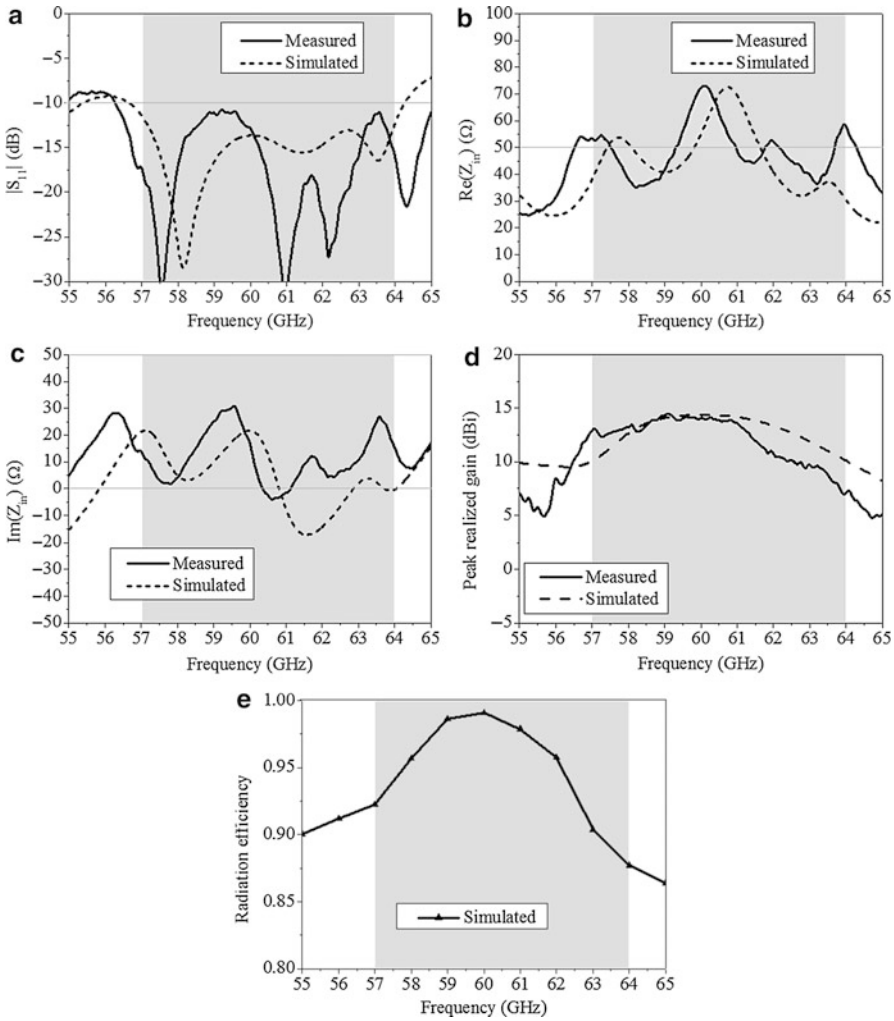


Fig. 10 Simulated and measured performance of the microstrip grid antenna array in the ball grid array package: (**a**) $|S_{11}|$, (**b**) input resistance, (**c**) input reactance, and (**d**) peak gain and efficiency as a function of frequency

the whole band of 57–64 GHz. By comparing with the simulated results in Fig. 10b, c, it is noted that the measured resonance frequencies shift down and the impedance becomes more inductive. This is attributed to the difference between the wave-port excitation in the simulation and the probe excitation in the measurement. In addition, the measured and calculated peak gain values in the main beam direction are

DOI 10.1007/978-981-4560-75-7_44-1

both 14.5 dBi as shown in Fig. 10d. It is achieved at 60 GHz in the simulation while 59 GHz in the measurement. Figure 10e also shows estimated efficiency $\eta > 88$ % over 57–64 GHz.

Figure 11 compares the simulated with the measured radiation patterns of the microstrip grid antenna array in the ball grid array package. They reveal that both xz- and yz-plane patterns have the main beam of radiations in the boresight directions with small cross-polarization radiation components. Note that the simulated and measured co-polar patterns at 60 GHz agree very well and both have small back radiations, while the back radiation levels increase at 57 and 64 GHz. This explains why the peak realized gain decreases at these frequencies as shown in Fig. 10d. Also note that the co-polar patterns on the xz-plane in the main radiation direction are more insensitive to the frequency than those on the yz-plane. This is because the grid antenna array has more meshes in the yz-plane making the control over phase synchronization of the far meshes from the source at different frequencies more difficult.

It should be mentioned that the performance of the microstrip grid antenna array in the ball grid array package landed on a system FR4 PCB with a top ground plane that was also simulated. It was found that the impedance matching performance is insensitive to the PCB and its further extended ground size. In addition, the peak realized gain showed a negligible decrease of 0.2 dB, while the gain performance over the whole bandwidth remains almost unchanged. Hence, one can conclude that the microstrip grid antenna array is an excellent antenna array candidate for 60-GHz radios.

Figure 12 shows the simulated insertion loss characteristics from 100 MHz to 5 GHz for 50-Ω controlled impedance signal traces. It can be observed that the signal loss is very low up to 5 GHz and the 3-dB insertion loss bandwidth is found to be much greater than 5 GHz, satisfying the 2-GHz bandwidth requirement easily. Figure 12 also shows the simulated return loss from 100 MHz to 5 GHz for 50-Ω controlled impedance signal traces. It is evident that the return loss values higher than 22 dB are

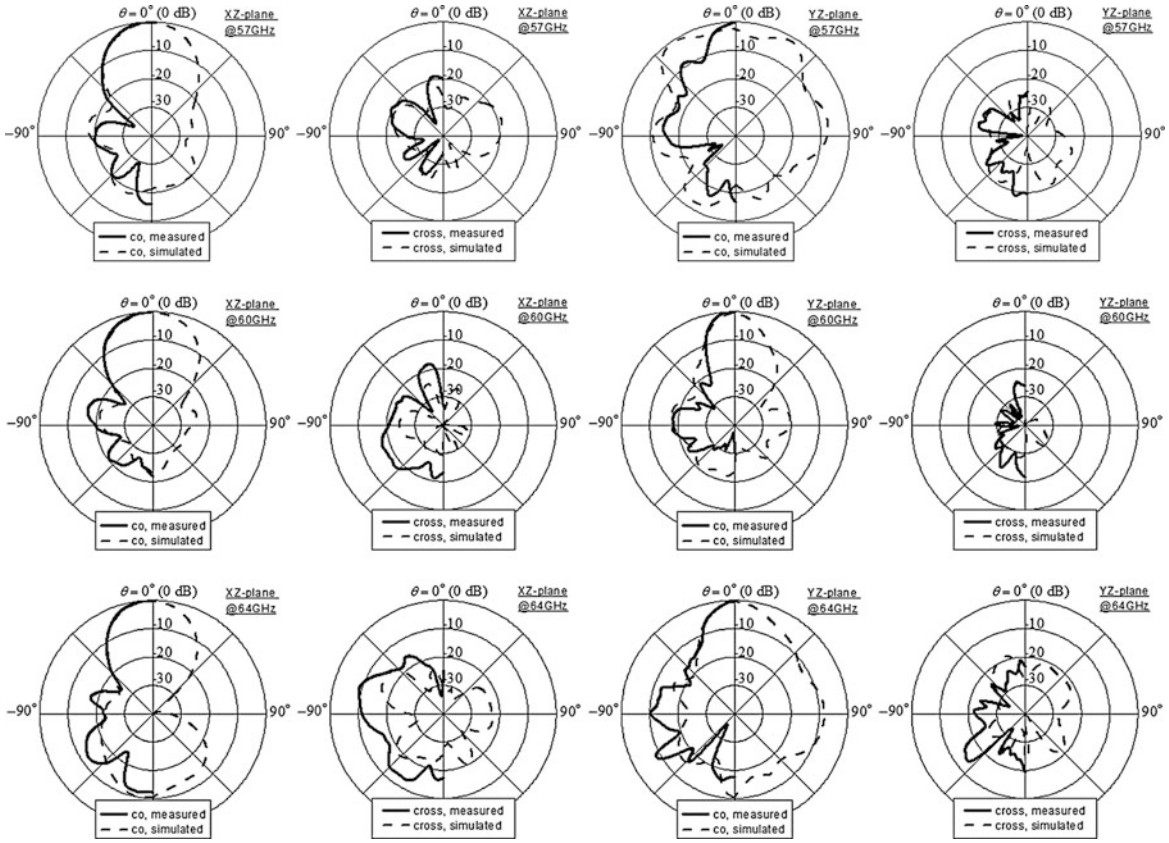


Fig. 11 Simulated and measured radiation patterns of the microstrip grid antenna array in the ball grid array package

DOI 10.1007/978-981-4560-75-7_44-1

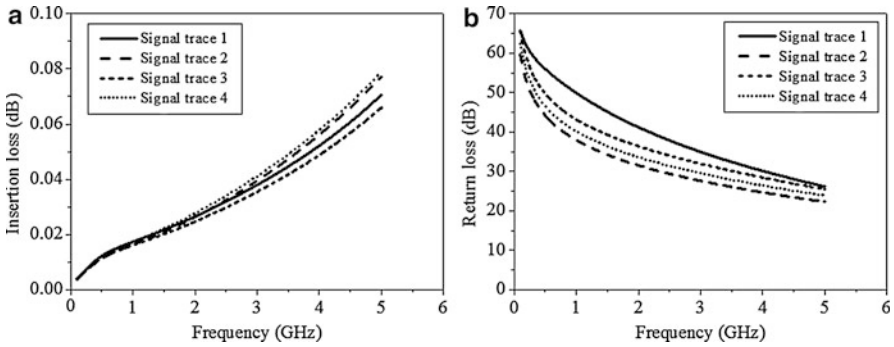


Fig. 12 (**a**) The insertion loss and (**b**) the return loss of the signal traces as a function of frequency

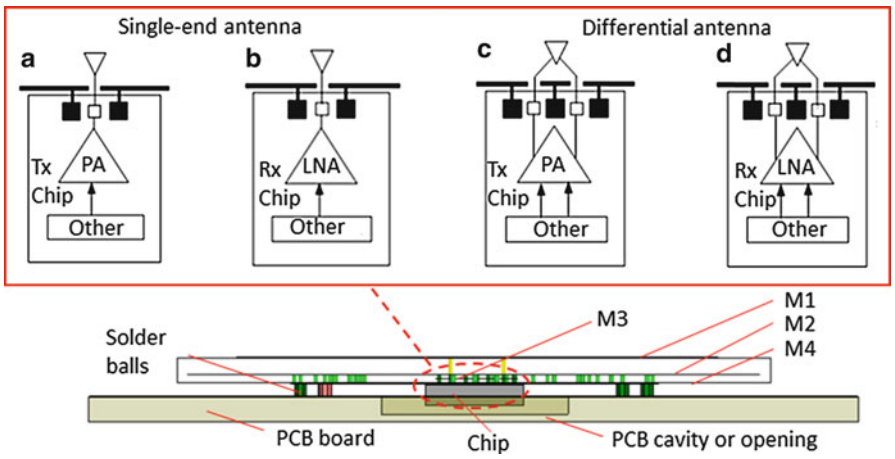


Fig. 13 The concept of assembling the highly integrated 60-GHz radio in the package to the system PCB as well as illustration of its (**a**, **b**) single-ended and (**c**, **d**) differential antenna-circuit interfaces

achieved over the 5-GHz bandwidth, indicating excellent wideband impedance matching to 50-Ω sources. Regrettably, no measurement was made to the frequency characteristics of these signal traces due to the limitation of our existing measurement setup and wire bonder. Nevertheless, the simulated insertion and return losses should be correct. They were obtained as each signal trace end was connected to a 50-Ω port while ground lines were short-circuited to the ground plane, and a single port was excited while other ports were tied to 50-Ω match loads.

Dual GAA Integrated in a Single Package

Figure 13 illustrates the concept of assembling the highly integrated 60-GHz radio in the chip-scale package to the system printed circuit board (PCB). A cavity or even an opening needs to be created in the PCB to house and protect the radio die. The lands on the chip package are soldered to the PCB to finish interconnect from the chip to the PCB through the package. This is believed to be a very cheap solution to highly integrated 60-GHz radios with the rather thin profile.

Figure 13 also illustrates the chip-antenna interface solutions for highly integrated radios. The highly integrated 60-GHz radio chip should be first discussed. Traditional commercialized 60-GHz radios have been designed as an assembly of several microwave monolithic integrated circuits (MMICs) in gallium arsenide (GaAs) semiconductor technology. They have been used for Gigabit Ethernet (1.25 Gb/s) bridges between local area networks (Maruhashi et al. 2001; Ohata et al. 2003). Recently, integrated

DOI 10.1007/978-981-4560-75-7_44-1

transmitter (Tx) and receiver (Rx) MMICs in 0.15-μm GaAs pHEMT and mHEMT processes have been realized to support data rates of several Gb/s for 60-GHz short-range applications (Gunnarsson et al. 2005, 2007). However, the 60-GHz radios in GaAs MMICs are expensive and bulky. In order for 60-GHz radio chips to have mass deployment and meet consumer marketplace requirements, the cost and size of any solution must be low and compact. That implies silicon, not GaAs as the better technology choice. In fact, designs toward low-cost, highly integrated 60-GHz radio chips have been realized in silicon technologies. For example, Floyd et al. have demonstrated a 60-GHz Tx and Rx chipset in a 0.13-μm silicon-germanium (SiGe) technology (Floyd et al. 2006) and Tanomura et al. in a 90-nm complementary metal oxide semiconductor (CMOS) technology (Tanomura et al. 2008). An examination of the above works and many other reported 60-GHz highly integrated radio chips in SiGe and CMOS reveals that two types of antenna-circuit interfaces as shown in Fig. 13 can be identified in the current solutions. The first type features the 50-Ω single-ended and the second type the 100-Ω differential antenna-circuit interfaces. For the first type, the 60-GHz on-chip input/output pads are designed as the ground-signal-ground (GSG) pads while for the second type the ground-signal-ground-signal-ground (GSGSG) pads. The GSG pads are bonded to an off-chip but in-package single-ended antenna while the GSGSG pads to a differential antenna with either flip-chip or wire-bonding techniques (Pfeiffer et al. 2006; Zhang et al. 2008a, 2009; Sun et al. 2008).

A single-chip solution of a 60-GHz radio transceiver (TRX) in CMOS has been attempted (Tomkins et al. 2009), where differential Tx and Rx are integrated on the same die. It is known that CMOS scaling improves amplifier noise performance and gain but exacerbates the difficulty of generating sufficient output power by the power amplifier (PA) at 60 GHz (Doan et al. 2004). Theoretically, a differential PA yields 3 dB more output power than a single-ended one does. Hence, the differential antenna-circuit interface in Fig. 13c is preferred to the single-ended antenna-circuit interface in Fig. 13a for the Tx integration of the TRX. Furthermore, the differential antenna-circuit interface in Fig. 13d is the better choice than the single-ended antenna-circuit interface in Fig. 13b for the Rx integration of the TRX because the differential low noise amplifier (LNA) can achieve higher linearity, lower offset, and better immunity to common-mode noise due to power supply variations or substrate coupling than the single-ended LNA does (Huang et al. 2006). The advantage of the fully differential architecture from the antenna to the circuit has been well understood for modern single-chip solutions of radio transceivers (Zhang et al. 2008b).

Regardless of the antenna-circuit interfaces, two antennas, one for transmission and the other for reception, are required for current highly integrated 60-GHz radios. The use of two antennas is not a problem as the antenna form factor at 60 GHz is on the order of millimeters. However, it may become a problem for the highly integrated 60-GHz radio that employs multiple antennas for beam steering to search the available path to enhance the link quality (Zhang and Liu 2009). This is because multiple antennas not only require corresponding multiple electrostatic discharge circuits that consume substantial die area but also make the whole radio bulky. It is known that the number of multiple antennas can be reduced to half by using transmit/receive (T/R) switches or circulators. Unfortunately, the T/R switches in CMOS in the 60-GHz band are still too lossy to be used (Tomkins et al. 2009; He and Zhang 2008) and information on circulators for highly integrated 60-GHz radios is unknown. In this section, the integration of the grid array antennas into a chip package in LTCC for highly integrated 60-GHz radio chipsets is described. As an example, a dual-feed grid antenna array is integrated for the differential Tx antenna-circuit interface and another single-feed grid antenna array for the single-ended Rx antenna-circuit interface.

As previously discussed, current highly integrated 60-GHz radios require dual antennas. Nakano et al. designed dual-grid array antennas in a double-layer structure (Nakano et al. 1995). A perpendicular orientation was arranged for the upper and lower grid array antennas. In this way, the

DOI 10.1007/978-981-4560-75-7_44-1

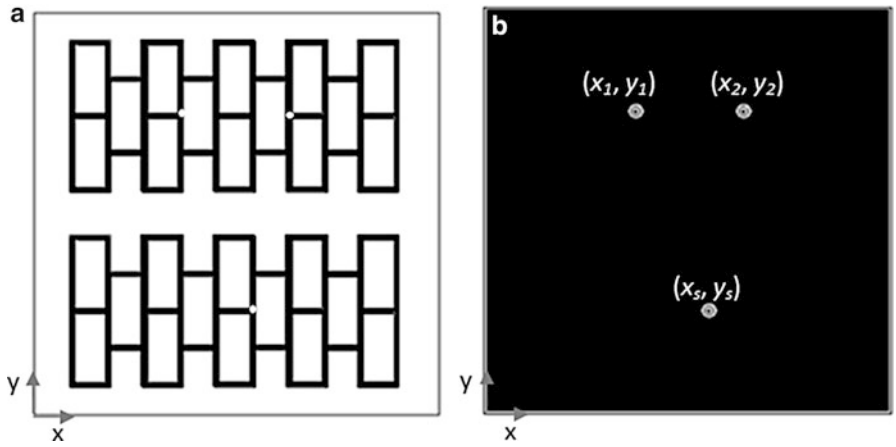


Fig. 14 The dual-grid array antennas: (**a**) top and (**b**) bottom views

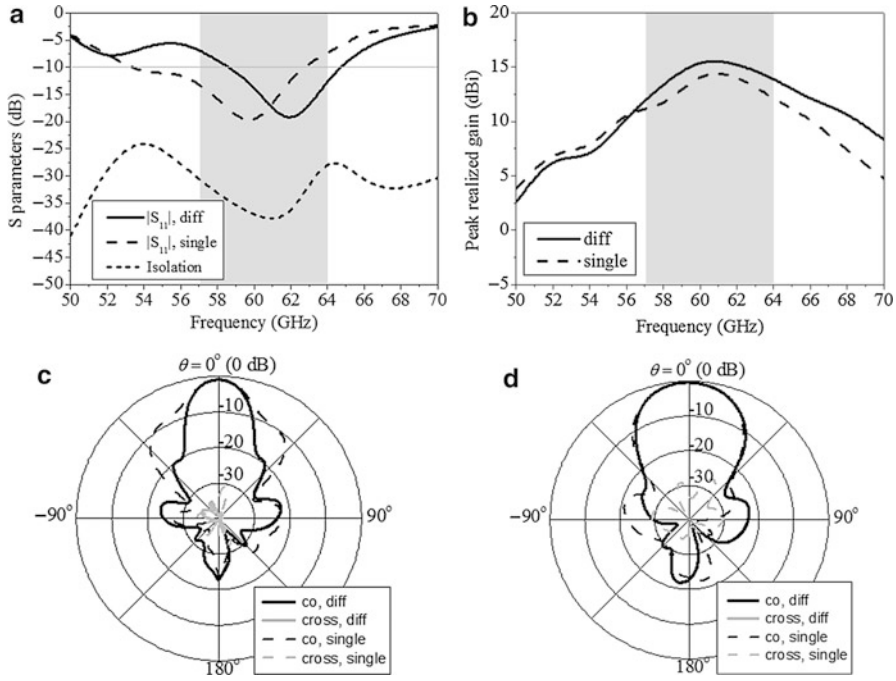


Fig. 15 The simulated results of the dual-grid array antennas with one for differential and the other for single-ended operations: (**a**) $|S_{11}|$ and isolation, (**b**) peak realized gain, and xz- and (**c**) yz-plane patterns at 61.5 GHz

upper grid antenna array radiates a horizontally polarized wave, while the lower grid antenna array does a vertically polarized one. A high isolation between both centrally located feeding terminals can be guaranteed. Figure 14 shows the dual-grid array antennas in a single-layer structure. It is formed by two grid array antennas of the basic structure that has a body size of $13.5 \times 13.5 \times 0.385$ mm^3. A parallel orientation is arranged for the dual-grid array antennas. Both grid array antennas radiate the wave of the same polarization. The single-layer structure simplifies the fabrication process. The parallel orientation reduces the outrage probability of 60-GHz radio links, which are usually deployed in line-of-sight environments, due to the polarization loss.

Figure 15 compares the simulated results of the grid antenna array with the dual feeds for differential operation and the grid antenna array with a single feed for single-ended operation. As this is a transitional

DOI 10.1007/978-981-4560-75-7_44-1

step, no performance enhancement is made and no patterned ground plane is used. When one grid antenna array is excited, the other grid antenna array acts as a parasitic element and vice versa. Figure 15a shows that the impedance bandwidth is 5.3 GHz (8.6 %) from 59 to 64.3 GHz for the differential antenna and 9.7 GHz (15.8 %) from 52.5 to 62.2 GHz for the single-ended antenna. With reference to Fig. 6c–e, one can find from Fig. 15b–d that the radiation patterns and the maximum gain of the single-ended antenna are improved more noticeably than those of the differential antenna by the parasitic element. For example, the beamwidth in the E-plane is narrowed, and the radiation of cross-polarization in the H-plane is suppressed, so the gain is improved to 15 dBi for the single-ended antenna. Figure 15a also shows the simulated isolation level between the dual-grid array antennas. For the sake of simulation simplicity, both grid array antennas are fed for single-ended operation. Note the isolation is high because of the large physical separation between the two feeding terminals.

Figure 16 illustrates the integration of dual-grid array antennas in a flip-chip package as sample A. Note that the package features standard flip-chip bonding, and there are three co-fired laminated ceramic layers for the package. The first ceramic layer is 0.385 mm thick, and the second to the third ceramic layers are both 0.095 mm thick. There are four metallic layers for the package. The top layer provides the metallization for the dual-grid array antennas, the first buried layer metallization for the patterned ground plane, the second buried layer the metallization for the antenna feeding traces, and the bottom exposed layer the metallization for the signal traces. The package has 48 input/outputs with a JEDEC standard pitch of 0.25 mm. The size of the whole package is $13.5 \times 13.5 \times 0.585$ mm^3.

Figure 16 also illustrates the zoom-in view of the feeding networks of the dual-grid array antennas of sample A. For the dual-feed one, it consists of such packaging elements as two quasi-coaxial cables cascaded first with two striplines, then another two quasi-coaxial cables, and finally vias through two apertures on the ground plane in a GSGSG arrangement. It is interesting to note that the differential

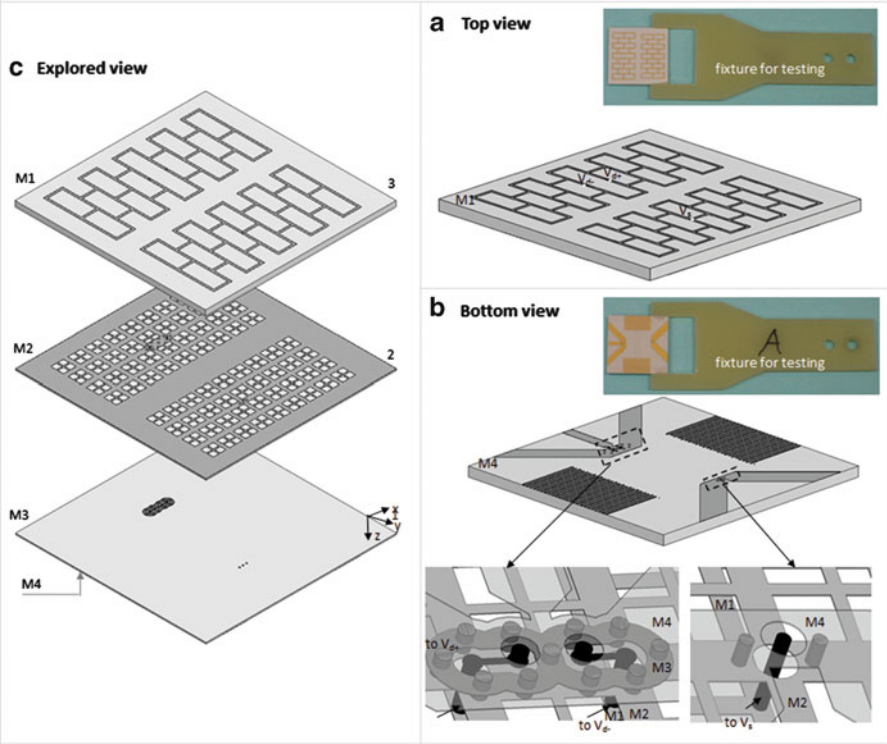


Fig. 16 Photos of the dual-grid antenna array A in the package with illustrations of **a** top view, **b** explored view, **c** bottom view with signal traces, as well as the zoom-in view of its feeding networks

DOI 10.1007/978-981-4560-75-7_44-1

feeding ports in Fig. 16 are brought closer to each other as compared with those in Fig. 14, due to the requirement of flip-chip bonding to the on-die GSGSG pads. The radiating element between the differential feeding ports is removed to enhance their isolation. For the single-feed one, it consists of a quasi-coaxial cable cascaded with via through one aperture on the ground plane in a GSG arrangement. It is known that the GSG and GSGSG arrangements not only minimize potential electromagnetic interference but also improve the feeding performance. The GSG and GSGSG feeding networks are designed together with the grid array antennas. Both GSG and GSGSG pads have a pitch of 0.25 mm.

Figure 17 shows the bottom view of the package sample B without the signal traces but with the integrated balun for testing the differential grid antenna array with the single-ended equipment. The package with the dual microstrip grid array antennas was fabricated in FERRO A6 LTCC in Singapore Institute of Manufacturing Technology. Figure 17 also illustrates the test fixture to hold the package for testing.

Figure 18 compares the simulated and measured results of the single-ended antenna for sample B. The return losses agree very well from 56 to 58.5 GHz. The agreement becomes poor for higher frequencies, due to the following reasons: dimension tolerance control, material property variation, and the difference between the wave-port excitation in simulation and the probe excitation in measurement. The measured return loss is higher than 10 dB from 56.4 to 61.7 GHz and 8 dB from 55 to 63.4 GHz indicating acceptable matching to 50-Ω sources at these frequencies. The simulated and measured radiation patterns are in close agreement at 61.5 GHz. The measured maximum gain is 13.5 dBi at 59 GHz with 3-dB gain bandwidth of 4.5 GHz. The simulated radiation efficiency is better than 85 %.

Figure 18 also compares the simulated and measured results of the differential antenna for sample B. The discrepancies between the simulated and measured return losses are due to the same reasons explained above for single-ended antenna. The measured return loss is higher than 8 dB from 56.2 to 63.2 GHz indicating acceptable matching to 50-Ω sources at these frequencies. Again, the simulated and measured radiation patterns at 61.5 GHz are in close agreement. The larger side lobes in the back side are caused by the balun. The measured maximum gain is 13.5 dBi at 57.5 GHz with 3-dB gain bandwidth of 5.3 GHz. If the additional loss of 1 dB from the integrated balun is de-embedded, the maximum gain of the differential antenna becomes 14.5 dBi. The simulated radiation efficiency is also better than 85 %. Generally, the measured results confirm the simulated ones. For instance, the differential grid antenna array exhibits a sharper radiation beam in the E-plane, a similar radiation pattern in the H-plane, and a higher gain as compared with those of the single-ended grid array antenna.

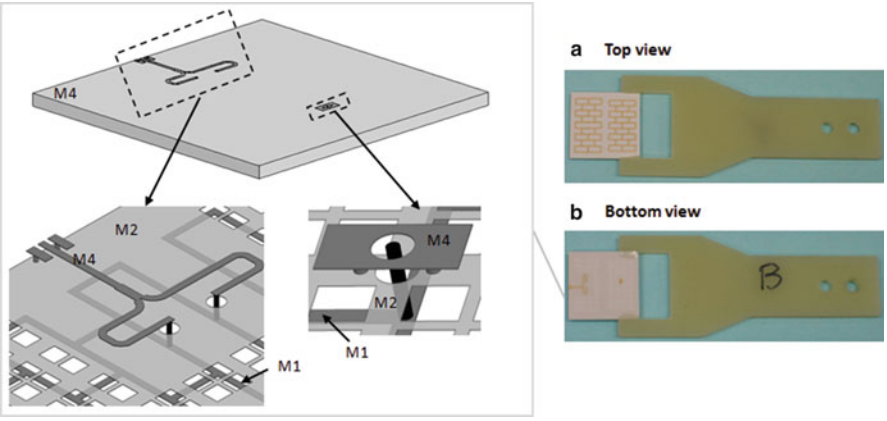


Fig. 17 Photos of the dual-grid antenna array B in the package with illustrations of (**a**) top view and (**b**) bottom view with integrated balun for testing the differential grid antenna array with the single-ended equipment

DOI 10.1007/978-981-4560-75-7_44-1

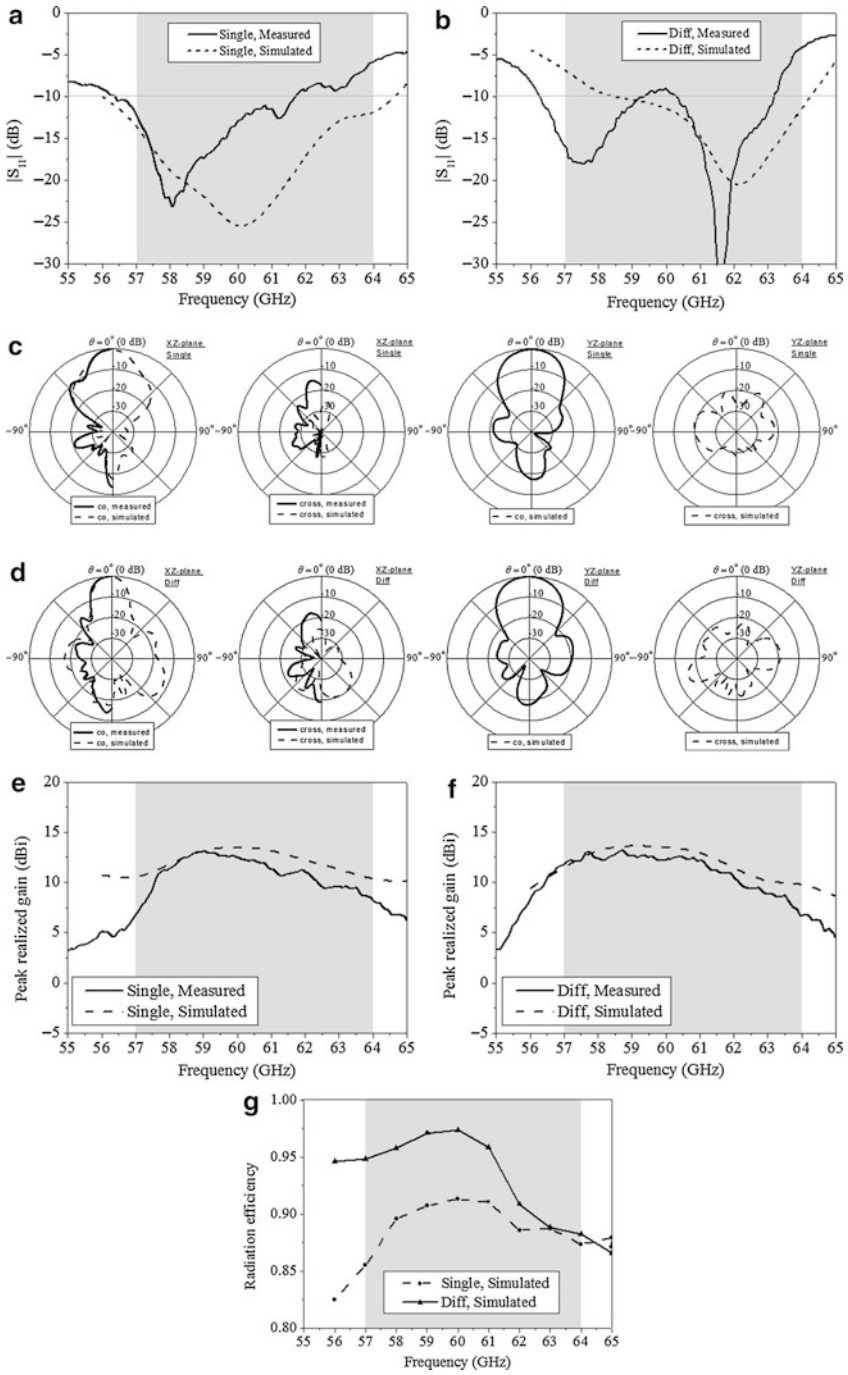


Fig. 18 The simulated and measured results of the single end and the differential antenna in the package sample B: (**a**, **b**) $|S_{11}|$, (**c**, **d**) patterns at 61.5 GHz, (**e**, **f**) peak realized gain, and (**g**) efficiency

Millimeter-Wave Applications

The literature review shows that GAA has wide applications in microwave frequencies. Zhang's group has made it an excellent antenna candidate for millimeter-wave "antenna-in-package" technology by providing various elegant package solutions especially at 60-GHz band as mentioned in the above sections. In addition, they successfully developed GAA for automotive radar sensor applications at 24 GHz (Zhang et al. 2011a, b, 2012) as mentioned in the introduction part.

DOI 10.1007/978-981-4560-75-7_44-1

Inspired by the work done by the Zhang's group, W. Menzel adopted the GAA for 79 GHz automotive radar (Bauer and Menzel 2011; Frei et al. 2011; Bauer et al. 2013), where the wide band GAA with a broadside beam is designed to cover 77–81 GHz of the sensor operating frequency. A total of four SIW-fed GAA are integrated in a LTCC package with a four-channel SiGe transceiver chip on the back, which offers four TRX channels with a very compact LTCC RF frontend size of only 23 × 23 mm. The radar measurement results are presented and the feasibility of the concept is verified in Bauer et al. (2013).

T. Zwick implemented D-band 120 GHz GAAs on liquid crystal polymer (LCP) substrate as a lid of a surface-mounted chip package (Beer et al. 2013). Two prototypes are fabricated to demonstrate the feasibility of the package concept. Prototype 1 consists of three parts: a multilayer LCP GAA antenna as a lid, a $2 \times 2 \times 0.254$ mm^3 MMIC dummy, and an $8 \times 8 \times 0.4$ mm^3 LTCC package base with a $5 \times 5 \times 0.3$ mm^3 cavity to hold the MMIC. The D-band RF signal from MMIC dummy is flip-chip interconnected to the top GAA lid for further radiation. The other low-frequency signals from dummy MMIC are flip-chip interconnected to the top GAA lid signal traces and then distributed to connect to the bond pads of the LTCC package base and finally connected to the outside circuits through signal vias and solder pads on the bottom side of the LTCC package base. The LTCC package base uses Hirai Precision LTCC process with a tape material CS71 from Japan. The MMIC dummy uses thin film technology with high permittivity Alumina material from Hightec MC AG, Switzerland, and its stud bumps with a challenging diameter of ~60 μm and height of 30 μm placed by Hybrid SA, Switzerland. The LCP substrate uses the challenging dimension with a line width of 60 μm and a spacing of 40 μm. The whole prototype is assembled using a Finetech Fineplacer Pico in the lab at Karlsruhe Institute of Technology (KIT), Germany. The cutout in the LCP substrate enables the MMIC feeding lines to be probe touched for testing. The $|S_{11}|$, antenna gain, patterns are finally measured showing the feasibility of the package concept. A Prototype with two GAA integrated in the same LCP lid is also demonstrated targeted to connect to a 122-GHz bistatic radar IC with both transmitter and receiver circuits in a single die (Beer et al. 2013).

Summary

It is concluded that GAA has wide applications from microwave to millimeter-wave frequencies. It demonstrates an excellent antenna candidate for millimeter-wave "antenna-in-package" technology as seen from the 60-GHz, 79-GHz, and 120-GHz application examples. Using multilayer millimeter-wave fabrication technologies, such as LTCC, LCP, etc., the GAA can provide various elegant package solutions with wide choice of differential feeding or single feeding, CP or LP radiation scheme, as well as good mechanical property resistant to the warpage.

Cross-References

- ▶ Circularly Polarized Antennas
- ▶ Holographic Antennas
- ▶ Low-Profile Antennas
- ▶ Metamaterials and Antennas
- ▶ Microstrip Patch Antennas
- ▶ Millimeter-Wave Antennas and Arrays
- ▶ Mm-Wave Sub-mm-Wave Antenna Measurement
- ▶ On-Chip Antennas

DOI 10.1007/978-981-4560-75-7_44-1

► Satellite Antennas on Vehicles
► Substrate-Integrated Waveguide Antennas

References

Bauer F, Menzel W (2011) A 79 GHz microstrip grid array antenna using a laminated waveguide feed in LTCC. In: Proceedings IEEE symposium on antennas and propagation, Spokane, WA, USA, pp 2067–2070

Bauer F et al (2013) A 79 GHz radar sensor in LTCC technology using grid array antennas. IEEE Trans Microw Theory Tech 61(6):2013–2521

Beer S et al (2013) D-band grid-array antenna integrated in the lid of a surface-mountable chip-package. In: Proceedings European conference on antennas and propagation, Gothenburg, Sweden, pp 1318–1322

Budka TP (2001) Wide-bandwidth millimeter-wave bond-wire interconnects. IEEE Trans Microw Theory Tech 49(4):715–718

Conti R et al (1981) The wire-grid microstrip antenna. IEEE Trans Antennas Propag AP-29(1):157–166

Doan C et al (2004) Design considerations for 60 GHz CMOS radios. IEEE Commun Mag 42:132–140

Floyd B et al (2006) A silicon 60 GHz receiver and transmitter chipset for broadband communications. ISSCC digest of technical papers, San Francisco, pp 184–185

Frei M et al (2011) A 79 GHz differentially fed grid array antenna. In: Proceedings European microwave conference, Manchester, UK, pp 1320–1323

Gunnarsson SE et al (2005) Highly integrated 60 GHz transmitter and receiver MMICs in a GaAs pHEMT technology. IEEE J Solid-State Circ 40(11):2174–2186

Gunnarsson SE et al (2007) 60 GHz single-chip front-end MMICs and systems for multi-Gb/s wireless communication. IEEE J Solid-State Circ 42(5):1143–1157

He J, Zhang Y (2008) Design and analysis of SPST and SPDT switches for 60-GHz applications in 65-nmCMOS. In: Proceedings Asia-Pacific microwave conference, Hong Kong and Macau, China, pp 1–4

Hildebrand L (1992) Then analysis of microstrip wire-grid antenna arrays. Master thesis

Huang D et al (2006) A 60 GHz CMOS differential receiver front-end using on-chip transformer for 1.2 volt operation with enhanced gain and linearity. In: Proceedings symposium on VLSI Circuits, Honolulu, pp 144–145

Iitsuka Y et al (2009a) Grid array antenna composed of V-shaped and rhombic elements for beam steering. In: Proceedings IEEE AP-S international symposium, Charleston, USA, pp 1–4

Iitsuka Y et al (2009b) Circularly polarized spiral grid array antenna for beam scanning. Proceedings International Symposium on Antennas and Propagation, Bangkok, pp 5–8

Iitsuka Y et al (2010) Circularly polarized grid array antenna composed of open-loop elements for beam steering. In: Proceedings IEEE AP-S international symposium, Toronto, pp 1–4

Iitsuka Y et al (2012) Grid array antenna radiating a circularly polarized wave. In: Proceedings International Symposium on Antennas and Propagation, Nagoys, pp 130–133

Kawano T, Nakano H (1998) Grid array antenna with c-figured elements. In: Proceedings IEEE Antennas and Propagation Society International Symposium, Atlanta, GA, USA, vol 2. pp 1154–1157

Kawano T, Nakano H (1999) Cross-mesh array antennas for dual LP and CP waves. In: Proceedings IEEE Antennas and Propagation Society International Symposium, Orlando, FL, USA, vol 4. pp 2748–2751

DOI 10.1007/978-981-4560-75-7_44-1

Kawano T, Nakano H (2000) Dual-polarized cross-mesh array antennas. In: Proceedings IEEE Antennas and Propagation Society International Symposium, Orlando, Salt Lake City, UT, USA, vol 2. pp 522–525

Kraus J (1964a) A backward angle-fire array antenna. IEEE Trans Antennas Propag 12(1):48–50

Kraus J (1964b) Wave velocities on the grid structure backward angle-fire antenna. IEEE Trans Antennas Propag 12:509–510

Kraus J (1966) Backward angle non-resonant wire mesh antenna array. US Patent 3209688

Lamminen AEI et al (2008) 60 GHz patch antennas and arrays on LTCC with embedded-cavity substrates. IEEE Trans Antennas Propag 56(9):2865–2874

Maruhashi K et al (2001) 60 GHz-band flip-chip MMIC modules for IEEE1394 wireless adapters. In: Proceedings 31th European microwave conference, London, England, pp 1–4

Nakano H, Kawano T (1997) Grid array antennas. In: Proceedings IEEE Antennas and Propagation Society International Symposium, Montreal, Quebec, Canada, vol 1. pp 236–239

Nakano H et al (1994) Numerical analysis of a grid array antenna. In: Proceedings international conference on communication system, Singapore, vol 2. pp 700–704

Nakano H et al (1995) Center-fed grid array antennas. In: Proceedings IEEE Antennas and Propagation Society International Symposium, Newport Beach, CA, USA, vol 4. pp 2010–2013

Nakano H et al (1996) Honeycomb wire antenna. Electron Lett 32(21):1937–1938

Nakano H et al (1997) The radiation characteristics of honeycomb antennas. In: Proceedings 10th international conference on antennas and propagation, Edinburgh, vol 1. pp 350–353

Nakano H et al (1998a) Meander-line grid array antenna. IEE Proc Microw Antennas Propag 145(4):309–312

Nakano H et al (1998b) Numerical analysis of honeycomb antennas with an electromagnetic feed system. IEE Proc Microw Antennas Propag 145(1):99–103

Nakano H et al (2001a) A cross-mesh array antenna. In: Proceedings 11th international conference on antennas and propagation, Manchester, vol 1. pp 327–330

Nakano H et al (2001b) Mesh antennas for dual polarization. IEEE Trans Antennas Propag 49(5):715–723

Nakano H et al (2001c) Analysis of a printed grid array antenna by a fast MoM calculation technique. In: Proceedings 11th international conference on antennas and propagation, Manchester, vol 1. pp 302–305

Nakano H et al (2005) A fast MoM calculation technique using sinusoidal basis and testing functions for a wire on a dielectric substrate and its application to meander loop and grid array antennas. IEEE Trans Antennas Propag 53(10):3300–3307

Nakano H et al (2007) A modified grid array antenna radiating a circularly polarized wave. In: Proceedings IEEE 2007 international symposium microwave antennas and propagation and EMC technologies for wireless communications, Hangzhou, pp 527–530

Nakano H et al (2013a) Rhombic grid array antenna. IEEE Trans Antennas Propag 61(5):2482–2489

Nakano H et al (2013b) Loop-based circularly polarized grid array antenna with edge excitation. IEEE Trans Antennas Propag 61(8):4045–4053

Ohata K et al (2003) 1.25 Gbps wireless gigabit Ethernet link at 60 GHz-band. In: IEEE MTT-S international microwave symposium digest, Philadelphia, pp 373–376

Palmer K, Cloete J (1997) Synthesis of the microstrip wire grid array. In: Proceedings 10th international conference on antennas and propagation, Edinburgh, vol 1. pp 114–118

Pfeiffer U et al (2006) A chip-scale packaging technology for 60-GHz wireless chipsets. IEEE Trans Microw Theory Tech 56(8):3387–3397

Sun M, Zhang Y (2008) Design and integration of 60-GHz grid array antenna in chip package. In: Proceedings Asia-Pacific microwave conference, Macau, pp 1–4

Sun Y et al (2006) An integrated 60 GHz transceiver front end for OFDM in SiGe: BiCMOS. Wireless World Research Forum 16, Shanghai
Sun M et al (2008) Integration of Yagi antenna in LTCC package for differential 60-GHz radio. IEEE Trans Antennas Propag 56(8):2780–2783
Sun M et al (2009a) Integration of grid array antenna in chip package for highly integrated 60-GHz radios. IEEE Antennas Wirel Propag Lett 8:1364–1366
Sun M et al (2009b) A ball grid array package with a microstrip grid array antenna for a single-chip 60-GHz receiver. IEEE Trans Antennas Propag 59(6):2134–2140
Tanomura M et al (2008) TX and RX front-Ends for 60 GHz band in 90 nm standard bulk CMOS. ISSCC digest technical papers, San Francisco, pp 558–559
Tiuri M et al (1974a) Chain antenna. In: Proceedings IEEE Antennas and Propagation Society International Symposium, Atlanta, Georgia, pp 274–277
Tiuri M et al (1974b) Chain antenna. US Patent 3806946
Tomkins A et al (2009) A zero-IF 60 Ghz 65 nm CMOS transceiver with direct BPSK modulation demonstrating up to 6 Gb/s data rates over a 2 m wireless link. IEEE J Solid-State Circ 44(8):2085–2099
Toth J et al (1983) Wire grid microstrip antenna. US Patent 4376938
Zhang Y, Liu D (2009) Antenna-on-chip and antenna-in-package solutions to highly-integrated millimeter-wave devices for wireless communications. IEEE Trans Antennas Propag 57(10):2830–2841
Zhang Y, Sun M (2008) Grid array antennas and an integration structure. US Patent 20110241969
Zhang B, Zhang Y (2010) A circularlypolarized microstrip grid array antenna for 60 GHz radios. In: Proceedings IEEE Asia-Pacific microwave conference, Yokohama, pp 2194–2197
Zhang B, Zhang Y (2011) Analysis and synthesis of millimeter-wave microstrip grid array antennas. IEEE Antennas Propag Mag 53(6):42–55
Zhang B, Zhang Y (2012a) Grid array antennas with subarrays and multiple feeds for 60-GHz radios. IEEE Trans Antennas Propag 60(5):2270–2275
Zhang B, Zhang Y (2012b) A microstrip array antenna for 60-GHz applications. In: Proceedings IEEE Aisa-Pacific conference antennas and propagation, Singapore, pp 88–89
Zhang B, Zhang Y (2013) A high-gain grid array antenna for 60-GHz antenna-in-package applications. In: Proceedings international symposium EM theory, Hiroshima, pp 195–198
Zhang B, Zhang Y (2014) Grid array antennas. Wiley Encyclopedia of electrical and electronics engineering, John Wiley & Sons, New York
Zhang Y et al (2008a) Integration of slot antenna in LTCC package for 60-GHz radios. Electron Lett 44(5):330–331
Zhang Y et al (2008b) Antenna and transmit/receive switch for single-chip radio transceivers of differential architecture. IEEE Trans Circ Syst I 55(11):3564–3570
Zhang Y et al (2009) Antenna-in-package design for wirebond interconnection to highly-integrated 60-GHz radios. IEEE Trans Antennas Propag 57(10):2842–2852
Zhang B et al (2011a) Design of low cost linearly-polarized microstrip grid array antenna for 24 GHz Doppler sensors. In: Proceedings 2011 I.E. international topical meeting microwave photonics, Singapore, pp 93–96
Zhang L et al (2011b) Microstrip grid and comb array antennas. IEEE Trans Antennas Propag 59(11):4077–4084
Zhang Y et al (2011c) Dual grid array antennas in a thin-profile package for flip-chip interconnection to highly integrated 60-GHz radios. IEEE Trans Antennas Propag 59(4):1191–1199

DOI 10.1007/978-981-4560-75-7_44-1

Zhang L et al (2012) A microstrip grid array antenna for 24 GHz Doppler sensors. In: Proceedings IEEE Asia-Pacific conference antennas and propagation, Singapore, pp 257–258

Zhang B et al (2013a) Integration of quadruple linearly-polarized microstrip grid array antennas for 60-GHz antenna-in-package applications. IEEE Trans Comp Packag Manuf Technol 3(8):1293–1300

Zhang B et al (2013b) A circularly-polarized array antenna using linearly-polarized sub grid arrays for highly-integrated 60-GHz radio. IEEE Trans Antennas Propag 61(1):436–439

Zhang B et al (2013c) A high-gain microstrip grid array antenna on low temperature co-fired ceramic for 60-GHz applications. In: Proceedings European conference on antennas and propagation, Gothenburg, pp 103–107

Zhang L et al (2013d) Integration of dual-band monopole and microstrip grid array for single-chip triband application. IEEE Trans Antennas Propag 61(1):439–443

Zwick T et al (2004) Probe based MMW antenna measurement setup. In: Proceedings IEEE Antennas and Propagation Society International Symposium, Monterey, California, USA, Vol 1. pp 747–750

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_45-1

Reflectarray Antennas

Shenheng Xu[a]* and Fan Yang[a]
[a]Department of Electronic Engineering, Microwave and Antenna Institute, Tsinghua University, Beijing, China
[b]Department of Electronic Engineering, Tsinghua University, Beijing, China

Abstract

The reflectarray is a novel type of high-gain antennas, which combines some of favorable features of reflectors and arrays. It consists of a space-feeding primary source and an array of reflecting elements with individually predesigned phases to collimate or shape the high-gain beam in the desired direction. In this chapter, the operating principles of reflectarray antennas will first be introduced. Then the design procedure and analysis methods of both reflectarray elements and systems will be explained in detail and some practical design considerations will be discussed. Finally, some latest developments and applications, such as broadband and multiband techniques, reconfigurable reflectarray designs, shaped-beam and multi-beam reflectarrays, etc., will be presented.

Keywords

Antenna; Array; Electronic beam scanning; Gain; Microstrip array; Pattern; Phased array; Radar; Reflectarray; Satellite communication; Spaceborne

Introduction

High-gain antennas are crucial components in satellite communications, radar detection and tracking, remote sensing, military and defense applications, and so on. The most widely used high-gain antennas are reflectors because of the high radiation efficiency and well-developed design and analysis procedures. However, their radiation performances are not versatile enough to meet the ever-increasing demands of modern applications. The slow beam scanning capabilities heavily rely on mechanical structures due to its severe off-focus degradation. The specifically curved metallic surfaces are difficult to fabricate at high frequencies and even become prohibitive in some shaped-beam designs. The bulky reflectors also render it inapplicable where mass, volume, and conformal structure are of great consideration.

Microstrip arrays have been successfully applied in many advanced applications due to its versatile radiation performances, planar and low profile, and lithographic etching technique for accurate fabrication. With the integration of phase shifters and transmit/receive (T/R) modules, phased arrays also excel at fast electronic beam scanning, shaped-beam, and multi-beam applications. The major disadvantages, on the other hand, are the considerable loss in the beam-forming network (BFN), especially for high-gain antennas at high frequencies. Its design also becomes formidably difficult for large-scale beam-forming networks. The added controlling modules are expensive, prohibiting it from massive adoption. In addition to the power inefficiency, they also increase the overall mass and volume and require specially designed heat ventilation structures in high-power systems.

*Email: shxu@tsinghua.edu.cn

DOI 10.1007/978-981-4560-75-7_45-1

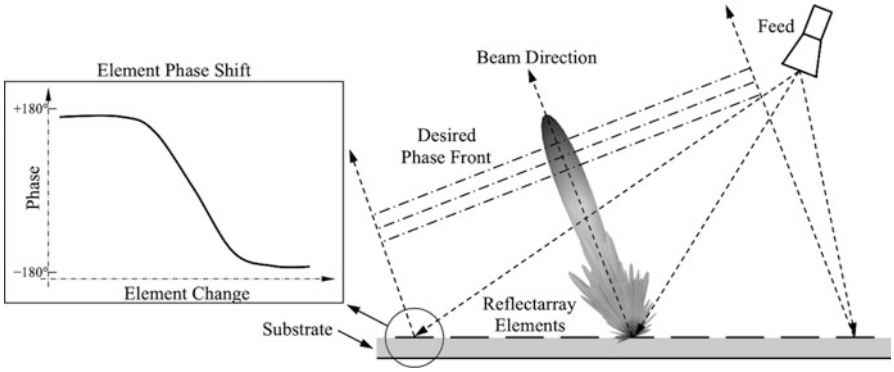


Fig. 1 General configuration of a reflectarray antenna

The reflectarray antenna (Huang and Encinar 2008; Shaker et al. 2014), as its name implies, is a hybrid of reflectors and arrays, which combines some favorable features of both antennas while mitigating their disadvantages. Its most general configuration is illustrated in Fig. 1, which consists of a space-feeding primary source (feed) and an array of reflecting elements with individually predesigned phases to collimate or shape the high-gain beam in the desired direction. The key parameter in a reflectarray is the phase shifts provided by the reflectarray elements arranged in array lattice, which are specifically designed to compensate for the differential spatial phase delays from the feed and form a planar (or shaped) phase front on the reflectarray aperture. The phase shift is usually achieved by varying one or more geometrical parameters of the element. By adding low-loss tuning devices or special substrate materials to each element, reconfigurable reflectarrays are also capable of generating electronically adjustable phase shifts and producing some of versatile functionalities as phased arrays.

As an alternative novel high-gain antenna configuration, the reflectarray technology has developed rapidly in recent years. Some favorable features, which successfully distinguish it among many candidates, are summarized as follows:

- Similar to reflectors, space-feeding mechanism is adopted so that the complicated and lossy beam-forming networks are avoided. Therefore, reflectarray antennas are usually able to achieve high radiation efficiency of more than 50 % (Huang and Encinar 2008).
- With proper analysis methods, the phase shift provided by each reflectarray element can be accurately designed. Similar to arrays, the phase adjustability of individual element allows the reflectarray to produce versatile radiation characteristics, including a high-gain beam in a desired direction, shaped-beam and multi-beam patterns, improved wide-angle beam scanning capability, etc.
- Combined with various phase synthesis or optimization methods, complex phase shift distributions can be readily designed and precisely achieved, thus producing very accurate shaped patterns. On the contrary to the difficulties in fabricating shaped reflectors with stringent accuracy, the fabrication process of shaped-beam reflectarrays is the same as simple single-beam reflectarrays without increasing the complexity or cost in fabrication. Furthermore, dual-linearly polarized elements can provide independent phase shifts for each polarization to obtain more flexible performance.
- The electronic phase adjustability can be achieved on the element level by incorporating low-loss and low-cost PIN diodes, RF MEMS switches, special substrate materials, or other tuning devices, in reflectarray element designs. Without using the lossy beam-forming network and expensive phase shifters (or T/R modules), electronically reconfigurable reflectarrays can provide fast electronic beam scanning capability with high efficiency and low cost.

DOI 10.1007/978-981-4560-75-7_45-1

- The array of reflectarray elements is usually fabricated on a grounded substrate using lithographic etching technology, which is a very matured process and ensures massive and low-cost manufacture with excellent accuracy. The reflectarray technology can readily cover the entire microwave and millimeter wave frequencies without significantly increasing fabrication complexity or cost.
- Reflectarrays are extremely low profile and light weighted. The reflectarray aperture is usually planar (or slightly curved or multifaceted), making it easily conformal to various mounting platforms.
- A large reflectarray aperture can be divided into small segments and assembled on site later, making fabrication, packaging, and shipment more convenient and keeping the cost low. Especially for spaceborne applications, compared to the conventional doubly curved reflectors, the planar structure of reflectarrays requires a much simpler and reliable folding mechanism to achieve a large-aperture deployable antenna with satisfactory surface accuracy. An inflatable reflectarray with elements printed on a large, flat, thin membrane can even further reduce its mass and stowing volume.

As a result, reflectarray antennas have been attracting growing interests in the recent years and a variety of research progresses have been accomplished. In this chapter, the fundamentals of reflectarray antennas, such as the operating principles, design procedure, and analysis methods, will be introduced. One of the main limitations of the reflectarray technology is its limited bandwidth. Hence, special attention will be paid to various broadband and multiband designs. And finally, some latest developments, innovations, and applications will be presented.

Reflectarrays: A Hybrid of Reflector and Array Antennas

The concept of reflectarray antennas was first conceived by Berry, Malech, and Kennedy in 1963, where short-ended waveguides with variable lengths illuminated by a horn feed were used to control the phase of the reflected wave (Berry et al. 1963). However, the structure was too bulky at lower microwave frequency so the related research was stalled for decades. With the development of microstrip arrays and lithographic etching technology, microstrip reflectarrays, in typical forms as seen nowadays, were introduced and intensively investigated in the late twentieth century (Huang 1991; Pozar and Metzler 1993; Pozar et al. 1997, 1999; Huang and Pogorzelski 1998). In the past 10 years or so, the increasing computational power enabled more rigorous analysis methods and phase synthesis/optimization approaches for reflectarray designs. Combined with driving demands of satellite communications, more diversified researches on reflectarrays were boosted and a variety of novel reflectarray technologies and designs have been accomplished.

As previously noted, a reflectarray is a hybrid of reflector and array antennas consisting of a feed and an array of many different elements printed on a grounded substrate. Considering the relatively complex structure, a systematic procedure is developed to accurately analyze and appropriately design a reflectarray antenna, as given in Fig. 2. It is usually carried out in several stages, and the two main stages are the element design and the system design, which will be discussed in detail in the following sections.

Reflectarray Elements

Similar to microstrip arrays, microstrip reflectarray elements are usually lithographically printed on a planar grounded substrate. Some typical reflectarray elements are illustrated in Fig. 3, which are patches with variable-length phase delay lines (Huang 1991), patches with variable sizes (Pozar and Metzler 1993), and split rings with variable rotation angles (Han and Chang 2003). They can take various forms and shapes but usually share the same property: the reflectarray elements are an array of reflecting microstrip elements arranged in a lattice configuration.

DOI 10.1007/978-981-4560-75-7_45-1

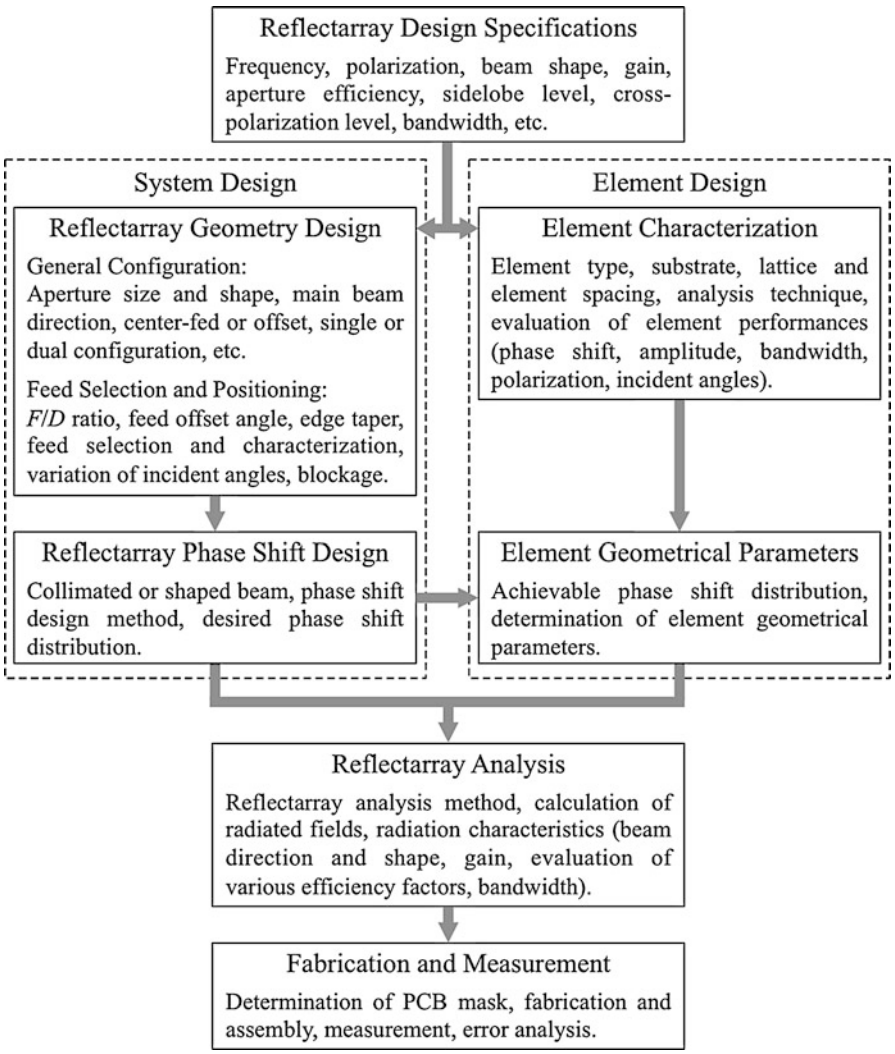


Fig. 2 Flowchart of the reflectarray antenna design and analysis procedure

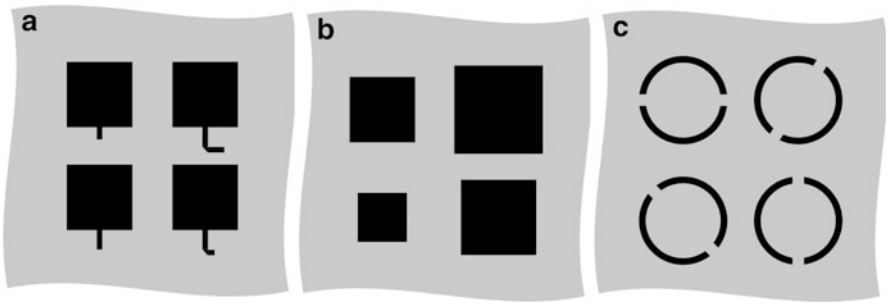


Fig. 3 Typical reflectarray elements: (**a**) patches with variable-length phase delay lines, (**b**) patches with variable sizes, (**c**) split rings with variable rotation angles

Unlike reflectors, which utilize specifically curved metallic surfaces to adjust the spatial phase delays and form the desired phase front, reflectarrays rely on the phase shift provided by individual element to form a specific phase front. Therefore, accurate characterization of a reflectarray element's response to the electromagnetic wave incident upon it is one of the most critical steps in the design procedure.

DOI 10.1007/978-981-4560-75-7_45-1

Element Analysis Techniques

The reflectarray element can be treated as an isolated element based on the assumption that each element works the same as an isolated element on a large grounded substrate and the mutual coupling between adjacent elements is negligible (Javor et al. 1995). However, it may not hold because the typical spacing between elements is so small that the mutual coupling effect can be strong enough to cause considerable errors. An improved surrounded-element approach is proposed, which considers the actual mutual coupling for a realistic configuration by including the surrounding nonidentical elements in element evaluations (Milon et al. 2007), but the enormous computational expenses it requires are hindering in practice.

Based on the observation that reflectarray elements form a quasiperiodic structure with similar elements surrounding each other, an infinite-array approach, which analyzes the reflection by an infinite array of identical elements with mutual coupling, is proposed (Pozar and Metzler 1993). This assumption is acceptable since the progressive phase distribution of the reflectarray aperture usually results in adjacent elements with similar shapes or sizes, except for those at the edges or with abrupt phase variation. Because of the similarity between the quasiperiodic structure of a reflectarray and the periodic structures of microstrip arrays and frequency selective surfaces (FSS), the analysis techniques from those fields can be directly borrowed and benefit the characterization of reflectarray elements. Specifically, by applying Floquet's theorem, the analysis of an infinite array with mutual coupling is dramatically reduced to only one unit element, thus resulting in extremely high efficiency and good accuracy (Huang and Encinar 2008). Therefore, the infinite-array approach has been widely adopted in reflectarray element analysis.

Various full-wave numerical methods, including the spectral domain method of moments (MoM), finite element method (FEM), and finite-difference time-domain (FDTD) method, have been employed to implement the infinite-array approach. Other than certain difference in computational cost, they all are able to yield accurate results. Some commercial full-wave electromagnetic simulators, such as Ansoft HFSS (HFSS 2014) and CST Microwave Studio (CST 2014), are also capable of fulfilling this purpose with built-in periodic boundaries and Floquet mode excitations. A crucial step in successfully applying these methods is to appropriately surround the unit cell with electric and magnetic walls (or periodic boundaries), which properly replicate it and span the desired infinite periodic structure. As an example, the numerical analysis model of a square microstrip patch with variable size using Ansoft HFSS is shown in Fig. 4. The unit cell is surrounded by the periodic boundaries. The port is set up with Floquet mode and de-embedded onto the patch surface. A plane wave with specific polarization and propagation direction is excited at the port, and by calculating the complex scattering parameter, the phase shift and amplitude response of the element can thus be achieved.

Other than the approaches mentioned above, the equivalent circuit method has also been used to speed up the calculation process (Bozzi et al. 2004; Shaker et al. 2014). For instance, a parallel *RLC* equivalent circuit model (Bozzi et al. 2004) can be utilized to characterize a lossy patch with variable size, as shown in Fig. 5, which provides useful insight into the physics of the reflectarray element. However, the construction of an equivalent circuit is usually based on a particular configuration, and its applicability is limited.

Evaluation of Element Performances

A reflectarray element reacts to the electromagnetic wave incident upon it and scatters it according to its design parameters. Its reflection characteristics directly affect the performances of the whole antenna. A careful evaluation of the element performances is important and critical, which helps determine the criteria for the selection of element configurations and optimization of element parameters to meet the design requirements. The characterization of a reflectarray element can be accomplished using one of the element analysis techniques mentioned above. Several key element performance parameters need to be

DOI 10.1007/978-981-4560-75-7_45-1

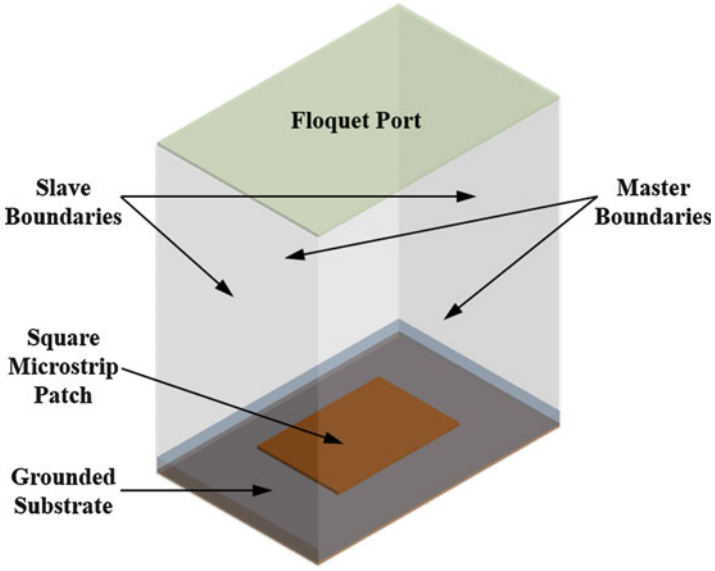


Fig. 4 Element analysis using Ansoft HFSS

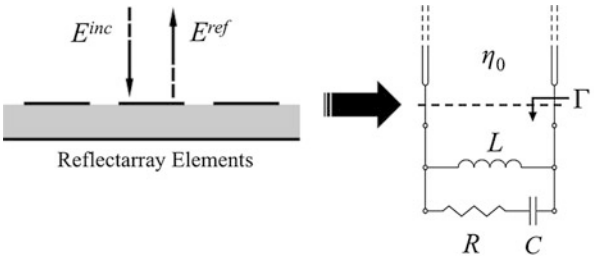


Fig. 5 Equivalent circuit model for a lossy reflectarray element

carefully evaluated, such as the phase shift, the reflection amplitude, the element bandwidth, and the element's responses to different polarization and different incident angles.

A detailed discussion on those parameters is carried out through a case study. As an example, a square microstrip patch with variable size at Ka-band is analyzed using the infinite-array approach in Ansoft HFSS, as shown in Fig. 4. The operating frequency is 32.5 GHz. The unit cell size is 4.6 mm in both directions, approximately half wavelength at the operating frequency. The thickness of the grounded substrate is 0.51 mm, with the dielectric constant ε_r of 2.55 and the loss tangent tanδ of 0.0019. A plane wave polarized in x direction is excited at the Floquet port and normally incident unto the patch. By de-embedding the port and calculating the complex scattering parameters, the reflection characteristics of the element are obtained, from which the element performance parameters can thus be achieved.

First and most importantly, the element phase shift is defined as the phase jump between the phases of the reflected and incident waves of the element, which equals the phase of the complex scattering parameter S_{11}. Different element phase shift is achieved by varying one or more geometrical parameters of the element. And in this case, it is the patch size. By changing it from 1 to 4 mm, the element phase shift versus the patch size can be obtained and plotted in Fig. 6a, which is known as the phase curve. Because of its typical S-shaped nonlinear variation, the phase curve is often called the S-curve as well. It is a measure of the phase adjustability of the element and is the fundamental ground on which the entire reflectarray concept is built. According to the required phase shift distribution of the reflectarray to form the desired phase front, the element sizes can be determined using the phase curve.

DOI 10.1007/978-981-4560-75-7_45-1

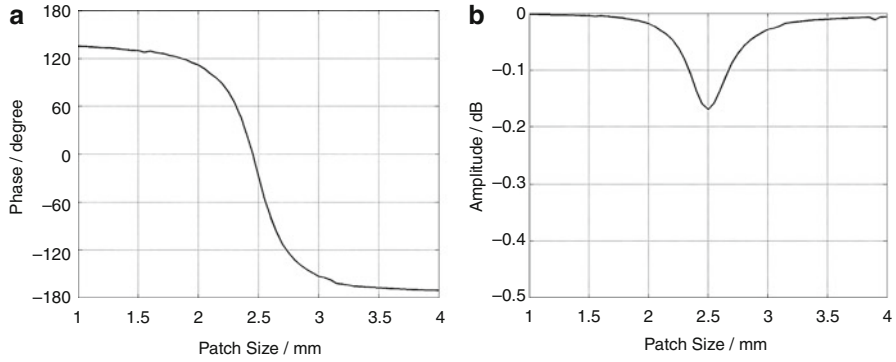


Fig. 6 (**a**) Phase curve and (**b**) reflection amplitude of square microstrip patches with variable sizes

The phase curve is one of the most crucial design curves in reflectarray designs and requires close attention. Two parameters are usually employed to examine the goodness of a phase curve. The first parameter is the maximum phase shift swing, which defines the phase range a set of reflectarray elements can possible cover. An ideal phase shift swing should cover at least 360°, and a smaller coverage inevitably results in phase errors. In Fig. 6a it is seen that the phase shift of the patches with variable sizes swings from −170° to +135°, so the phase range is approximately 305°. Although less than ideal, this coverage is often considered acceptable in practice and the resultant phase error is within an allowed range. There is no standard guideline on the minimum phase shift swing, and it is up to a designer's empirical knowledge and the trade-off among different performance parameters. The second one is the slope of the phase curve when it crosses 0°. It represents the resonance of the infinite-array structure and can be considered as a representative of the element bandwidth. A more gradual slope leads to a wider element bandwidth. Moreover, a steep slope causes the element phase response overly sensitive to the element change, and a minor fabrication error can cause considerable phase error.

The element reflection amplitude is plotted in Fig. 6b, which equals the amplitude of the complex scattering parameter S_{11}. Due to the existence of the ground plane, all the incident energy is scattered back, except for the part dissipated in the lossy substrate and nonideal conductors. The dielectric and conductor loss reduces radiation efficiency. The element amplitude response exhibits a maximum dip at the resonance of the structure where the maximum current flow on the element is excited and loss is maximized. The width of the dip also matters. A wider dip corresponds to a larger group of elements with high loss, leading to inefficiency of the reflectarray.

The back-scattered energy can be further divided into three components: the reradiated field due to the resonant activity of the patch, the specular reflected field due to the ground plane, and the scattered field due to the nonresonant structure of the patch. When the reradiated field does not coincide with the specular reflections, the latter components can cause higher sidelobes and reduced efficiency (Huang and Encinar 2008).

Both the phase and amplitude responses of the elements can be profoundly affected by the material properties. Firstly, as the thickness of the substrate increases, the element phase shift swing reduces as shown in Fig. 7a, leading to large phase error and reduced efficiency. The increased loss in thicker lossy dielectric substrates may also reduce efficiency. On the other hand, a thicker substrate reduces the Q-factor of the resonant structure. Hence, the slope of the phase curve becomes more gradual, improving the element bandwidth. Secondly, when the dielectric constant increases, the resonance of the structure shifts towards smaller patches and the slope of the phase curve becomes steeper, negatively impacting the element bandwidth, as shown in Fig. 7b. Thirdly, it is obvious that the increased loss tangent causes higher dielectric loss and reduced efficiency. And lastly, the finite conductivity of nonideal metallic materials

DOI 10.1007/978-981-4560-75-7_45-1

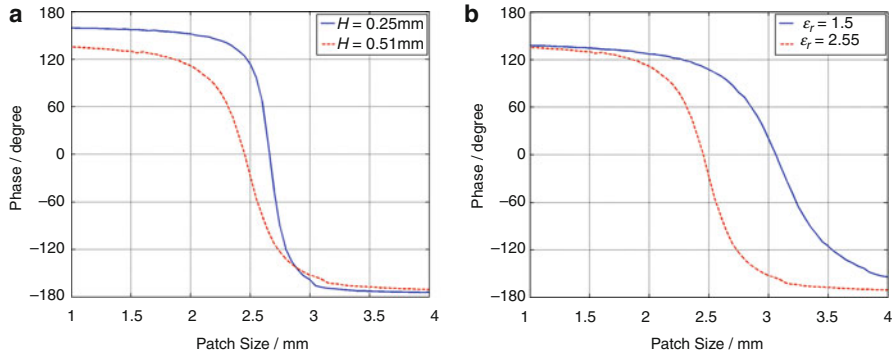


Fig. 7 Phase curves for (**a**) different thicknesses and (**b**) different dielectric constants of the substrate

increases the conductor loss (An et al. 2014a). Although mostly negligible at microwave frequencies, it becomes an important issue at terahertz (THz) and optical frequencies and requires special attention (Yang et al. 2012).

The unit cell size, or the element spacing in the quasiperiodic structure, is also an important design parameter of a reflectarray element. As a rule of thumb, it should be properly chosen to avoid grating lobes, which can be described as

$$\frac{d}{\lambda} \leq \frac{1}{1 + \sin\theta} \tag{1}$$

where d is the element spacing, λ is the wavelength in free space, and θ is either the incident angle or the main beam tilt angle from the broadside direction, whichever is larger (Huang and Encinar 2008). In practice, an element spacing of half wavelength is commonly employed. More recently, subwavelength element spacing is also adopted, where electrically small elements are closely spaced with element spacing less than half wavelength (Pozar 2007; Nayeri et al. 2010, 2011). The phase and amplitude responses for different unit cell sizes in this case study are presented in Fig. 8. It is observed that a coupled resonance is excited by the closely spaced elements. As the element spacing decreases to $\lambda/3$ and $\lambda/4$, both the slope of the phase curve and the amplitude loss improve, but the phase shift swing is considerably reduced. The subwavelength element spacing can substantially improve the bandwidth performance of the reflectarray antenna. However, the small patches and gaps also tighten the fabrication tolerance.

Another important element performance parameter is the element bandwidth. Reflectarray elements are intrinsically narrowband. As an example, the microstrip patch element with variable size generally has a bandwidth of 3–5 %. Its phase curves at different frequencies are shown in Fig. 9. Generally speaking, an element can achieve better bandwidth performance if the phase curves are smooth with gradual slopes and the variation between them is similar with element change. As briefly discussed above, careful tuning of element parameters can help improve the element bandwidth, such as using thick substrate with low dielectric constant and adopting subwavelength element spacing. For instance, using a thick layer of air or foam as substrate is a common way in practice. More recently, a lot of efforts have been devoted to this issue and various innovative broadband and multiband designs and techniques have been proposed.

A reflectarray element should also be able to properly respond to a polarized incident wave with correct phase and amplitude. Different element designs have different polarization performances, such as single- or dual-linear polarization and circular polarization, and appropriate selection of elements is critical. For instance, a square patch with variable size is suitable for single-linear polarization or circular polarization. Its side lengths can also be independently tuned so that the rectangular patch element can provide dual-linear polarization performance.

DOI 10.1007/978-981-4560-75-7_45-1

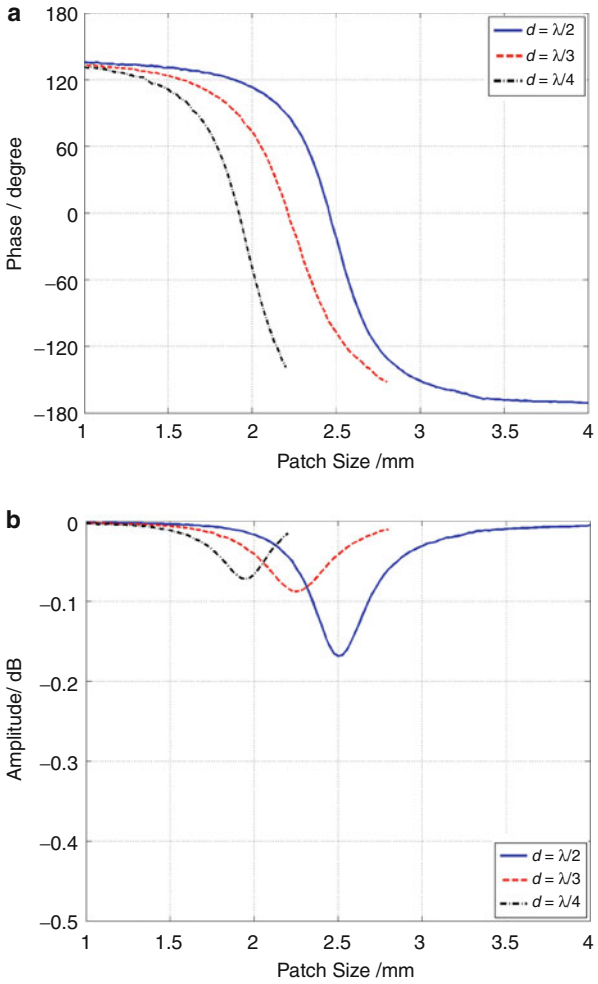


Fig. 8 (**a**) Phase and (**b**) amplitude responses for different unit cell sizes

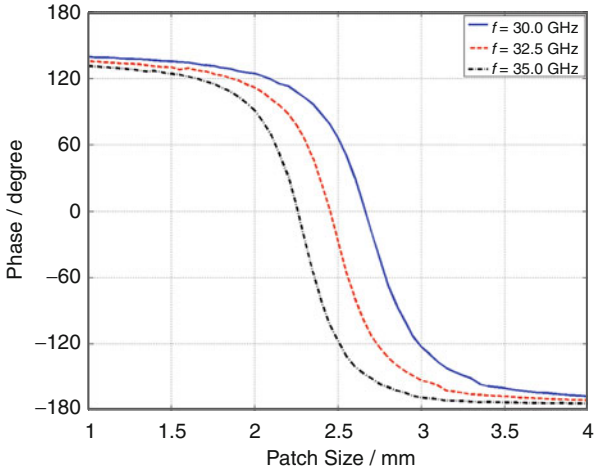


Fig. 9 Phase curves at different frequencies

DOI 10.1007/978-981-4560-75-7_45-1

Lastly, a reflectarray element scatters differently the incident wave from different directions. The incident angle of the element can be determined in receive mode or transmit mode, depending on how the reflectarray operates (Rengarajan 2009). Using a single pencil-beam reflectarray as an example, whose general configuration is depicted in Fig. 1, in receive mode a plane wave is incident from the main beam direction. The incident angle of the element is simply the main beam tilt angle and all elements have the same incident angle. In transmit mode, it becomes more complicated. The spherical wave originated from the feed impinges on the element at a certain angle. Assuming the element is located in the far-field region of the feed, then the incident wave can be locally considered as a plane wave by revoking the local plane wave approximation, and the incident angle of the element is determined by its location with respect to the feed. In this case, all elements have different incident angles. This is a unique feature of reflectarrays and requires close attention in the element analysis and reflectarray design process.

The phase curves of square patch elements for different incident angles are plotted in Fig. 10. It is observed that the element phase shift changes with the incident angle. Ideally, the reflectarray element should be analyzed for each incident angle in order to accurately evaluate its phase characterization and minimize the potential phase error in reflectarray designs. Considering the large number of elements in a typical high-gain reflectarray, however, this process becomes extremely time-consuming and even prohibitive. On the other hand, the difference between the phase curves for normal and oblique incidences is relatively small within a certain angular range, i.e., 25° in this case. By appropriately designing the reflectarray geometry, the incident angles of the well-illuminated elements can be limited in that range. Therefore, only the normal incidence case is analyzed and all element parameters are determined using the phase curve for normal incidence. It is a common practice in most reflectarray designs, and the phase errors caused by this approximation are considered acceptable. However, in some advanced applications where optimal aperture efficiency for high gain or precise phase contour for shaped beam is required, an element-by-element analysis process should be adopted.

Representative Types of Elements

Various reflectarray elements have been developed to better meet different requirements of reflectarray designs, such as radiation efficiency, polarization, bandwidth, fabrication accuracy and cost, etc. Each has its own advantages and limitations. According to adopted phasing mechanism, most microstrip reflectarray elements are categorized into three representative types, namely, the resonant elements, the elements loaded with phasing structures, and the circularly polarized elements with variable rotation angles. Some designs are introduced and their performances are briefly discussed.

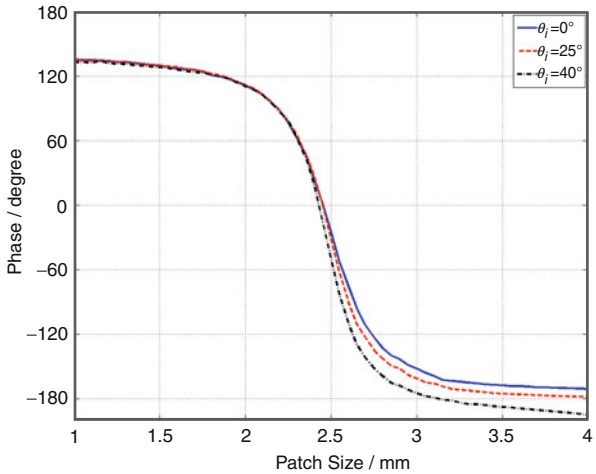


Fig. 10 Phase curves for different incident angles

DOI 10.1007/978-981-4560-75-7_45-1

The resonant elements employ the variable-size phasing mechanism to achieve the required phase adjustability. A patch element in an array environment is first designed to resonate at the operating frequency, and a small change in patch size results in substantial phase variation. The square patch with variable size, as shown in Fig. 3b, is the most widely used one in this category. Different shapes and forms, such as dipoles, square rings, and circular patches and rings, have also been investigated (Bialkowski and Sayidmarie 2008). They all have very simple configurations with a few geometrical parameters and can be easily fabricated using lithographic etching technology. The typical phase curve of the resonant elements is similar to that shown in Fig. 6a. The total phase shift swing is usually less than 360°, and the unobtainable phase range causes certain phase error and inefficiency. Because of the intrinsic resonant nature of the element, it exhibits a strong nonlinear variation with steep slope near the resonance and little change for extreme patch sizes. It limits the element bandwidth and makes it more sensitive to fabrication error. A large number of patches are off resonant, which may cause specular reflections, leading to higher sidelobes and reduced efficiency. More recently, a variety of multi-resonant elements have been proposed. By combining or stacking two or more simple resonant elements to create multiple resonance of the structure, they can effectively smooth the phase curve and substantially improve the element bandwidth performance (Encinar 2001; Chaharmir et al. 2009).

A different phasing mechanism is to load a resonant microstrip patch with variable-length phase delay lines, as shown in Fig. 3a (Huang 1991; Javor et al. 1995). The incident wave is coupled into the element, travels along the loading stub until it reflects back at the open end, and finally reradiates with a phase delay. The amount of phase delay varies approximately linearly with the length of the microstrip line, which can be calculated using an empirical formula (Javor et al. 1995). The stubs are usually bent to be able to fit within the limited space, which may cause spurious radiation in manifestation of higher cross-polarization and sidelobe levels. The long microstrip lines also increase dissipative loss in the elements. In order to mitigate these drawbacks, a multilayered slot-coupled patch element is developed (Bialkowski and Song 2001), and improved bandwidth performance can be achieved by using true-time delay lines (Carassco et al. 2008). In addition, slots can also be employed as loading structures (Chaharmir et al. 2003a). It is worthwhile to mention that the extra room in these elements allows the integration of tuning devices to achieve reconfigurable reflectarray designs (Rajagopalan et al. 2008; Carassco et al. 2012).

The circularly polarized elements with variable rotation angles produce desired phase shifts by physically rotating identical elements, and the phase shift is twice the angle of rotation because of the reflection process. This phasing mechanism can be applied to various circularly polarized elements, such as elements with loading stubs (Huang and Pogorzelski 1998), split rings as shown in Fig. 3c (Han and Chang 2003), single split rings (Yu et al. 2009), etc. A variation of this technique is to rotate the gaps of a square loop instead of the whole element itself in order to minimize the effectively occupied area by individual elements (Yu et al. 2012). The element rotation technique can achieve a full 360° phase shift swing. It also benefits the reflectarray performances in terms of sidelobe and cross-polarization levels because of the pseudorandomly arranged elements and improves radiation efficiency due to the suppression of specular reflection (Huang and Encinar 2008). It is crucial to note this technique is only applicable to circular polarization. And unlike the polarization reversal observed in other elements, the elements with variable rotation angles produce a same-sensed polarization response.

The above discussion only presents a commonly used method of categorization. Many novel reflectarray elements have been developed since the inception of reflectarray technology and may use one or a combination of various phasing mechanisms in its design. Other design parameters, such as the heights of the block elements (Nayeri et al. 2014; Deng et al. 2014) and the dielectric constant of the substrate material (Hu et al. 2008), can also be utilized to achieve the desired phase shifts. An appropriate selection of reflectarray element is the foundation of a reflectarray design.

DOI 10.1007/978-981-4560-75-7_45-1

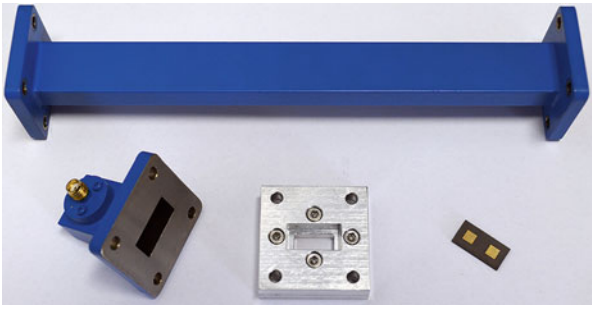


Fig. 11 Element measurement using a waveguide simulator

Element Measurement and Diagnosis

The phase shift and dissipative loss of a reflectarray element can be measured in a waveguide simulator (Hannan and Balfour 1965). An example of a Ku-band waveguide simulator is presented in Fig. 11, which includes a Ku-band standard waveguide, a coax-to-waveguide adapter, and a back plate. The element sample is inserted into the rectangular waveguide with the help of the back plate. The waveguide walls mirror the inserted sample and span an infinite array with specific element spacing. The opening on the back plate can be either sealed with conductive tape to ensure a total reflection or used for applying control signals in reconfigurable elements. The fundamental TE_{10} mode of the waveguide is excited, and the angle of incidence θ_0 of the impinging plane wave on the infinite array can be determined by

$$\sin\theta_0 = \frac{\lambda}{2a} \tag{2}$$

where λ is the free-space wavelength at the operating frequency and a is the longer dimension of the waveguide. The complex reflection coefficient is then measured using a vector network analyzer, from which the element phase and amplitude responses can be extracted. This measurement technique is easy to implement; however, there are several critical limitations that must be considered. The plane wave can only propagate in the H-plane with a fixed incident angle of θ_0. Only linear polarization can be excited with electric field parallel to the shorter waveguide walls. The element size and spacing are limited by the waveguide dimensions. And it may cause certain errors if the element responses in an actual configuration deviate from the quasi-infinite-array approximation.

It is sometimes desirable to check the accuracy of reflectarray element designs especially when the adjacent elements have abrupt phase variations or stringent performance requirements are demanded. Therefore, a visualization approach (Rajagopalan et al. 2012a) is proposed which can potentially assist the design, analysis, and diagnosis of reflectarrays. By properly interpreting the radiation mechanism of a reflectarray, the actual phase responses of an entire reflectarray are achieved through full-wave simulations and individual element performance in the actual configuration can be closely examined. Fine-tuning of element parameters can then be performed to achieve more accurate phase shifts as designed. However, the required computational expenses may limit its application in practice.

Reflectarray Design and Analysis

Reflectarrays utilize the same space-feeding mechanism as reflectors. Resemblance between them justifies that various theories and design guidelines from reflectors can be readily borrowed to facilitate the design and analysis of reflectarrays. On the other hand, reflectarrays rely on the phase adjustability of the array of elements, rather than specifically curved metallic surfaces in reflectors, to achieve the phase transformation of a spherical wavefront originated from the feed into a planar wavefront. This unique

DOI 10.1007/978-981-4560-75-7_45-1

feature gives more flexibility in reflectarray designs, but it also requires particular attention in analyzing its radiation characteristics.

Reflectarray Geometry Design

Similar to reflectors, reflectarray geometries can take various forms, such as a single reflectarray with a prime focus feed or an offset feed, in order to meet specific design requirements. For example, an offset geometry can avoid the feed blockage in a center-fed system and is widely adopted in practice. Moreover, unlike an offset reflector whose projected aperture is always smaller than its physical size, the main beam of an offset reflectarray can be aligned in the broadside direction so that the whole aperture is fully used. On the other hand, the variation of incident angles on the elements of an offset reflectarray is larger than that of a center-fed system, and some extremely large incident angles may occur for far-end edge elements. Under this circumstance, the actual element phase shift may considerably deviate from the phase curve for normal incidence, which is commonly assumed in reflectarray element analysis, resulting in phase errors and inefficiency. Meanwhile, less energy is intercepted and reradiated by those elements due to the element beamwidth effect (Huang and Encinar 2008), which may further reduce the radiation efficiency. Special attention needs to be paid to the design and analysis of those elements, making it more difficult to design offset reflectarrays. Therefore, a compromise can be made by reducing the feed offset angle while tilting the main beam away from the broadside. A blockage-free offset design with small feed offset angle is possible at the price of a reduced effective aperture size.

The reflectarray aperture is usually planar but can also be slightly curved for conformal designs. It can be circular, square, polygonal, or in other special shapes. The aperture's electrical size is determined by the system gain requirement. Generally speaking, the radiation performance in terms of sidelobe and cross-polarization levels and efficiency is better for reflectarrays of larger sizes, because the specular reflection, edge diffraction, and feed blockage effects become less noticeable compared to its desirable radiation from the elements.

A crucial design parameter F/D, the focal length to diameter ratio, must be properly chosen. A larger F/D usually can achieve better radiation performances in terms of cross-polarization level, bandwidth, and scan capability, but it may also increase the feed size and overall volume of the reflectarray antenna. A smaller F/D significantly increases the variation of spatial phase delays to the center and edge elements, which becomes a critical factor in bandwidth limitation. It may also result in extremely large incident angles for the edge elements. These are two distinctive differences compared to its reflector counterpart and require particular attention. In practice an F/D from 0.5 to 1 is commonly adopted.

The edge taper (ET) on the reflectarray aperture is another important parameter, which profoundly affects its radiation characteristics, such as beamwidth, sidelobe level, and efficiency, in the same way as it acts on reflectors. Generally speaking, the taper efficiency deteriorates and the spillover efficiency improves as the edge taper decreases, and optimal radiation efficiency can be achieved when the edge taper is around -10 dB. Therefore, the feed must be properly designed according to the chosen F/D so that it provides appropriate illumination on the reflectarray. Detailed derivations and calculations of these efficiency parameters will be presented in the following discussions.

Dual configurations have also been developed to fold the optics (Arrebola et al. 2008; Almajali et al. 2012; Tienda et al. 2012; Rajagopalan et al. 2013). Dual reflectarrays, as well as a combination of reflectarray and reflector, can be employed. Due to the unique phase controlling capability of reflectarray elements, some specific purposes, such as beam scanning (Hu et al. 2009), reflector surface distortion compensation (Xu et al. 2009; Rajagopalan et al. 2012b), and shaped beam for direct broadcast satellite (DBS) applications (Tienda et al. 2013), can be achieved. The complex structure, however, significantly increases the complexity in designing and analyzing a dual system.

DOI 10.1007/978-981-4560-75-7_45-1

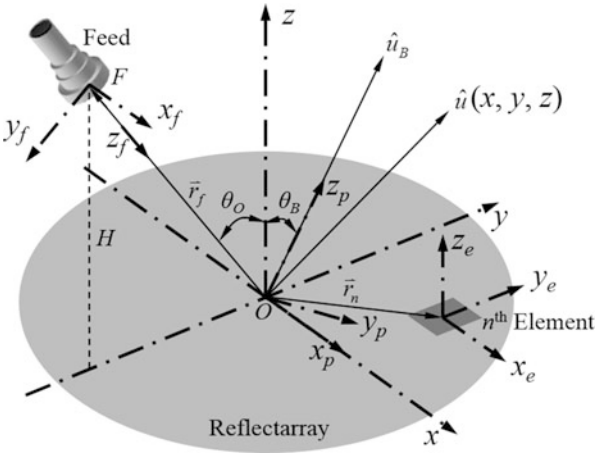


Fig. 12 The coordinate systems and configuration parameters of a reflectarray

Reflectarray Phase Shift Design

After the reflectarray geometry is determined, the phase shift distribution of the reflectarray required to produce the desired radiation characteristics can be obtained using a ray tracing method or a phase-only synthesis method. For a single high-gain beam design, the former is widely employed.

As illustrated in Fig. 12, four sets of coordinate systems are usually utilized to facilitate the analysis of a planar reflectarray with N elements. (x, y, z) refers to the reflectarray coordinates with its origin located at the center of the aperture O, and $\hat{u}$ is the observation direction. The one with subscript f refers to the feed coordinates with its origin located at the focus F. The phase center of the feed is positioned at the focus F, and the feed pattern is pointed toward the reflectarray aperture along the z_f axis. H is the feed height and θ_O is the feed offset angle. $\vec{r}_f$ is the position vector of the feed. The main beam is aligned in the $\hat{u}_B$ direction, which is also the z_p axis in the pattern coordinate system, and θ_B is the main beam tilt angle. The subscript e denotes the element coordinates whose origin is located at the center of each element, and $\vec{r}_n$ is the position vector of the nth reflectarray element. A coordinate transformation between different coordinate systems can be easily carried out using the Eulerian angles (Rahmat-Samii 1979).

Considering a reflectarray operating in transmit mode and assuming the reflectarray elements are in the far-field region of the feed, then the electromagnetic wave emanated from the feed can be considered as a spherical wavefront. The incident field on each element can be locally approximated by a plane wave with a phase proportional to the distance from the phase center of the feed to each element. To collimate the beam in the $\hat{u}_B$ direction, a uniform phase planar wavefront perpendicular to $\hat{u}_B$ is required. By using a simple ray tracing method, the desired phase shift φ_n of each element can be calculated as

$$\varphi_n = k\left(\left|\vec{r}_n - \vec{r}_f\right| - \vec{r}_n \cdot \hat{u}_B\right) + \Delta\varphi_0 + 2n\pi, n = 0, \pm 1, \pm 2, \cdots \tag{3}$$

where k is the free-space wave number, $\Delta\varphi_0$ is a constant reference phase, and the integer n is used to keep φ_n within 360° bound. It is critical to note that the phase shift relative to each other, rather than the absolute phase value of each element, is needed to fulfill the phase transformation. Hence, an arbitrary constant reference phase $\Delta\varphi_0$ may be added to all the elements. It is also observed that the main beam direction $\hat{u}_B$ is a design parameter independent of the feed position or the aperture geometry. For instance, a tilted pencil beam off the broadside can be achieved by a center-fed reflectarray. This is an example of the flexibility offered by reflectarrays that does not have an equivalent in reflectors.

Besides the reflectarrays that mimic parabolic reflectors to convert a spherical wavefront into a plane wave, the ray tracing method has also been successfully utilized to derive the closed-form expressions of

the required phase shifts of a reflectarray used as a hyperboloidal or ellipsoidal subreflector in a dual system (Almajali et al. 2012; Rajagopalan et al. 2013). Both full-wave simulations and measurements demonstrate that the reflectarrays are able to convert an incident spherical wave originated from one focus into another spherical wave from the secondary focus.

In reflectarrays, as opposed to active phased arrays with controllable phase and amplitude, only the phase shift of each element can be adjusted and its amplitude is determined by the illumination of the feed. Therefore, direct phase-only synthesis techniques from array designs can be borrowed, and various methods have been employed in advanced reflectarray designs, for example, the intersection approach for shaped-beam designs for DBS applications (Encinar et al. 2006, 2011), the alternating projection method (APM) for a quad-beam design (Nayeri et al. 2012), and the particle swarm optimization (PSO) method for improving scanned beam performance (Nayeri et al. 2013b). Generally speaking, these approaches are able to achieve satisfactory results, but the phase synthesis process is usually complicated and computationally extensive. Good understanding and *a prior* knowledge of the desired phase shift distribution of the reflectarray are needed to improve its effectiveness.

Reflectarray Analysis Methods

Assuming that the performances of the reflectarray elements can be accurately evaluated and the desired phase shift distribution or its closest values can be realized by the chosen element design, then the whole reflectarray can be analyzed to obtain its radiation characteristics such as the radiation pattern, gain, sidelobe and cross-polarization levels, bandwidth, etc. Three reflectarray analysis methods are widely adopted, namely, the array theory method, the aperture field method, and the full-wave analysis method.

The most straightforward approach to calculate the radiation pattern of a reflectarray is to compute the direct summation of the field radiated by each element using the array theory method (Chang and Huang 1995; Huang and Pogorzelski 1998; Nayeri et al. 2013a). To simplify the calculation, both the feed and reflectarray elements are modeled by an ideal $\cos^q(\theta)$-type point source with no azimuth dependence (Rahmat-Samii et al. 1981), and q_f and q_e are the matching q-factors controlling the feed and element patterns, respectively. Being located in the far-field region of the feed, the electric filed on each element originated from the feed can be expressed as

$$E^f = \frac{\cos^{q_f}\left(\theta_{f,n}\right)}{\left|\vec{r}_n - \vec{r}_f\right|} e^{-jk\left|\vec{r}_n - \vec{r}_f\right|} \tag{4}$$

where $\theta_{f,n}$ is the spherical coordinate of the nth element in the feed coordinate system. Taking into account the element beamwidth effect, the incident field intercepted by the element is modified as

$$E^i = \cos^{q_e}\left(\theta_{e,n}\right) \cdot E^f \tag{5}$$

where $\theta_{e,n}$ is the spherical coordinate of the feed in the nth element coordinate system. Assuming that by using a proper element analysis technique, the element's amplitude and phase responses are Γ_n and φ_n, respectively, the element excitation can then be written as

$$I_n = \frac{\cos^{q_f}\left(\theta_{f,n}\right)}{\left|\vec{r}_n - \vec{r}_f\right|} e^{-jk\left|\vec{r}_n - \vec{r}_f\right|} \cos^{q_e}\left(\theta_{e,n}\right) \Gamma_n e^{j\varphi_n} \tag{6}$$

Combined with the element pattern

DOI 10.1007/978-981-4560-75-7_45-1

$$A_n = (\theta, \phi) = \cos^{q_e}(\theta) e^{jk\left(\vec{r}_n \cdot \hat{u}\right)} \tag{7}$$

the radiated field of the reflectarray can be calculated by using a direct summation of the contributions from all the elements

$$E(\theta, \phi) = \sum_{n=1}^{N} \cos^{q_e}(\theta) \cos^{q_e}\left(\theta_{e,n}\right) \Gamma_n \frac{\cos^{q_f}\left(\theta_{f,n}\right)}{\left|\vec{r}_n - \vec{r}_f\right|} e^{-jk\left(\left|\vec{r}_n - \vec{r}_f\right| - \vec{r}_n \cdot \hat{u}\right) + j\varphi_n} \tag{8}$$

Obviously, Eq. 8 produces a uniform phase wavefront in the plane perpendicular to the $\hat{u}_B$ direction if the element phase shift φ_n is designed using Eq. 3, and a broadside radiation can be obtained if $\hat{u}_B$ coincides with the normal direction of the reflectarray aperture. It is also worthwhile to point out that if an appropriate element analysis technique, for instance, the infinite-array approach, is employed to obtain the element phase shift φ_n, the mutual coupling between elements is included in this calculation.

In general, the array theory method is able to yield a good prediction of some important radiation characteristics such as beam direction, main beamwidth, and general pattern shape of the first few sidelobes. The formulation of this approach is very simple and easy to implement. It requires little computational resources and computational time, making it suitable for integration with optimization techniques in phase synthesis applications. A 2-D inverse digital Fourier transform (IDFT) can be employed to further expedite the calculation (Huang and Encinar 2008). On the other hand, there also exist several noticeable limitations. The cross-polarization of the reflectarray cannot be obtained in this calculation, because the polarizations of the feed and elements are not accounted for in the simplified $\cos^q(\theta)$-type models. The effects of finite ground plane and edge diffraction are not taken into account, and the back radiations of the feed and reflectarray elements are ignored. Moreover, a careful loss analysis is required to improve the accuracy of gain calculation.

Based on the equivalence principle, the aperture field method has been developed to calculate the radiation pattern of a reflectarray from the tangential component of the reflected electric field on the reflectarray aperture (Pozar et al. 1997; Huang and Encinar 2008; Nayeri et al. 2013a). First, the incident field on the reflectarray aperture is obtained while considering the polarization of the feed. The radiated fields of x- and y-polarized feeds are given by Eqs. 9 and 10 (Rahmat-Samii 1988), respectively,

$$\vec{E}_x^f\left(\theta_f, \phi_f\right) = A_0 \left[\hat{\theta}_f C_E\left(\theta_f\right) \cos\phi_f - \hat{\phi}_f C_H\left(\theta_f\right) \sin\phi_f\right] \frac{e^{-jkr_f}}{r_f} \tag{9}$$

$$\vec{E}_y^f\left(\theta_f, \phi_f\right) = A_0 \left[\hat{\theta}_f C_E\left(\theta_f\right) \sin\phi_f + \hat{\phi}_f C_H\left(\theta_f\right) \cos\phi_f\right] \frac{e^{-jkr_f}}{r_f} \tag{10}$$

where A_0 is a complex constant and the subscript f denotes the feed coordinates. C_E and C_H are the E- and H-plane radiation patterns of the feed, respectively. They are typically modeled as $\cos^q(\theta)$-type functions as well

$$C_E\left(\theta_f\right) = \cos^{q_E}\left(\theta_f\right) \tag{11}$$

$$C_H\left(\theta_f\right) = \cos^{q_H}\left(\theta_f\right) \tag{12}$$

where q_E and q_H are the matching q-factors determined from the simulated or measured pattern data of the feed. Note that the above feed model could be further expanded to include cross-polarization fields.

DOI 10.1007/978-981-4560-75-7_45-1

Using the coordinate transformation (Rahmat-Samii 1979), the x- and y-components of the incident electric field on the reflectarray aperture can then be computed from these equations. Alternatively, the incident fields can be directly obtained from the full-wave simulation or measurement data of the feed. In this case, the near-field effect of the feed radiation may also be considered and the accuracy of reflectarray analysis may be improved.

Using the infinite-array approach for the element analysis, the generalized scattering matrix $[\boldsymbol{\Gamma}]$ relating the Cartesian components of the incident and scattered fields of the nth element in a periodic environment can be obtained

$$\begin{bmatrix} E^s_{x,n} \\ E^s_{y,n} \end{bmatrix} = \begin{bmatrix} \Gamma_{xx} & \Gamma_{xy} \\ \Gamma_{yx} & \Gamma_{yy} \end{bmatrix} \cdot \begin{bmatrix} E^i_{x,n} \\ E^i_{y,n} \end{bmatrix} \tag{13}$$

The radiated field of each element can then be readily calculated using the vector potentials (Balanis 2005), and the reflectarray pattern is achieved by summing the contributions from all the elements.

In general, the tangential electric field must be extended to an infinite plane containing the reflectarray aperture. However, it is always limited to the size of reflectarray aperture, and the field outside of its boundary is assumed to be zero. Another important approximation in this calculation is that the tangential components are assumed to be constant within each periodic cell, so the integration over the reflectarray aperture is approximated by a summation. Discontinuities at the border of each unit cell may result in erroneous radiation patterns. Only the fundamental Floquet harmonics are used in the element analysis, and the higher-order modes are ignored. It is also worthwhile to mention that to simplify the calculations, usually only the electric field is used in the computation. This corresponds to the second principle of equivalence, which implies that the reflectarray aperture is immersed in a perfect electric conductor. A more accurate computation is possible by using both tangential electric and magnetic fields on the reflectarray aperture, i.e., the first principle of equivalence, but the computational expenses also increase.

The aperture field method uses a more accurate feed model and takes into account both the co- and cross-polar components from the feed, the geometrical projection, and the elements. It allows an accurate prediction of the radiation pattern in a wider angular range as well as the cross-polarization level. The shortcomings of this approach are the significantly more complicated formulation and analysis program development. The computational time is also increased.

With the ever-increasing computational power nowadays, the full-wave analysis of the whole reflectarray also becomes possible. Some general-purpose commercial electromagnetic simulation software, such as Ansoft HFSS (HFSS 2014), CST Microwave Studio (CST 2014), FEKO (2014), etc., can be employed to design and analyze reflectarrays. They are able to provide accurate results by taking into account all approximations in reflectarray element design and mutual coupling, in addition to edge diffraction effects. Although having already been widely used in characterizing the element performances in an infinite periodic array, the applications of full-wave analysis methods are limited to small- or medium-size reflectarrays. The largest reflectarray size that can be handled by the software obviously depends on the available computational resources and the complexity of the software. Alternatively, useful information on the performance of the reflectarray can be obtained without simulating the whole structure. For instance, a good estimation of the reflectarray gain bandwidth is achieved by calculating the scattering from the center row of a large reflectarray surrounded by perfectly magnetic conductor (PMC) walls (Chaharmir et al. 2010).

Other approaches have also been developed to improve the accuracy or efficiency of the aforementioned reflectarray analysis methods. For instance, an improved technique based on the equivalent currents from a continuous plane wave spectrum and the spectral dyadic Green's function is proposed to account for the effect of finite ground plane (Zhou et al. 2011). An appropriate selection of reflectarray

DOI 10.1007/978-981-4560-75-7_45-1

analysis method must be carried out according to the requirements of reflectarray designs and the availability of computational resources.

Gain and Efficiencies of Reflectarrays

The ultimate goal of the design and analysis of a reflectarray is to obtain its important radiation characteristics, such as the radiation pattern, main beam direction, gain, aperture efficiency, sidelobe and cross-polarization levels, and bandwidth, to meet the design requirements. With the reflectarray geometry and element parameters determined, the radiation pattern of the reflectarray can be computed using the selected reflectarray analysis method, from which main beam direction, sidelobe, and cross-polarization levels can be readily extracted. Note that the radiated field of the reflectarray is usually expressed in the reflectarray coordinate system where the z axis is along the normal direction of the reflectarray aperture. A coordinate transformation to the pattern coordinate system whose z_p axis is along the main beam direction is needed to plot its far-field radiation patterns in two principal planes.

Among all characteristics, the antenna gain is of special importance. As an aperture radiator, the gain of a reflectarray antenna is defined by the classical formula

$$G = \frac{4\pi A_a}{\lambda^2}\eta_a \tag{14}$$

where A_a is the aperture area and η_a is the aperture efficiency. The aperture efficiency is a key parameter that dictates the effectiveness of a reflectarray's capability of converting transmitted or received power. It is usually expressed by the product of several efficiency factors. Some of them share similar definitions as their counterparts in a conventional reflector antenna, while others are unique to reflectarrays. And their calculations may be simplified due to the planar array structure of a reflectarray. Good understanding of these efficiency factors can greatly facilitate the design and optimization of a reflectarray's radiation performance.

The taper efficiency (or illumination efficiency) is a measure of the uniformity of the amplitude distribution I on the reflectarray aperture. Its mathematical expression is given by (Yu et al. 2010b)

$$\eta_t = \frac{1}{A_a}\frac{\left|\int_A I dA\right|^2}{\int_A |I|^2 dA} \tag{15}$$

where A denotes the reflectarray aperture. Specifically, to simplify the calculations in reflectarrays, the integrals can be replaced by summations

$$\eta_t = \frac{1}{A_a}\frac{\left(\sum_{n=1}^{N}|I_n|\right)^2}{\sum_{n=1}^{N}|I_n|^2} \tag{16}$$

where I_n is the element excitation given in Eq. 6.

For a reflectarray illuminated by a simple feed, its amplitude distribution usually has a peak in the central area and monotonically decays towards the rim, if the element loss is ignored. The edge taper (ET) is then defined as the ratio of the element amplitude at the rim to the maximum value over the reflectarray aperture. It is well known that as the edge taper decreases, the taper efficiency decreases, and the radiation pattern tends to exhibit a slightly broadened main beam with lowered sidelobe levels. It is worthwhile to note that on the contrary to reflectors, where the illumination taper is solely determined by the feed pattern and spatial path lengths, the amplitude distribution of a reflectarray also depends on the

radiation pattern of the elements. As implied in Eq. 5, the intercepted power by the element decreases as the incident field from the feed shifts away from its broadside. Therefore, the edge taper is further attenuated by the element pattern, which may negatively affect the taper efficiency, especially for reflectarrays with F/D smaller than 0.5. The amplitude loss of the elements adds variation to the amplitude distribution as well, but it is usually much less significant compared to the feed illumination.

The spillover efficiency defines the percentage of the radiated power from the feed that is intercepted by the reflectarray aperture. Its classical formula is

$$\eta_s \frac{\int_\sigma \vec{P}(\vec{r}) \cdot d\vec{S}}{\int_\Sigma \vec{P}(\vec{r}) \cdot d\vec{S}} \tag{17}$$

where $\vec{P}(\vec{r})$ is the Poynting vector from the feed, Σ denotes the entire spherical surface centered at the feed, and σ denotes the solid angle subtended to the reflectarray aperture. The direct approach to calculate the spillover efficiency is to perform the integrals over the specified areas analytically or numerically. In practice, however, it is not straightforward to determine the solid angle for σ, especially for an offset reflectarray with a square aperture. Alternatively, the integral in the numerator is performed over the reflectarray aperture

$$\eta_s = \frac{\int_A \vec{P}(\vec{r}) \cdot d\vec{S}}{\int_\Sigma \vec{P}(\vec{r}) \cdot d\vec{S}} \tag{18}$$

which is calculated in a coordinate system that accommodates the shape of the reflectarray aperture and the position of the feed (Yu et al. 2010b). For a center-fed reflectarray with circular aperture illuminated by a simplified $\cos^q(\theta)$-type feed, the spillover efficiency in a closed form is given by (Huang and Encinar 2008)

$$\eta_s = 1 - \cos^{2q_f+1}(\psi_0/2) \tag{19}$$

where ψ_0 is the subtended angle from the feed to the reflectarray rim.

Similar to reflectors, the spillover power from the feed decreases and the spillover efficiency increases as the edge taper decreases. The product of the taper efficiency and the spillover efficiency is the main factor that dictates the aperture efficiency. The spillover efficiency, taper efficiency, and their product versus the q-factor of the feed for a representative reflectarray are plotted in Fig. 13. An optimal aperture efficiency can be achieved if the feed provides an edge taper around -10 dB.

The element loss efficiency accounts for the dielectric and conductor loss in reflectarray elements. In the case of reflectors, only curved metallic surfaces are used where conductor loss is negligible at microwave frequencies. In reflectarrays, however, the dielectric loss is inherent due to the existence of lossy dielectric substrate and the resonance nature of the element design, as shown in Fig. 6b. Moreover, the finite conductivity of nonideal metallic materials increases the conductor loss (An et al. 2014a), and it becomes more pronounced at terahertz (THz) and optical frequencies (Yang et al. 2012). Both the dielectric and conductor loss are included in the element amplitude response, and the element loss efficiency can be readily quantified as

DOI 10.1007/978-981-4560-75-7_45-1

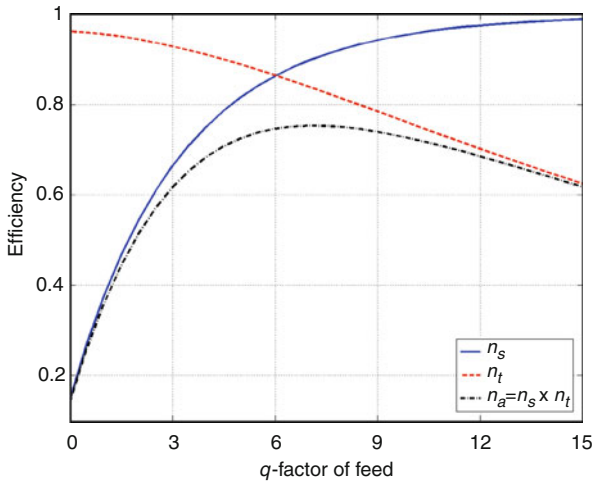


Fig. 13 Reflectarray efficiencies versus the q-factor of the feed

$$\eta_l = \frac{\sum_{n=1}^{N} |I_n|^2}{\sum_{n=1}^{N} |I_n/\Gamma_n|^2} \tag{20}$$

where Γ_n is the amplitude response of the nth element. Although Γ_n is present in the element excitation given in Eq. 6 and included in the calculation of taper efficiency in Eq. 16, it is canceled out in the numerator and denominator, especially when Γ_n is the same for all the elements. Hence, it is necessary to incorporate Eq. 20 in the product of efficiency factors and take into account the dielectric and conductor loss of elements.

The phase efficiency is a measure of the deviation from the uniform phase distribution (or the specifically shaped phase front) on the reflectarray aperture. The phase errors cause additional gain loss and degraded radiation pattern, which can be attributed to several factors. First and most importantly, it stems from the design and analysis of reflectarray elements. An infinite-array structure is usually assumed in the element analysis. The actual phase realized by each element in the actual reflectarray configuration may be different, especially where an abrupt phase variation occurs and adjacent elements have very different configurations. The simple element design process only based on the phase curve for normal incidence also introduces phase errors if the angle of incidence on the elements varies considerably. And for some element designs, the achievable phase shift swing is less than 360°; an approximate phase shift closest to its desired value is selected, leading to phase errors. Although it may be somehow mitigated by adopting more sophisticated element design and analysis technique, the phase error in the element phase response is usually inevitable in a practical reflectarray design. Secondly, the actual radiated field from the feed may deviate from the spherical wavefront, which is an important assumption commonly used in reflectarray element analysis and phase shift design. It can be circumvented by incorporating the actual incident field from the simulation or measurement data of the feed. Thirdly, the phase center of the feed wanders with changes in frequency. An axial defocusing effect is expected especially in broadband and multiband designs. This issue can be addressed by carefully choosing the feed or tuning the element phase responses at different frequencies to compensate for the movement of the phase center of the feed. Lastly, the fabrication error leads to deviation from desired phases. It is similar to the phase errors caused by the surface roughness in reflectors. However, the lithographic etching technique in reflectarray fabrication usually yields much better accuracy compared to those for reflectors. The phase deviation due to

DOI 10.1007/978-981-4560-75-7_45-1

fabrication error at Ka-band or lower frequencies is almost negligible and may be further alleviated if the element phase curve is more gradual.

The polarization efficiency represents the polarization purity of the reflectarray, and for a x-polarized antenna, it can be calculated as

$$\eta_x = \frac{1}{A_a} \frac{\left(\sum_{n=1}^{N} \left|E_{x,n}^{ref}\right|\right)^2}{\sum_{n=1}^{N} \left(\left|E_{x,n}^{ref}\right|^2 + \left|E_{y,n}^{ref}\right|^2\right)} \tag{21}$$

It measures the percentage of the power that is coupled into the co-polar component of the reflectarray radiation and is mostly affected by the polarization performances of both the feed and the element performance. A low cross-polarization performance can be achieved by adopting an appropriate element design such as dipoles with variable lengths to suppress the cross terms in Eq. 13. Therefore, an offset reflectarray is able to produce superior cross-polarization performance than its reflector counterpart as a result of the polarization selectivity of its elements.

The blockage efficiency accounts for the blockage effects of the feed assembly, subreflector, and other structural components on the aperture radiation. A simple estimation is to compute the percentage of unblocked aperture area projected in the main beam direction. Just like reflectors, an offset configuration is able to eliminate the feed blockage. Moreover, with the agile phase adjustment of reflectarray elements, the main beam can be properly set in an off-broadside direction to avoid the feed assembly, even for a center-fed system.

The return loss efficiency of the feed is mostly neglected in the computation of aperture efficiency and gain of a reflectarray. It is acceptable since the horn antenna is widely used as the feed and its reflection coefficient is usually below −20 dB. For reflectarrays with no feed blockage, the interference of the reflected wave from its aperture is also negligible.

In summary, various efficiency factors must be carefully evaluated. The aperture efficiency of a reflectarray can thus be computed by multiplying them altogether and its gain can be obtained using Eq. 14.

With the gain performance acquired at the frequency band of interest, the gain bandwidth of the reflectarray is simply defined as the frequency range of allowed gain drop from its designed value at the operating frequency. In practice, a gain drop of 1 dB or 3 dB is commonly used. The 1-dB gain bandwidth usually ranges from 3 % to 8 %, depending on the element design and the electrical size of the reflectarray aperture, and 10 % bandwidth, up to over 20 %, is possible in some broadband designs. The port impedance matching is usually very good if a horn feed is employed. Hence, the impedance bandwidth is not considered in reflectarrays.

An Example of Reflectarray Design

As an example, an offset single-beam reflectarray antenna operating at 32.5 GHz is designed, as shown in Fig. 14.

The reflectarray aperture is circular with a diameter of 516 mm. A moderate F/D value is selected to balance the feed illumination and the overall volume of the antenna. An offset configuration is utilized to mitigate the feed blockage effect. The x-polarized feed is located at 400 mm above the aperture, tilted toward the center of the aperture with a feed offset angle of 15°. By matching the $\cos^q(\theta)$-type function to the −10 dB beamwidth of the feed pattern, q_f is determined to be 5.5. In order to further reduce the blocked area, the main beam is tilted 15° as well, opposite to the feed offset direction.

DOI 10.1007/978-981-4560-75-7_45-1

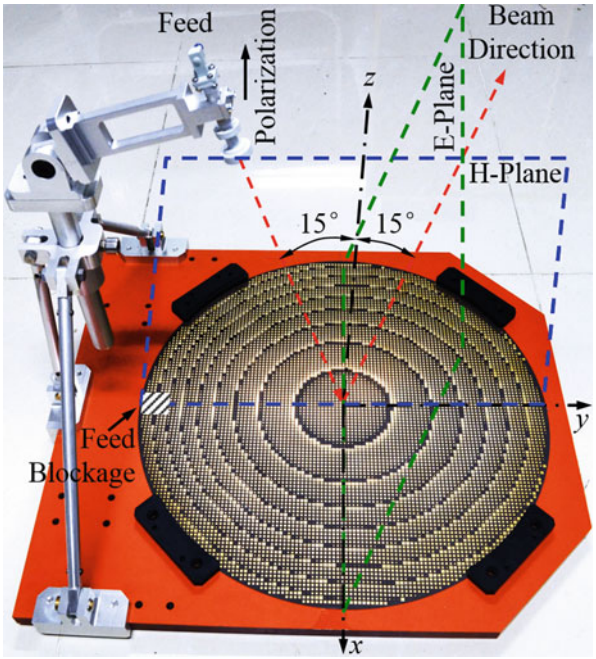


Fig. 14 Prototype of an offset reflectarray antenna operating at 32.5 GHz

Square microstrip patches with variable sizes are employed and arranged in a square lattice with element spacing of 4.6 mm. The thickness of the grounded substrate is 0.51 mm, with the dielectric constant ε_r of 2.55 and the loss tangent tanδ of 0.0019. The element is analyzed using the infinite-array approach. Its phase shift and amplitude response for the normal incidence are shown in Fig. 6. Based on the ray tracing method, the desired phase shift of each element φ_n can be obtained using Eq. 3 and plotted in Fig. 15a, where the phase shift is limited to one 360° cycle and the constant reference phase $\Delta\varphi_0$ is set to be zero. Because the achievable phase shift swing is less than 360°, the desired phase shift beyond this range is approximated by its closest value. The maximum angle of incidence on the far-end element is 42.4°; therefore, several phase curves for different incident angles are obtained and the element sizes are determined accordingly. The amplitude of the element excitation given in Eq. 6 is plotted in Fig. 15b. The edge taper is about −10.5 dB, well within the realm for optimized aperture efficiency. Note that a small area blocked by the feed is still present, although both the feed and main beam are tilted. However, its impact on the gain and pattern of the reflectarray is expected to be negligible because the blocked area is relatively small and the elements within the shadowed area are well under-illuminated.

The radiated field of the reflectarray is then computed using Eq. 8, and its calculated far-field radiation patterns in the pattern coordinate system are plotted in Fig. 16. The feed blockage effect is approximated by excluding those elements located within the shadowed area from the summation in Eq. 8. The reflectarray is also fabricated and measured in a compact-range anechoic chamber. The measured patterns are plotted and presented in Fig. 16 as well for comparison. It is observed that the main beam direction is well aligned in the designed direction. The main beam and first few sidelobes agree well, except the measured sidelobe levels are higher due to phase errors in element designs and system assembly. The array theory method predicts an overall pattern shape with lower sidelobe levels. The measured cross-polarization level is comparable to that of its reflector counterpart because the chosen elements are insensitive to polarizations.

The calculated and measured gains are 43.4 dBi and 43.0 dBi, respectively, which correspond to aperture efficiency of 70.9 % and 64.7 % using Eq. 14. It is worthwhile to point out that due to the tilted main beam, its effective aperture area is reduced by the amount of cos(θ_B). Taking into account this

DOI 10.1007/978-981-4560-75-7_45-1

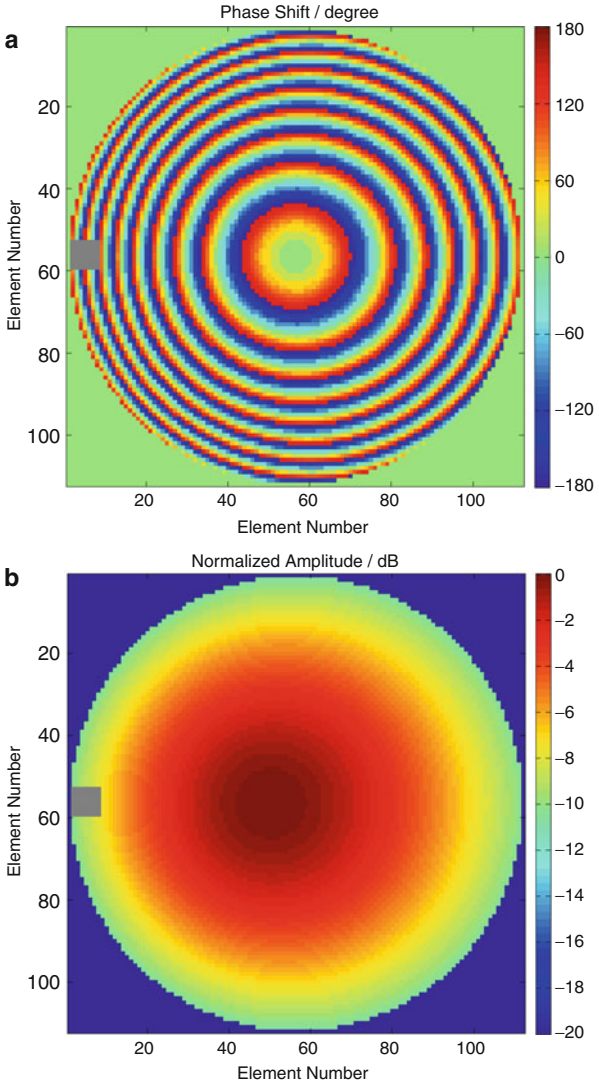


Fig. 15 (**a**) Desired phase shift distribution and (**b**) amplitude distribution of the reflectarray. The shadowed area represents the feed blockage

reduction, the calculated and measure aperture efficiency are modified to be 73.4 % and 67.0 %, respectively. This revision agrees with the convention in offset reflectors where the area of the projected aperture, instead of its physical aperture, is commonly used as the basis for aperture efficiency calculation. The simulation and measurement results are in good agreement. Although the achieved aperture efficiency is slightly suboptimal compared to that of reflectors, this example successfully demonstrates the superior radiation performance of a reflectarray as an efficient high-gain antenna.

The measured gain performance in the frequency band of interest is shown in Fig. 17. The ripple in the gain curve is mainly due to the imperfect frequency response of the anechoic chamber. The range of gain over 42 dBi is from 31.9 to 33.6 GHz, corresponding to a 1-dB gain bandwidth of 5.2 %. The moderate bandwidth performance is within expectation, given the simple element design and the large electrical size of the reflectarray. Improved bandwidth can be achieved if broadband techniques are adopted in the reflectarray design.

DOI 10.1007/978-981-4560-75-7_45-1

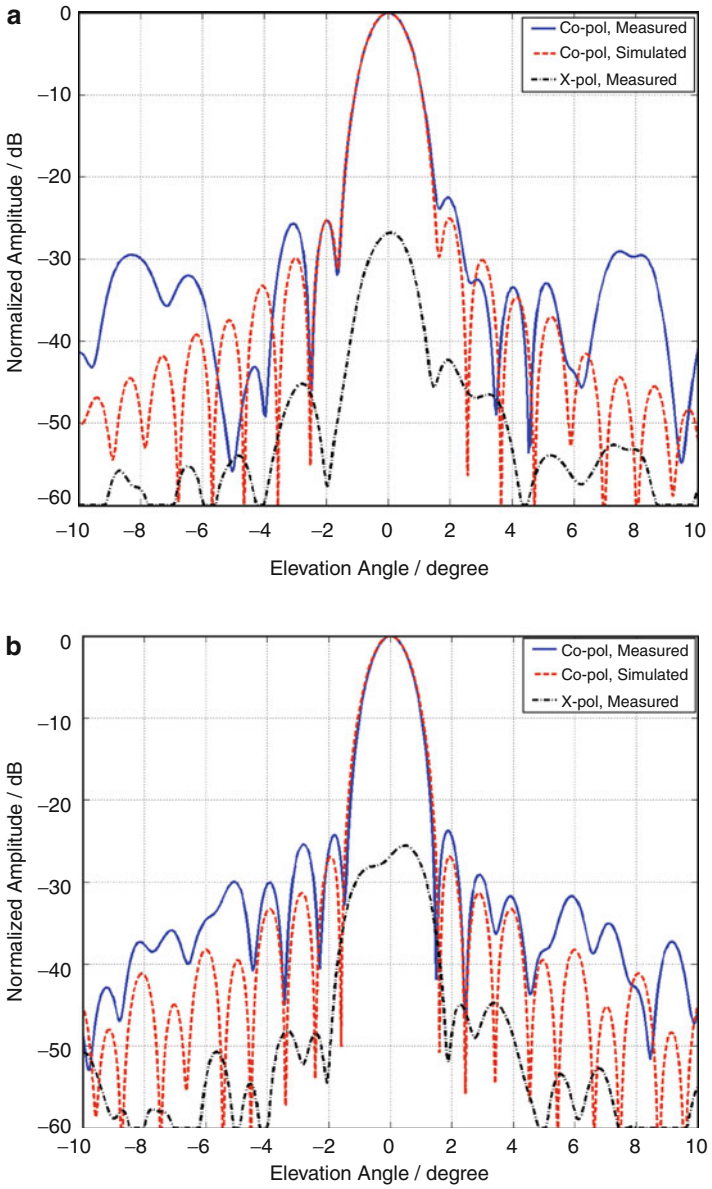


Fig. 16 Simulated and measured radiation patterns: (**a**) H-plane (offset plane), (**b**) E-plane

Latest Developments and Applications

With the development of numerical analysis tools, the integrated circuits and solid-state tuning devices, new materials, and fabrication technologies, a variety of novel reflectarrays have been investigated to satisfy the ever-increasing demands in satellite communications, spaceborne applications, defense and military areas, etc. More sophisticated elements and systems have been designed to fully exploit the feasibility and versatility provided by reflectarrays. Some latest developments and applications are briefly introduced in this section.

Broadband and Multiband Techniques

The most critical limitation in reflectarray technology is its limited bandwidth. It is attributed to two main factors. The first is the element bandwidth due to the intrinsic narrowband nature of reflectarray elements as previously discussed. It plays a more crucial role in the bandwidth performance of small- and medium-

DOI 10.1007/978-981-4560-75-7_45-1

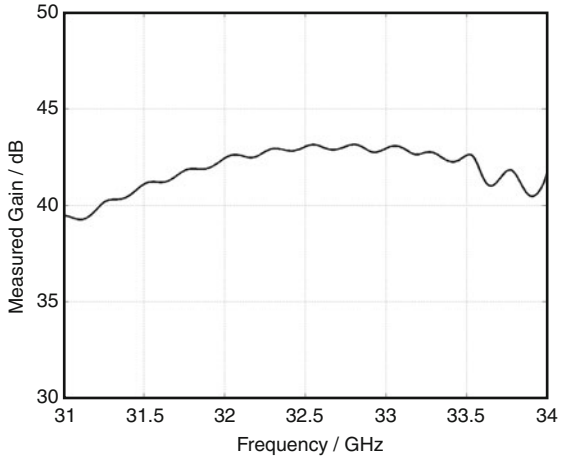


Fig. 17 Measured gain performance in the frequency band of interest

size reflectarrays with moderate or large F/D. The second is the differential spatial phase delay due to the different path lengths from the feed to the reflectarray plane where the desired phase front is located. It is obviously seen in Eq. 3 that the desired phase shift to collimate a beam is proportional to frequency. The reflectarray elements can no longer effectively compensate for the spatial phase delays if the realized element phase shift does not adapt to the change in frequency, leading to considerable phase errors. Its adverse impact on the bandwidth is more pronounced in large reflectarrays as well as those with small F/D where the variation in path lengths is large.

Multi-resonant structures are widely used to improve the bandwidth performance of reflectarray elements. One approach is to stack two or more layers of microstrip patches with variable sizes (Encinar 2001; Encinar and Zornoza 2003). Compared to the conventional single-layer patch elements with variable sizes, the additional layer introduces a second resonance to the structure, resulting in a smoother phase curve with more gradual slope. The achievable phase shift swing is extended beyond a full 360° range, and the element phase responses at different frequencies are considerably improved. A significant drawback of this design is its increased complexity in element structure, leading to a more complex and expensive fabrication process. The sizes of the elements on both layers change simultaneously, which also limits the flexibility in reflectarray designs.

Recently, more efforts have been devoted to develop single-layer multi-resonant elements. It is usually obtained by combining two or more simple structures with appropriate shapes, patches, rings, and crossed dipoles, for instance, in one element design. Various elements, such as double square rings (Chaharmir et al. 2009), double cross loops (Chaharmir et al. 2006), triple rectangular loops (Mohammadirad et al. 2012), the phoenix cell (Moustafa et al. 2011), etc., have been introduced in the literature. Alternatively, a secondary resonant structure such as a slot can be embedded in a simple patch element (Cadoret et al. 2005; De Vita et al. 2007) to obtain extra freedom in phase controlling.

Generally speaking, the multi-resonant elements can significantly broaden the element bandwidth, and a 1-dB gain bandwidth of more than 10 %, up to over 20 % (Chaharmir et al. 2009; Mohammadirad et al. 2012), may be achieved in a reflectarray design. The complex structures and controlling geometrical parameters in these elements considerably increase the complexity in element design and analysis as well as reflectarray fabrication. It is worthwhile to note that other design techniques such as a thick substrate with low dielectric constant, subwavelength element spacing (Pozar 2007; Nayeri et al. 2010, 2011), and optimization of feed position (Mohammadirad et al. 2012) are often adopted at the same time to further improve the element bandwidth.

DOI 10.1007/978-981-4560-75-7_45-1

To overcome the second issue in bandwidth limitation, the differential spatial phase delay, the desired phase shift, and the element phase response within the entire frequency band, rather than at a single operating frequency, need to be properly designed and analyzed. As observed in Eq. 3, if a frequency-dependent phase shift of the element described by

$$\frac{\varphi_n(f)}{f} = \frac{\varphi_n(f_0)}{f_0} \tag{22}$$

can be realized, where f_0 and f refer to the center and off-center frequencies, respectively, the same phase transformation can be achieved by the reflectarray within the entire frequency band, significantly improving its bandwidth performance. An ideal solution to compensate for frequency dispersion is to use the true-time delay element (Carassco et al. 2008). An aperture-coupled patch element is utilized and the length L of a U-shaped true-time delay line is adjusted to produce a phase shift of approximately $-2\beta L$, where β is the propagation constant at frequency f. A 1.5-dB gain bandwidth of more than 26 % can be achieved even for large reflectarrays. The only drawbacks are higher ohmic loss due to the long line covering $3 \times 360^\circ$ phase range and some distortion of the element phase response at extreme frequencies.

By using an appropriate phase optimization method, multi-resonant elements can be properly designed to alleviate the adverse impact of frequency dispersion as well (Encinar and Zornoza 2003; Chaharmir et al. 2010). It is achieved by taking into account the overall phase error between the desired and realized phase shifts of the element at the center frequency and two extreme frequencies at the same time. The flexibility in multi-resonant element configurations allows a number of elements to provide the same phase shift at the center frequency but different frequency dispersion characteristics. An optimization routine is then employed to minimize it by selecting the most suitable one from a pool of elements. A measured 1-dB gain bandwidth of 12 % is reported for a 1.2-m Ku-band reflectarray (Chaharmir et al. 2010), which successfully validates the effectiveness of the optimization method.

More recently, a novel phase synthesis approach for broadband reflectarray designs is proposed (Mao et al. 2012). The impact of selecting the constant reference phase $\Delta\varphi_0$ in Eq. 3, which has been overlooked since the inception of reflectarrays, is taken into account, and the sum of the phase errors of all the elements at two frequencies is minimized by an optimization algorithm. This phase synthesis process can be graphically described in the phase plane as in Fig. 18. By adjusting the constant reference phases at two frequencies independently, the straight line that represents the desired phase shifts at those frequencies is shifted in the phase plane to have a better matching with the achievable phase shifts provided by the elements, effectively reducing the overall phase errors of the reflectarray. A 1.5-dB gain bandwidth of 16.7 % is achieved by a reflectarray with simple square loop elements. Note that this approach can also be applied to reflectarrays with multi-resonant elements, whose achievable phase shifts can be optimized simultaneously to further improve the bandwidth performance.

An alternative approach to reduce the variation of path lengths in reflectarrays is the multifacet design. The reflectarray configuration is made of several small planar reflectarray panels arranged to approximate a parabolic surface (Huang and Encinar 2008). Several designs from simple one-dimensional arrangements to more complex circumferential structures have been reported. The complexity in fabrication and assembly as well as the deployment structure, however, is significantly increased. Hence, this technique is usually limited to some specific applications.

Several techniques in broadband reflectarray designs can be utilized to obtain multiband performances as well. Two sets of elements can be arranged on two stacked layers, and each one of them corresponds to one operating band (Encinar 1996; Han et al. 2004, Huang and Encinar 2008). The element configurations must be properly designed and arranged in order to minimize the mutual interference and blockage between the two layers of elements. Single-layer multiband designs have also been intensively

DOI 10.1007/978-981-4560-75-7_45-1

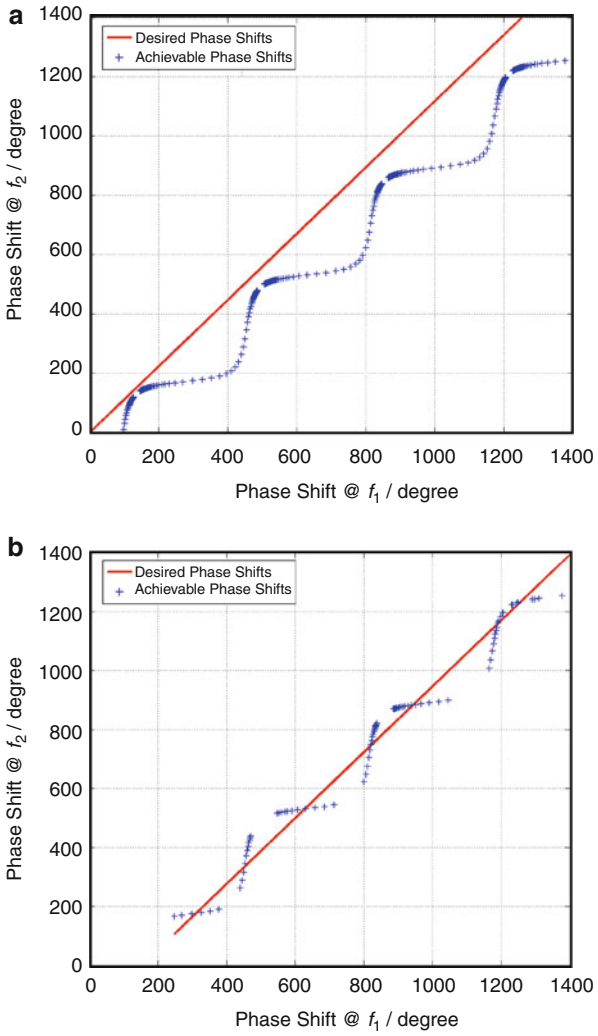


Fig. 18 Graphical illustration of the phase synthesis approach: (**a**) before and (**b**) after the optimization of the reference phases

investigated, which successfully circumvent the complex structure and blockage effects in multi-layer designs. For two widely separated frequencies, two sets of simple elements of appropriate forms are usually alternatively spaced without physically interfering with each other (Huang and Encinar 2008). Each set can be independently designed at one frequency with the presence of the other set of elements in the element analysis. Triple-band designs are also feasible (Yang et al. 2008; Yu et al. 2010a), but it becomes more challenging to choose the elements that are able to fit in the limited space. For two closely separated frequencies, multi-resonant elements such as a concentric dual split-loop (Smith et al. 2013) and a patch with embedded slots (Malfajani and Atlasbaf 2014) can be employed, with each of its substructure corresponding to one resonant frequency. Because the phase shifts at both frequencies are realized by the same element, the realized phase shift and amplitude loss of the element at one frequency may be interfered by the change in phase controlling parameters at the other frequency. A thorough element analysis is thus required to minimize the phase errors caused by this interference (Smith et al. 2013). The main drawback of single-layer multiband designs is the lack of flexibility in selecting element configuration, element spacing, and the thickness of substrate, which thus may not be optimized for all the elements.

DOI 10.1007/978-981-4560-75-7_45-1

Especially, the phase synthesis approach (Mao et al. 2012) is well suitable for dual-band designs. It inherently provides a dual-frequency element response. Moreover, in a single-layer multiband reflectarray, simple elements such as crossed dipoles and single loops are usually chosen in order to avoid physical interference between elements. Each set of elements typically exhibits narrow bandwidth due to its limited element performance. With no limitation on the selection of elements, this approach can be applied separately to optimize the bandwidth performance of each set of simple elements so that a broadband multiband reflectarray design becomes feasible.

Reconfigurable Reflectarrays

Beam scanning is a desired feature in many applications such as satellite communications, radar, remote sensing, etc. The most straightforward way is to mechanically rotate a fixed beam reflectarray as in reflectors. Although being low cost and with no loss in radiation performance, it suffers from the low beam scanning speed, which potentially limits its applicability in advanced applications. The vibration due to its mechanical rotation and the maintenance requirement may also adversely affect its reliability.

Owing to the independent phase adjustability of reflectarray elements as in phased arrays, electronic beam scanning becomes feasible (Shaker et al. 2014; Hum and Perruisseau-Carrier 2014). Especially, in the past 10 years or so, a lot of efforts have been devoted to the developments of a variety of reconfigurable reflectarrays. These designs can potentially provide exceptional radiation performance of high efficiency and fast beam scanning with low loss, low power consumption, and low cost. In addition, agile radiation characteristics such as shaped beam for coverage modification and anti-interference, multi-beam for multiple target detection, and dynamic beamwidth control can be achieved.

According to the adopted phase adjustment mechanism, reconfigurable reflectarrays can be categorized into three types. The most intensively investigated one is the electronically reconfigurable reflectarrays. It is typically accomplished by integrating solid-state tuning devices such as PIN diode, varactor diode, and RF MEMS switch into the element design (Hum et al. 2007; Perruisseau-Carrier and Skrivervik 2008; Rajagopalan et al. 2008; Sorrentino et al. 2009; Kamoda et al. 2011; Carassco et al. 2012). Various element configurations such as patches, dipoles, loops, and aperture-coupled elements have been studied, where the tuning devices can be integrated in the microstrip radiators, the embedded coupling slots, or the phase lines. The tuning devices vary in terms of loss, cost, speed, power consumption, power handling, mounting and biasing complexity, and technology readiness (Sorrentino et al. 2009). And the element performance also varies in terms of the achievable phase range, the number of phase states (or the number of bits), loss, polarization, etc. Generally speaking, these designs can achieve very fast beam scanning capability and reliable performance. Monolithic fabrication of the whole reflectarray along with the tuning devices and biasing network is also feasible (Perruisseau-Carrier and Skrivervik 2008). It is worthwhile to mention that a crucial issue for the element with PIN diodes or RF MEMS switches, which are controlled digitally, is the achievable number of phase states. It is shown that a simple 1-bit design typically causes over 3 dB gain loss due to quantization errors, while an average gain reduction of 0.2 dB with better beam pointing accuracy can be obtained for a 3-bit design (Wu et al. 2008). Particular attention is also required to properly design the biasing network, which may become overwhelming for an element design with a large number of tuning devices. A trade-off between the reflectarray performance and the complexity and cost of the system must be made according to the design requirements.

An alternative approach to design electronically reconfigurable reflectarrays is to use tunable materials such as liquid crystals (Hu et al. 2008; Gaebler et al. 2009), photo-induced plasma (Chaharmir et al. 2006), and ferroelectric materials (Romanofsky 2007) as the substrate or radiating elements. Under different biasing voltage or illumination intensity, the material properties such as permittivity, permeability, and conductivity change, resulting in a corresponding change of element phase. These designs are typically able to provide analog phase variation. However, it usually takes some time for the material to change its

DOI 10.1007/978-981-4560-75-7_45-1

properties so the respond speed is also relatively slow. More critically, the considerable loss in these materials may limit its wide application.

The phase reconfigurability may be obtained by mechanical methods as well. Different than the conventional mechanical rotation in reflectors, miniaturized actuators or motors are integrated into each element, so that the element phase is individually adjusted to produce the desired phase front (Gianvittorio and Rahmat-Samii 2006; Fusco 2005; Phillion and Okoniewski 2008; Cabria et al. 2009). Other methods such as a fluidic loading mechanism (Long and Huff 2011) and a three-state beam switching design (Chaharmir et al. 2003b) have also been developed. The typical loss and cost in tuning devices and tunable materials can be avoided, but the respond speed is the slowest among the three types of reconfigurable reflectarrays.

Some novel reconfigurable reflectarray designs with more flexibility have also been proposed in literature. Active amplifiers are incorporated in reflectarray elements, so the incident signal on each element can be amplified before being retransmitted (Bialkowski et al. 2002; Kishor and Hum 2012). Reconfigurable elements with capabilities of operating at variable frequencies (Rodrigo et al. 2013) and variable polarizations (Perruisseau-Carrier 2010) are also investigated. These designs further demonstrate the powerful flexibility that can be accomplished by the reflectarray technology.

Novel Applications of Reflectarrays

As a hybrid of reflector and array antennas, the reflectarray exhibits numerous favorable features and exceptional flexibility and versatility, making it suitable for a variety of advanced applications. With the latest developments in reflectarray technology, better designs in terms of radiation performance or fabrication complexity and cost may be achieved for some existing applications, while some previously unimagined scenarios now become feasible.

A large reflectarray aperture can be divided into small segments and assembled on site later, making fabrication, packaging, and shipment more convenient and keeping the cost low. In addition, its planar, low-profile, and light-weighted structure makes it more suitable for spaceborne applications. Compared to the conventional doubly curved reflectors, the planar structure of reflectarrays requires a much simpler and reliable folding mechanism to achieve a large-aperture deployable antenna with satisfactory surface accuracy. An inflatable reflectarray with elements printed on a large, flat, thin membrane can even further reduce its mass and stowing volume (Huang and Encinar 2008).

Specifically, reflectarrays with shaped beams have great potentials in direct broadcast satellite (DBS) applications (Encinar et al. 2006, 2011; Tienda et al. 2013). With various phase synthesis or optimization methods, complex phase shift distributions can be designed and precisely achieved, thus producing very accurate shaped patterns. Dual-linearly polarized elements can provide independent phase shifts for each polarization to obtain dual coverage with a single reflectarray. On the contrary to the difficulties in fabricating shaped reflectors with stringent accuracy, the fabrication process of shaped-beam reflectarrays is the same as simple single-beam reflectarrays without increasing the complexity or cost in fabrication.

The solar panels and high-gain antennas are usually the major components on a satellite platform with largest apertures. In order to reduce the volume, mass, cost, and complexity of a satellite, it is desirable to integrate them into one. The advent of planar reflectarray technology makes this integration feasible. By directly printing reflectarray elements with small footprint on solar cells (Huang and Zawadzki 2000; Dreyer et al. 2014; An et al. 2014b), the solar panels provide the required large aperture for high-gain antennas as well as solar energy harvesting for the satellite. It even becomes possible to install a high-gain antenna on a small satellite, which is previously prohibited due to the limited payload space and mass on this type of satellite.

The flexibility in element designs allows a reflectarray to provide independent phase shift distributions at different frequencies for different polarizations using the same aperture. It can effectively separate the

DOI 10.1007/978-981-4560-75-7_45-1

main beams and feeds, making the arrangement of a reflectarray system more convenient. The element phases can be optimized to produce multiple independent high-gain beams at the same time when being illuminated by a single feed (Nayeri et al. 2012). This capability makes it suitable for multiple target detection. More interestingly, this technique shows exceptional prospect in spatial power distributing/combining (Bialkowski and Song 2002), especially at high frequencies where the loss in traditional devices becomes prohibitively large.

Reconfigurable reflectarrays are able to provide similar functionalities as in phased arrays, for instance, electronic beam scanning, agile radiation pattern, and multiple target detection. Moreover, when monolithic fabrication of reconfigurable reflectarrays becomes available, better performance can be achieved with the reflectarray technology in terms of radiation efficiency, system complexity, fabrication cost, and power consumption. Wide applications of reconfigurable reflectarrays in satellite communications, radar, military and defense areas, etc., are hence expected.

Lastly, the planar quasiperiodic structure and the accuracy in fabrication make reflectarrays more suitable at high frequencies. The research has been extended to terahertz (THz) and optical frequency range (Ginn et al. 2007; Ahmadi et al. 2010; Yang et al. 2012; Niu et al. 2012; Carassco et al. 2013; Zainud-Deen et al. 2013; Nayeri et al. 2014; Deng et al. 2014). Potential applications in THz imaging, controlled beam forming, and power distributing/combining are some of the most cutting-edge research frontiers of the reflectarray technology.

Conclusion

The reflectarray antennas exhibit exceptional performance in both radiation characteristics and mechanical structures. As a hybrid of reflector and array antennas, it performs a phase transformation to produce a desired radiation pattern by properly designing the phase shifts of an array of quasiperiodic elements. Many important aspects of the reflectarray technology have been explored and great potentials in advanced applications have been well demonstrated, thanks to the flexibility and versatility arising from independent phase adjustability of reflectarray elements. Equipped with the fundamentals of reflectarrays and instructive design and analysis procedures presented in this chapter, the readers are encouraged to explore deeper into this novel and promising research frontier and to apply the reflectarray technology in practical engineering applications.

Cross-References

- ▶ Antennas in Radio Telescope Systems
- ▶ Antennas in Radiometer Systems including Passive Imaging Systems
- ▶ Applications of Phased Array Feeders in Reflector Antennas
- ▶ Broadband and Multiband Planar Antennas
- ▶ Conformal Array Antennas
- ▶ Frequency Selective Surfaces
- ▶ Fresnel Zone Plate Lens Antenna
- ▶ Lens Antennas
- ▶ Low-Profile Antennas
- ▶ Metamaterials and Antennas
- ▶ Millimeter-Wave Antennas and Arrays
- ▶ Multibeam Antenna Arrays

▶ Optical Nanoantennas
▶ Phased Arrays
▶ Photomixer Antennas at Terahertz
▶ Radio Frequency Beamforming for Scanned and Multibeam Antenna Systems
▶ Reconfigurable Antenna Arrays for Wireless Communications
▶ Reconfigurable Passive and Active Antennas/Tunable Antennas
▶ Reflector Antennas
▶ Satellite Antennas on Vehicles
▶ Space Antennas including Terahertz Antennas
▶ Terahertz Antennas and Measurement

References

Ahmadi A, Ghadarghadr S, Mosallaei H (2010) An optical reflectarray nanoantenna: the concept and design. Opt Express 18:123–133

Almajali E, MaNamara D, Shaker J, Chaharmir MR (2012) Derivation and validation of the basic design equations for symmetric sub-reflectarrays. IEEE Trans Antennas Propag 60:2336–2346

An W, Xu S, Yang F (2014a) A metal-only reflectarray antenna using slot-type elements. IEEE Antennas Wirel Propag Lett 13:1553–1556

An W, Xu S, Yang F (2014b) A Ka-band reflectarray antenna integrated with solar cells. IEEE Trans Antennas Propag 62:5539–5546

Arrebola M, De Haro L, Encinar JA (2008) Analysis of dual-reflector antennas with a reflectarray as subreflector. IEEE Antennas Propag Mag 50:39–51

Balanis CA (2005) Antenna theory: analysis and design, 3rd edn. Wiley, Hoboken

Berry DG, Malech RG, Kennedy WA (1963) The reflectarray antenna. IEEE Trans Antennas Propag 11:645–651

Bialkowski ME, Sayidmarie KH (2008) Investigations into phase characteristics of a single-layer reflectarray employing patch or ring elements of variable size. IEEE Trans Antennas Propag 56:3366–3372

Bialkowski ME, Song JH (2001) Dual linearly polarized reflectarray using aperture coupled microstrip patches. Proc IEEE Antennas Propag Symp 3:486–489

Bialkowski ME, Song HJ (2002) Investigations into a power-combining structure using a reflect array of dual-feed aperture-coupled microstrip patch antennas. IEEE Trans Antennas Propag 50:841–849

Bialkowski ME, Robinson AW, Song HJ (2002) Design, development, and testing of X-band amplifying reflectarrays. IEEE Trans Antennas Propag 50:1065–1076

Bozzi M, Germani S, Perregrini L (2004) A figure of merit for losses in printed reflectarray elements. IEEE Antennas Wirel Propag Lett 3:257–260

Cabria L, Garcia JA, Gutierrez-Rios J, Tazon A, Vassallo J (2009) Active reflectors: possible solutions based on reflectarrays and Fresnel reflectors. Intl J Antennas Propag 2009:653952

Cadoret D, Laisne A, Gillard R, Legay H (2005) A new reflectarray cell using microstrip patches loaded with slots. Microw Opt Tech Lett 44:270–272

Carassco E, Encinar JA, Barba M (2008) Bandwidth improvement in large reflectarrays by using true-time delay. IEEE Trans Antennas Propag 56:2496–2503

Carassco E, Encinar JA, Barba M (2012) X-band reflectarray antenna with switching-beam using PIN diodes and gathered elements. IEEE Trans Antennas Propag 60:5700–5708

Carassco E, Tamagnone M, Perruisseau-Carrier J (2013) Tunable graphene reflective cells for THz reflectarrays and generalized law of reflection. Appl Phys Lett 102:104103
Chaharmir MR, Shaker J, Cuhaci M, Sebak A (2003a) Reflectarray with variable slots on ground plane. IEE Proc Microw Antennas Propag 150:436–439
Chaharmir MR, Shaker J, Cuhaci M, Sebak A (2003b) Mechanically controlled reflectarray antenna for beam switching and beam shaping in millimetre-wave applications. Electron Lett 39:591–592
Chaharmir MR, Shaker J, Cuhaci M, Ittipiboon A (2006a) Broadband reflectarray antenna with double cross loops. Electron Lett 42:65–66
Chaharmir MR, Shaker J, Cuhaci M, Sebak A (2006b) Novel photonically-controlled reflectarray antenna. IEEE Trans Antennas Propag 54:1134–1141
Chaharmir MR, Shaker J, Cuhaci M, Ittipiboon A (2009) Wideband reflectarray research at the communications research centre Canada. Proc Intl Symp Antenna Tech Appl Electromagn 1−4
Chaharmir MR, Shaker J, Gagnon N, Lee D (2010) Design of broadband, single layer dual-band large reflectarray using multi open loop elements. IEEE Trans Antennas Propag 58:2875–2883
Chang DC, Huang MC (1995) Multiple-polarization microstrip reflectarray antenna with high efficiency and low cross-polarization. IEEE Trans Antennas Propag 43:829–834
CST (2014) www.cst.com, Accessed on 16 March 2015
De Vita P, Freni A, Dassano GL, Pirinoli P, Zich RE (2007) Broadband element for high-gain single-layer printed reflectarray antenna. Electron Lett 43:1247–1248
Deng R, Yang F, Xu S (2014) Design and comparison of four different reflectarray antennas towards THz applications. Proc IEEE Antennas Propag Symp 1019−1020
Dreyer P, Morales-Masis M, Nicolay S, Ballif C, Perruisseau-Carrier J (2014) Copper and transparent-conductor reflectarray elements on thin-film solar cell panels. IEEE Trans Antennas Propag 62:3813–3818
Encinar JA (1996) Design of a dual frequency reflectarray using microstrip stacked patches of variable size. Electron Lett 32:1049–1050
Encinar JA (2001) Design of two-layer printed reflectarray using patches of variable size. IEEE Trans Antennas Propag 49:1403–1410
Encinar JA, Zornoza JA (2003) Broadband design of three-layer printed reflectarrays. IEEE Trans Antennas Propag 51:1662–1664
Encinar JA, Datashvili LS, Zornoza JA, Arrebola M, Sierra-Castaner M, Besada-Sanmartin JL, Baier H, Legay H (2006) Dual-polarization dual-coverage reflectarray for space applications. IEEE Trans Antennas Propag 54:2827–2837
Encinar JA, Arrebola M, De La Fuente LF, Toso G (2011) A transmit-receive reflectarray antenna for direct broadcast satellite applications. IEEE Trans Antennas Propag 59:3255–3264
FEKO (2014) www.feko.info, Accessed on 16 March 2015
Fusco VF (2005) Mechanical beam scanning reflectarray. IEEE Trans Antennas Propag 53:3482–3484
Gaebler A, Moessinger A, Goelden F, Manabe A, Goebel M, Follmann R, Koether D, Modes C, Kipka A, Deckelmann M, Rabe T, Schulz B, Kuchenbecker P, Lapanik A, Mueller S, Haase W, Jakoby R (2009) Liquid crystal-reconfigurable antenna concepts for space applications at microwave and millimeter waves. Intl J Antennas Propag 2009:876989
Gianvittorio JP, Rahmat-Samii Y (2006) Reconfigurable patch antennas for steerable reflectarray applications. IEEE Trans Antennas Propag 54:1388–1392
Ginn JC, Lail BA, Boreman GD (2007) Phase characterization of reflectarray elements at infrared. IEEE Trans Antennas Propag 55:2989–2993
Han C, Chang K (2003) Ka-band reflectarray using ring elements. Electron Lett 39:491–493

Han C, Rodenbeck C, Huang J, Chang K (2004) A C/Ka dual-frequency dual-layer circularly polarized reflectarray antenna with microstrip ring elements. IEEE Trans Antennas Propag 52:2871–2876
Hannan PW, Balfour MA (1965) Simulation of phase array antennas in waveguides. IEEE Trans Antennas Propag 13:342–353
HFSS (2014) www.ansys.com, Accessed on 16 March 2015
Hu W, Cahill R, Encinar JA, Dickie R, Gamble H, Fusco V, Grant N (2008) Design and measurement of reconfigurable millimeter wave reflectarray cells with nematic liquid crystal. IEEE Trans Antennas Propag 56:3112–3117
Hu W, Arrebola M, Cahill R, Encinar JA, Fusco V, Gamble HS, Alvarez Y, Las-Heras F (2009) 94 GHz dual-reflector antenna with reflectarray subreflector. IEEE Trans Antennas Propag 57:3043–3050
Huang J (1991) Microstrip reflectarray. Proc IEEE Antennas Propag Symp 612−615
Huang J, Encinar JA (2008) Reflectarray antennas. Wiley, New York
Huang J, Pogorzelski RJ (1998) A Ka-band microstrip reflectarray with elements having variable rotation angles. IEEE Trans Antennas Propag 46:650–656
Huang J, Zawadzki M (2000) Antennas integrated with solar arrays for space vehicle applications. Proc Intl Symp Antenna Propag Electromagn Theory 86−89
Hum SV, Perruisseau-Carrier J (2014) Reconfigurable reflectarrays and array lenses for dynamic antenna beam control: a review. IEEE Trans Antennas Propag 62:183–198
Hum SV, Okoniewski M, Davies RJ (2007) Modeling and design of electronically tunable reflectarrays. IEEE Trans Antennas Propag 55:2200–2210
Javor RD, Wu X-D, Chang K (1995) Design and performance of a microstrip reflectarray antenna. IEEE Trans Antennas Propag 43:932–939
Kamoda H, Iwasaki T, Tsumochi J, Kuki T, Hashimoto O (2011) 60-GHz electronically reconfigurable large reflectarray using single-bit phase shifters. IEEE Trans Antennas Propag 59:2524–2531
Kishor KK, Hum SV (2012) An amplifying reconfigurable reflectarray antenna. IEEE Trans Antennas Propag 60:197–205
Long SA, Huff GH (2011) A fluidic loading mechanism for phase reconfigurable reflectarray elements. IEEE Antennas Wirel Propag Lett 10:876–879
Malfajani RS, Atlasbaf Z (2014) Design and implementation of a dual-band single layer reflectarray in X and K bands. IEEE Trans Antennas Propag 62:4425–4431
Mao Y, Wang C, Yang F, Elsherbeni AZ (2012) A single-layer broad-band reflectarray design using dual-frequency phase synthesis method. Proc Asia Pac Microw Conf 64−66
Milon MA, Cadoret D, Gillard R, Legay H (2007) Surrounded-element approach for the simulation of reflectarray radiating cells. IET Microw Antennas Propag 1:289–293
Mohammadirad M, Komjani N, Chaharmir MR, Shaker J, Sebak AR (2012) Impact of feed position on the operating band of broadband reflectarray antenna. IEEE Antennas Wirel Propag Lett 11:1104–1107
Moustafa L, Gillard R, Peris F, Loison R, Legay H, Girard E (2011) The phoenix cell: a new reflectarray cell with large bandwidth and rebirth capabilities. IEEE Antennas Wirel Propag Lett 10:71–74
Nayeri P, Yang F, Elsherbeni AZ (2010) Broadband reflectarray antennas using double-layer sub-wavelength patch elements. IEEE Antennas Wirel Propag Lett 18:19–29
Nayeri P, Yang F, Elsherbeni AZ (2011) Bandwidth improvement of reflectarray antennas using closely spaced elements. Prog Electromagn Res 9:1139–1142
Nayeri P, Yang F, Elsherbeni AZ (2012) Design and experiment of a single-feed quad-beam reflectarray antenna. IEEE Trans Antennas Propag 60:1166–1171
Nayeri P, Elsherbeni AZ, Yang F (2013a) Radiation analysis approaches for reflectarray antennas. IEEE Antennas Propag Mag 55:127–134

Nayeri P, Yang F, Elsherbeni AZ (2013b) Bifocal design and aperture phase optimizations of reflectarray antennas for wide-angle beam scanning performance. IEEE Trans Antennas Propag 61:4588–4597

Nayeri P, Liang M, Sabory-Garcia RA, Tuo M, Yang F, Gehm M, Xin H, Elsherbeni AZ (2014) 3D printed dielectric reflectarrays: low-cost high-gain antennas at sub-millimeter waves. IEEE Trans Antennas Propag 62:2000–2008

Niu T, Withayachumnankul W, Ung B, Menekse H, Bhaskaran M, Sriram S, Fumeaux C (2012) Design and implementation of terahertz reflectarray. In: 37th international conference on infrared, millimeter terahertz waves, Wollongong, pp 1–2

Perruisseau-Carrier J (2010) Dual-polarized and polarization-flexible reflective cells with dynamic phase control. IEEE Trans Antennas Propag 58:1494–1502

Perruisseau-Carrier J, Skrivervik AK (2008) Monolithic MEMS-based reflectarray cell digitally reconfigurable over a 360° phase range. IEEE Antennas Wirel Propag Lett 7:138–141

Phillion RH, Okoniewski M (2008) Improving the phase resolution of a micromotor-actuated phased reflectarray. Proc Microsyst Nanoelectron Res Conf 169–172

Pozar DM (2007) Wideband reflectarrays using artificial impedance surfaces. Electron Lett 43:148–149

Pozar DM, Metzler TA (1993) Analysis of a reflectarray antenna using microstrip patches of variable size. Electron Lett 29:657–658

Pozar DM, Targonski SD, Syrigos SD (1997) Design of millimeter wave microstrip reflectarrays. IEEE Trans Antennas Propag 45:287–296

Pozar DM, Targonski SD, Pokuls R (1999) A shaped-beam microstrip patch reflectarray. IEEE Trans Antennas Propag 47:1167–1173

Rahmat-Samii Y (1979) Useful coordinate transformations for antenna applications. IEEE Trans Antennas Propag 27:571–574

Rahmat-Samii Y (1988) Reflector antennas. In: Lo YT, Lee SW (eds) Antenna handbook: theory, applications, and design. Van Nostrand Reinhold, New York

Rahmat-Samii Y, Cramer PJ, Woo K, Lee SW (1981) Realizable feed-element patterns for multibeam reflector antenna analysis. IEEE Trans Antennas Propag 29:961–963

Rajagopalan H, Rahmat-Samii Y, Imbriale WA (2008) RF MEMS actuated reconfigurable reflectarray patch-slot element. IEEE Trans Antennas Propag 56:3689–3699

Rajagopalan H, Xu S, Rahmat-Samii Y (2012a) On understanding the radiation mechanism of reflectarray antennas: an insightful and illustrative approach. IEEE Antennas Propag Mag 54:14–38

Rajagopalan H, Xu S, Rahmat-Samii Y (2012b) Reflector antenna distortion compensation using sub-reflectarrays: simulations and experimental demonstration. IEEE Antennas Propag Mag 54:235–246

Rajagopalan H, Xu S, Rahmat-Samii Y (2013) Experimental demonstration of reflectarrays acting as conic section subreflectors in a dual reflector system. IEEE Trans Antennas Propag 61:5475–5484

Rengarajan SR (2009) Reciprocity considerations in microstrip reflectarrays. IEEE Antennas Wirel Propag Lett 8:1206–1209

Rodrigo D, Jofre L, Perruisseau-Carrier J (2013) Unit cell for frequency-tunable beamscanning reflectarrays. IEEE Trans Antennas Propag 61:5992–5999

Romanofsky RR (2007) Advances in scanning reflectarray antennas based on ferroelectric thin-film phase shifters for deep-space communications. Proc IEEE 95:1968–1975

Shaker J, Chaharmir MR, Ethier J (2014) Reflectarray antennas: analysis, design, fabrication, and measurement. Artech House, USA

Smith T, Gothelf U, Kim OS, Breinbjerg O (2013) Design, manufacturing, and testing of a 20/30-GHz dual-band circularly polarized reflectarray antenna. IEEE Antennas Wirel Propag Lett 12:1480–1483

Sorrentino R, Gatti RV, Marcaccioli L (2009) Recent advances on millimetre wave reconfigurable reflectarrays. Proc 3rd Euro Conf Antenna Propag 2527−2531

Tienda C, Arrebola M, Encinar JA (2012) Recent developments of reflectarray antennas in dual-reflector configurations. Intl J Antennas Propag 2012:125287

Tienda C, Encinar JA, Arrebola M, Barba M, Carrasco E (2013) Design, manufacturing and test of a dual-reflectarray antenna with improved bandwidth and reduced cross-polarization. IEEE Trans Antennas Propag 61:1180–1190

Wu B, Sutinjo A, Potter ME, Okoniewski M (2008) On the selection of the number of bits to control a dynamic digital MEMS reflectarray. IEEE Antennas Wirel Propag Lett 7:183–186

Xu S, Rahmat-Samii Y, Imbriale WA (2009) Subreflectarrays for reflector surface distortion compensation. IEEE Trans Antennas Propag 57:364–372

Yang F, Ang Y, Elsherbeni AZ, Huang J (2008) Single-layer multi-band circularly polarized reflectarray antenna: concept, design, and measurement. Proc URSI Gen Assem 1−4

Yang F, Nayeri P, Elsherbeni AZ, Ginn JC, Shelton DJ, Boreman GD, Rahmat-Samii Y (2012) Reflectarray design at infrared frequencies: effects and models of material loss. IEEE Trans Antennas Propag 60:4202–4209

Yu A, Yang F, Elsherbeni AZ, Huang J (2009) A single layer broadband circularly polarized reflectarray based on the element rotation technique. Proc IEEE Antennas Propag Symp 1−4

Yu A, Yang F, Elsherbeni AZ, Huang J (2010a) Experimental demonstration of a single layer tri-band circularly polarized reflectarray. Proc IEEE Antennas Propag Symp 1−4

Yu A, Yang F, Elsherbeni AZ, Huang J, Rahmat-Samii Y (2010b) Aperture efficiency analysis of reflectarray antennas. Microw Opt Tech Lett 52:364–372

Yu A, Yang F, Elsherbeni AZ, Huang J, Kim Y (2012) An offset-fed X-band reflectarray antenna using a modified element rotation technique. IEEE Trans Antennas Propag 60:1619–1624

Zainud-Deen SH, Malhat HAE, Gaber SM, Awadalla KH (2013) Perforated nanoantenna reflectarray. Prog Electromagn Res 29:253–265

Zhou M, Sorensen SB, Jorgensen E, Meincke P, Kim OS, Breinbjerg O (2011) An accurate technique for calculation of radiation from printed reflectarrays. IEEE Antennas Wirel Propag Lett 10:1081–1084

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_47-1

Terahertz Antennas and Measurement

Xiaodong Chen[a]* and Xiaoming Liu[b]
[a]EM and Antenna Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK
[b]School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, China

Abstract

Terahertz technology has been receiving expanding interest in the recent years. With the help of newly invented terahertz sources, terahertz systems have been developed for various applications. One important technology for terahertz systems is the design of efficient antennas for terahertz wave transmission and receiving. Accordingly, the terahertz antenna measurement technology is of equal importance as terahertz antenna design. This chapter gives an overview of the state-of-the-art terahertz antenna design and measurement for a range of applications including the photoconductive antennas and radioastronomy/remote sensing antennas.

Keywords

Terahertz; Antenna; Antenna Measurement; Photoconductive Antenna; Horn Antenna; Reflector Antenna; Far-field; Near-field; CATR

Introduction

The terahertz (THz) band is usually defined as the frequency range between 300 GHz and 10 THz. The wavelength of a THz wave is in the range of 1,000–30 μm (Lee et al. 2006; Grade et al. 2007). In the literature, an expanded range of 0.1–10 THz (Liu et al. 2008; Hoffmann and Fülöp 2011) may also be used for the THz band. The location of the THz band on the spectral diagram is shown in Fig. 1. It is interesting to see that below the THz band, electronic technologies can be employed to generate high-power RF/microwave radiation, while above the THz band, photonic methods are efficient to produce IR/visible waves. It had been a long period before the effective THz source was invented, leaving a gap (so-called THz gap) between microwave and light wave to be bridged.

An overview of the THz-power performance of different sources around the THz gap is given by Tonouchi (2007). Microwave methods encompass Gunn diode, resonant tunnel diode, and multiplexer impact ionization avalanche transit time diode. Quantum cascade laser and lead-salt laser are grouped as photonic methods. It has been observed that the output power level of the conventional RF/microwave techniques drops sharply down to 10 μW when the operating frequency approaches 1 THz. It is also true when optical methods are applied to generate THz waves. Such an awkward situation was not changed before the advent of the first photoconductive THz source (Auston and Smith 1983; Smith et al. 1988). Various compact THz systems spring up (Zhang 2002; Cai et al. 1997) of which the THz time domain spectroscopy (THz-TDS) is particularly worth mentioning (Schmuttenmaer 2004). More details of the THz-TDS system will be discussed in subsequent sections.

*Email: xiaodong.chen@qmul.ac.uk

DOI 10.1007/978-981-4560-75-7_47-1

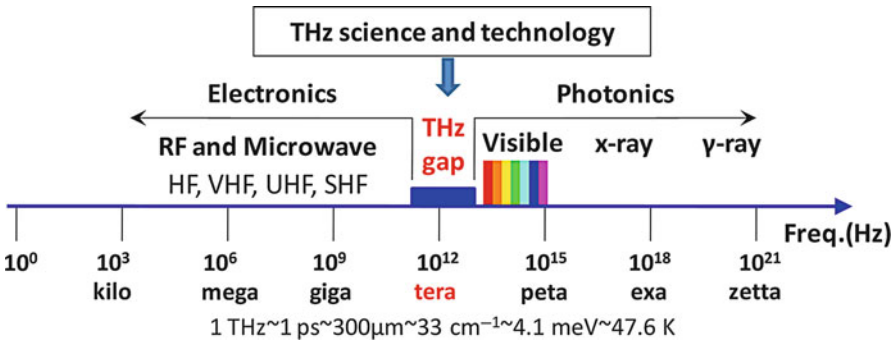


Fig. 1 Illustration of the location of the THz band. Below the THz band, electronic technologies dominate, while beyond the THz band, it is the photonic phenomenon that has to be considered

The reason that people invested huge efforts into the THz band lies in its rich scientific and practical potentials. For instance, about 98 % of the photons from the Big Bang are in the THz range (Blain et al. 2002); a variety of semiconductor phonons resonate in the THz range (Cho et al. 1999); macrobiomolecules have unique fingerprint lines in the THz range (Sushko et al. 2013); THz waves are less radioactive and do not cause ionization hazardous to human beings; THz waves provide better temporal resolution compared with RF/microwave and have larger penetration depth compared to IR/visible light, which bears the advantages in the applications of imaging and nondestructive inspection (Henry et al. 2013). These facts are self-evident in that THz is emerging as a promising field, spanning a wide range of disciplines.

As a matter of fact, an antenna is a key component for transmitting and receiving THz waves in the existing THz systems. In the conventional modality, the source and the antenna are usually considered as two parts. The distinct difference of a THz system is that the source and the radiation part may need to be considered as a whole, which is particularly true for a THz-TDS system. Nevertheless, one has to bear in mind that THz can still be generated by the conventional microwave ways such as multiplexer. Therefore, THz antennas are very versatile such as the photoconductive antenna developed in recent decades, the conventional horn antenna, and the reflector antenna. Also, terahertz antennas benefit from the development of nanotechnology (Bareiss et al. 2011) and metamaterials (Zhang et al. 2014). For photoconductive antennas, the major limitation is their power conversion efficiency, which though can be improved through many ways, while for horn and reflector antennas working in the THz band, the main technological issue is the fabrication precision.

Apart from the design and fabrication of efficient antennas, antenna measurement in the THz band is also a challenge that cannot be circumvented. For the photoconductive antenna scheme, antenna measurement is essentially the detection of THz waves. And therefore, photoconductive antennas are frequently utilized for THz antenna measurement (Zhang and Xu 2009). Details will be presented in the coming parts. It is, however, problematic for systems working on upconverting technology. There are many issues already identified associated with existing methods for antenna measurement. For instance, far-field method requires a long distance between the radiation source and antenna under test in the THz range (Balanis 2005). Such a condition is to ensure the phase uniformity in the test zone, which however brings in large signal losses, say path loss and atmospheric absorption loss. The path loss in the THz range is significant, as shown in Fig. 2. Near-field scanning method can be conducted in shielding environment, minimizing the possible issues in outdoor environment. Unfortunately, near-field method requires demanding stability of both electrical and mechanical systems. Moreover, the scanning time for

DOI 10.1007/978-981-4560-75-7_47-1

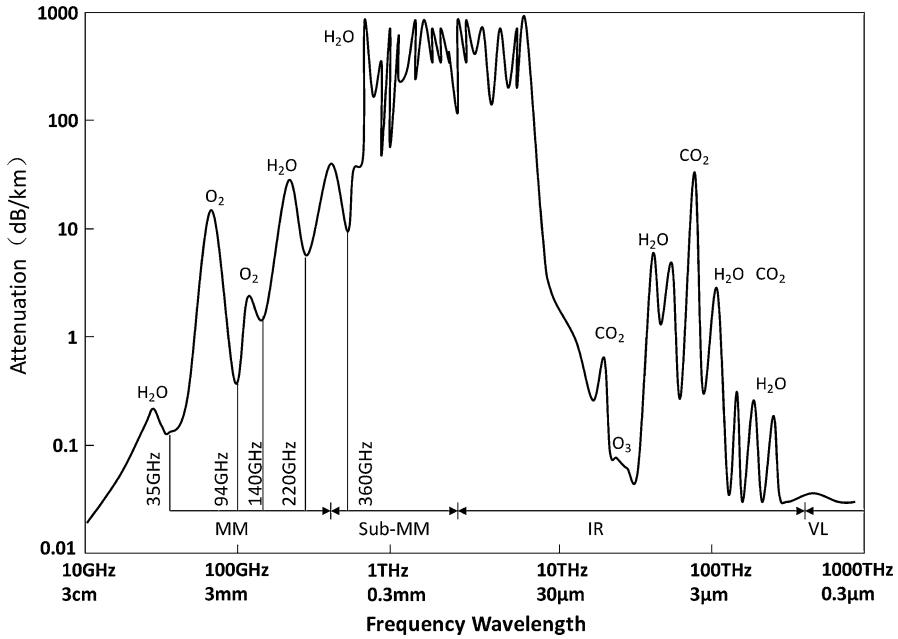


Fig. 2 Atmospheric absorption window (plotted by using the model in Klein 2000). MM stands for millimeter wave, VL denotes visible light

electrically large systems is in most cases extremely long (Tuoviizen 1993), sometimes maybe unacceptable. Recently, people have resorted to compact antenna test range (CATR) techniques aiming at providing a feasible approach to THz antenna measurement (Liu et al. 2011a, b; Rieckmann et al. 1999; Hirvonen et al. 1997). A CATR system transforms the field of a feed horn to a regional plane wave or pseudoplane wave in a very short distance, normally in several or several tens of meters, mimicking the far-field method effectively. Unlike conventional far-field method, a CATR works in an anechoic chamber, providing controlled all-whether measurement conditions and eliminating large path loss and atmospheric absorption loss. Compared with near-field method, measurement using CATR can be conducted in much shorter period. Many types of CATR are currently under investigation such as trireflector CATR, holographic CATR, and so on.

In addition to the aforementioned methods, many technologies are under development to enable on-orbit measurement (Forma et al. 2009). Also, precision measurement techniques have been introduced into antenna measurement to minimize the influence of noise or scattering effects (Paquay and Marti 2005; Paquay et al. 2007; Appel-Hansen 1979). Furthermore, mechanical–electrical combined method is also utilized (Dominic et al. 2009). Even more, simulation techniques are employed in combination with measurement for some cases where current technology cannot provide feasible measurement (Luis et al. 2010).

In this chapter, it is aimed to provide an overview of THz antennas and measurement. THz technology has been undergoing fast-evolving state in the past decades, when many novel types of THz antennas have been reported. The authors apologize to any contributors whose work is not included in this chapter due to limited space. The following parts are organized as follows: section "THz Antennas" introduces the fundamentals of THz antennas such as photoconductive antenna, horn antenna, and reflector antenna; section "THz Antenna Measurement" discusses different measurement technologies with four common methods (photoconductive detection method, far-field measurement, near-field measurement, and

DOI 10.1007/978-981-4560-75-7_47-1

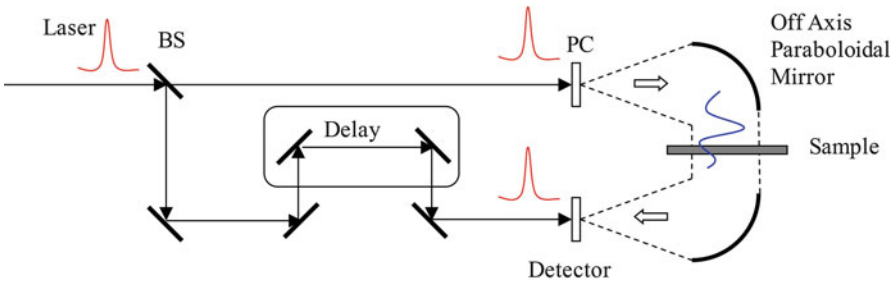


Fig. 3 A diagram of the THz-TDS. *BS* beam splitter, *PC* photoconductive antenna. A femtosecond laser is divided into two beams, with one incident on the photoconductive antenna and the other diverted to the detector. Terahertz wave is generated by the photoconductive antenna, and then was focused by off axis paraboloidal mirrors before going through the sample or arriving at the detector

compact antenna test range) and two infrequently used methods (defocusing method and electrical–mechanical combined method).

THz Antennas

Photoconductive Antennas

Photoconductive antennas are possibly the most used form in the THz-TDS. This method was first introduced by Auston and his colleagues in the early 1980s (Auston and Smith 1983; Smith et al. 1988). Later on, the photoconductive antennas were widely utilized in THz systems. A diagram of the THz-TDS is shown in Fig. 3. An 800 nm laser is directed to a beam splitter to divide the laser to a pump pulse and a probe pulse. The pump pulse shines on the photoconductive antenna to produce THz radiation, which is then refocused by an off-axis paraboloidal reflector. After passing through the sample, which can be biological species, semiconductor wafers, and other imaging objects, the THz wave is again refocused to the detector. The other path, the probe pulse, is delayed by a pair of delay mirrors to produce time delay to the pump pulse. Generally speaking, the laser pulse is much narrower in comparison with the THz wave, with the former on the order of femtosecond and the latter on the order of picosecond, ensuring sufficient time-resolved sampling rate.

It is clear that the photoconductive antenna is a key part of the whole system. As shown in Fig. 4a, a photoconductive antenna consists of several parts: a semiconductor substrate, two electrodes, and a bias voltage. It is well known that when a photon is incident on a semiconductor substrate, an electron may transit from the valence band to the conduction band if the bandgap E_g is smaller compared to the energy of the photon, as shown in Fig. 4b. In this way, an electron–hole pair is formed, with an electron in the conduction band and a hole in the valence band. If there is no further treatment, recombination of the newly created electron–hole pair is possible to happen. Therefore, a bias voltage is applied across the two electrodes, forming an electric field, which can efficiently separate the electron–hole pair. Therefore, an instant current takes place, producing THz radiation. In the literature, photoconductive antennas are also referred to as photoconductive switchers since the free carriers can only be generated when the photonic energy is higher than the bandgap.

It has been derived based on a dipole radiation model (Lee 2009), i.e., assuming the radiation of THz wave from a photoconductive antenna resembles the radiation field of an electric dipole, that the electric field is proportional to the second-order time derivative of the polarization (or dipole moment) within the substrate such that

DOI 10.1007/978-981-4560-75-7_47-1

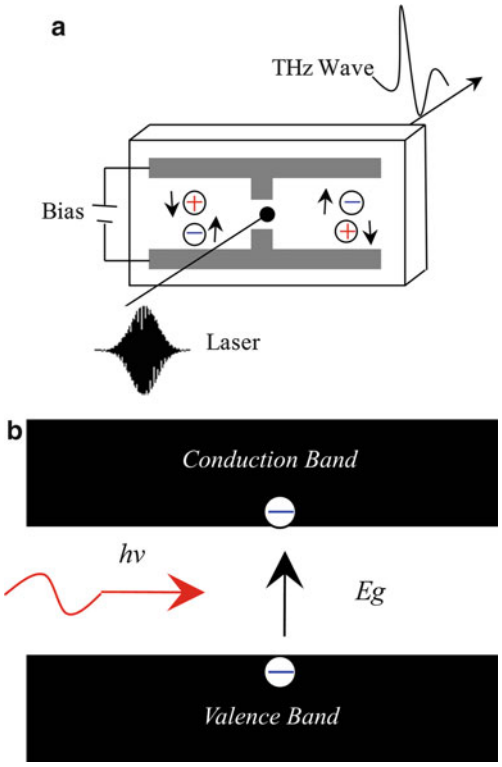


Fig. 4 The principle of a photoconductive antenna. Laser with photonic energy larger that the band gap of the substrate creates electron-holes pairs, which are accelerated by the bias voltage, and in this way producing terahertz wave

$$\vec{E}_{THz}(t) = \frac{\mu_0}{4\pi}\frac{\sin\theta}{r}\frac{d^2}{dt^2}\left[p\left(t-\frac{r}{c}\right)\right]\hat{\theta}, \tag{1}$$

where $\vec{E}_{THz}(t)$ is the time-dependent THz electric field, r is the distance from the dipole to the observation point, p is the polarization or dipole moment, and c is the light speed in the free space. It needs to be mentioned that the retarded time effect $(t-r/c)$ has to be involved when calculating the radiated field. As a matter of fact, under the condition of one-dimensional transport of the carriers, the photoconductive current can be written as

$$I_{PC} = \frac{dp(t)}{w_0 dt}, \tag{2}$$

where w_0 is the spot size of the laser beam. This immediately gives

$$\vec{E}_{THz}(t) = \frac{\mu_0 w_0}{4\pi}\frac{\sin\theta}{r}\frac{d}{dt}\left[I_{PC}\left(t-\frac{r}{c}\right)\right]\hat{\theta}, \tag{3}$$

showing that the radiated field is linearly proportional to the time derivative of photocurrent. The photocurrent can also be represented by the carrier density and carrier mobility as follows:

$$I_{PC} = e(n_e v_e + n_h v_h), \tag{4}$$

DOI 10.1007/978-981-4560-75-7_47-1

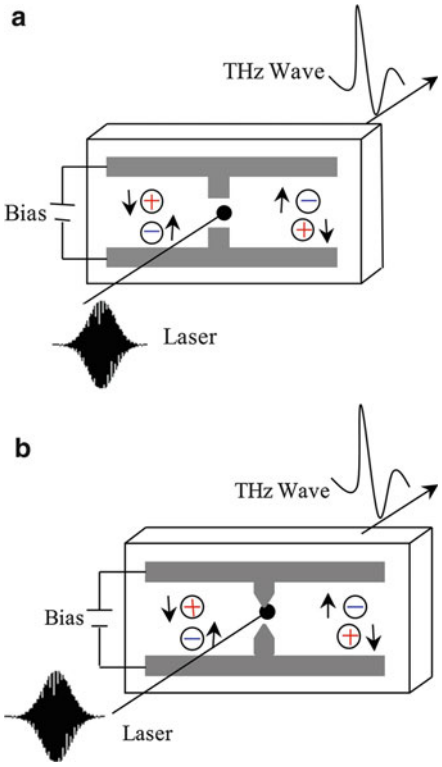


Fig. 5 Schematic diagram of (**a**) stripline and (**b**) sharp dipole structures

where n_e and n_h are the densities of electrons and holes respectively, while v_e and v_h are the average velocities of electrons and holes respectively. Normally, the carrier density under a pulse excitation can be modeled as

$$n = n_0 e^{-\frac{t}{\tau}}, \tag{5}$$

which implies that the carrier density is dependent on the carrier lifetime τ. Carrier lifetime and the velocity (also called mobility) significantly affect the radiated THz electric field. The mobility of the holes is approximately one order of magnitude lower than that of electrons. Therefore, the electrons dominate THz generation in photoconductive antennas (PCAs). A wide range of semiconductors can be employed in PCA fabrication. The main considerations include response time, carrier lifetime, and so forth. Unlike conventional microstrip antennas, where only the macroscopic dielectric property and loss tangent are considered, the design of PCAs needs to consider the microscopic property of semiconductors.

Of equal importance, the structure of PCAs is intensively investigated to create efficient THz radiation. Indeed, since the mobility of electrons is higher than that of holes, in a stripline structure, the laser spot is normally placed intentionally closer to the anode, shown in Fig. 5a. The distance between two electrodes is several tens of micrometers. It has been found that stripline structure produces narrower pulse than dipole structure (Fig. 4) under the condition of the same parameters for both structures. Subsequently, dipole structure has narrower bandwidth compared to stripline structure. However, dipole structure can produce higher radiation power under the same laser illumination and bias voltage. In general, the THz generation efficiency is very low, less than 0.01 %. Various structures have been investigated to enhance the efficiency, for instance, by sharpening the ends of the dipole, as shown in Fig. 5b. It has been reported that an average THz radiation power of 2–3 μW from the structure has been obtained under a 60-V bias voltage and a 20-mW optical excitation with a Ti:sapphire femtosecond oscillator (Cai et al. 1997).

DOI 10.1007/978-981-4560-75-7_47-1

Fig. 6 The geometry of photoconductive antennas common dipole and sharp dipole

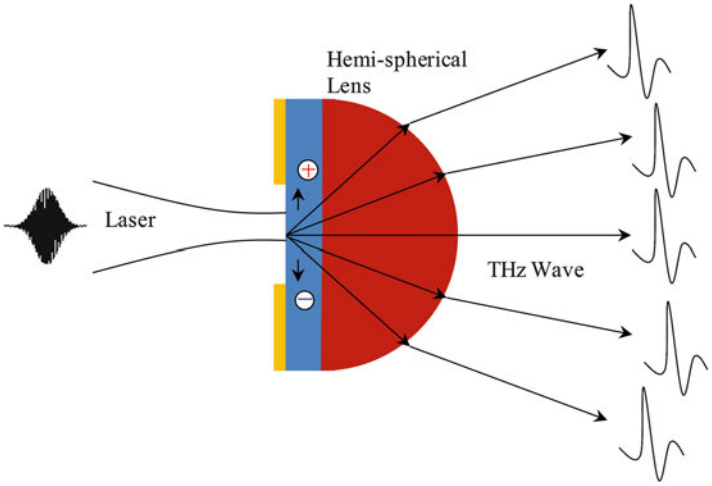


Fig. 7 Illustration of lens structure of a photoconductive antenna. With a hemi-spherical lens mounted on the substrate side, the terahertz wave would be more concentrated around the boresight direction of the photoconductive antenna, and therefore increase its directivity

Li and Huang (2006) theoretically investigated this phenomenon by comparing a dipole structure and a sharp dipole structure, see Fig. 6. The electric field generated by line charges can be expressed as

$$E_d = \frac{Q}{2\pi\varepsilon_0 r\sqrt{4r^2 + w^2}}, \tag{6}$$

where Q is the total charge of the dipole, w is the width of the dipole, as shown in Fig. 7, and r is the distance from the charge to the observation point. The electric field formed by point charge can be written as

$$E_s = \frac{Q}{4\pi\varepsilon_0 r^2}. \tag{7}$$

If one uses the same antenna substrate and bias voltage, it is reasonable to assume that the total charge Q for both structures is the same. Consequently, one has

DOI 10.1007/978-981-4560-75-7_47-1

$$\frac{E_s}{E_d} = \frac{\frac{Q}{4\pi\varepsilon_0 r^2}}{\frac{Q}{2\pi\varepsilon_0 r}} = \frac{\sqrt{4r^2 + w^2}}{2r}. \tag{8}$$

If the observation point is in the gap between the two electrodes, r is of the same order as the gap g. In this circumstance,

$$\frac{E_s}{E_d} = \frac{\sqrt{4g^2 + w^2}}{2g}. \tag{9}$$

In an experiment carried out by Cai et al. (1997), the width $w = 20$ μm, the gap $g = 5$um, which results in an electric field enhancement of

$$\frac{E_s}{E_d} = 2.25. \tag{10}$$

The measured results gave an enhancement of 2.34, showing great agreement with the theoretical prediction within an accuracy of 5 %.

Generally speaking, a dipole antenna provides relatively larger beamwidth. To narrow the beamwidth and increase the directivity, a hemispherical antenna is mounted on the surface of the photoconductive antenna and normally on the substrate side. To decrease the multireflection between the substrate and the lens, the material for the lens is so chosen that the refractive index of the lens is almost the same as that of the substrate. For instance, the refractive index of silicon in the terahertz band is approximately 3.4, which is nearly the same as that of GaAs. On the interface of lens and air, the curved surface makes the terahertz rays less divergent by the refraction law, leading to an increased gain. Such a structure is crucial to low-power-level THz applications. It has been reported that by using a silicon lens, the gain can be increased by 10 dB (Liu et al. 2013). A schematic diagram of lens-mounted THz antenna is shown in Fig. 7.

A log-periodic structure has been reported by (Mendis et al. 2005) where the outer diameter is 1.28 mm, as shown in Fig. 8. Generally speaking, a periodic structure provides relatively wide bandwidths. It is interesting, however, to find that the bandwidth of a log-periodic structure may be quite limited (Mendis

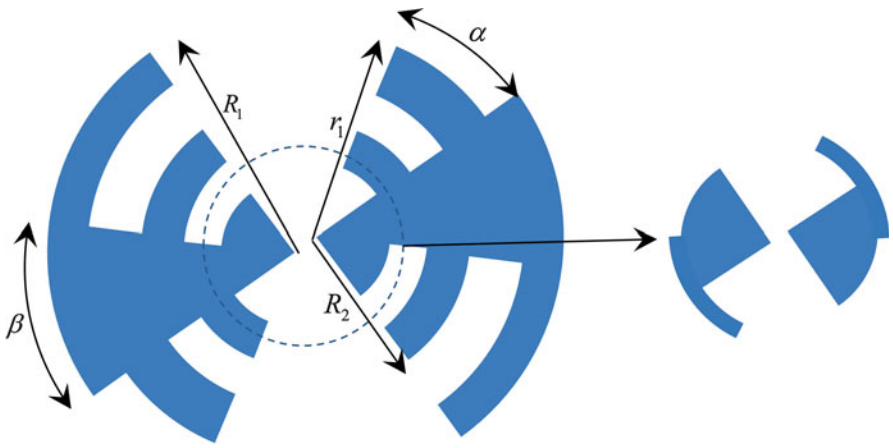


Fig. 8 The configuration of a log-periodic antenna. When working on high frequency side, a log-periodic antenna resembles a bow-tie antenna (Adapted from Mendis et al. 2005)

DOI 10.1007/978-981-4560-75-7_47-1

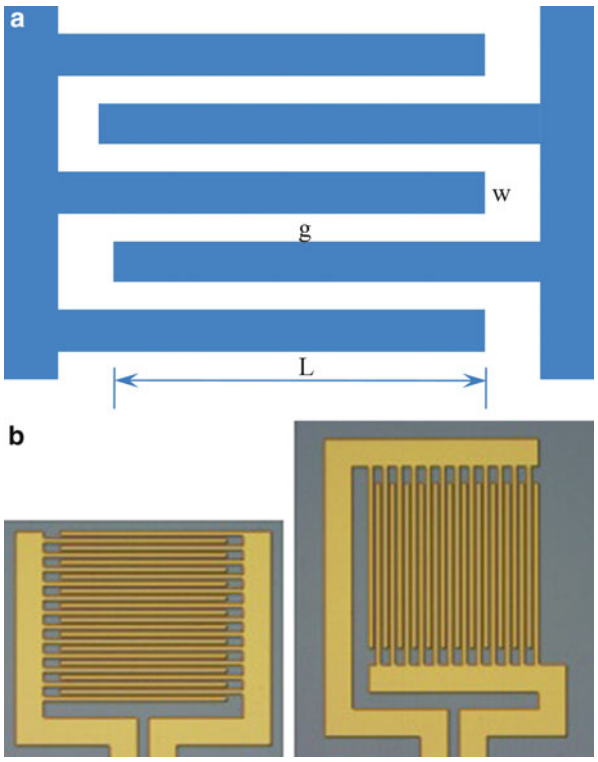


Fig. 9 Interdigital array. (**a**) Diagram; (**b**) Photography (By courtesy of Del Mar Photonics)

et al. 2005). A conventional strip structure may have a bandwidth from several fractions of THz to 2 THz, while the reported log-periodic structure works below 1 THz. Actually, a log-periodic antenna may provide wider bandwidth if properly designed. It is the outer radius that determines the low-resonant frequency. For higher frequency, a log-periodic antenna may act as a bowtie antenna.

The impedance matching is another issue with photoconductive antennas. Normally, the input impedance for a photomixer may be as high as 10 kΩ. However, the impedance for a dipole antenna may be much smaller. Impedance matching may be achieved through proper design of the structure. A Yagi-Uda antenna has been reported providing input impedance of 2.6 kΩ, which significantly improved impedance matching (Han et al. 2010). By further optimizing the driving unit, the input impedance may be increased to some 4–5 kΩ. Also, such a structure has relatively narrower beamwidth and hence higher directivity. The azimuthal plane has a 3-dB beamwidth of 49°, and the 3-dB beamwidth is 58°in the elevation plane. It is possible to increase the directivity by adding more director elements.

To increase radiation power, lens array has been investigated by many research groups, with some of the structures already commercialized. One lens array is based on interdigital structure, as shown in Fig. 9. Such structure provides higher radiation power than a conventional dipole structure. The interdigital structure forms an array of photoconductive antennas. An improved version may be achieved by mounting a silicon lens array on top of the interdigital structure.

Horn Antennas

A horn antenna is still one of the most popular types of antennas in THz bands. In most of the cases, a horn antenna is used in combination with reflector antennas, which will be discussed in later a section. Conventional horn antennas include pyramidal horn antenna, circular horn antenna, and corrugated horn antenna, as shown in Fig. 10.

DOI 10.1007/978-981-4560-75-7_47-1

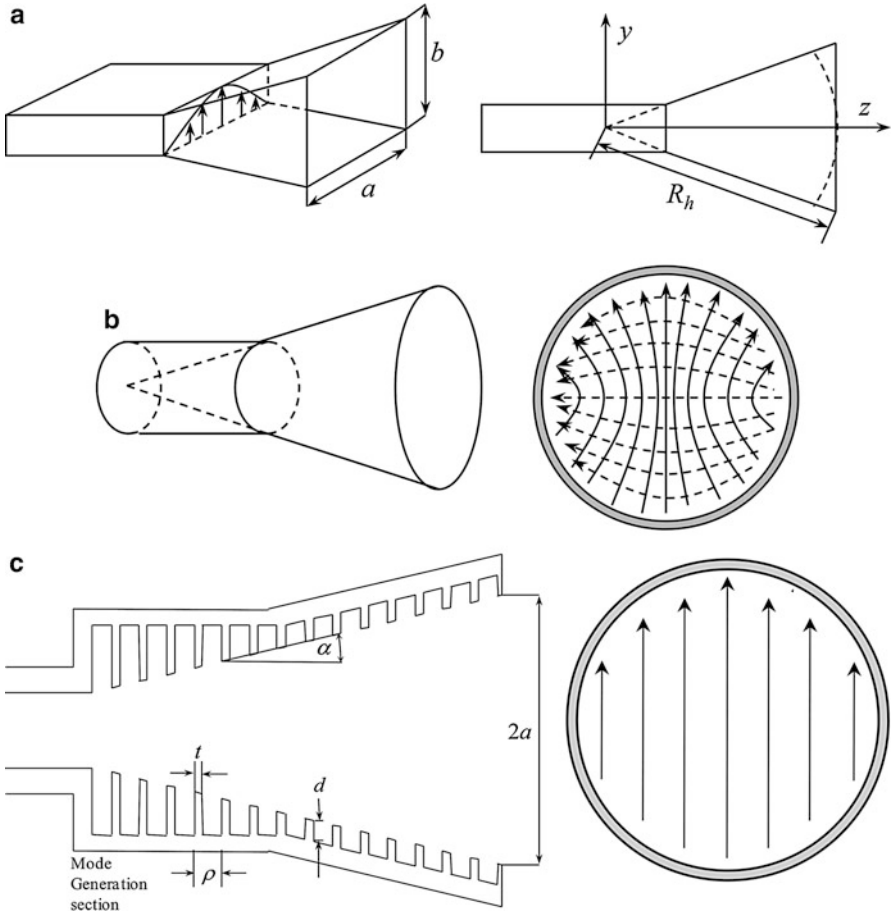


Fig. 10 Diagram of horn antennas. (**a**) Rectangular horn antenna; (**b**) Circular horn antenna; (**c**) corrugated horn antenna (Balanis 2005; Goldsmith 1997)

The most popular feed horns may be corrugated horns for their purity of polarization and scalar field distribution. The principle of operation of corrugated horns can be explained by considering the way in which the corrugated walls alter the field pattern in a corrugated waveguide to achieve symmetric main beam and low cross-polarization levels, which can only be met by an aperture with nearly linear-field distribution, as shown in Fig. 10c. Unfortunately, a linear electric field cannot be produced by waveguide supporting pure transverse-electric (TE) or transverse-magnetic (TM) modes. Therefore, in general cases, the corrugation is designed to generate a hybrid HE11 mode with almost linear electric field on the output aperture. Since the field distribution is independent on the azimuth angle, the radiation pattern will be symmetric with extremely good polarization purity.

The design of a corrugated horn has been reviewed by many researchers, and the interested readers are directed to a monograph on microwave horns (Olver et al. 1994). A well-designed corrugated horn can provide cross-polarization levels lower than −40 dB and first side lobe better than −40 dB as well as excellent fidelity to Gaussian distribution. The transverse component of the near field of a Gaussian beam propagating along z-axis can be described by

$$E_t(r,z) = \left(\frac{2}{\pi w^2}\right)^{0.5} \exp\left(\frac{-r^2}{w^2} - jkz - \frac{j\pi r^2}{\lambda R} + j\phi_0\right), \tag{11}$$

DOI 10.1007/978-981-4560-75-7_47-1

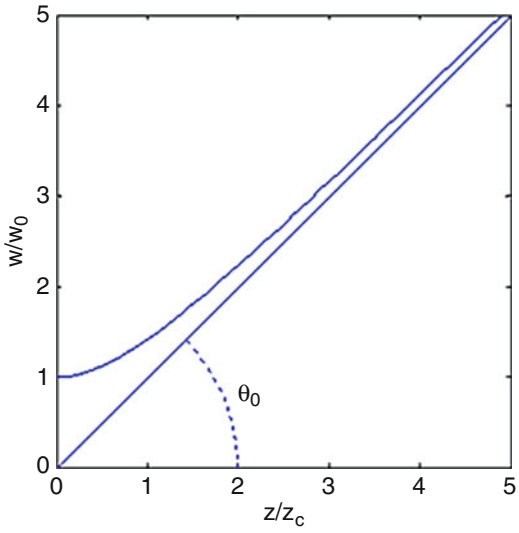


Fig. 11 The divergence angle of a corrugated horn antenna

where

$$R = z + \frac{1}{z}\left(\frac{\pi w_0^2}{\lambda}\right), \tag{12}$$

$$w = w_0\left[1 + \left(\frac{\lambda z}{\pi w_0^2}\right)^2\right]^{0.5}, \tag{13}$$

$$\tan\phi_0 = \frac{\lambda z}{\pi w_0^2}. \tag{14}$$

The far field can be written as

$$\frac{E_t(\theta)}{E_t(0)} = \exp\left(-\left(\frac{\theta}{\theta_0}\right)^2\right), \tag{15}$$

where

$$\theta_0 = \lim_{z \gg z_c}\left[\tan^{-1}\left(\frac{w}{z}\right)\right] = \tan^{-1}\left(\frac{\lambda}{\pi w_0}\right). \tag{16}$$

As shown in Fig. 11, the far field can be represented by the divergence angle. High-precision ultra-Gaussian beam horn antennas provide better than 40 dB cross-polarization isolation due to its scalar field distribution property and high Gaussian beam profile fidelity down to −40 dB (McKay et al. 2013; Adel et al. 2009). Such techniques enable one to calculate the far field and gain analytically, facilitating the prediction of the far field of a reflector antenna system. Horn antennas utilized as feeds are seen in ALAMA array and Planck telescope (Gonzalez and Uzawa 2012; Paquay et al. 2007), both arranged in a focal plane array with more than a dozen feeds working at different frequencies. Planck array horns

DOI 10.1007/978-981-4560-75-7_47-1

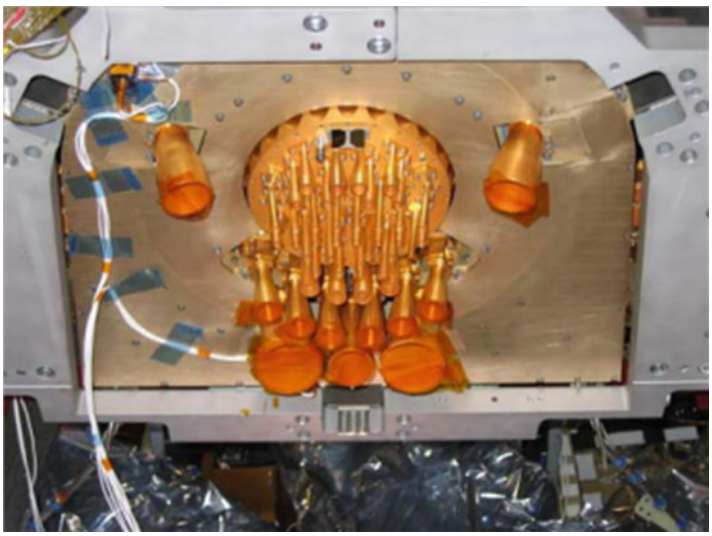


Fig. 12 Planck array horns covering the frequency range of 27.5–870 GHz, centered at 30, 44, 70, 100, 143, 217, 353, 545, 857 GHz (Paquay et al. 2007, by courtesy of ESA)

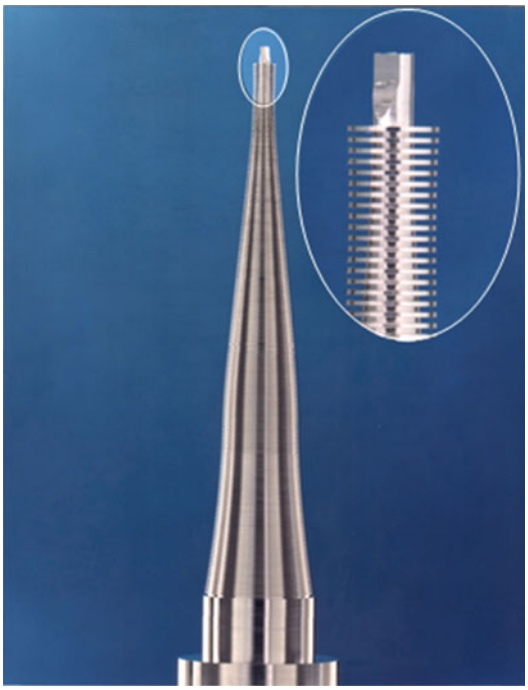


Fig. 13 The mandrel of a corrugated horn antenna up to 2 THz (Alderman et al. 2008) (Reproduced by permission from Rutherford and Appleton Laboratory)

covering the frequency range of 27.5–870 GHz are centered at 30, 44, 70, 100, 143, 217, 353, 545, 857 GHz (Paquay et al. 2007). The feed array of the Planck satellite is shown in Fig. 12.

The difficulty of fabricating such horn antennas is again the accuracy. Note well that the corrugation of such antennas has a depth of a quarter of a wavelength and the teeth have a width of less than a quarter of a wavelength. The fabrication of such corrugated horns starts with fabricating a mandrel followed by electroforming on the surface of the mandrel a proper layer of copper. The mandrel was later on washed off by using strong base solution such as NaOH solution. Generally speaking, the mandrel is made of

DOI 10.1007/978-981-4560-75-7_47-1

aluminum, which reacts with NaOH solution. Several error sources can be identified during the fabrication. Firstly, the mandrel needs precision fabrication. Its precision directly affects the depth of slots and teeth of the horn. Secondly, the electroforming process is a time-consuming process, which is very likely to introduce the *fill-in error*, i.e., the imperfection when filling copper to the slots of the mandrel. Lastly, in the washing-off stage, residual aluminum may exist that distorts the radiation pattern or increases the reflection coefficient. A magnified mandrel fabricated in Rutherford and Appleton Laboratory is shown in Fig. 13.

Actually, THz horns are normally connected with waveguides, the previous international standards specifying the dimensions of which only to a maximum frequency of 330 GHz (Hesler et al. 2007). The newly approved standard rectangular metallic waveguide for THz applications has extended the operating frequency to 5 THz, shown in Table 1. It clearly shows that the physical dimensions of the waveguides working in the THz range lies in the micrometer range, which surely will impose challenges on fabrication.

Reflector Antennas

A reflector antenna is deployed in a wide scope of applications ranging from microwave communication to optical telescope. Most satellite communication systems are deployed with reflector antennas. Radioastronomy cannot have been so developed if a reflector antenna had not yet been invented. When the working frequency is high up to the millimeter wave range, the size of resonate type of antenna starts to decrease, leading to decrease in the effective receiving area. Reflector antennas are normally electrically large, from several wavelengths in diameter to tens of hundred wavelengths. Such electric size usually

Table 1 IEEE standard P1785.1/D3: frequency bands and waveguide dimensions (IEEE Standard P1785.1/D3)

Waveguide name	Aperture width (μm)	Aperture height (μm)	Minimum frequency (THz)	Maximum frequency (THz)
WM-710	710	355	0.26	0.40
WM-570	570	285	0.33	0.50
WM-470	470	235	0.40	0.60
WM-380	380	190	0.50	0.75
WM-310	310	155	0.60	0.90
WM-250	250	125	0.75	1.10
WM-200	200	100	0.90	1.40
WM-164	164	82	1.10	1.70
WM-130	130	65	1.40	2.20
WM-106	106	53	1.70	2.60
WM-86	86	43	2.20	3.30
WM-71	71	35.5	2.60	4.0
WM-57	57	28.5	3.30	5.0

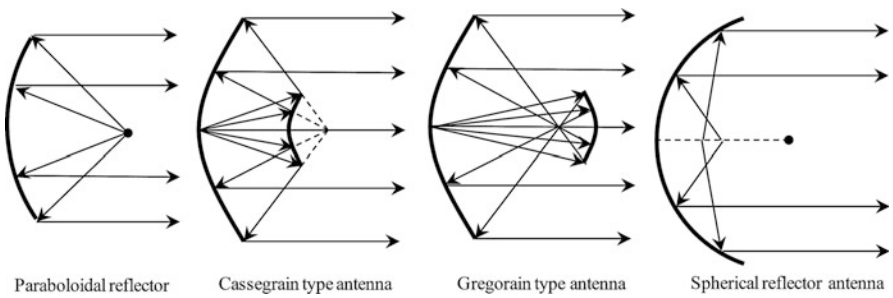


Fig. 14 Various reflector antennas used for THz radiometers and payloads

DOI 10.1007/978-981-4560-75-7_47-1

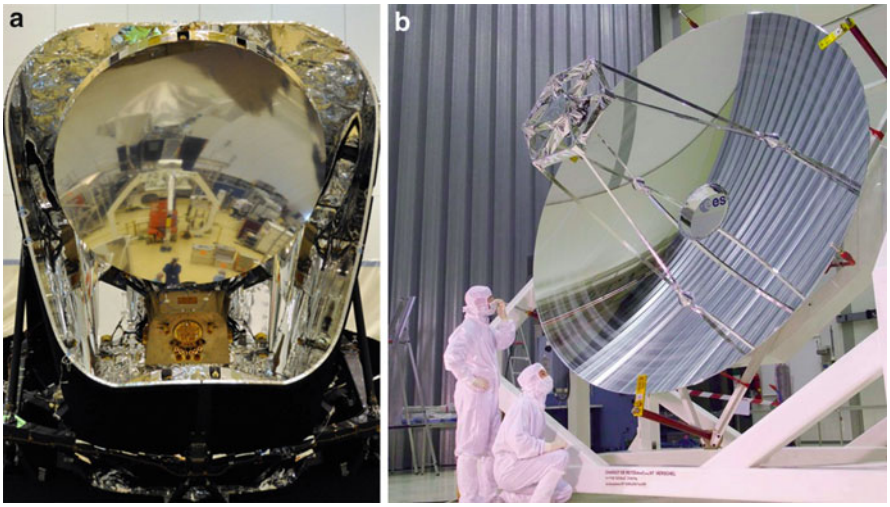


Fig. 15 The telescope of (**a**) Planck with a main reflector 1.5 m in diameter, and (**b**) Herschel with a main reflector of 3.5 m in diameter (Reproduced by permission of ESA)

bears high gain or high effective receiving area. In general, electrically large antenna provides very narrow beam, usually the 3 dB beam width being a couple of degrees or even narrower. Such a narrow beam provides one with high gain since the energy is very concentrated, which is very critical for remote sensing applications.

The surface profile of a reflector antenna is very versatile such as paraboloid, spheroid, ellipsoid, and hyperboloid surfaces, as shown in Fig. 14. The most popular one may be paraboloidal reflector. Launching from the focal point of a paraboloidal reflector, optical rays become parallely collimated after being reflected by the paraboloidal reflector. This attribute ensures equal phase on the output aperture plane. An improved structure is so-called Cassegrain antenna, where a hyperboloidal subreflector is employed to increase its effective focal length. An astronomy telescope may also in the form of spherical reflector receive signals from different incident angles. One example is the Five hundred meter Aperture Spherical Telescope (FAST) under construction in China (Nan and Li 2013).

Single paraboloidal reflector antenna used for THz application is rarely reported; one example is the main dish equipped on the payload of AMSU-B (Martin and Martin 1996). Examples of single reflectors, however, can be found in the microwave and millimeter wave range. In the THz range, paraboloidal reflectors are utilized in Cassegrain or Gregorian configurations. Offset structure was employed in Planck telescopes, while the Herschel telescope adopted a front-feed one. The Planck telescope has a main dish of diameter of 1.5 m, with its operating frequency high up to 870 GHz (Dominic et al. 2009). The Herschel telescope, in contrast, has a much larger main dish of 3.5 m in diameter (Dominic et al. 2009), shown in Fig. 15. Such electrically large antennas enable the detection of cosmic background radiation and deep space signal with gain up to 70 dB (Dominic et al. 2009). Ground base observatory also adapts electrically large-aperture reflector antenna or reflector array. The ALAMA antenna arrays have single antennas of diameter of 7 m or 12 m (Gonzalez and Uzawa 2012). Such array forms an interferometer that can detect signals from deep space with an angular resolution of 0.1 acres. The working frequency of the ALAMA array covers the range of 31–950 GHz. Spherical main dishes are also used such as the main reflectors of FAST under construction in China and Arecibo Observatory in Chili.

In quasioptical systems, paraboloid, spheroid, ellipsoid, and hyperboloid reflectors are all possibly employed for refocusing purpose. The equivalent focal length of an elliptical of hyperbolical reflector can be written as

DOI 10.1007/978-981-4560-75-7_47-1

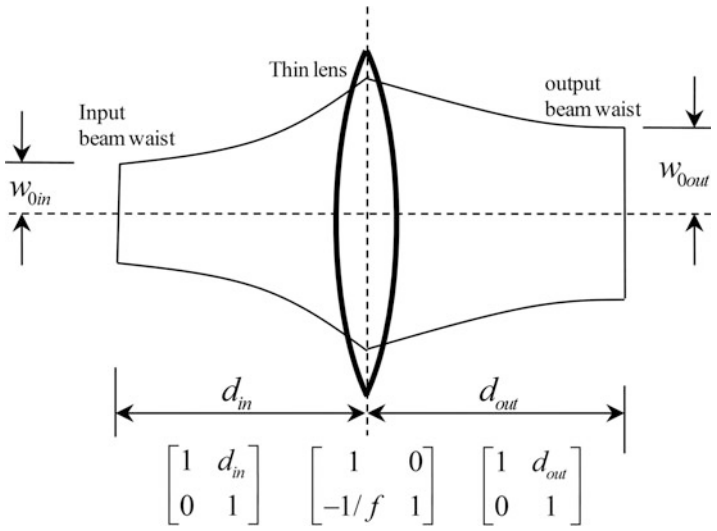


Fig. 16 Transform a Gaussian beam by thin lens

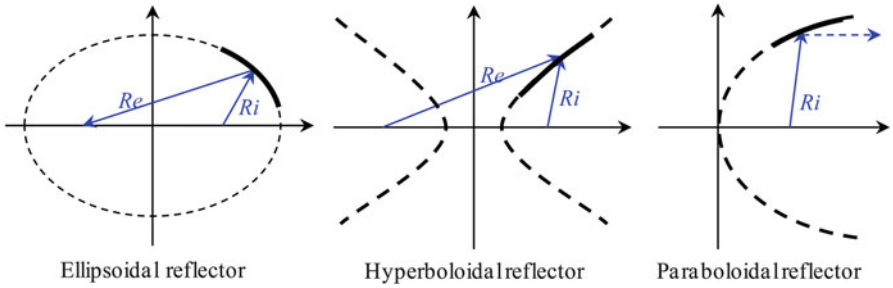


Fig. 17 Reflectors for quasi-optical systems. *Ri* stands for the distance from the incident focal point to the optical center of the reflector, and Re for the distance from the optical center to the emergency focal point

$$f = \frac{R_i R_e}{R_i + R_e}, \tag{17}$$

while the equivalent focal length for a parabolic reflector is

$$f = R_i. \tag{18}$$

These reflectors act as lenses transforming an incoming Gaussian beam to another one of different beam radius (Fig. 16). Such a transform can be implemented through ABCD matrix method for a preliminary design to determine the output beam profile. For practical applications, the size of the reflectors is normally four times o the beam radius to ensure 99.8 % of the power intersected by the reflectors (Goldsmith 1997). Normally, in a quasioptical system, the reflector is a section of an ellipsoidal, hyperboloidal, or paraboloidal surface, see Fig. 17. Such reflector surfaces are only a second-order approximation of thin lenses, suggesting that distortion may be introduced to the output beam. The distortion has been analyzed by Murphy (1987). Essentially, it is the coupling to higher Gauss-Hermit modes that cause distortion to the fundamental Gaussian mode. To characterize the distortion, a distortion parameter is defined and found as $U = 2^{-1.5} w_m \tan\theta_i f^{-1}$, showing a linear proportion to the tangent of the incident angle θ_i and the beam width at the mirror w_m and inversely proportional to the focal length of the mirror. Interestingly, the cross-polarization is also dependent of these factors and obeying the same rules.

DOI 10.1007/978-981-4560-75-7_47-1

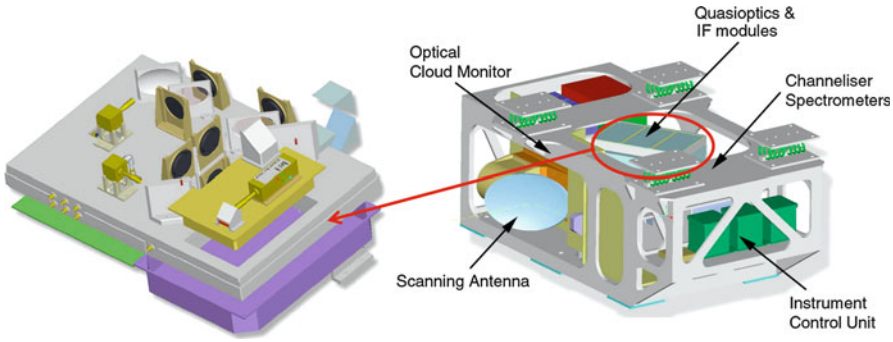


Fig. 18 An assembly model for the quasi-optical system equipped on MARSHAL limb-sounding unit (Reproduced with permission from Rutherford and Appleton Laboratory)

Such observations clearly indicate that to minimize the distortion and suppress the cross-polarization, the incident angle shall be as small as possible. Also, the beam width at the mirror should be controlled as small as possible. At last, one could choose long-focal length reflectors to ensure good performance. Nevertheless, such ideal conditions may not be met in a practical design. For instance, small incident angles may introduce blockage from other reflectors in a multichannel system. In this sense, the design of an optimized quasioptical system is a trade-off between many factors.

Since the designs of the reflectors are based on Gaussian beam theory, which is frequently encountered in a laser propagation textbook, the term *quasioptical technique* is often utilized to indicate its intermediating property between microwave and optical technologies. In the literature, quasioptical techniques are also referred to as beam waveguide techniques. Quasioptical techniques have been utilized in the MARCHAL limb-sounding unit (as shown in Fig. 18) and the receiver front end of the Herschel satellite.

Generally speaking, the fabrication of reflector antennas is not a concern. However, with the increase of operation frequency, the surface accuracy is a critical factor that affects the performance of an antenna working in the THz range. To ensure good performance, the surface accuracy in terms of root measure square (RMS) shall be better than $\lambda/20$ (18°), best $\lambda/100$ (3.6°). But such high accuracy is a challenge for reflector fabrication. Talking about 1THz, the corresponding wavelength is on the order of 300 μm; the requirement of $\lambda/20$ and $\lambda/100$ surface RMS accuracy would lead to 15 and 3 μm, respectively. For medium-sized systems, for example, the fabrication of quasioptical systems is less difficult. A high-precision numerical controlled machine can provide surface accuracy better than 5 μm. Two shaped reflectors of 300 mm in diameter have been machined in Thomas Keating (TK), UK, for an antenna measurement system in Beijing University of Posts and Telecommunications, with a surface accuracy better than 3 μm (Yu et al. 2013). However, for electrically large systems, say 3.5 m diameter reflectors, the accuracy is a critical issue to be considered.

In the fabrication of the Herschel telescope, a complicated procedure has been customized. The reflector of Herschel telescope is fabricated using silicon carbide. Due to the fact that the main dish is too large, the main dish was constructed by piecing up 12 separate petals (Tauber et al. 2005). It has to be mentioned that this is the first monolithic space mirror fabricated in a segmented fashion. The fabrication of such a big reflector is very complicated, involving SiC manufacturing (supplied by Boostec Industries, Tarbes, France), polishing of the main and secondary dishes (by Opteon, Finland, for the main dish and Zeiss, Germany, for the secondary dish), coating (Calar Alto observatory, Spain), and optical testing (Center Spatiale de Liege, Belgium) (Toulemont et al. 2004). Initially, it is intended to build the reflector as an all-composite (CFRP) telescope, which is however canceled due to funding issues. The final performance was eventually verified by measuring the wavefront error, which was measured to

DOI 10.1007/978-981-4560-75-7_47-1

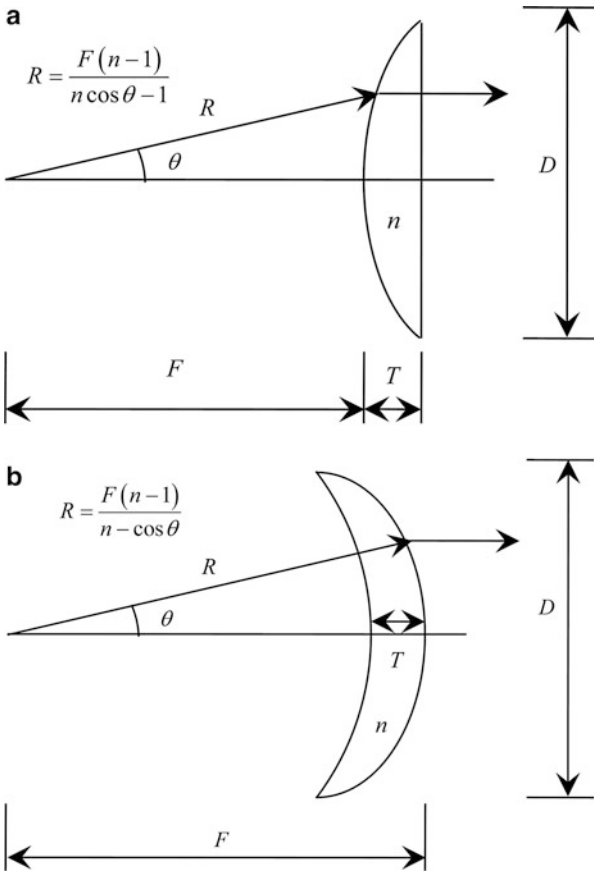


Fig. 19 Lens-based antennas. (**a**) Convex-planar structure; (**b**) concave-convex structure

be 5.5 μm at the edge of the reflector, better than the requirement, i.e., 6 μm (Dominic et al. 2009). Surface accuracy measurement seems too expensive, which however is possible using photogrammetry and interferometry techniques. These optical methods can be used in association with simulation and measurement techniques to predict in-orbit RF performance of electrically large antennas (Luis et al. 2010).

The Planck telescope is in aplanatic off-axis configuration, with both of the two reflectors made of sandwiched thermally stable carbon fiber-reinforced plastic (CFRP). To increase the reflectivity of the reflector, the front faces of both reflectors are coated with Al + Plasil. Such a fabrication technique leads to low coefficient of thermal expansion and high stiffness as well as low gross weight, 25 kg for the primary and 12 kg for the secondary. It is noted that in a reflector made of aluminum with diameter 1 m, the gross weight can be as high as 300 kg (Tauber et al. 2005; Toulemont et al 2004).

Other Antennas

The lens antennas have been widely utilized in the millimeter wave range. Lens antenna provides circular symmetrical radiation pattern due to its circular symmetrical structure. Particularly, this type of antenna is capable of generating Gaussian beam pattern and for this reason is frequently employed in a QO system. Two types of lens antenna are illustrated in Fig. 19. A lens antenna is usually made of polymer or silicon, correcting a spherical wave front to a planar one, at the same time calibrating the path length, i.e., ensuring equal phase distribution on the output plane.

In the THz range, a lens antenna is used in combination with photoconductive antenna. Or it can be used along in millimeter wave wireless communication system. One reported laboratory system working

DOI 10.1007/978-981-4560-75-7_47-1

at 300 GHz uses two polyethylene lens antennas of 5 cm in diameter, 12 cm focal length to compensate free-space path loss (Jastrow et al. 2008). Unavoidably, the fabrication error of a lens antenna produces uncertainty to the performance of its far field (Uvarov et al. 2007). In most cases, the side lobe level increases and the gain decreases as the surface accuracy decreases. For polymers, the machining of the curved surface with high precision is a challenge work. Also, temperature fluctuation makes the radiation pattern easily distorted. In addition, the reflection on the air-polymer surface decreases the efficiency depending on the refractive index.

With the development of nanotechnology, nanostructure has been investigated for THz wave radiation enhancement in photoconductive antennas (Gao et al. 2009; Park et al. 2012; Jornet and Akyildiz 2013). In most cases, THz radiation from nanostructure is attributed to surface Plasmon resonance (SPR). Strictly speaking, the THz radiation enhancement is due to field enhancement in nanostructure. It has been reported that by aid of nanostructure, the THz emission power can be doubled compared to a conventional photoconductive antenna (Jornet and Akyildiz 2013).

Also, metamaterials are also utilized for THz antennas (Zhang et al. 2014; Cong et al. 2012; Kyoung et al. 2011). Such structures can be designed for many uses such as polarization conversion, enhancement of radiation gain, and filtering. Due to the limitation of the space, this kind of antenna will not be covered in this chapter. Actually, there are other types of THz antennas not covered in this chapter for the same reason.

THz Antenna Measurement

Photoconductive Antenna Method

The measurement of a photoconductive antenna is usually conducted using the photoconductive antenna (PCA) scheme, as shown in Fig. 20.

The laser beam undergoing time delay will arrive at the photoconductor at Δt relative to the THz wave acting as the applied voltage, Fig. 20. In some THz-TDS systems, the induced current proportional to the magnitude of the THz wave is measured by a galvanometer. In this way, the THz wave is mapped out through delaying the reference laser by Δt relative to the THz wave, as illustrated in Fig. 21. To obtain the spectral information such as absorption fingerprint lines of biomolecules, Fourier transform can be applied. A frequently used scheme for measuring the current signal is electro-optical (EO) sampling by using an EO crystal, ZnTe for instance. EO crystal normally has different refractive index along different axes, which is referred to as birefringence phenomenon. In the EO sampling scheme, the THz field

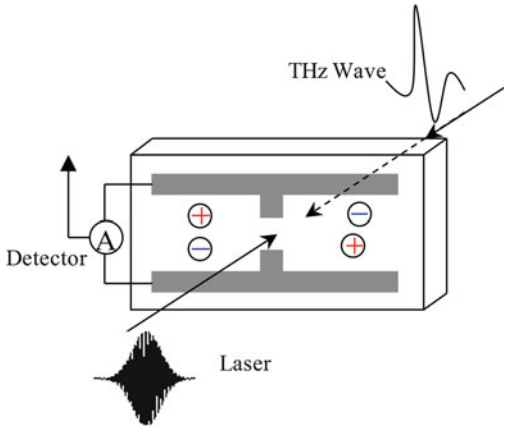


Fig. 20 Photoconductive antenna for THz wave detection

DOI 10.1007/978-981-4560-75-7_47-1

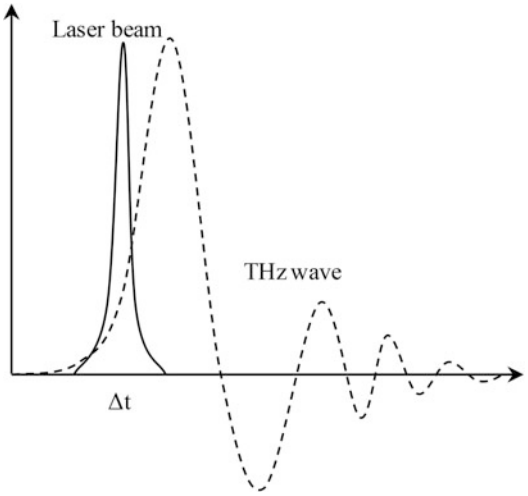


Fig. 21 The detection of the THz wave. The laser beam is at Δt relative to the THz wave. By changing Δt, the THz pulse can be detected

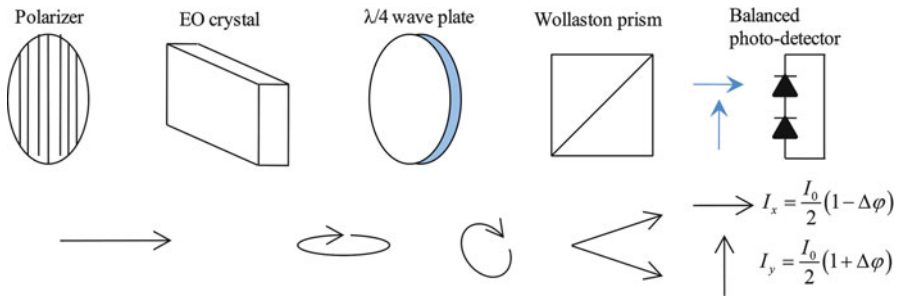


Fig. 22 Balanced detection method using EO crystal

changes the birefringence of the EO crystal. A linear polarized THz wave will be converted to a slight ellipsoidal polarized one. After going through the quarter-wave plate and a Wollaston prism, the THz wave will be eventually separated to a horizontally and a vertically polarized component, which can be detected using a balanced photodetector. The EO crystal detection scheme is illustrated in Fig. 22.

The radiation pattern measurement of a photoconductive antenna resembles a near-field imaging system. In general, the images are obtained via raster scanning. Many methods have been developed to achieve high-resolution imaging; the interested readers are referred to (Adam 2011). Those methods aim at achieving subwavelength resolution. One commercial product worth mentioning is a freely positionable photoconductive probe tip manufactured by the group of Nagel in Germany. The gap between the two electrodes is only 1.5 μm, achieving a resolution of 6 μm for the bandwidth of 0.1–3THz. Such a probe in theory can also be utilized for far-field measurement (Wächter et al. 2009).

Far-Field Method

Ideally, for antenna measurement, the incident wave to the antenna under test (AUT) shall be a plane wave with uniform amplitude and phase. An approximate approach is placing the AUT far away from the radiating source. Normally, it is in the far field of the test and source antennas. In the far field of an antenna, the radiation wave has a spherical phase front. To ensure accurate antenna measurement, a rule of thumb

DOI 10.1007/978-981-4560-75-7_47-1

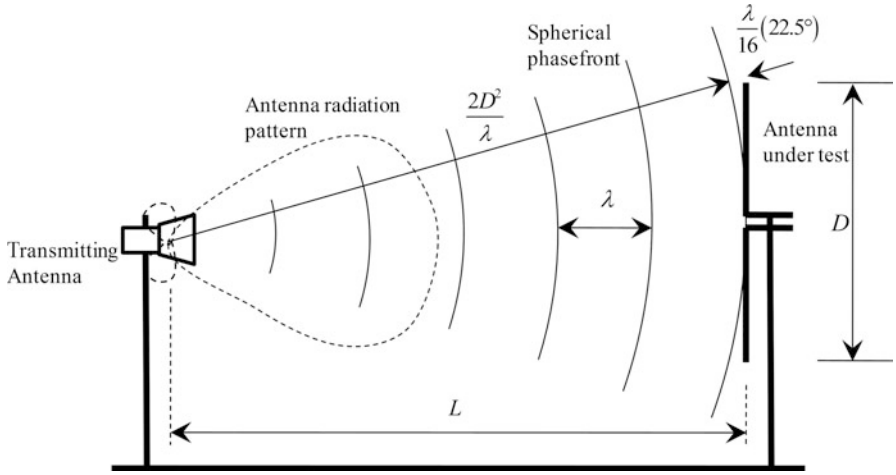


Fig. 23 Illustration of far-field measurement. The test antenna has to be far enough from the source antenna to ensure that the phase difference of any part of the test antenna shall be less than 22.5°

is that the phase difference of the phase front to a planar surface of diameter D is less than 22.5°, corresponding to λ/16 path difference. As shown in Fig. 23, the separation distance between the source and test antennas is L. The phase difference between the edge of the planar surface and the midpoint is

$$\Delta L = \sqrt{L^2 + \left(\frac{D}{2}\right)^2} - L \leq \frac{\lambda}{16}. \tag{19}$$

If L is much larger than D, the far-field condition becomes

$$L \geq \frac{2D^2}{\lambda}. \tag{20}$$

This is the minimal distance for far-field antenna measurement. However, it is a problem for most THz antennas since the distance may be as long as several tens of kilometers for electrically large antennas. For instance, a 1.5 m diameter antenna working at 0.3 THz requires L to be larger than 4.5 km.

One has to conduct this kind of far-field measurement outdoors. However, it is seen from Fig. 2 that the atmospheric attenuation beyond 0.3 THz is almost more than 10 dB/km, which imposes more than 45 dB absorption loss if the path length is 4.5 km. The atmospheric attenuation is even more pronounced in the range of 1–10 THz, more than 100 dB/km. For this reason, the far-field method is not feasible for electrically large antenna measurement.

The far-field method is only feasible for the measurement of small-aperture THz antennas, such as corrugated horns. A corrugated horn with center frequency 275 GHz has been measured using the far field in Queen Mary University of London with the measured results shown in Fig. 24. The aperture of the corrugated horn is around 8 mm in diameter. It is seen from the results that the far field of the corrugated horn has very good Gaussian profile, down to −40 dB, with high symmetry of the two cuts. The side lobe level of this corrugated horn is down to −36 dB, showing excellent Gaussian purity.

Near-Field Method

The near-field method is another technique frequently used for the measurement of millimeter antennas. The near-field method was introduced in the 1980s. In such a method, the near-field distribution over a

DOI 10.1007/978-981-4560-75-7_47-1

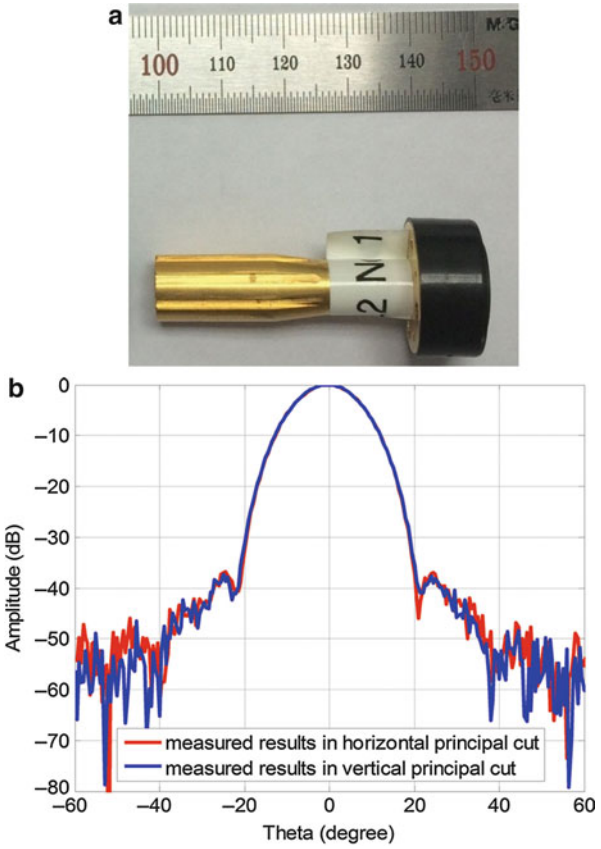


Fig. 24 Measurement of a 275 GHz corrugated horn using far-field method. (**a**) Photograph, (**b**) The radiation pattern

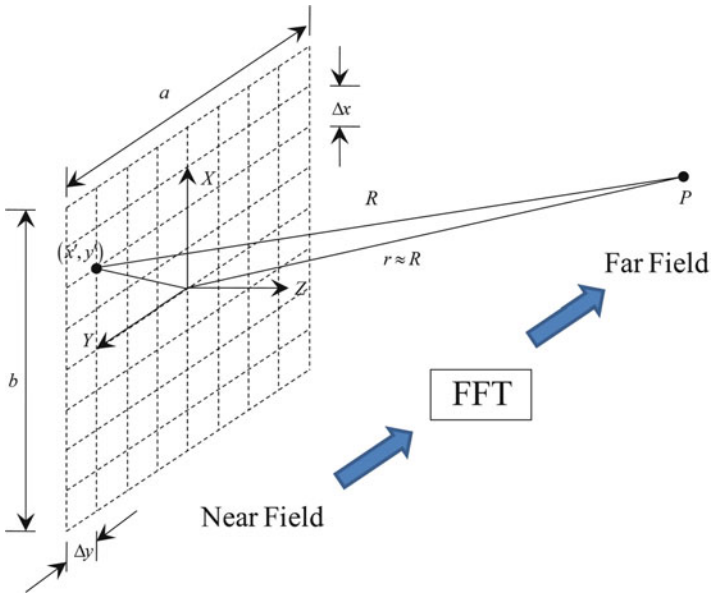


Fig. 25 The near-field scanning schemes

DOI 10.1007/978-981-4560-75-7_47-1

planar, cylindrical, or spherical surface is firstly picked up by using a probe. Then, the near-field-to-far-field (NF/FF) transform method is applied to obtain far-field radiation characteristics.

Generally speaking, planar scanning is suited for electrically large and high-gain reflector antennas. Data acquisition over such a planar scanning is usually conducted by raster scanning, as shown in Fig. 25. The maximal scanning interval for Δx and Δy is half a wavelength, which is dedicated by Niquist sampling rate.

Near-field transform to far-field transform is essentially Fourier transform. For planar scanning scheme, far-zone fields can be expressed as

$$\vec{E}\ (x,y,z) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\vec{f}\ (k_x,k_y)e^{-j\vec{k}\cdot\vec{r}}dk_xdk_y, \tag{21}$$

where

$$\begin{cases}\vec{f}\ (k_x,k_y) = f_x(k_x,k_y)\hat{x} + f_y(k_x,k_y)\hat{y} + f_z(k_x,k_y)\hat{z}\\ \vec{k} = k_x\hat{x} + k_y\hat{y} + k_z\hat{z}\\ \vec{r} = x\hat{x} + y\hat{y} + z\hat{z}\end{cases}. \tag{22}$$

The spectral components f_x and f_y are related to the near field by

$$\begin{cases} f_x(k_x,k_y) = \int_{-b/2}^{b/2}\int_{-a/2}^{a/2} E_x(x',y',z'=0)e^{j(k_xx+k_yy)}dx'dy' \\ f_y(k_x,k_y) = \int_{-b/2}^{b/2}\int_{-a/2}^{a/2} E_y(x',y',z'=0)e^{j(k_xx+k_yy)}dx'dy' \end{cases}, \tag{23}$$

where E_x and E_y are electric components along x and y axes. The far field can be approximated by using asymptotic evaluation to be

$$\begin{cases} E_\theta(r,\theta,\phi) \approx j\frac{ke^{-jkr}}{2\pi r}\cos\theta\left(f_x\cos\phi + f_y\sin\phi\right) \\ E_\phi(r,\theta,\phi) \approx j\frac{ke^{-jkr}}{2\pi r}\cos\theta\left(-f_x\sin\phi + f_y\cos\phi\right) \end{cases}. \tag{24}$$

Near-field scanning method has been employed to measure antennas of many satellite payloads or ground-based observation stations. Up to 950 GHz system has been measured by using near-field scanning. For instance, JPL Earth Observing System Microwave Limb Sounder working at 650–660 GHz (Slater et al. 2001) was measured using near-field method. And NSI's most recent submillimeter wave scanner was built in support of the Atacama Large Millimeter Array (ALMA) program at 950 GHz (Van Rensburg and Hindman 2008). One example of the near field of a corrugated horn working in the W band is shown in Fig. 26.

Although near-field method is successful for THz antenna radiation pattern measurement, the limitation of this method is also obvious in several aspects. The first aspect is the data acquisition time for electrically large antennas. By the Niquist sampling theorem, the maximal sampling interval shall be half a wavelength. For an aperture antenna, the scanning area would be slightly larger than the antenna aperture

DOI 10.1007/978-981-4560-75-7_47-1

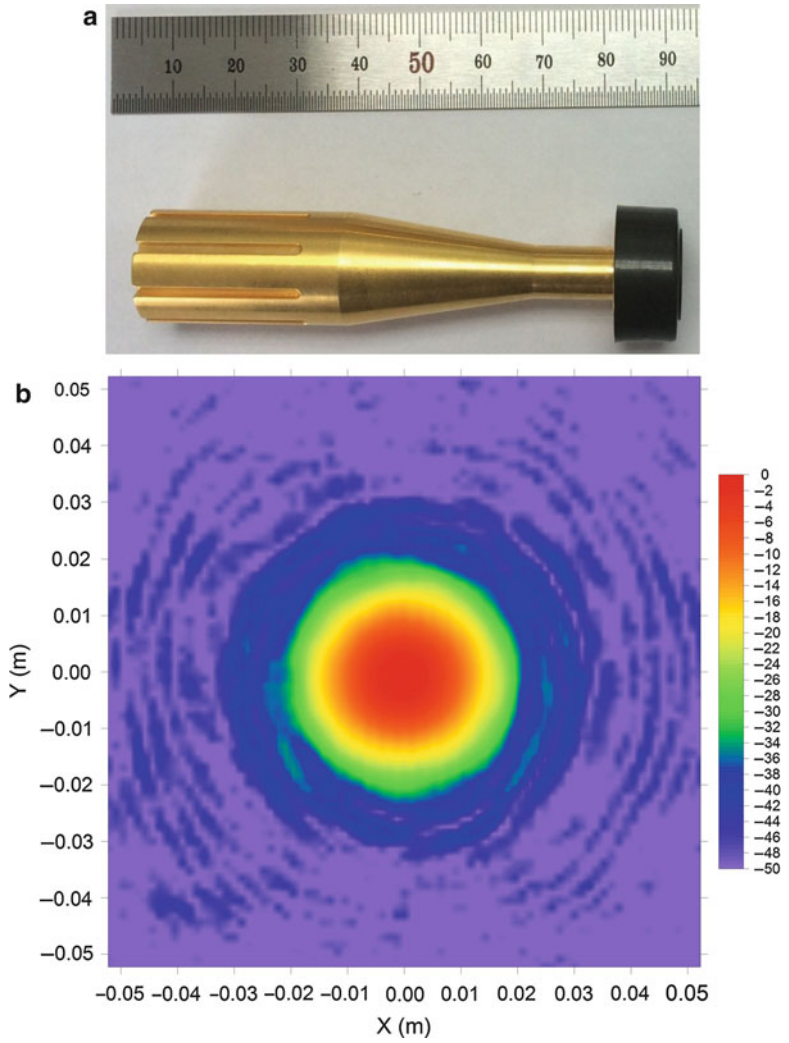


Fig. 26 Near-field measurement of a W-band corrugated horn using a NSI near-field scanning system in Queen Mary University of London. (**a**) Photograph, (**b**) The amplitude of the near-field

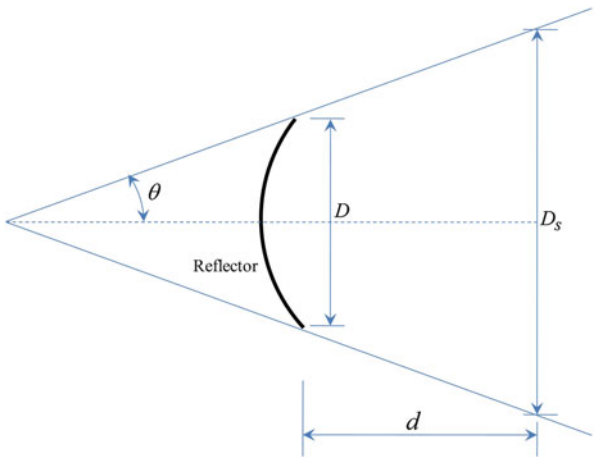


Fig. 27 The required measurement diameter in a near-field scanning system

DOI 10.1007/978-981-4560-75-7_47-1

depending on the far-field angle θ to be measured, as shown in Fig. 27. Suppose the diameter of the reflector is D, the distance between the probe and the aperture of the antenna is d, then the scanning area has a diameter larger than

$$D_s = D + 2d \tan\theta. \tag{25}$$

The number of scanning points N is calculated by

$$N = \left[\frac{2D_s}{\lambda}\right]^2, \tag{26}$$

where the bracket denotes the integer part of a number. For instance, for a 1 m diameter antenna working at 1 THz and 30° far-field angle, D_s is 2.15 m if d is 1 m. Even if the far-field angle is 10°, D_s will be 1.35 m. Then, the number of scanning points will be on the order of 10^8. Such a massive data collection needs tens or even hundreds of hours scanning. To reduce the data acquisition time, nonredundant data recovery techniques have been developed. Such a method can effectively reduce the sampling points by using postsignal processing techniques.

The second aspect is the planarity of the scanning structure. Working up to 1 THz, the wavelength will be on the order of 300 μm. The reported RMS planarity by NSI is 3 μm, corresponding to λ/100 or 3.6° phase error. Mechanical planarity of such is almost to its extreme. The reported working frequency of the NSI system is 950 GHz. Systems operating at higher will require better planarity, imposing a great challenge to mechanical system fabrication and installation.

The third aspect is the stability of the mechanical structure and the electronic systems. Inevitably, the mechanical structure is vibrating with the moving of the scanning arms. In addition, temperature variation causes thermal effect on the mechanical structure, which also causes stability issue to long-time scanning measurement. Electronic systems themselves are often noise sources. The signal stability over the whole period of measurement is also a source of error.

A fourth aspect is RF cable phase instability. This aspect may be one of the most concerned effects in the scanning method. With the movement of the scanning, the flexible cable causes extra phase fluctuation, which is difficult to overcome. Many techniques have been developed such as differential phase measurement method (Tuovinen et al. 1991) and pilot signal-based real-real time method (Saily et al. 2003).

The fifth aspect is probe translation during polarization rotation. Also, probe compensation has to be considered in the postprocessing stage (Leach et al. 1974). To overcome the phase issue, phaseless techniques have been proposed for near-field methods to recover the far field (Isernia et al. 1991; McCormack et al. 1990).

Compact Antenna Test Range

Compact antenna test range is the most feasible technique for THz antenna measurement. The reasons are obvious: its indoor nature making all-weather measurement in controlled environment possible; its system stability being better than near-field scanning; and its measurement time being relatively short.

The diagram of the working principle of a CATR system is demonstrated in Fig. 28. The incoming plane wave is generated through focusing elements such as lens and mirrors within a short distance. The plane wave is actually not a perfect one in a practical CATR but a regional pseudoplane wave where the antenna under test (AUT) is placed. The far field of the AUT can be measured by rotating the turntable by using a position controller.

DOI 10.1007/978-981-4560-75-7_47-1

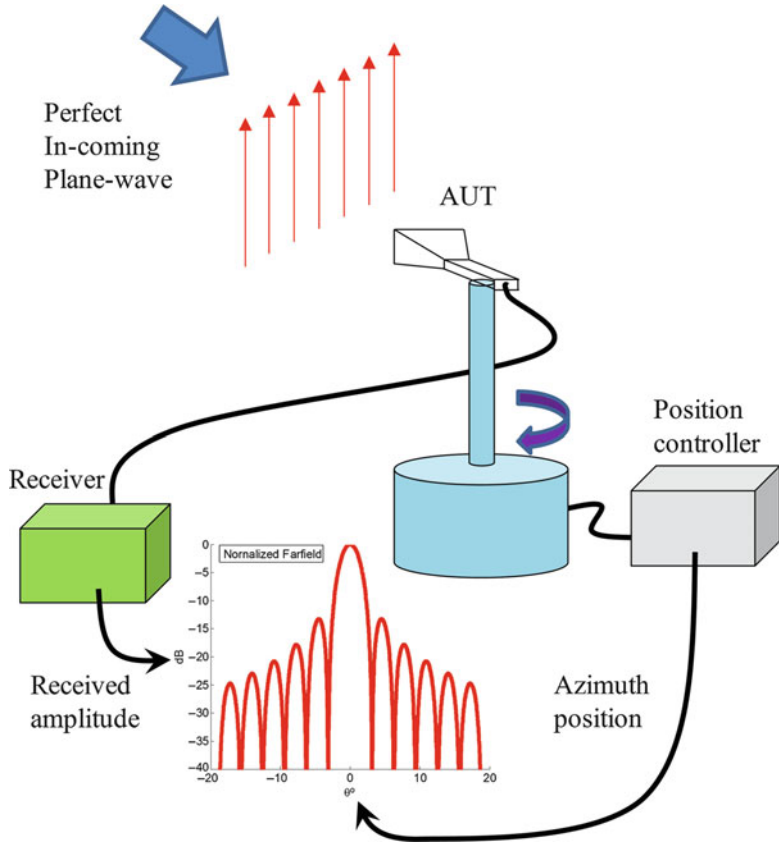


Fig. 28 The operating principle of a CATR system

The simplest CATR is a parabolic reflector antenna, see Fig. 29a, giving a 30 % quiet zone size of the dimension of the reflector. The quiet zone is the region where an EM wave can be approximated as a plane wave, with normally ±0.5 dB amplitude ripple and ±5°phase ripple. Generally speaking, ±0.5 dB amplitude ripple and ±5°phase ripple are the industrial convention, which is not an international standard. If the edge of the reflector is treated with serrated edge, the quiet zone can be increased to 50 %. Such a single-reflector scheme has disadvantage of low aperture usage ratio. To increase the aperture usage, a dual-reflector system has been developed in ASTRIUM, so called CPTR also, as shown in Fig. 29b. Such system can increase the aperture usage ratio to 60 % and even to 70 %. However, its drawback is that the two reflectors have to be of the same order of dimension.

Recently, a novel trireflector system has been proposed and developed in Queen Mary University of London, with the working frequency up to 325 GHz. The configuration of such is illustrated in Fig. 30a, with simulated quiet zone performance shown in Fig. 30b, c for amplitude and phase ripple, respectively. An improved version is being developed in Beijing University of Posts and Telecommunications in order to push the working frequency to 550 GHz. The trireflector CATR consists of a spherical main reflector with two-shaped relatively small subreflectors. Such a configuration gives one freedom manipulating both amplitude and phase distribution in the quiet zone. This novel CATR can achieve aperture usage ratio up to 80 %, a significant increase compared to single-reflector CATR. In addition, the fabrication cost is dramatically decreased due to the increase of the aperture usage ratio since a much smaller dish is needed. Note well that the fabrication cost of a reflector is not a linear function of the area but might be an exponential relationship. Furthermore, spherical mirrors are frequently used in telescope systems with a relatively lost fabrication cost. For the two-shaped subreflector, the surface is not a conic function but a numerically defined one, i.e., defined by a set of discrete points. The coordinates of these points are

DOI 10.1007/978-981-4560-75-7_47-1

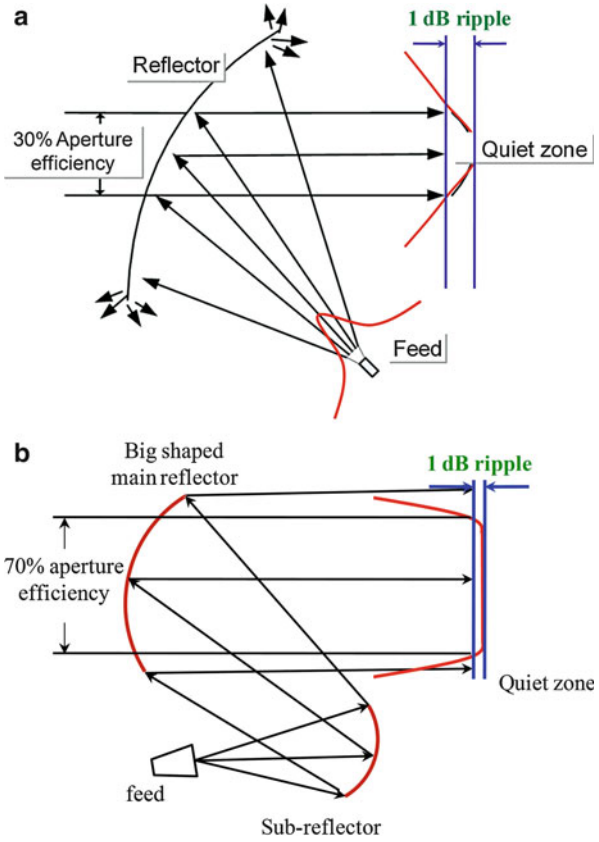


Fig. 29 Reflector based CATR, (**a**) Single reflector CATR; (**b**) dual-reflector CATR

synthesized by dynamic and kinematic ray tracing method (Kildal 1990) by gracefully incorporating both equal optical path condition and the law of conservation of energy to obtain uniform phase and amplitude distribution in the quiet zone, respectively. The size of the two subreflectors can be significantly smaller compared with the main dish, usually 10 % of the area of the main dish.

The configuration of trireflector CATR can be versatile, having at least four combination options, double-Cassegrain, Cassegrain-Gregorian, Gregorian-Cassegrain and double-Gregorian. The most useful two are Cassegrain-Gregorian and double-Gregorian configurations (Rieckmann et al. 2005). The other two configurations have the disadvantage of blockage from the first subreflector. Cassegrain-Gregorian normally produces smaller ripple in the quiet zone, while double-Gregorian has relatively better cross-polarization isolation. By adding diffraction stop in the caustic region, it is possible to reduce the amplitude of the ripple.

Ideally, a reflector CATR is expected to work over very wide bandwidth. However, two factors, i.e., electrical size and surface accuracy of dishes, determine the operation frequency range. The low frequency end is limited by the dish electrical size due to the edge diffraction while the high frequency end is determined by the dish surface accuracy for the reasons of scattering and phase error. The low-frequency limitation occurs when the size of the reflector is in the range of 25–30 wavelengths (Schluper et al. 1987). The dish surface accuracy is required to be better than 1/20 of the wavelength for high-frequency operation as a rule of thumb (Balanis 2005).

Another novel method in building CATRs for THz antenna measurement is a so-called hologram-based CATR (Hakli et al. 2005), as shown in Fig. 31. The hologram CATR is based on Fourier optics using a

DOI 10.1007/978-981-4560-75-7_47-1

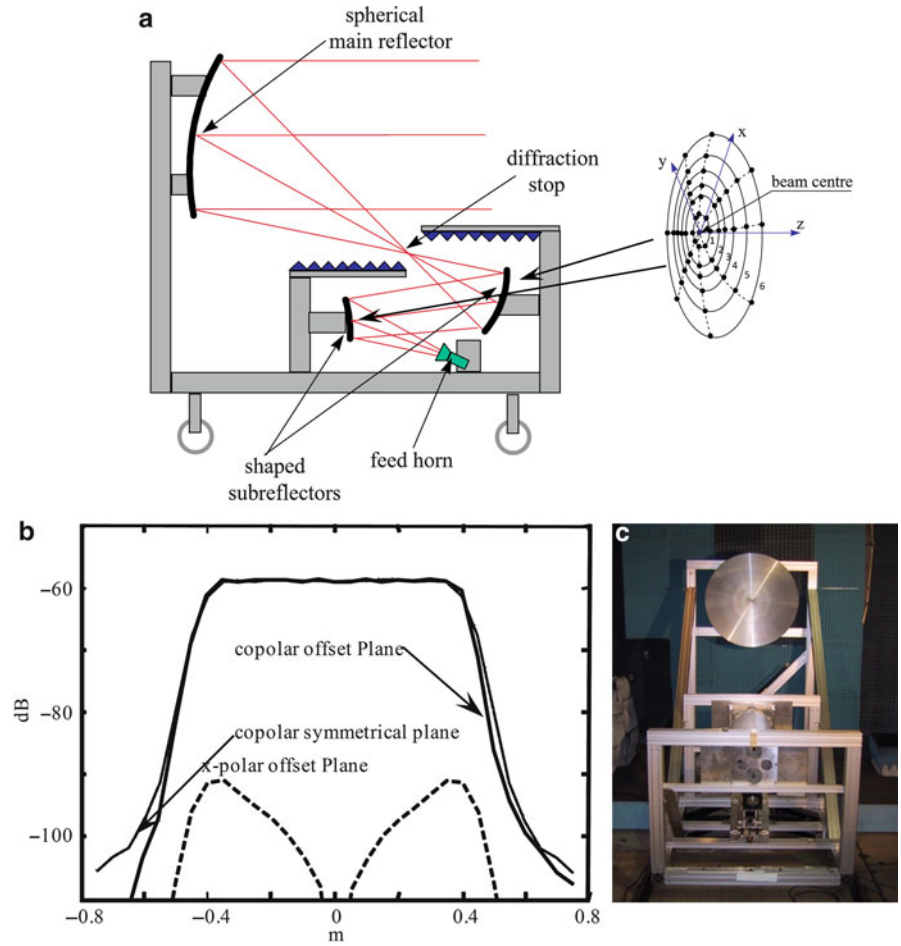


Fig. 30 Tri-reflector CATR. (**a**) configuration, (**b**) amplitude distribution in the quiet zone in the quiet zone at 200 GHz (Liu et al. 2011b), (**c**) Photograph of the tri-reflector CATR in Queen Mary University of London (Rieckmann et al. 2006)

hologram plane to transform spherical wave to plane wave, as illustrated in Fig. 31. The core technical part of a hologram CATR is the synthesis and fabrication of the hologram plane. It has been reported that a 650 GHz hologram-based CATR has been fabricated to measure a 1.5 m antenna operating at 650 GHz. The 3.16 m large hologram plate is pieced up by three plates to form a quiet zone of ±0.4 dB ripple in amplitude and ±6°ripple in phase (Karttunen et al. 2009). Unfortunately, such specifications were not achieved due to many factors such as the fabrication error, piecing up error, and so on. Such a structure, due to its nature, produces an asymmetrical quiet zone with low cross-polarization isolation (only −24 dB achieved). Also, the operating bandwidth of a hologram CATR is quite narrow. In addition, the chamber accommodating such a structure is relatively larger than a reflector-based CATR.

In the CATR measurement, one important procedure is the calibration of its quiet zone. Unless one uses near-field scanning to sample the whole area of the quiet zone, amplitude and phase ripple in the quiet zone is difficult to assess. But this again falls into near-field measurement. Alternative methods have been proposed (Griendt et al. 1996) such as using a standard antenna with much smaller electric size. Measuring such an electrical antenna is not an issue even using near-field measurement. But this method only gives one a ballpark estimation of the CATR, no details of the amplitude or phase ripples. One established method may evaluate the test zone (or quiet zone) of a CATR using a radar cross section (RCS) flat plate, which has been demonstrated at the compact payload test range at ESTEC, Noordwijk, the Netherlands.

DOI 10.1007/978-981-4560-75-7_47-1

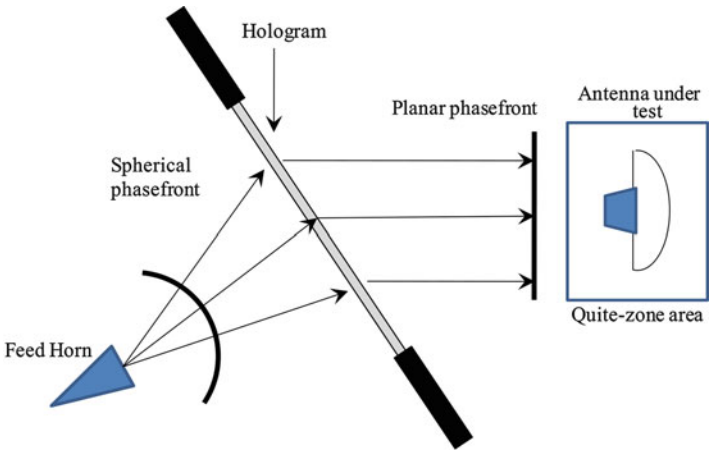


Fig. 31 Hologram based CATR

The underpinning principles are as follows. When a planar plate is illuminated by a nonplanar wave, the scattered field from this plate is a function of azimuth and elevation angle. The scattered field can therefore be represented by

$$E_s(\alpha,\beta) = \iint_A E_i(x,y)e^{2jk(x\sin\alpha + y\cos\alpha\sin\beta)}dxdy, \tag{27}$$

where $E_s(\alpha, \beta)$ is the scattered field, α and β are the azimuth and elevation angle, respectively, $E_i(x, y)$ is the incident angle, and A is the area of the flat plate. If one takes

$$\begin{cases} u = 2k\sin\alpha \\ v = 2k\cos\alpha\sin\beta \end{cases}. \tag{28}$$

then one has

$$E_s(\alpha,\beta) = \iint_A E_i(x,y)e^{j(xu+yv)}dxdy. \tag{29}$$

The inversion of the scattered field gives

$$E_i(x,y) = 4\pi^2\iint_{u,v} E_s(u,v)e^{-j(xu+yv)}dudv. \tag{30}$$

This method is particularly suitable for THz CATR evaluation since the scanning of such a large area is almost an impractical mission over a range of frequencies. In addition, electrically large plate only produces very large value of RCS in very narrow area of view of angle. It suffices to measure the RCS in a range of $(-20°, 20°)$. To further improve the measurement accuracy, the calibration of the RCS measurement could be implemented through single-reference, three-target calibration measurement technique (Wiesbeck and Kähny 1991).

DOI 10.1007/978-981-4560-75-7_47-1

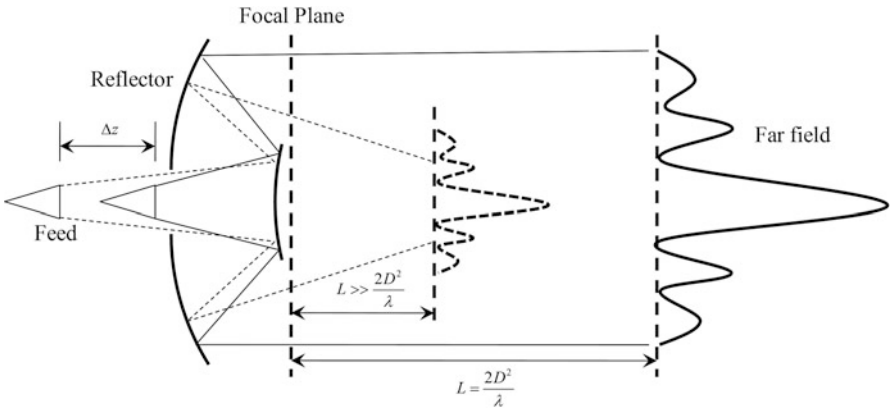


Fig. 32 Defocusing method

Other Methods

The far field of an antenna can be moved closer to its aperture plane by moving the feed out of the focal plane, as shown in Fig. 32 (Hohnsn et al. 1973). By using the thin-lens formula,

$$\frac{1}{d_2} = \frac{1}{F} - \frac{1}{d_1}, \tag{31}$$

where d_2 is the far-field distance, d_1 is the distance of the feed to the focal point, and F is the focal length of the antenna. If the feed is located at the focal point, then the far field will be infinite, while if the feed is displaced by a certain distance, the far field will be much closer to the antenna. Despite that this method enables one to measure the far field at the same distance of near field, the drawbacks are also apparent. For the THz systems, the movement of the feed may not be possible. In addition, it assumes that the defocusing of the feed affects the phase only. In fact, the amplitude may also need to be considered.

Actually, the measurement of antenna deployed on a THz system is a challenging task. To assess the performance of a THz antenna in the operating environment, one has to use an electrical and mechanical combined method. One example is the measurement of the antenna deployed on the Planck telescope. The test strategy employed is to measure the surface dimension of the antenna under the in-orbit condition using the optical methods photogrammetry and interferometry. Then, the measured dimension data are fed to GRASP for simulation at a relatively low frequency (320 GHz). The electrical measurement of the antenna is conducted using the near-field scanning at 320 GHz. Correlating the simulated and measured antenna performance at 320 GHz, the performance at even higher frequency is estimated using simulation alone. Though not ideal, this combined method has managed to predict the radiation pattern of the Planck telescope antenna in orbit.

There are several methods to measure the surface of a reflector antenna. A most straightforward method is using a three-coordinate gauging machine. Such a method is of course very time consuming for electrically large antenna and may have accuracy issues as well. The photogrammetry and interferometry techniques, however, provide an efficient and accurate method for electrically large reflector antenna measurement. In general, photogrammetry measures large-scale to medium-scale deformation, while interferometry measures small-small deformation. This kind of two-phase technique allows one to derivate the global shape as well as align the reflector. The cost of this method is also apparent. (1) Large arrays of cameras have to be employed. For instance, 250 camera stations were utilized in the Planck telescope verification; (2) Surface reconstruction has to be implemented based on a large set of data; (3) RF characterization using conventional antenna measurement techniques has to be conducted at low frequency; (4) Simulation using physical optical method is normally computational resource

DOI 10.1007/978-981-4560-75-7_47-1

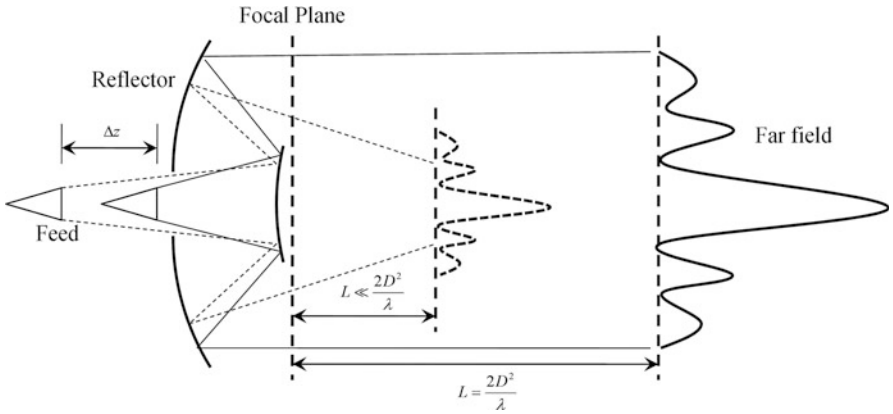


Fig. 33 The process of electrical–mechanical combined method

demanding. Nevertheless, before any other approaches are accessible to the THz RF verification of electrically large antenna, such a method after all provides a possibility. The procedures are illustrated in Fig. 33.

Summary

The antennas operating in the THz band and their associated measurements have been reviewed in this chapter with a focus on the basic concepts and operating principles.

In the THz range, the design of an antenna often has much more to consider than its lower-frequency counterpart such as sources and fabrication issues. The photoconductive antenna, though being called an antenna, is an integrated THz signal generator. Various methods have been explored in order to increase its directivity and the radiation power. Other high-gain antennas such as horn antennas and reflector antennas often have fabrication issues in the THz range due to its short wavelength.

Another challenge arises from the THz antenna measurement. The near-field and CATR measurement techniques are possible but not free of problems in the frequency range up to 1 THz. However, there is no feasible RF measurement method in the frequency range higher than 1 THz. For example, the 3.5 m diameter antenna on the Herschel telescope operating up to 5 THz wasn't measured electrically at all but was only simulated based on the surface shape measurement.

In the near future, the antennas and their measurement in the THz band remain as a technical challenge. It is hoped this challenge will be overcome with the development of high-power THz sources and advanced fabrication process in the long run.

Acknowledgment

The tri-reflector CATR presented in this Chapter is part of the work of the pre-research program (Funded by the Chinese Ministry of Industry and Information) led by Prof. Junsheng Yu with Beijing University of Posts and Telecommunications.

Cross-References

► Antennas in Radio Telescope Systems

DOI 10.1007/978-981-4560-75-7_47-1

▶ Mm-Wave Sub-mm-Wave Antenna Measurement
▶ Near-Field Antenna Measurement Techniques
▶ Radiometer Antennas
▶ Reflector Antennas
▶ Space Antennas Including Terahertz Antennas

References

Adam AJL (2011) Review of near-field terahertz measurement methods and their applications-how to achieve sub-wavelength resolution at THz frequencies. J Infrared Milli Terahz Waves 32:976–1019

Adel PA, Wylde RJ, Zhang J (2009) Ultra-Gaussian horns for CLOVER – a B-mode CMB experiment. In: The proceedings of 20th international symposium on space terahertz technology, J Charlottesville, 20–22 Apr, vol 1, Charlottesville, Virginia, USA, pp 128–137

Alderman B, Matheson D, Ellison BN (2008) Milimetre-wave technology at the Rutherford Appleton Laboratory. In: The proceeding of the China-UK/Europe workshop on millimeter wares and terahertz technologies, vol 1, 20th–22th October, 2008, Chengdu, China, pp 3–10

Appel-Hansen J (1979) Accurate determination of gain and radiation patterns by radar cross section measurements. IEEE Trans Antennas Propag 27:640–646

Auston DH, Smith PR (1983) Generation and detection of millimeter waves by picosecond photoconductivity. Appl Phys Lett 43:631–633

Balanis CA (2005) Antenna theory: analysis and design, 3rd edn. Wiley, New York

Bareiss M, Tiwari BN, Hochmeister A, Jegert G, Zschieschang U, Klauk H, Fabel B, Scarpa G, Koblmuller G, Bernstein GH, Porod W, Lugli P (2011) Nano antenna array for terahertz detection. IEEE Trans Microwave Theory Tech 59:2751–2757

Blain AW, Smail I, Ivison RJ, Kneib JP, Frayer DT (2002) Submillimetre galaxies. J Phys Rep 369:111–176

Cai Y, Brener I, Lopata J, Wynn J, Pfeiffer L, Federici J (1997) Design and performance of singular electric field terahertz photoconducting antennas. Appl Phys Lett 71:2076–2078

Cho GC, Rensselaer P, Troy, Han PY, Zhang XC (1999) Time-resolved THz phonon spectroscopy in semiconductors. In: The proceedings of lasers and electro-optics, 1999.CLEO/Pacific Rim '99. The Pacific Rim conference on August 30 - September 3, 1999, vol 3, Renaissance Seoul Hotel, Korea, pp 789–790

Cong LQ, Cao W, Tian Z, GuJ Q, Han JG, Zhang WL (2012) Manipulating polarization states of terahertz radiation using metamaterials. New J Phys 14:115013

Del Mar Photonics http://www.dmphotonics.com. Accessed Sept 2014

Dominic D, Göran P, Jan T (2009) The Herschel and Planck space telescopes. Proc IEEE 97:1403–1411

Forma G, Dubruel D, Marti CJ, Paquay M, Crone G, Tauber J, Sandri M, Villa F, Ristorcelli I (2009) Radiation-pattern measurements and predictions of the PLANCK RF qualification model. IEEE Antennas Propag Mag 51:213–219

Gao YH, Chen M-K, Yang C-E, Chang Y-C, Yin S, Hui RQ, Ruffin P, Brantley C, Edwards E, Luo C (2009) Analysis of terahertz generation via nanostructure enhanced plasmonic excitations. J Appl Phys 106:074302

Goldsmith PF (1997) Quasioptical systems: Gaussian beam quasioptical propogation and applications. Wiley, New Jersey

Gonzalez A, Uzawa Y (2012) Tolerance analysis of ALMA band 10 corrugated horns and optics. IEEE Trans Antennas Propag 60:3137–3145

Grade J, Haydon P, van der Weide D (2007) Electronic terahertz antennas and probes for spectroscopic detection and diagnostics. Proc IEEE 95:1583–1591

Griendt MAJ, Vokurka VJ, Reddy J, Lemanczyk J (1996) Evaluation of a CPTR using an RCS flat plate method. In: The proceedings of XVIII AMTA symposium, Seattle/Washington. Annual meeting and symposium- antenna measurement techniques association, vol 1, pp 329–334

Hakli J, Ala-Laurinaho J, Koskinen T, Lemanczyk J, Lonnqvist A, Mallat J, Raisanen AV, Saily J, Tuovinen J, Viikari V (2005) Sub-mm antenna tests in a hologram-based CATR. IEEE Antennas Propag Mag 47:237–240

Han K, Nguyen TK, Han H, Park I (2010) Yagi-Uda antennas for terahertz photomixer. In: The proceedings of: 2010 international workshop on antenna technology (iWAT), Lisbon, Portugal 1–3 March 2010

Henry SC, Zurk LM, Schecklman S (2013) Terahertz spectral imaging using correlation processing. IEEE Trans Terahertz Sci Technol 3:486–493

Hesler JL, Kerr AR, Grammer W, Wollack E (2007) Recommendations for waveguide interfaces to 1 THz. In: Karpov A (ed) Proceedings of the eighteenth international symposium on space terahertz technology, vol 1, 21–23 March 2007, Pasadena, California, USA, pp 1–7

Hirvonen T, Ala-Laurinaho JPS, Tuovinen J, Raisanen AV (1997) A compact antenna test range based on a hologram. IEEE Trans Antennas Propag 45:1270–1276

Hoffmann MC, Fülöp JAF (2011) Intense ultrashort terahertz pulses: generation and applications. J Phys D Appl Phys 44:083001

Hohnsn RC, Ecker HA, Hollis JS (1973) Determination of far-field antenna patterns from near-field measurements. Proc IEEE 61:1668–1694

IEEE standard P1785.1/D3 (2012) IEEE Approved Draft Standard for Rectangular Metallic Waveguides and Their Interfaces for Frequencies of 110 GHz and Above. Part 1: frequency bands and waveguide dimensions

Isernia T, Pierri R, Leone G (1991) New technique for estimation of far-field from near-zone phaseless data. Electron Lett 27:652–654

Jastrow C, Münter K, Piesiewicz R, Kürner T, Koch M, Kleine-Ostmann T (2008) 300 GHz transmission system. Electron Lett 44:213–215

Jornet JM, Akyildiz IF (2013) Graphene-based plasmonic nano-antenna for terahertz band communication in nanonetworks. IEEE J Sel Areas Commun 31(Suppl Part 2):685–694

Karttunen A, Ala-Laurinaho J, Vaaja M, Koskinen T, Häkli J, Lönnqvist A, Mallat J, Tamminen A, Viikari V, Räisänen AV (2009) Tests with a hologram-based CATR at 650 GHz. IEEE Trans Antennas Propag 57:711–720

Kildal PS (1990) Synthesis of multireflector antennas by kinematic and dynamic ray tracing. IEEE Trans Antennas Propag 38:1587–1599

Klein M (2000) Nadir sensitivity of passive millimeter and submillimeter wave channels to clear air temperature and water vapor variations. J Geophys Res 105:17481–17511

Kyoung JS, Seo MA, Koo SM, Park HR, Kim HS, Kim BJ, Kim HT, Park NK, Kim DS, Ahn KJ (2011) Active terahertz metamaterials: nano-slot antennas on VO2 thin films. Phys Status Solidi 8:1227–1230

Leach W, Joy EB, Paris DT (1974) Probe compensated near-field measurements: basic theory, numerical techniques accuracy, application of probe-compensated near-field measurements. In: The proceedings of antennas and propagation society international symposium, vol 12, June 10th–12th, 1974, Atlanta, Georgia, USA, pp 155–157

Lee Y-S (2009) Principles of THz science and technology. Springer, New York

Lee AWM, Williams BS, Kumar S, Hu Q, Reno JL (2006) Real-time imaging using a 4.3-THz quantum cascade laser and a 320 240 microbolometer focal-plane array. IEEE Photon Technol Lett 18:1415–1417

DOI 10.1007/978-981-4560-75-7_47-1

Li D, Huang Y (2006) Comparison of terahertz antennas. In: The proceedings of EuCAP 2006, first European conference on antennas and propagation, 6–10 November 2006, Nice, France

Liu HC, Luo H, Song CY, Wasilewski ZR, SpringThorpe AJ, Cao JC (2008) Terahertz quantum well photodetectors. IEEE J Sel Top Quantum Electron 14:374–377

Liu XM, Mai Y, Su HS, Li DH, Chen XD, Donnan R, Parini C, Liu SH, Yu JS (2011a) Design of tri-reflector Compact Antenna Test Range for millimetre/sub-millimetre wave and THz antenna measurement, Antenna Technology (iWAT). In: The proceeding of international workshop, vol 1, 7–9 March 2011, Hong Kong, China, pp 144–147

Liu X, Mai Y, Su H, Li D, Chen X, Donnan R, Parini C, Liu S, Yu J (2011b) Tri-reflector compact antenna test range design at high frequency. J Terahertz Sci Technol 4:1–6

Liu H, Yu J, Huggard P, Alderman B (2013) A multichannel THz detector using integrated bow-tie antennas. Int J Antennas Propag 2013:417108

Luis FR, Maurice HP, Robert JD, Jafar AP, Dominic D, Peter M (2010) Terahertz antenna technology and verification: Herschel and Planck-A review IEEE. Trans Microwave Theory Tech 58:2046–2063

Martin RJ, Martin DH (1996) Quasi-optical antennas for radiometric remote-sensing. Electron Commun Eng J 8:37–48

McCormack JE, Junkin G, Anderson AP (1990) Microwave metrology of reflector antennas from a single amplitude. IEE Proc Microwaves Antennas Propag 137:276–284

McKay JE, Robertson DA, Cruickshank PAS, Bolton DR, Hunter RI, Wylde R, Smith GM (2013) Compact wideband corrugated feedhorns with ultra-low side lobes for very high performance antennas and quasi-optical systems. IEEE Trans Antennas Propag 61:1714–1721

Mendis R, Sydlo C, Sigmund J (2005) Spectral characterization of broadband THz antennas by photoconductive mixing: toward optimal antenna design. IEEE Antennas Wirel Propag Lett 4:85–88

Murphy JA (1987) Distortion of a Simple Gaussian Beam on Reflection from off-axis Ellipsoidal Mirrors. International Journal of Infrared and Millimeter Waves 8:1165–1187

Nan R, Li D (2013) The five-hundred-meter aperture spherical radio telescope (FAST) project. IOP Conf Ser Mater Sci Eng 44:012022

Olver AD, Clarricoats PJB, Kishk AA, Shafai L (1994) Microwave Horns and Feeds. IEEE Press, New York

Paquay M, Marti CJ (2005) Pattern measurement demonstration of an untouchable antenna. In: The proceedings of annual meeting and symposium – antenna measurement techniques association, vol 1, 30 October – 4 November 2005, Newport, Rhode Island, USA, pp 20–25

Paquay M, Dubruel D, Forma G, Marti-Canales J, Wylde R, Rolo L (2007) Quasi optical instrumentation for the planck fm telescope rf alignment verification measurements at 320 GHz. In: The proceedings of annual meeting and symposium – antenna measurement techniques association, vol 1, 4–9 November 2007, St Louis, Missouri, pp 315–320(A07-0047)

Park S, Jin KH, Yi M, Ye JC, Ahn J, Jeong K (2012) Enhancement of terahertz pulse emission by optical nanoantenna. J ACS Nano 6:2026–2031

Rieckmann C, Rayner MR, Parini CG (1999) Optimisation of cross-polarisation performance for tri-reflector CATR with spherical main reflector. Electron Lett 35:1403–1404

Rieckmann C, Parini CG, Donnan RS, Dupuy J (2005) Experimental validation of the design performance for a spherical main mirror tri-reflector CATR operating at 90GHz. In: The proceedings of 28th ESA antenna workshop on space antenna systems and technologies, vol 1, 31 May–3 June 2005, WPP-247, pp 395–400

Rieckmann C, Dupuy J, Donnan RS, Parini CG, Moynar B, Oldfield M, Matherson D, de Maagt P (2006) Tri-reflector CATR antenna pattern verification of the MARSCHALS airborne millimetre-wave

limb-sounder at 300, 325 and 345 GHz. In: The proceedings of 4th ESA workshop on millimetre wave technology and applications, vol 1, 15–17 Feb 2006, Espoo, Finland, pp 443–448

Saily J, Eskelinen P, Raisanen AV (2003) Pilot signal-based real-time measurement and correction of phase errors caused by microwave cable flexing in planar near-field tests. IEEE Trans Antennas Propag 51:195–200

Schluper HF, Van Damme VJ, Vokurka VJ (1987) Optimized collimators-theoretical performance limits. In: AMTA proceedings, P.131, Seattle, Oct 1987

Schmuttenmaer CA (2004) Exploring dynamics in the far-infrared with terahertz spectroscopy. J Chem Rev 104:1759–1779

Slater DA, Stek P, Cofield R, Dengler R, Hardy J, Jarnot R, Swindlehurst R (2001) A large aperture 650 GHz near-field measurement system for the earth observing system microwave limb sounder. In: The proceedings of 23rd annual AMTA meeting and symposium, Denver

Smith PR, Auston DH, Nuss MC (1988) Subpicosecond photoconducting dipole antennas. IEEE J Quantum Electron 24:255–260

Sushko O, Dubrovka R, Donnan RS (2013) Terahertz spectral domain computational analysis of hydration shell of proteins with increasingly complex tertiary structure. J Phys Chem B 117:16486–16492

Tauber JA, Chambure D, Crone G, Daddato RJ, Martí CJ, Hills R, Banos T (2005) Optical design and testing of the Planck satellite. In: The proceedings of the XXVIIIth URSI general assembly in New Delhi, 23rd–29th, Oct, 2005, New Delhi, India, JB1.3:01148

Tonouchi M (2007) Cutting edge terahertz technology. Nat Photonics 1:97–105

Toulemont Y, Passvogel T, Pilbratt GL; Chambure D, Pierot D, Castel D (2004) A ø 3.5 m SiC telescope for Herschel Mission, "The 3.5-m all-SiC telescope for HERSCHEL". Proceedings of the SPIE 5487, optical, infrared, and millimeter space telescopes, 1119, 12 Oct, http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1317163

Tuoviizen J (1993) Methods for testing reflector antennas at THz frequencies. IEEE Antennas Propag Mag 35:7–12

Tuovinen J, Lehto TA, Raisanen A (1991) A new method of correcting phase errors caused by flexing of cables in antenna measurements. IEEE Trans Antennas Propag 39:859–861

Uvarov AV, Shitov SV, Uzawa Y, Vystavkin AN (2007) Tolerance analysis of THz-range integrated lens antennas. In: The proceedings of 2007 international symposium on antennas and propagation (ISAP 2007), Niigata, report POS2-3

Van Rensburg DJ, Hindman G (2008) An overview of near-field sub-millimeter wave antenna test applications. In: The proceedings of COMITE 2008. 14th conference on, 23–24 Apr 2008

Wächter M, Nagel M, Kurz H (2009) Tapered photoconductive terahertz field probe tip with subwavelength spatial resolution. Appl Phys Lett 95:041112

Wiesbeck W, Kähny D (1991) Single reference, three target calibration and error correction for monostatic polarimetric free space measurements. Proc IEEE 79:1551–1558

Yu JS, Liu XM, Yao Y, Yang C, Lu ZJ, Wylde R, Sebek G, Chen X, Parini C (2013) The design and manufacture of a high frequency CATR. In: The proceedings of millimeter waves and THz technology workshop (UCMMT), the 6th Europe/UK-China workshop, 2013, Rome, Italy

Zhang XC (2002) Terahertz wave imaging: horizons and hurdles. Phys Med Biol 47:3667–3677

Zhang XC, Xu J (2009) Introduction to THz wave photonics. Springer, New York

Zhang Q-L, Si L-M, Huang Y, Lv X, Zhu W (2014) Low-index-metamaterial for gain enhancement of planar terahertz antenna. AIP Adv 4:037103

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_48-3

Small Antennas

Kyohei Fujimoto[a]* and Zhinong Ying[b]
[a]Institute of Applied Physics, University of Tsukuba, Tsukuba, Japan
[b]Research and Technology, Sony Mobile Communication, Lund, Sweden

Abstract

This chapter first provides basic treatment of small antennas (SAs), beginning with overview; definition, giving four types; discussions of the limitation related with the size and Q or the bandwidth; and significance of SA, in which a short history is included. Then, principles and methods of making antennas small based on those principles are described for four types of SA. Representative examples of practical SAs are introduced for the four types individually by referencing recent technical journals. Finally, the future prospective of SA is discussed.

Keywords

Electrically small antenna (ESA); Principles and methods of creating small antenna (SA); Definition of SA; Limitation in SA; Use of slow-wave structure; Metamaterial applied in SA; Integrated and composite SA; Space effectively used by SA

Introduction

With emergence of various wireless systems, including mobile phones, broadband systems, high-rate data transmission systems, radio-frequency identification (RFID) systems, body-centric communication systems, and wearable systems, requirements for small antennas (SAs) are rather urgent, and many SAs have so far been newly developed and practically applied to terminals of wireless systems. Because most wireless systems use small terminals, antennas mounted on such terminals are inevitably SAs. Practically, many small terminals require ESA; however, SAs used so far are not only those having small dimensions but also some functions to implement a variety of performances, including wideband, multiband, reconfigurable operation and other sophisticated operations (FSA). Practical SAs with low-profile structure (PCSA) and having simply physically small dimensions without regarding the operating frequency (PSA) are also significant in small wireless terminals. Each of these four types of SA is appropriately designed and applied to related systems.

This chapter intends to provide engineers and researchers in the fields of antennas and communications with basic concepts, fundamental treatment, methods, and techniques to make antennas small based on the principles; multiple examples are introduced showing practical modern SAs, which can guide design and development of newly required SAs.

This chapter consists of five parts: (1) general principles of SA, (2) short history of SA, (3) principles and methods of producing SA, (4) practical examples of SA, and (5) future perspectives for SA.

Kyohei Fujimoto has retired

*Email: hqm11446@flute.ocn.ne.jp

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_48-3

Small Antenna Generalizations

Overview of Modern Small Antennas

Many recent small antennas (SAs) are not only types of electrically small antenna (ESA), having much smaller dimensions with respect to the operating wavelength, but also other types that can be classified based on functions and dimensions with different aspects than the wavelength. All of those SAs have been used practically in various wireless systems not only for communications but also radio control, radio identification, data and video transmission, sensors, and so forth, as the indispensable component in the systems. Among those wireless systems, the most popular and significant wireless applications are mobile systems that deal with voice, data, and image transmission, radio identification, and wearable systems. Typical voice transmission mobile systems are cellular phones, which have presently advanced to smartphones and data tablets for transmitting information, images (both still and moving), large-quantity data, NFC (near-field communication) systems, and wearable systems including body-centric communication and biomedical systems.

Those wireless systems require small antennas, since the size of their equipment is generally small and the space to install antennas is restricted to a limited narrow area. In addition, antennas for some mobile terminals not only should be small but should facilitate wideband or multiband capability, because the relentlessly increasing trend in such mobile terminals is to equip them with multiband and other systems such as GPS, WLAN (wireless local area network), Wi-Fi (Wireless Fidelity), etc. as well as telephony. Although multiple operation is required, antennas must still necessarily be small.

Furthermore, the wireless world should expect emergence of a new system in the near future, cognitive radio, by which effective and flexible use of radio spectrum in wide frequency bands will be made feasible (Dudley et al. 2014). This system will urge further needs of ultra-wideband small antennas.

Nevertheless, small antennas play very significant roles in the wireless systems for various situations, as they are always indispensable components.

Design and development of small antennas need special considerations, especially when the dimensions become smaller and designers encounter difficult situations to conquer trade-offs between practical requirements and the fundamental limitation of SA. What the fundamental limitation addresses is relations of antenna size with the bandwidth in transmitting antennas and, in a similar way, relation of antenna size with the noise factor or SNR (signal-to-noise ratio) as well as the bandwidth in receiving antennas.

The fundamental limitation of SA was first treated by H. A. Wheeler in 1947 (Wheeler 1947). The discussion on the limitation is still continuing up through today to obtain reasonable conclusions.

The essential issues to realize a small antenna are to overcome difficulties based on the limitation and create a small antenna which can exhibit performance approaching the limitation as closely as possible. Nevertheless, modern small antennas have shown significant progress. Various novel functional and intelligent antennas, yet with small dimensions, have been developed. Design techniques have also advanced along with emergence of new wireless systems, for which novel antennas are required.

Small antennas have so far been used in numerous wireless systems in various fields; however, the importance of them may be overlooked as many of them are invisible, since they are installed inside the small units of wireless equipment.

Definition of Small Antenna

The term electrically small antenna (ESA), which is defined in terms of the operating wavelength, generally has been used as regards SA. However, there are numerous other types of small antennas defined in different aspects from the wavelength, that is, with respect to function and physical dimensions. When an antenna is constructed to have enhanced performance, with the size kept unchanged, the antenna

DOI 10.1007/978-981-4560-75-7_48-3

is considered effectively small, since the performance exhibited by the antenna may be that produced by a larger antenna. This type of antenna is defined in regard to the function and named as functionally small antenna (FSA). Other types of antennas made in low-profile or thin structure, the thickness of which is much smaller than the wavelength, can be defined from a dimensional point of view as physically constrained small antenna (PCSA), as a part of the antenna is constrained to be small. Typical example of PCSA is MSA (microstrip antenna).

Another type is an antenna having physically small dimensions, but not necessarily small in relation to the wavelength, for example, a physically small antenna that can be put on a human palm, even though it could be electrically large. This type of antenna is defined as physically small antenna (PSA). There is no exact physical meaning in PSA as there is with other types of small antennas. A small millimeter-wave horn antenna is one of the PSAs when the operating frequency is in the SHF region and the physical size of the antenna aperture is a few cm, which is a few wavelengths and not electrically small.

The classification of small antenna is not unique for an antenna, because one antenna can be included in two or more types of category; for instance, a PCSA can be an ESA, if the whole structure of a low-profile antenna is much smaller than the operating wavelength. ESA is defined from another aspect by taking account of a dimensionally numerical value, that is, when the whole antenna structure is put into a sphere of the radius a which satisfies $ka \leq 0.5$ ($k = 2\pi/\lambda$) (Fujimoto and Morishita 2013; Volakis et al. 2010a). This sphere is the minimum-sized one that encloses the whole structure of an antenna and is referred to the "Chu sphere."

Another example is a small multiband mobile terminal antenna, which is classified into an FSA and can be at the same time an ESA, if the maximum size of the antenna structure can be embraced by a sphere with radius a and ka is smaller than 0.5.

There are many small antennas which are not only electrically small but also functional, defined as FSA and another type, PCSA, when the thickness is very thin.

Limitation in Small Antennas

When the size of an antenna is made smaller, the bandwidth tends to grow narrower with reduction of the size. In turn, antenna Q is increased as the antenna size is reduced. As Q becomes large, the bandwidth becomes inversely proportional to Q, implying that the reduction of the antenna size gives rise to increased Q and makes the bandwidth narrower.

Now, it should be noticed that increase in the antenna Q with the antenna size reduction is not limitless but bounded by some value. This is the limitation inherently observed in small antennas; that is, the antenna Q or the bandwidth cannot take excess value over the limit value that corresponds to the antenna size. In other words, for a small antenna, a greater bandwidth or a lower Q that violates the limit value depending on the size cannot be achieved.

Wheeler first discussed limitations of small antennas in 1947 and showed that Q of a small antenna for the most limiting case was $Q = 1/(ka)^3$ where k is the wave number $2\pi/\lambda$ and a is the radius of the minimum-sized sphere which encloses the whole structure of an antenna. This indicates Q is inversely proportional to the three-dimensional antenna size, and hence the antenna size imposes a fundamental limitation on the bandwidth B, as when $Q > 1$, the fractional bandwidth B is considered approximately inversely proportional to the Q.

Chu followed Wheeler's work (Chu 1947) and studied the limitation by using spherical vector wave functions, from which the minimum possible Q for an antenna enclosed in the Chu sphere along with the maximum G/Q (G, gain) was derived.

Limitation of the small antenna was studied by many researchers with considerations on the antenna configuration, current distributions, gain, impedance, exciting modes, and polarizations (Volakis et al. 2010b; Fujimoto and Morishita 2013). Some representative expressions are McLean's (McLean

DOI 10.1007/978-981-4560-75-7_48-3

1996) and Thal's (Thal 2006. 2009). McLean presented the limitation of Q for a Chu antenna (antenna enclosed by the Chu sphere) as

$$Q \leqq (1/M)\left[1/(ka)^3 + M/(ka)\right] \tag{1}$$

where M refers to the excitation mode (linear polarization). M = 2 means equal excitation of both TE and TM modes (either linear or circular polarization), while M = 1 indicates when only either TE or TM mode (linear polarization) is excited. When M = 2, the absolute lowest Q is obtained by an antenna of a given size. Thal gave the minimum Q as $1.5/(ka)^3$ for TM_{10} mode radiation, $3/(ka)^3$ for TE_{10} mode radiation, and $1/(ka)^3$ for both TE_{10} and TM_{10} mode radiation assuming surface current distributions on the Chu sphere surface.

Hansen has shown a new formula for Q of a Chu antenna (Hansen and Collin 2009) as

$$Q \leqq 3/2(ka)^3 + 1/\sqrt{2}\,(ka) \tag{2}$$

These results can be compared with Wheeler's limitation,

$$Q \leqq 1/(ka)^3 \tag{3}$$

When the antenna size becomes smaller, Eq. 1 approaches to Eq. 3 and Eq. 2 becomes Thal's results.

Nevertheless, as Wheeler's limitation is understood as the lowest achievable Q for an antenna of given size, design of ESA is a sort of challenge to realize antenna having Q or bandwidth which can approach the limitation as closely as possible.

Significance of Small Antenna

Small antennas are crucial devices and recognized as being indispensable for any wireless systems. This understanding continues up through today since the earliest days of radio communications when ESAs began to be used. Antennas used first in communication systems were ESA, although the antenna dimensions were physically large, as the operating frequency bands were LF and MF. The communication range was gradually extended along with the progress in the communication systems from telegraphy to voice transmission. This progress was attributed greatly to the advancement of the antenna technology. Dimensions of antennas had gradually reduced as the frequency allocation progressed to higher bands like HF and VHF, by which long-range communication, even covering worldwide areas, was realized. In those days, antennas practically applied were not only ESA but also other types of small antennas, and though small, their role in communication systems was important. Advancement in antenna technology contributed to progress in communication systems, and in turn communication systems inspired progress in the advancement of antennas.

Recent communication systems cover wide ranges of transmission for voice, image, data, and radio control, radio identification, sensors, and wideband signal transmissions, including both moving and still images, wearable systems, and so forth. Most of those systems have mobile applications, where equipment is generally small, and antennas to be installed are inevitably small and mostly built into the equipment. In any wireless system, a small antenna is naturally indispensable. There may be cases, where small antennas would determine the system performance, when the RF front end of the wireless system is designed optimally, and to further improve the performance as is required is almost impossible.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_48-3

The latest mobile systems widely used in public are smartphones, handheld tablets, and wireless broadband systems. Other mobiles are radio sensors, RFID (radio-frequency identification), and wearable systems, including body-centric communications. Types of antennas used for those systems can be any of ESA, FSA, PCSA, and PSA. There is growing need for mobile systems that operate multiband or wideband, as systems tend to include more than two subsystems like GPS, WLAN (wireless local area network), Wi-Fi (Wireless Fidelity), etc. that operate in different frequency bands. Thus, antennas for those systems are designed to perform multi- or wideband, corresponding to the systems, without significant change in the dimensions. Furthermore, some systems require antennas to have functions such as frequency or pattern reconfigurable operation, diversity, and so forth, keeping the antenna size increase as small as possible. In any case, importance of small antennas does not vary depending upon the system and will continuously grow with further prevalence and advancement of mobile systems.

The latest topic in the wireless world is expectation of practical cognitive radio in the near future, which will require small UWB (ultra-wideband) antennas. Antenna engineers who design small UWB antennas will overcome the fundamental limitation regarding antenna size related with bandwidth and efficiency/gain.

Each era has been lively with innovative antenna development along with progress in new social systems decade to decade. Obviously small antenna technologies will advance along with the progress in the wireless systems; in turn, progress of the small antenna technologies will inspire advancement of the wireless systems. Significance of small antennas will further be augmented generation to generation.

Short History of Small Antenna

Antenna history started with a small antenna in 1846, which was inaugurated by J. Henry, who experimented with spark-gap transmission and reception by using a small loop loaded with an induction coil (Fig. 1). Many pioneers like Edison, Hertz, and Lodge did experiments of radio-wave radiation and reception by using electrically small antennas (ESAs), although the dimensions were physically large, as the frequency bands used were LF and MF. Types of antennas used in those days were thin long wires, monopoles, biconical dipoles, loops, and so forth (Rumsey 1981). By replacement of the early

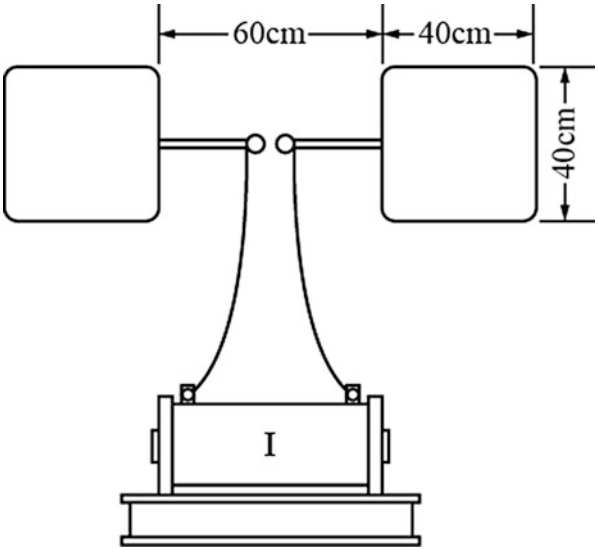


Fig. 1 Loop antenna used by Marconi for his first wave transmission

DOI 10.1007/978-981-4560-75-7_48-3

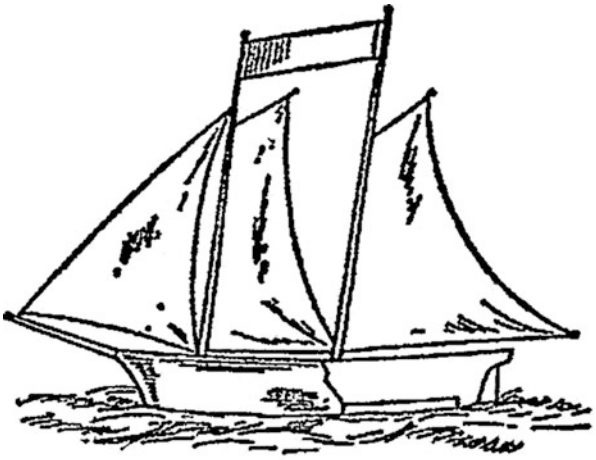


Fig. 2 Wire invented - L antenna used in ship

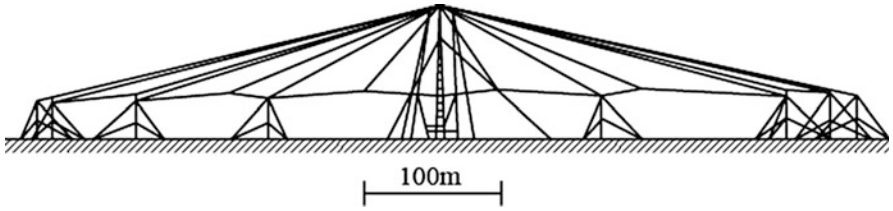


Fig. 3 Top- Loading technique applied umbrella type antenna

radiotelegraphy spark-gap transmitters, longer-distance communication was promoted, and various antenna technologies such as grounding of one antenna terminal to the earth, resonance by using LC circuits, and impedance matching were developed.

The world's first radio communication accomplished by Marconi in 1901 used also types of ESA, although they were physically quite large; the antenna at the England site was the fan type having dimensions of 48-m wide and 68-m high, but that is only several tenths of the operating wavelength about 3000 m (820 KHz). Since then, radio communications over the Atlantic Ocean spread to be used between ships and coasts in the early 1900s. Ship antennas were mostly thin long wires or wired inverted-L types (Fig. 2).

In 1906, the frequency spectrum for the radio communications was restricted to LF and MF bands for ships, long-distance land communications, and military use. Since then, many large, low-frequency antennas were developed. To shorten the long-wire antennas, top-loading technique was invented. Later, by applying the top-loading technique, a new flat-top antenna like an umbrella (Fig. 3) was developed. Since radiation efficiency of those antennas was very low, multiple tuning techniques were invented for improving the efficiency. Antenna technology had further advanced to achieve directional radiation by arraying antenna elements. The directional radiation was designed by using another technique, the use of a traveling wave on a long wire placed in parallel to the ground with low height. Other antenna types practically used in those days were (a) T shape, (b) inverted L, (c) harp shape, (d) rhombus, (e) flat top (Fig. 4), and so forth; all of those were ESA. In the late 1930s, slot antennas and ferrite antennas were put into practice.

With progress in generation, operating frequencies tended to be higher, and in the mid-twentieth century, HF was the major band, and aeronautical radio and marine radio were made feasible. Remarkable advancement of small antenna technology can be attributed to World War I (1914–1918) and World War II

DOI 10.1007/978-981-4560-75-7_48-3

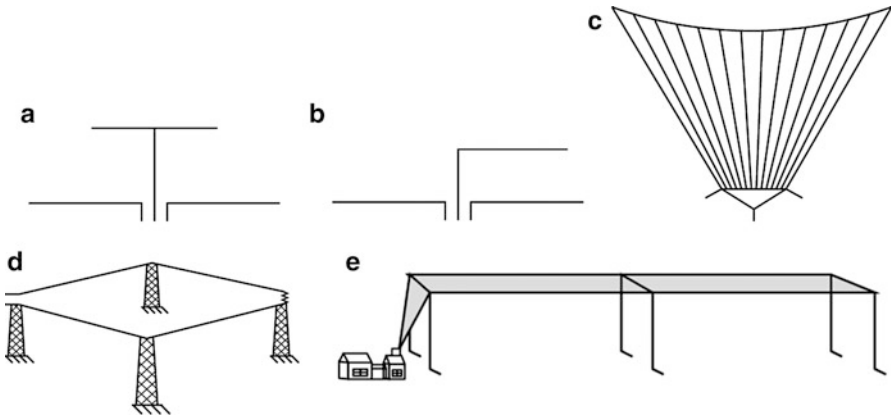


Fig. 4 Top- loading technique applied antennas: a. T-shape, b. inverted -L, c. harp -shape, d. Rhombus and e. flap-top

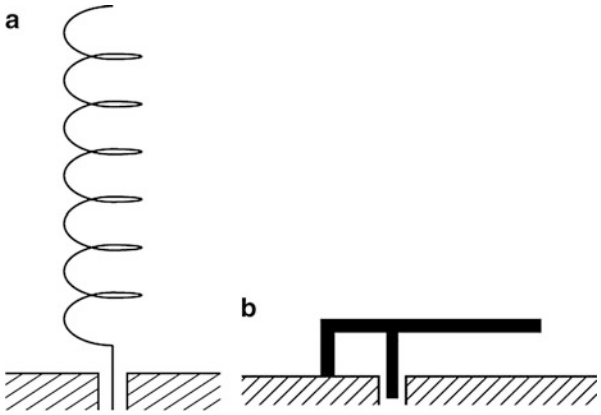


Fig. 5 a. Normal mode helical antenna (NMHA) and b. inverted -F antenna (IIFA)

(1939–1945) due to the needs for military communications for airplanes and airships, as well as for land systems. The type of antennas used in such communication systems during World War I was mainly a thin $\lambda/4$ wire monopole, which was installed on the equipment body (metal) that was considered as the ground. Significant progress seen in World War II was mobile systems, particularly for land mobiles used for military force and vehicles.

System advancement was accelerated by introduction of higher-frequency bands, VHF, and FM technology. After the war, application of mobile systems was further expanded to various new mobile systems for the public: telephones for personal use and on the airplane. Another notable subject in the 1940s was study of the fundamental limitation of small antennas, initiated by Wheeler in 1947. The discussion on the limitation has still been active in the present days.

In the 1950s, the UHF band was opened to the public and design concepts of antennas for mobile systems changed. The body of the small radio unit (metal) traditionally considered as the ground was changed to be treated as a part of the radiator, to which the ground was included. This result gave rise to improving the antenna performances. In the mid-1950s and later, spiral antennas were developed and applied to various wideband systems. By using spiral structures, design of a small antenna having wideband or UWB (ultra-wideband) characteristics has become feasible. Other antenna types used in those days were newly developed NMHA (normal-mode helical antenna), IFA (inverted-F antenna) (Fig. 5a, b), and various composite antennas, for instance, combination of a dipole and a loop.

DOI 10.1007/978-981-4560-75-7_48-3

In the 1960s, systems underwent changes through time with emergence of various new systems and services, which deployed according to the social needs; they were radar, space communication, TV broadcast, mobile systems, and so forth. Along with advancement of systems, antenna technologies progressed.

It should be noted that the design techniques of small antennas changed remarkably to treat an antenna with a comprehensive concept in which wireless system, propagation, and environmental conditions were included, and so the design concept of small antennas has advanced to treat an antenna as a "system." There were some occasions where the antenna performance, in particular with a small antenna, could finally dominate the system performance. Considerations of the environmental conditions, including wireless unit body and hardware nearby the antenna element within the unit, are vital in the design of a small antenna, because all of those factors affect system performance. When the system unit is a handset, influence of the unit operator's hand, head, and body affects antenna design as one of the crucial parameters.

The concept of integrated antennas was introduced in the late 1960s, which contributed to miniaturization of an antenna and to produce functional antennas. These are generally called IASs (integrated antenna systems) (Fujimoto 2005), since they work as a system, where their structures, being composed of passive or active devices, are dealt in total with systematic concepts. Among them, AIAS (active integrated antenna system) is devoted to small-sizing antennas, particularly in application to the matching. Very small antennas exhibit very large reactive components but very small resistive components in their input impedances, making proper matching difficult. An NIC (negative impedance converter) (Linvill 1953), which is one of the representative AIASs, is used for overcoming the matching problem in very small antennas. The NIC can be used at either the antenna load terminals or in the near field of a radiator. The latter is known as space matching (Stuart and Pilwerbetsky 2006; Billoti et al. 2000; Fujimoto and Morishita 2013), which is a recently introduced technology.

In the 1970s, public mobile phones called cellular phones were first introduced. That technology has progressed to the fourth-generation system from the first systems in the 1980s, and antenna systems used in mobile terminals (handsets) evolved from a wire monopole to planar types such as PIFA, MSA, and parallel plate, and they were at last entirely built into the equipment body. In some phones, functional performances such as diversity, frequency reconfigurable control, and pattern control were required and so facilitated. Those antennas were classified into the type of FSA.

In the 2000s, an increasing demand for access to information anywhere, anytime led to an explosive growth of wireless technologies and accelerated the emergence of various wireless broadband systems, such as Wi-Fi, WiMAX (Worldwide Interoperability for Microwave Access), WLAN, and the advanced mobile system LTE (Long-Term Evolution) evolved from cellular phones, which can deal with high-data-rate signals including both still and moving images, large amounts of digital data, and high-quality sounds as well as voice. Antennas for those wireless broadband systems had made remarkable progress along with the newly developed systems, where wideband operation covering wide frequency ranges are needed – several tens of MHz and UWB (ultra-wideband).

Most of the latest practical antennas have been no longer simple like just a dipole, a monopole, or a loop but constructed in composite structure with the shape transformed to perform wideband or multiband, integrated types to facilitate function, and combined arrangement of two or more types of antenna to operate adaptive control, even with the small dimensions as ESA. In the integrated antennas, active devices are used as AIAS (active integrated antenna system), which of course contributes to antenna small-sizing. NIC (negative impedance converter) is a representative AIAS. For a very small antenna, which exhibits difficulty in matching because of highly reactive and very small resistive impedance, an NIC can be applied to ease this difficulty, as it can compensate highly reactive components.

Contribution of electromagnetic materials to advance antenna technology is notable. Dielectric or magnetic materials have long been applied to antennas for making the dimensions small. In addition,

DOI 10.1007/978-981-4560-75-7_48-3

recently introduced metamaterials which exhibit either negative mu, negative epsilon, or both play an important role for creating such versatile antennas as miniaturized ones; functionalized ones; ones provided with enhanced performances, including reconfigurable control; and so forth. Those antennas can be any of ESAs, FSAs, PCSAs, and PSAs.

Principle and Method for Small-Sizing an Antenna

Principle in General

The fundamentals of small-sizing an antenna are to lower the resonance frequency, while the size of the antenna is kept much smaller than the operating wavelength. Another aspect is to either improve or enhance the performance of the antenna without varying its size significantly. The former is for producing ESA while the latter is for creating FSA. PCSA can be realized by lowering the profile of an antenna. The simplest way is to thin down the thickness of antenna much smaller than the operating wavelength. PSA is simply to compose an antenna physically small, regardless of the operating wavelength.

Following are methods of small-sizing antennas.

Methods of Small-Sizing an Antenna

Methods of small-sizing an antenna differ depending on the type of antenna.

ESA

The major principles for realizing an ESA are to:

(a) Lower the resonance frequency
(b) Extend/increase the current paths on the surface of the antenna element
(c) Occupy antenna space more effectively
(d) Increase number of radiation modes
(e) Make current distributions on the antenna element uniform

These are to be accomplished while the antenna size is kept unchanged. The following are practical methods for producing an ESA:

Lower the Antenna Resonance Frequency

Lowering the resonance frequency of an antenna while keeping the antenna size unchanged implies effectively small-sizing of an antenna, since the lower resonance frequency corresponds to that of a larger antenna.

There are various methods to lower the resonance frequency of an antenna without changing the antenna dimensions.

The use of a slow-wave structure

This is the typical method for lowering the resonance frequency of an antenna. Slow wave is a wave that travels a medium with slower phase velocity v_p than that of free space c; that is, $\mathbf{c} > v_p$. By applying the slow-wave concept to an antenna structure, the resonance frequency f_p of the antenna can be lowered. It can be explained by noting since $v_p = \boldsymbol{\omega}_p/\boldsymbol{\beta}_p$ ($\boldsymbol{\omega}_p = 2\pi\ f_p$ and $\boldsymbol{\beta}_p$ is the phase constant), $\boldsymbol{\beta}_p = 2\pi f_p/v_p$, and $\mathbf{c} = \boldsymbol{\omega}_0/k$ ($\boldsymbol{\omega}_0 = 2\pi f_0$ and k is free-space phase constant), then $v_p/\mathbf{c}$, and $f_p/f_0 < 1$. From this $f_p < f_0$, meaning that the resonance frequency of the antenna is lowered without changing the size by means of slow wave.

DOI 10.1007/978-981-4560-75-7_48-3

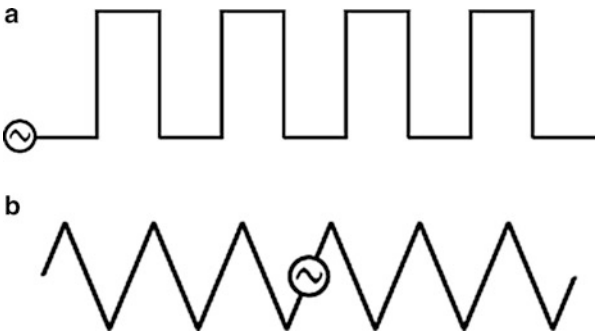


Fig. 6 Slow wave (SW) structures; a. Meander-line and b. Zigzag-line

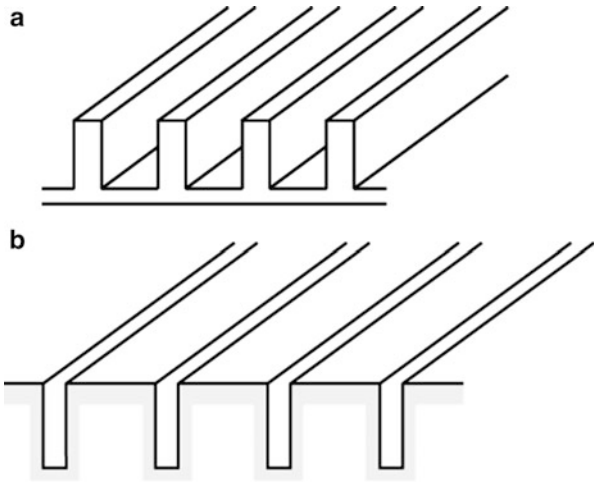


Fig. 7 3D SW structures: conductor with a. corrugation and b. trough on the surface

Practical examples of slow-wave structure are as follows:

1. Periodical structure

   The slow-wave structure is attained by periodical arrangement of wire or planar elements and solid material. Meander line, zigzag line, etc. (Fig. 6 (a) meander line, (b) zigzag) are examples of wire-type elements. Examples of solid structures or 3D type are those having corrugation or troughs on the surface of the materials (Fig. 7(a) Corrugation, (b) trough). Conducting sheet with holes on the surface is an example of planar type (Fig. 8).
2. Electromagnetic materials

   Electromagnetic (EM) materials such as dielectric or magnetic materials convey slow-wave properties and hence can be used for reducing antenna size. Within the EM materials with permittivity $\boldsymbol{\varepsilon_r}$ or permeability $\boldsymbol{\mu_r}$, an EM wave travels with phase velocity $v_p = \mathbf{c}/\sqrt{\boldsymbol{\mu_r}\boldsymbol{\varepsilon_0}}$ or $\mathbf{c}/\sqrt{\boldsymbol{\varepsilon_r}\,\boldsymbol{\mu_0}}$.

   In addition to ordinary EM materials, metamaterials (MMs), which exhibit properties of negative mu/epsilon, transport an EM wave with slower velocity v_p than that of light **c**, meaning that MMs have slow-wave properties and are useful for producing a small antenna.

DOI 10.1007/978-981-4560-75-7_48-3

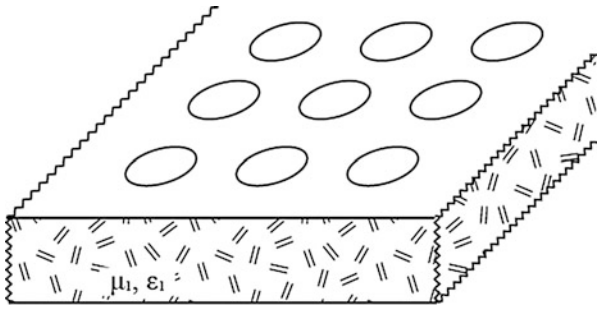


Fig. 8 2D SW structures: conducting sheet with holes on the surface

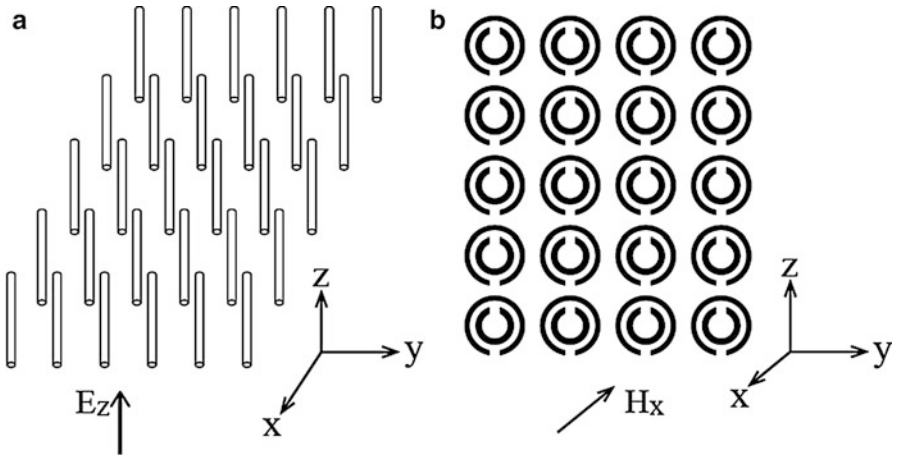


Fig. 9 Meta materials: a. dense array of wires exhibiting negative ε property and b. array of split- ring resonators (SRR) exhibiting negative μ property

The MMs are known as materials that are not available in nature, as they have uncommon properties. In practice, equivalent materials which have negative $\boldsymbol{\varepsilon}$ can be realized by using a dense array of wires (Fig. 9a) (Pendry et al. 1998), while a negative $\boldsymbol{\mu}$ property can be achieved by an array of SRRs (split-ring resonators) (Fig. 9b) (Pendry et al. 1999).

Another way to compose artificial MM structures is the use of a transmission line (TL), which can be devised to exhibit negative constitutive parameters ($-\boldsymbol{\varepsilon}$ and $-\boldsymbol{\mu}$), by arranging TL parameters (Lai et al. 2004). This TL is said to be a medium having left-handed (LH) property (Caloz and Itoh 2009) and is characterized by the property of negative propagation constant –k ($k = 2\pi/\lambda$) by the attributes of $-\boldsymbol{\varepsilon}$ and $-\boldsymbol{\mu}$. The term LH is originated from the concept of right-hand (RH) rule (Johnk 1975), derived from the use of a right hand to indicate the directions of the wave vector k and the Poynting vector $S = E \mathbf{x} H$ (E, the electric field; H, the magnetic field) as shown in Fig. 10a, b. In the LH media, the direction of the wave vector –k is opposite to that of S which is illustrated by Fig. 10c.

LH media can be realized equivalently by a transmission line (TL). A TL, on which a TEM wave propagates, is expressed by a model comprised of periodical connections of equivalent circuits with unit length *l*, consisting of a series per unit length impedance Z and a shunt per unit length admittance Y as shown in Fig. 11a, b, where dispersion characteristics on the β-ω coordinate are given. The

DOI 10.1007/978-981-4560-75-7_48-3

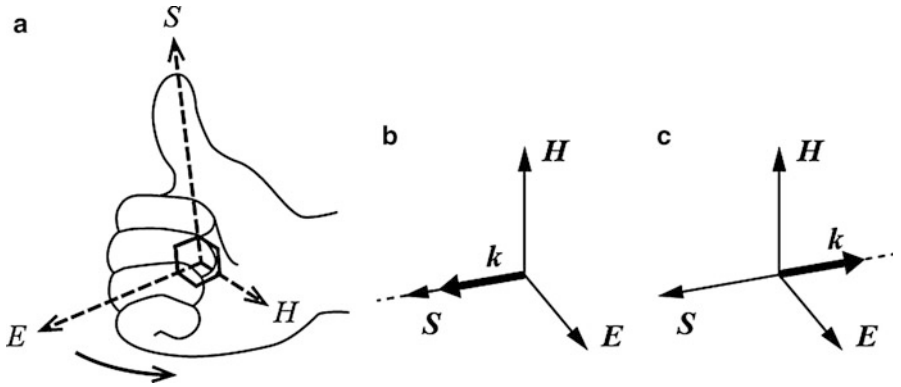


Fig. 10 (a) Explanation of Right -hand (RH) rule and direction of vectors, (b) RH case and (c) left-Hand (LH) case

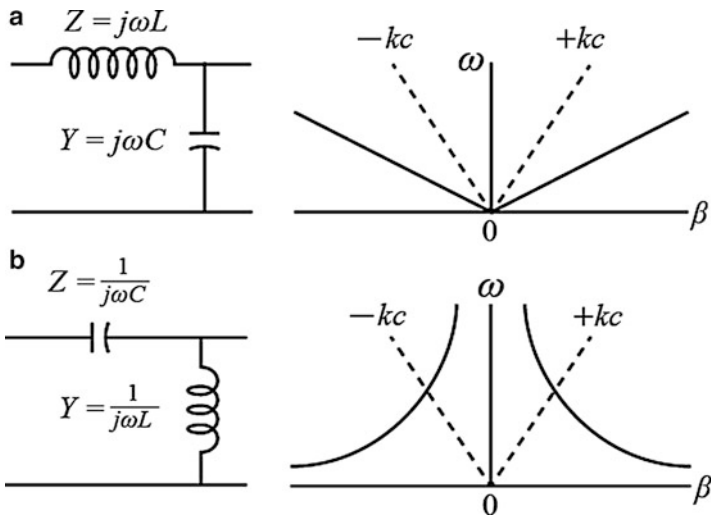


Fig. 11 Equivalent circuit of a transmission line unit cell consisting with a series impedance Z and a shunt admittance Y. (a) LC circuit and (b) CL circuit and dispersive characteristics of each circuit

immitances Z and Y, respectively, are given by using constitutive parameters $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ of the propagation medium as

$$\mathrm{Z}(\boldsymbol{\omega}) = \mathrm{j}\boldsymbol{\omega}\boldsymbol{\mu}$$

$$\mathrm{Y}(\boldsymbol{\omega}) = \mathrm{j}\boldsymbol{\omega}\boldsymbol{\varepsilon} \tag{4}$$

The immitances Z and Y in Fig. 11 are expressed by lumped circuit parameters C and L. By using these parameters, the effective constitutive parameters $\boldsymbol{\varepsilon}_{\mathbf{eff}}$ and $\boldsymbol{\mu}_{\mathbf{eff}}$, respectively, of the circuit (a) are given by

$$\boldsymbol{\varepsilon}_{\mathbf{eff}} = \mathrm{C}/l$$

$$\boldsymbol{\mu}_{\mathbf{eff}} = \mathrm{L}/l \tag{5}$$

and in the circuit (b), they are expressed by

DOI 10.1007/978-981-4560-75-7_48-3

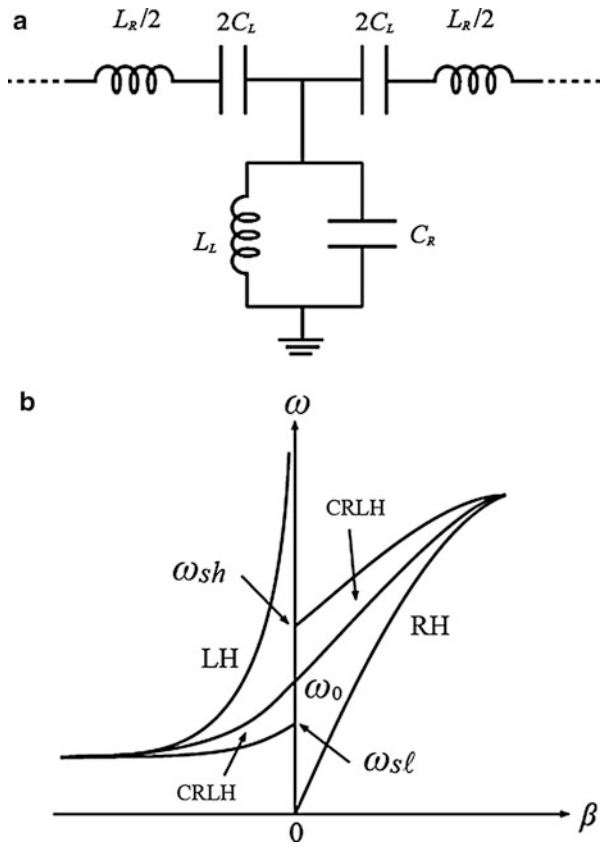


Fig. 12 (a) Unit of TL expressed by a T- circuit and (b) dispersive characteristic

$$\boldsymbol{\varepsilon}_{\mathbf{eff}} = -1/\left(\boldsymbol{\omega}^2 \mathrm{L}/l\right)$$

$$\boldsymbol{\mu}_{\mathbf{eff}} = -1/\left(\boldsymbol{\omega}^2 \mathrm{C}/l\right) \tag{6}$$

The TL with constitutive parameters given by Eq. 6 is considered an LH media as they are negative. The LH media has property of negative propagation constant –k, the attribute of $-\boldsymbol{\varepsilon}$ and $-\boldsymbol{\mu}$. The circuit model shown by Fig. 11b is characterized by negative phase constant –β as is observed in the dispersion characteristics given in Fig. 11. In the dispersion diagrams, dotted line gives k**c**, and the region below this dotted line is that where β is greater than k, indicating that in this media the phase velocity v_p is smaller than **c** (free-space velocity), meaning the media supports slow waves.

In practice, the complete equivalent circuit including a parasitic series inductance and a parasitic parallel capacitance is depicted by a T circuit composed of series reactance L_R and C_L and shunt reactance L_L and C_R as is shown in Fig. 12a, and the dispersion characteristic in the β-ω coordinate is also illustrated in Fig. 12b. In the figure, ω_{sh}, $\omega_{s\ell}$, and ω_0 are series resonance frequency, parallel resonance frequency, and the transition frequency, respectively. They are given by $\omega_{sh} = 1/\sqrt{L_R' C_L'}$, $\omega_{s\ell} = 1/\sqrt{L_L' C_R'}$, and $\omega_0 = \sqrt{\omega_{sh}\omega_{s\ell}}$, respectively. Prime (′) indicates parameters per unit length and times unit length. Which is higher ω_{sh} or ω_{se} depends on the circuit parameters. The circuit behavior appearing in the –β region in the diagram shows LH region, while that in the +β region expresses RH region. The CRLH line gives the circuit behavior having both LH and RH properties, and ω_0 is the frequency where the circuit property transfers from LH to RH or vice versa. This type of TL exhibits

DOI 10.1007/978-981-4560-75-7_48-3

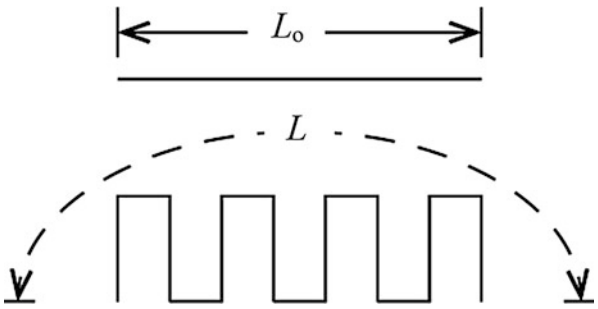


Fig. 13 Extension of current path by a meander-line compared with a straight line

behavior of either LH or RH or both depending on the parameters and is referred to as CRLH (composite RH/LH) TL (Caloz and Itoh 2005). The equivalent circuit shown in Fig. 12 is considered as the most general form of a CRLH TL structure of unit cell with LH attributes. By arranging the circuit parameters, the TL having either or both LH and RH behaviors can be designed.

By using the LH and RH behavior of a TL, the versatile application to develop novel antennas can be expected. For example, backward radiation can be realized by using the LH behavior, and beam sweeping can be obtained by switching RH and LH behaviors arranged on a TL.

3. Compose antenna by combining with different types of antenna

Resonance frequency can be lowered by composing an antenna with different types of antenna elements having different resonance frequencies. Modification of antenna geometry to a different shape is also a useful way to lower the resonance frequency. By modifying antenna geometry, the number of radiation current paths on the surface of the antenna increases so that resonance frequency is lowered. There is another modification of current paths with change in the direction of the current flows that will result in widening the bandwidth.

Extend/Increase the Path of Current Flows

Extension of current path on the surface of an antenna element makes the resonance frequency lower.

(a) Use of periodical structures

Previously introduced meander line, zigzag line, etc. are typical ones, as their length (extended length L) is longer than the length L_0 of the linear line (Fig. 13). In case of the planar antennas, the simplest way is to put slots/slits on the antenna surface so the current flow takes roundabout ways around the slot/slit (Fig. 14) so that current paths become longer and the resonance frequency becomes lower.

(b) Varying antenna geometry

Making the area of a planar antenna wider, the thickness of a linear element thicker, or antenna geometry to some different shapes will make the paths of current distributions longer or increase the number of paths so that the resonance frequency becomes lower or frequency bandwidth wider.

Efficient Use of Antenna Space

When an antenna structure occupies three-dimensional space fully where the antenna is located, the antenna may have a wider bandwidth compared with the case where the space is not effectively used by the antenna (Wheeler 1958). When an antenna can be made wider bandwidth without changing its size, the antenna is said to be a smaller one as the wider bandwidth is that of a larger antenna.

DOI 10.1007/978-981-4560-75-7_48-3

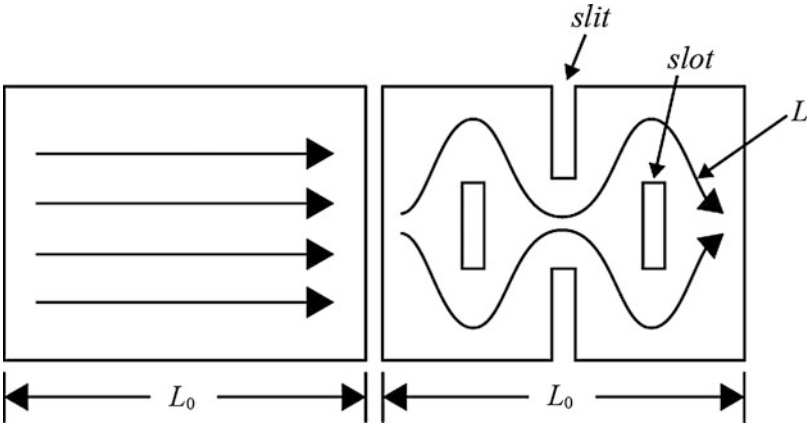


Fig. 14 Roundabout current path around the slits and slots composed with a plane surface

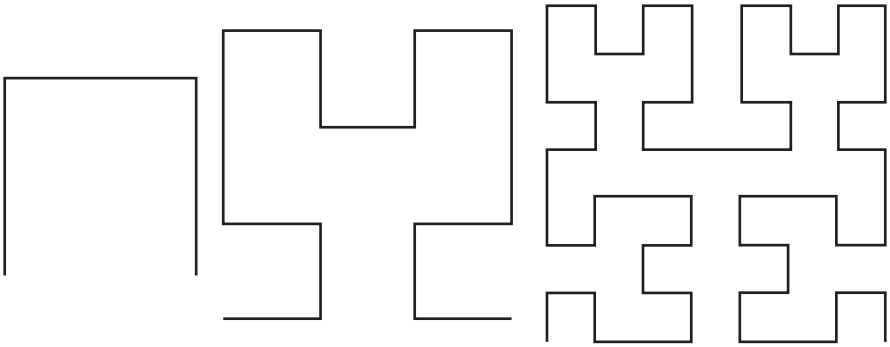


Fig. 15 Fractal geometry: Peano shape

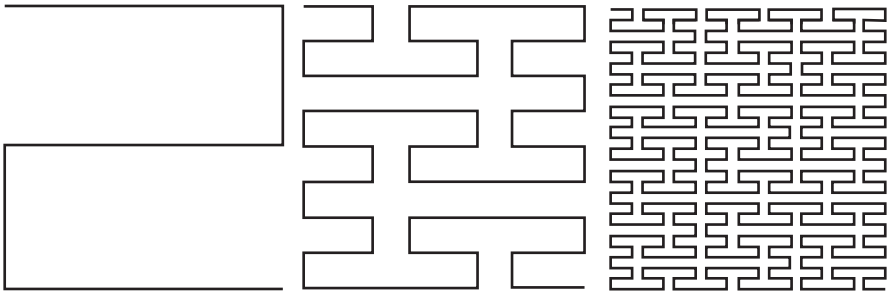


Fig. 16 Fractal geometry: Hilbert shape

To use space effectively, antenna geometry is arranged to occupy the space as much as possible. Easy ways of making antenna structures use the space effectively are applications of fractal geometries (Werner 2003) such as Peano shape (Fig. 15) (Sagan 1994; Zhu et al. 2004), Hilbert shape (Fig. 16) (Sagan 1994; Zhu et al. 2004), Minkowski (Fig. 17) (Gianvittoro 2002), and Koch (Fig. 18) (Baliarda et al. 2000). Examples of a dipole using Peano and Hilbert lines are shown in Figs. 19 and 20, respectively, where a dot is the feeding point.

A three-dimensional antenna, a sphere, in which an antenna is enclosed, is taken into consideration, and the antenna structure is composed to occupy space inside the sphere fully or as much as possible. The most

DOI 10.1007/978-981-4560-75-7_48-3

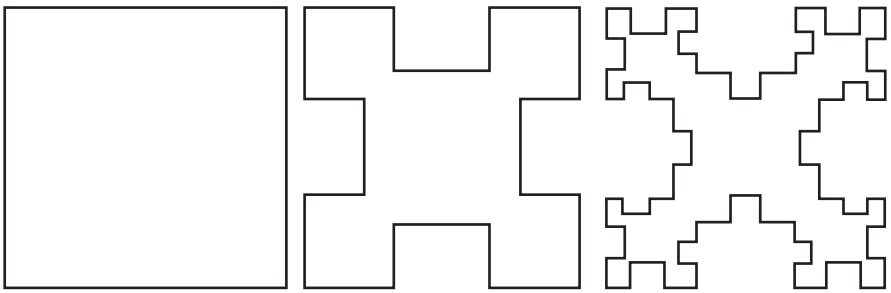


Fig. 17 Minkowsky Fractal shape

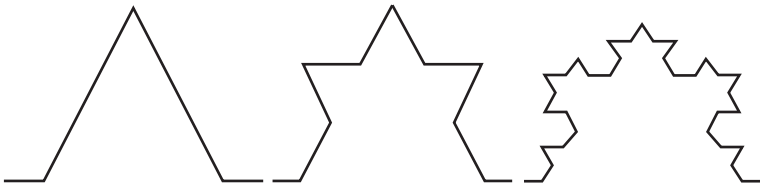


Fig. 18 Koch Fractal shape

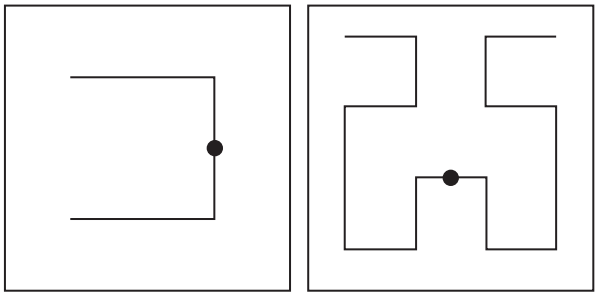


Fig. 19 Dipole using Peano shape

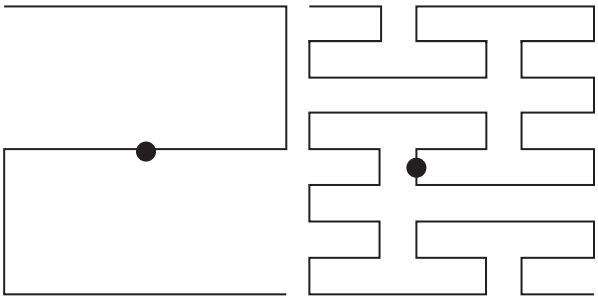


Fig. 20 Dipole using Hilbert shape

typical example is a spherical helix. To occupy the sphere periphery efficiently, a spherical helix is placed on the outer surface of the sphere as shown in Fig. 21.

DOI 10.1007/978-981-4560-75-7_48-3

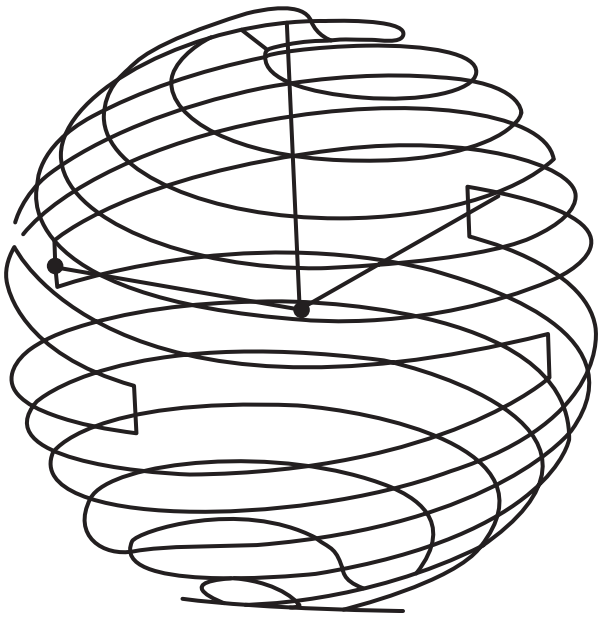


Fig. 21 Spherical helix occupying a sphere periphery © IEEE 2015

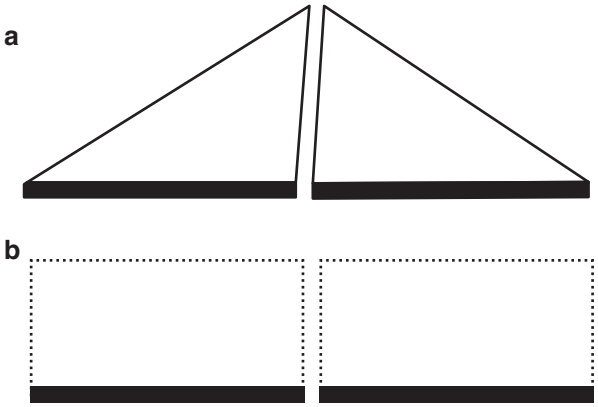


Fig. 22 (a) Triangular current distribution on a very small dipole and (b) uniform distribution

Increase of Radiation Modes

By combining an antenna with another type of antenna – for example, a TE-mode antenna and a TM-mode antenna combined – modes of radiation are increased, and antenna performance can be enhanced. Gain can be increased or bandwidth may be made wider. Enhancement of gain, bandwidth, etc. without significant increase in the dimensions from a single antenna qualifies the antenna as small sized.

Practical methods to increase radiation modes are to combine TE- and TM-mode antenna elements, for example, a dipole and a loop or the use of self-complementary concept (Mushiake 1996a) and conjugate structure. By the self-complementary structures, wideband performance – theoretically infinite if the ground plane is infinite – can be achieved; meanwhile by the conjugate structure, matching process becomes easier for a small-sized antenna, usually having very low resistive and very high capacitive input impedance, making the impedance matching difficult.

DOI 10.1007/978-981-4560-75-7_48-3

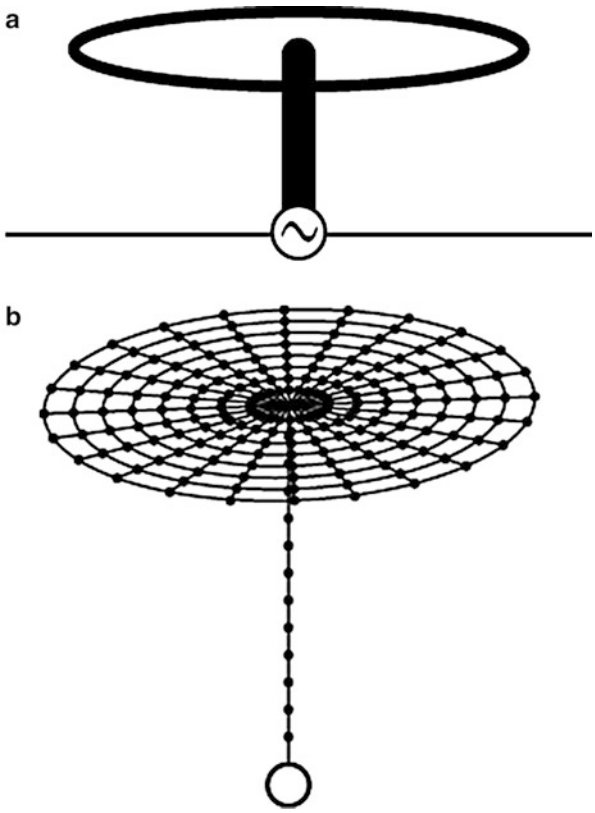


Fig. 23 Top-loaded monopole with (a) circular disk and (b) a circularly-wound disk on monopole top

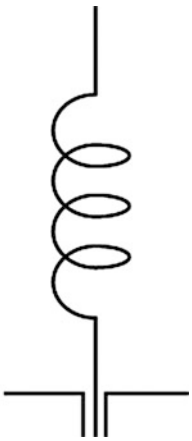


Fig. 24 A dipole loaded with an inductance in the middle

Make the Current Distributions on the Antenna Element Uniform

When the current distribution on the antenna element is uniform, the gain will be maximum with a given size of antenna. Practically, current distribution on a very small dipole is treated as triangular (Fig. 22a). To obtain maximum gain by that size of small dipole, current distribution on the dipole should be made uniform (Fig. 22b). To realize uniform current distribution on a small dipole, loading technique is often applied. The loading is done either on the top or the middle of the dipole. For example, a circular disk or circularly wound wire is used on the top loading (Fig. 23a, b), and a reactance component like an

DOI 10.1007/978-981-4560-75-7_48-3

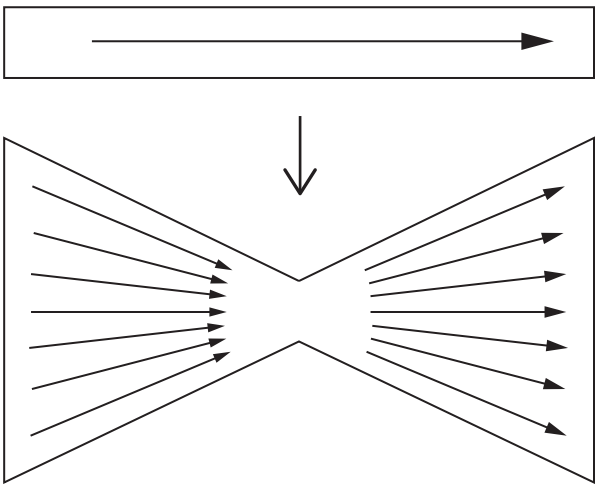


Fig. 25 Enlarging an antenna geometry to enhance the bandwidth in the figure a missing line on the top is added

inductance is loaded in the middle of the dipole (Fig. 24). Increased gain means effectively small-sizing of an antenna, because the higher gain may be obtained equivalently to a larger antenna.

FSA

To produce an FSA, essentials are either enhancement of antenna performance with the dimensions of the antenna unchanged or addition of some function by integrating either passive or active devices into the antenna structure. Typical methods for enhancing antenna performance are (1) modification of antenna geometry, (2) composing an antenna with different types of antenna elements, (3) use of HIS (high-impedance surface) (Sievenpiper et al. 1989), and (4) integration of an electronic device, either active or passive or both, into the antenna structure. Application of metamaterials (MMs) is also a useful means to enhance the antenna performance.

Among the wideband performances, UWB (ultra-wideband) is a specific one, which requires particular considerations to realize very wide bandwidth in small-sized antennas of inherently narrow bandwidth.

Modification of Antenna Geometry

Typical modification of antenna geometry is to augment the antenna surface or thickness. By doing so, the number of current flow paths on the surface increases so that the bandwidth is widened.

Figure 25 illustrates examples of modification of antenna geometry, and enlarging an antenna area to increase current flow paths on the antenna surface is depicted.

Composite Antenna

Combination of TE- and TM-mode antenna elements is one of the examples. As a practical example, composing a monopole (TE mode) with a slot having self-complementary structure (Fig. 26) to the monopole (TM mode) will result in an antenna of very wide bandwidth. With the slot monopole on the infinite ground plane, the bandwidth is theoretically infinite; however, the practical ground plane cannot be infinite but only finite, and hence the bandwidth is finite. Although it is finite, the bandwidth will still be very wide when the ground plane has effectively very large size, at least a few wavelengths.

Another type of composite antenna is constituted by using an antenna with its conjugate antenna. An antenna with capacitive impedance can be combined with an antenna having inductive impedance to achieve resonance within the antenna structure. A TM-mode antenna may be made conjugate against the TE-mode antenna. A combination of a short monopole having capacitive impedance with a slot antenna

DOI 10.1007/978-981-4560-75-7_48-3

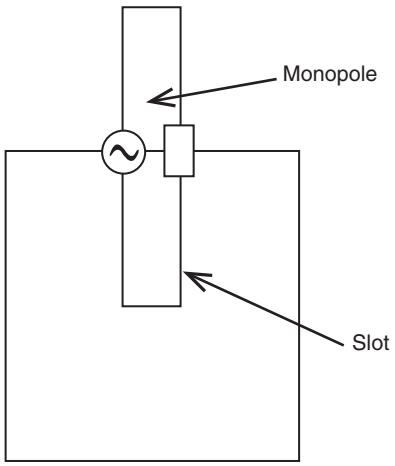


Fig. 26 An example of a combined TE and TM mode antenna: an L-shaped monopole (TE) and its complementary L- shaped slot (TM)

having inductive impedance is an example. Conjugate combination of two antennas, one of which is capacitive while the other is inductive, provides conjugate reactance corresponding to those antennas loaded to each of the two antennas.

Integrated Antennas

By integrating components, devices either passive or active, circuitry, and materials, including metamaterials, into an antenna structure – that is IAS – various notable modifications in the antenna structure can be expected. Some of those are improvement or enhancement of antenna performance, size reduction of antennas, and addition of functions into antenna structures that will create novel FSA. For instance, increase in gain is the result of constituting an antenna with different radiation modes that would add radiation power to that of the original antenna. Meanwhile, bandwidth can be increased by combining an antenna with another one having slightly different resonance frequency. The size reduction is an attribute of composing an antenna with slow-wave structure that lowers the resonance frequency.

Integration of active components into an antenna structure will produce various types of FSA. An antenna, to which oscillating circuitry is integrated, will act as a transmitting antenna with increased power compared to the original passive antenna. However, in designing the integrated antenna, noise produced by the circuit and nonlinear characteristics that may cause intermodulation can deteriorate the system performance. Factors like these should be carefully evaluated and excess noise and intermodulation should be eliminated.

By using passive components like capacitive or inductive reactance, antenna impedance can be varied, resulting in resonance in an antenna structure, and by using a varactor diode, which varies the capacitance by the bias voltage supplied to the diode, the resonance frequency of the antenna can be varied, and by utilizing this property, a frequency reconfigurable antenna is realized.

Application of Metamaterial (MM)

MM can be applied to an antenna structure or near field of the radiating antenna element for small-sizing antennas. Integration of MM, which exhibits negative reactance ($-$C or $-$L), into an antenna structure is a strong candidate for producing small antennas, wideband antennas, functional antennas, and other various sophisticated antennas. MM can be equivalently composed by various ways. The examples are the use of transmission line resonators constructed by split-ring structure (split-ring resonator, SRR), its complementary structure (complementary split-ring resonator, CSRR), and so forth.

DOI 10.1007/978-981-4560-75-7_48-3

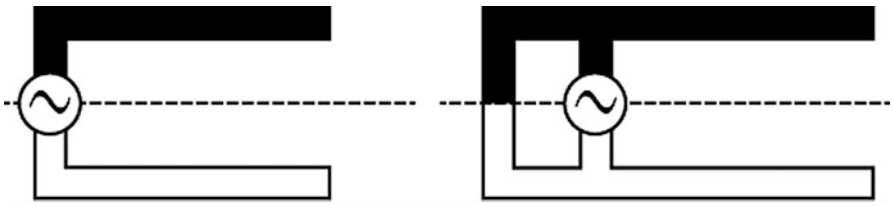


Fig. 27 An L-shape and inverted F shape antennas and their images beneath the PEC ground plane

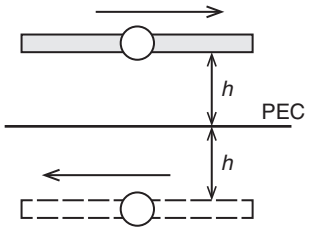


Fig. 28 Negative phase current on the image antenna beneath the PEC ground plane

PCSA

PEC Ground Plane

To realize a PCSA, an antenna is fabricated in low-profile structure with low height or narrow thickness. The thickness of PCSA is necessarily much smaller than the operating wavelength, whereas other antenna dimensions may be arbitrary. A TE-mode antenna like an inverted L or inverted F can be placed on the PEC ground plane in parallel. With the ground plane, the antenna will have an image of the original antenna symmetrically under the ground plane as shown in Fig. 27, where arrows indicate currents on the elements. A TM-mode antenna like a dipole will produce a negative-phase image element beneath the PEC ground plane when it is placed on the PEC ground plane in parallel as shown by Fig. 28.

The PEC ground plane should be infinite or equivalently infinite size, having at least a few wavelengths in one side of the plane from the feed point to the end. When the size of the ground plane is finite, some radiation current would flow on its surfaces, and the antenna performance will differ from that which is placed on the infinite ground plane due to the effect produced by the currents on the ground plane. In other words, antenna design should consider the effect of the ground plane as well as the antenna element when the size of the ground plane is finite.

Use of High-Impedance Surface (HIS)

The ground plane is not necessarily PEC, but HIS (Sievenpiper et al. 1989) can be used, with which a negative image will not be produced. A dipole antenna can be located close to the HIS ground plane with lower height h as shown in Fig. 29 and so the dipole can be a low profile antenna. In this case antenna impedance will be varied and radiation efficiency may be enhanced.

The input impedance Z_d of an antenna placed on the HIS is given by

$$Z_d = Z_{11} + Z_{12}e^{j\theta} \tag{7}$$

where Z_{11} is input impedance of the original antenna, Z_{12} is the mutual impedance between that of the original antenna and the image antenna, and θ is the reflection phase angle, which is 180° for the PEC surface.

DOI 10.1007/978-981-4560-75-7_48-3

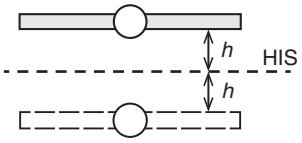


Fig. 29 A dipole antenna closely placed on the HIS ground plane with low height h

The surface impedance of the HIS is high so that the current flow on the surface can be suppressed. This will result in reducing the current flow to the rear side of the ground plane that will result in enhancing the radiation into the space and decreasing the radiation beneath the ground plane, contrary to the case of finite PEC ground plane. It also reduces the mutual coupling between two antennas placed on the HIS ground plane. Hence, two antennas can be placed with closer distance than the ordinary arraying space. By utilizing this effect, several antennas can be placed in a small limited area without higher mutual influence on each other. Practically this is applied to antenna settings for MIMO (multiple-input and multiple-output) systems, which employ several antennas in a limited small area, often in small mobile terminals.

Application of Electromagnetic (EM) Materials

EM materials, typically dielectrics, are often used as the substrate to assist thinning the thickness of an antenna. MSA (microstrip antenna) is the representative one.

Another example of EM material application is the use of HIS and EBG (electromagnetic band gap) (Rahmat-Samii and Mosallaei 2001), by which not only thinning of the antenna structure but also other various ways to create characteristic antennas can be brought out (Engheta and Ziolkowsy 2006). The HIS or EBG can be realized by periodical structure and geometrical patterns that exhibit resonance for certain frequency bands. A typical example of HIS is a metal slab with quarter-wavelength deep corrugations, shown previously in Fig. 8. This corrugation structure is considered as an array of quarter-wavelength transmission lines which exhibit a high-impedance boundary condition for electric fields polarized perpendicular to the grooves, because of the short circuit at the bottom of the grooves, whereas low impedance appears for parallel electric fields. This type of structure is known to have soft and hard surfaces (Kildal 1990) which depend on the polarization and direction of the propagating wave. A practical example of EBG material is the mushroom type shown in Fig. 30, which illustrates an external appearance on the top, cross section of several segments on the middle, and an equivalent circuit for one segment of the cross section on the bottom. In equivalent circuit, C is an equivalent capacitance due to a gap between the two surface plates and L is an inductance produced by a current flowing through the surface plate, the via, and the ground plane.

PSA

Microwave (MW) or Millimeter-Wave (MMW) Antennas

MW and MMW antennas are inherently small size as the frequency is very high. Design of those antennas is not specific, but recent antenna technology in MM and MMW regions has progressed to designing by using integration techniques, and antennas are constructed in a sophisticated structure, operating with some functions such as multi-mode, adaptive control, and so forth, keeping the dimensions small.

DOI 10.1007/978-981-4560-75-7_48-3

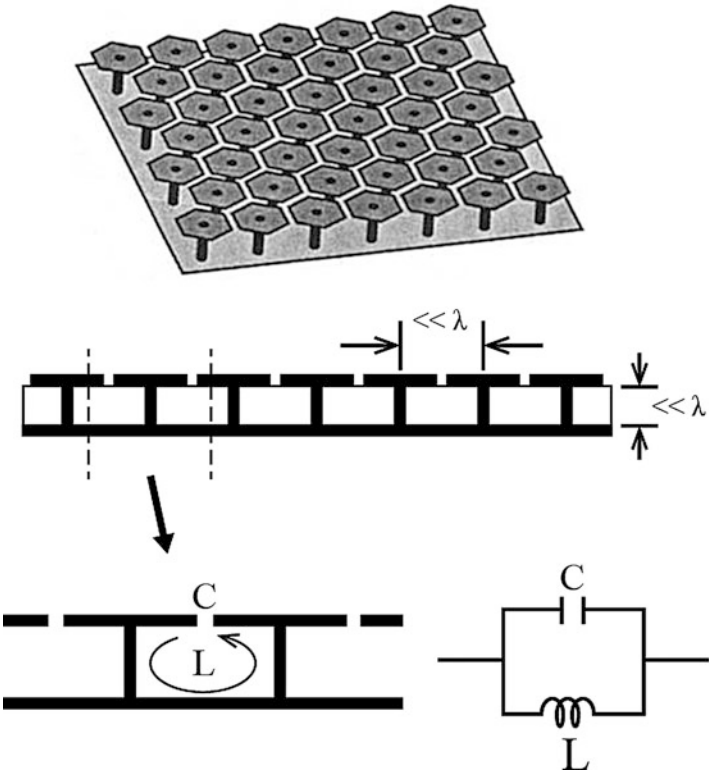


Fig. 30 An example of EBG material using mushroom structure - top: pictorial; bottom: cross-sections and equivalent circuit

Physically Small Antennas (PSAs)

Any type of antennas having physically small dimensions is called PSA. Other types of antennas, ESA, FSA, and PCSA, can be included in PSA, when their size is physically small. The term physically small means here that the size is smaller than 30 cm or so or something roughly comparable to the size of a human hand.

In recent wireless systems, antennas used are mostly PSA which are practically used in various wireless systems. Antennas applied to small RFID systems are the representative ones. Small wireless terminals used for RFID, NFC, and wearable systems, including body-centric communication and so forth, generally contain small antennas having physically small size, as the RFID terminals are generally small. Types of antennas belonging to PSA are not specific, but generally linear, planar, and three dimensional.

Of importance is antenna impedance matching in any case. When equipment is very small, any antenna installed in it is necessarily very small. There may be cases where the antenna is not properly matched to the load due to disturbance of environmental conditions near the antenna element. This kind of problem is often encountered when the evaluation of the antenna impedance is difficult because of no allowance of space for performing the measurement. The other case may be an awkward situation in placing a matching circuit. However, whatever the condition may be, matching should be properly made for any antenna and never omitted.

Taking an example in a small RFID case without appropriate matching, although the unit may be practically usable, transmitting power might be wastefully large, required sensitivity might be excessive,

DOI 10.1007/978-981-4560-75-7_48-3

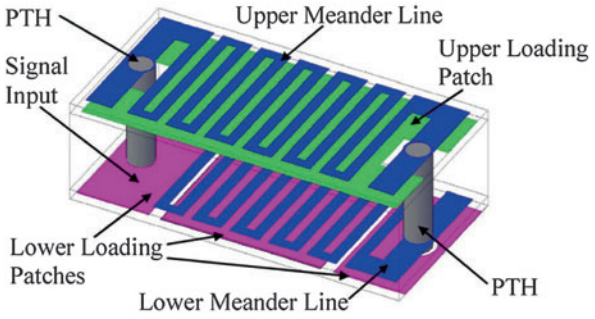


Fig. 31 Perspective view of the folded meander-line chip monopole antenna © IEEE 2015

the battery could require greater capacity than necessary, or its size may be larger than the normal one. On the contrary, if matching is properly done, the RFID unit could work with smaller power, lower sensitivity, and smaller capacity or smaller-sized battery. Or with the same power, the sensitivity might be higher or the working area broader. Also the battery capacity or size can be allowed to be smaller.

Practical Small Antennas

Practical examples of small antennas will be introduced in the next section, based on the four categories: ESA, FSA, PCSA, and PSA.

ESA

Slow-Wave Structure

Typical example is a use of meander line for an antenna element as Fig. 31 illustrates (Lee et al. 2014). The antenna element is constituted with a meander line, folded to make the size half of a normal dipole-type antenna, by which the impedance becomes a quarter of that of a dipole-type antenna. As a result, the antenna becomes very small and can be constructed as a chip. The physical dimensions of the antenna are only $3.2 \times 1.6 \times 0.83$ mm^3. The antenna is composed of a three-layered structure, two prepreg layers and one copper core layer. Side view of the antenna is shown (The exploded view is depicted) in Fig. 32. The upper meander-line element is placed on the top copper layer, and the lower one is folded inside the antenna block over the bottom copper layers. They are connected by two plated-through-hole (PTH) vias. The signal is fed into one end of the bottom meander line, and the top meander-line element is excited through the PTH vias. Reflection coefficient S_{11} measured for a prototype antenna when matched at 2.5 GHz is shown in Fig. 33, where the bandwidth is observed to be from 2.40 GHz to 2.59 GHz (7.6 %), although the antenna size is very small.

Extension/Increase of Current Paths

A small modified PIFA using a circular disk for the planar element and some meandered slots on the planar element (Fig. 34) is introduced (Kiourti and Nikita 2012). The use of the meandered slots increases the path lengths of the current on the surface of the planar element and leads to lowering the resonance frequency, achieving small-sizing of the antenna. The antenna is applied for telemetry operating at MICS (medical implant communications service) band (402.0–405.0 MHz) and ISM (industrial, scientific, and medical) band (433.1–434.8, 868–868.6, and 902.8–908 MHz). The antenna is composed of three thin circular plates, one with 6-mm-radius ground plane, and two of 5-mm radius with meander lines, as shown in Fig. 35, which gives from left to right, the bottom plate as the ground plane, the second plate and the

DOI 10.1007/978-981-4560-75-7_48-3

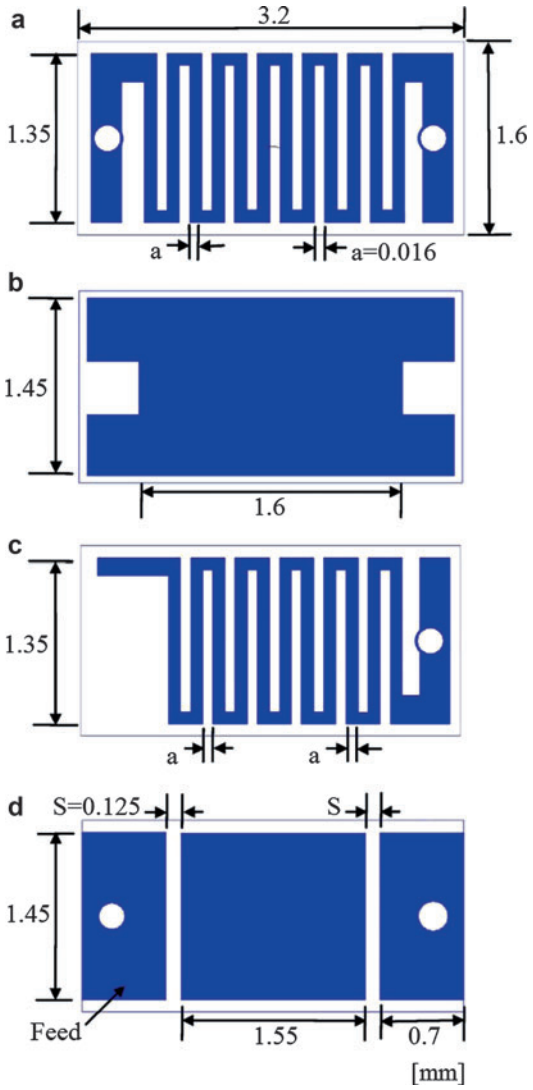


Fig. 32 Exploded view of the folded meander-line chip monopole antenna (a) the top view, (b) the second layer, (c) the third layer and (d) the bottom layer © IEEE 2015

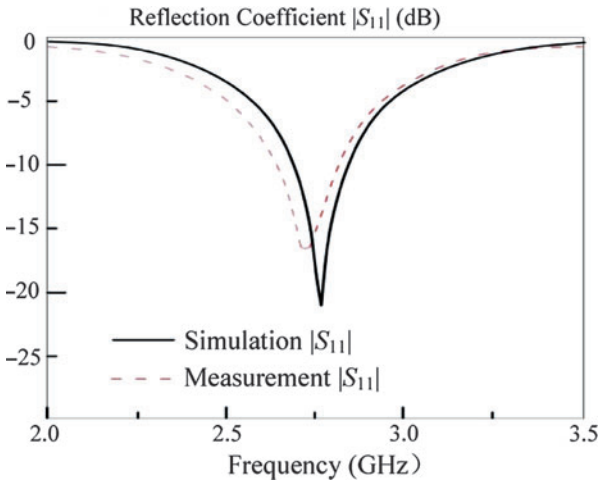


Fig. 33 Measured and simulated S_{11} Of the proposed antenna © IEEE 2015

DOI 10.1007/978-981-4560-75-7_48-3

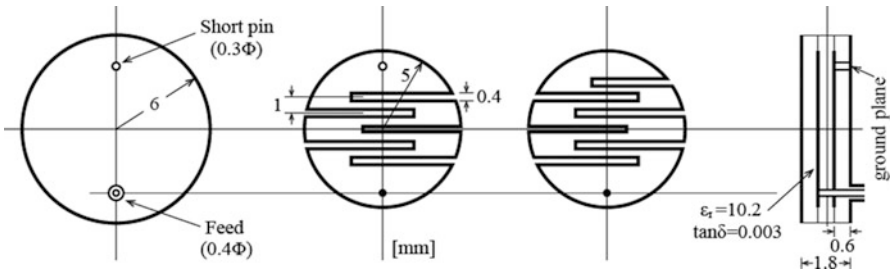


Fig. 34 Parametric model of proposed PIFA: from left to right; a ground plane, lower patch, upper patch and side view © IEEE 2015

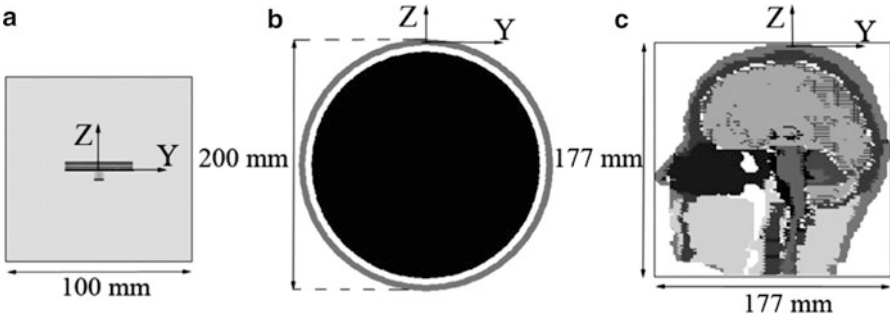


Fig. 35 setup for simulation of the PIFA implanted inside (a) a skin-issue simulation cube and the scalp of (b) a three layer spherical and (c) an anatomical human head model © IEEE 2015

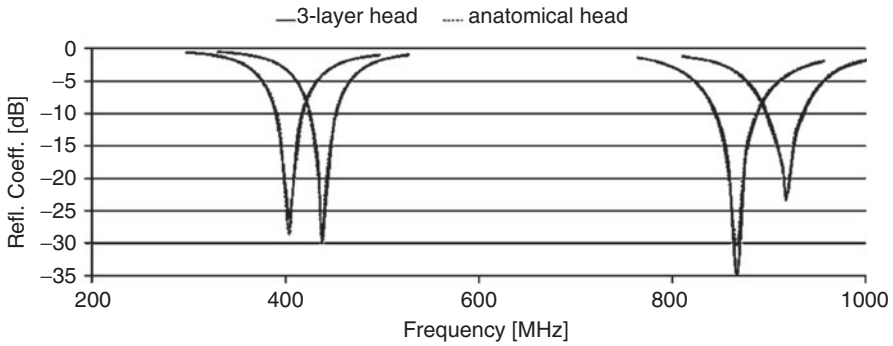


Fig. 36 Reflection coefficient of the optimized PIFA placed inside the scalp model shown in Fig. 36 (b) and (c) © IEEE 2015

third plates. The three plates are vertically stacked as shown in the extremely right. The second plate has the feed point, and one end is shorted to form a PIFA planar element, and the third plate constitutes the additional upper planar element of the PIFA. Patches are plated on 0.6-mm-thick substrate layers. Between the two planar elements, dielectric substrates ($\boldsymbol{\varepsilon_r} = 10.2$ and $\tan \boldsymbol{\delta} = 0.003$) are used. The thickness of the antenna is 1.2 mm.

The antenna is in practice applied on a human head by being implanted in it and used as a medical implant device.

The antenna design is optimized for the implantation scenario under consideration. The optimization of the scalp-implantable antenna is performed so as to fit the PIFA for implantation inside the scalp of a 100-mm-radius three-layer spherical human head model consisting of (a) a 10-cm skin tissue cube, a scalp of (b) three-layer spherical human head model consisting of skin (scalp), cortical bone (skull), and gray

DOI 10.1007/978-981-4560-75-7_48-3

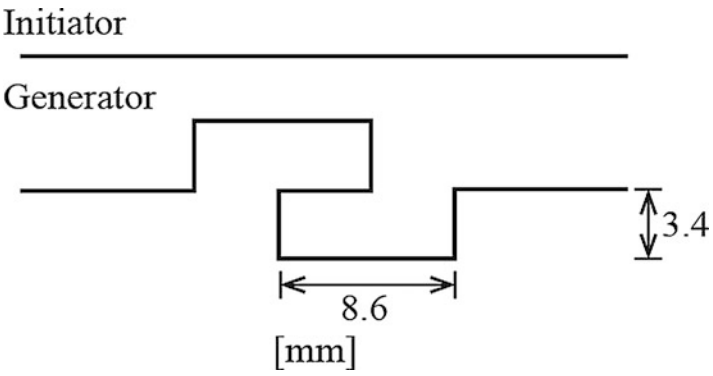


Fig. 37 Initiator and generator of the Peano shape © IEEE 2015

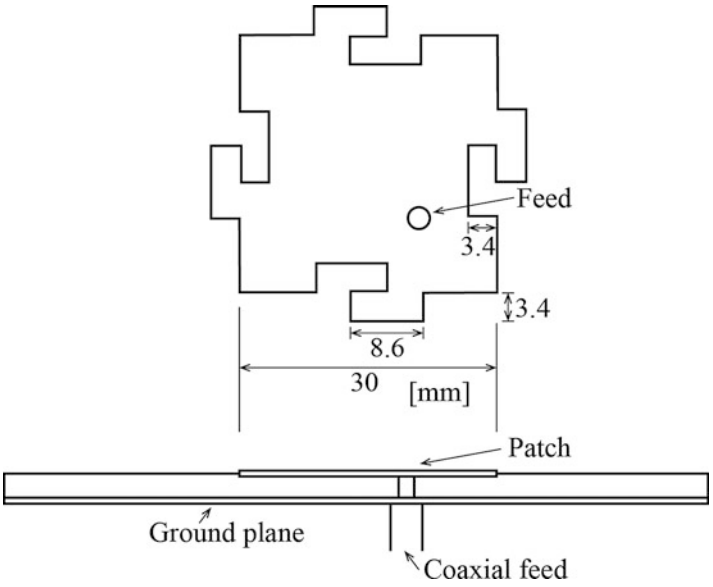


Fig. 38 Proposed Peano shape applied to a square patch antenna

matter (brain) tissue as Fig. 35 illustrates. Figure 35c depicts the anatomical human head model, a scalp consisting of 13 tissues, for which antenna performances are also investigated. Simulated reflection coefficients for the optimized antenna are shown in Fig. 36, which gives data for the three-layer head model and the anatomical head model. To provide realistic results, radiation performance of an optimized PIFA is evaluated for the case of the anatomical head scenario. Radiation patterns are studied; SAR (specific absorption rate), which should be within the specified value for safety when the device is allied to a transmitter, is evaluated; and communication links, either bidirectional or half duplex to be established between the optimized PIFA and exterior antennas, are discussed.

Efficient Use of Space

Geometrical patterns like Hilbert, Peano, and fractal shapes are applied to construct an antenna, which occupies a space efficiently, leading to small-sizing the antenna (Chen et al 2003). Here is an example of using Peano shape to modify an antenna geometry and realize a small antenna remaining yet broadband with higher gain compared with other antennas having the same area (Oraizi and Hedayati 2012).

The initiator (a line) and the generator of the Peano shape are depicted with its dimensions in Fig. 37. A proposed antenna, where the first-iteration Peano-shaped line is applied to the edge of a square patch, is

DOI 10.1007/978-981-4560-75-7_48-3

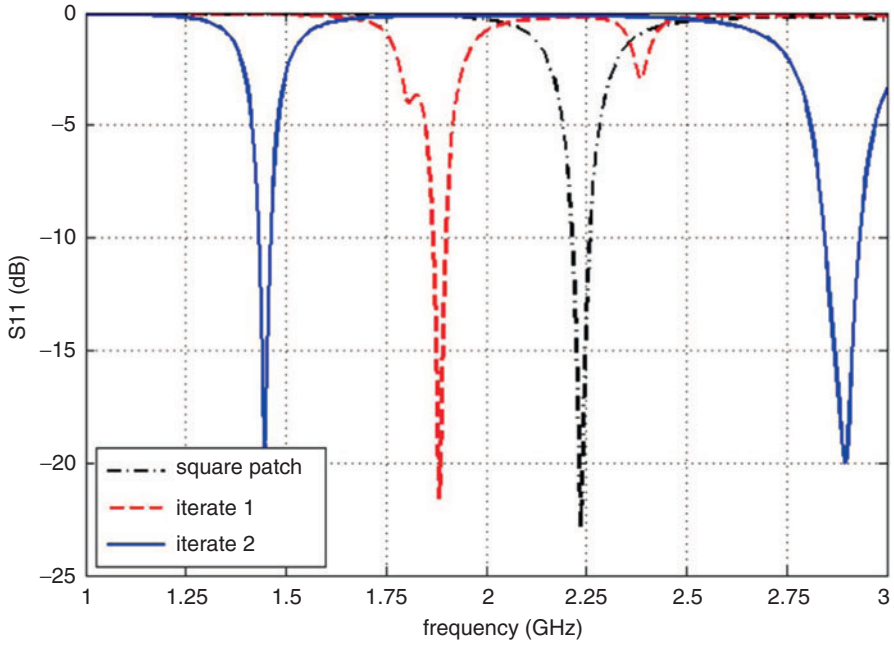


Fig. 39 Simulated S_{11} of the first and second iteration of Peano shape applied to square patch © IEEE 2015

shown in Fig. 38. Simulated reflection coefficient S_{11} for antennas with first and second iteration applied is shown in Fig. 39, in which S_{11} for an antenna of zeroth iteration (square patch) is also provided. The antennas of zeroth and first iteration have only the first-mode resonance frequency, whereas the antenna of second iteration produces the second-mode resonance frequency at 2.9 GHz, by which the antenna may be used for dual-band operation. The resonance frequency of the first-iteration antenna is about 400 MHz lower than that of the square patch, and the second-iteration antenna has about 800 MHz lower resonance frequency. From these results, dimensions of antennas, to which Peano shapes are applied, are made effectively smaller than the square patch, implying that small antennas are achieved by applying Peano shapes.

A Peano-shaped antenna is compared with antennas of other fractal shapes such as triangular Koch, square Koch, tree-type fractal, and Sierpinski fractal (Fig. 40) in terms of resonance frequency, bandwidth, and perimeter, keeping the area constant at 900 mm^2. Table 1 provides a comparison of characteristics of those shapes. Peano shape has the longest perimeter, while the resonance frequency is the lowest, meaning that the Peano shape is the most effective one to realize the smallest antenna compared to other Fractal types.

Use of Metamaterials (MMs)

A very small patch antenna using a MM concept is developed (Ouedraogo et al. 2012). A MM structure is realized by using CSRR (complementary split-ring resonator) (Baena et al. 2005; Lee et al. 2007). Resonance frequency of a patch antenna can be lowered by applying the SRR concept, which increases effective capacitance and inductance depending on the multiple concentric split rings; however, by adopting the concept of duality and complementarity to the SRR, CSRR is composed, and the resonance frequency can be further lowered. The effective permittivity of the medium can be designed to a desirable value when the medium is excited by an electric field polarized along the axis of the CSRR. In a patch antenna, the CSRR medium is placed between the radiating patch and the ground plane as shown by Fig. 41, in which illustrates the side view of, a circular patch antenna is illustrated, where the radiating patch is placed above a circular dielectric substrate backed by a circular ground plane with the same diameter as the substrate. The CSRR is contained in the circular copper disk and the geometry is depicted in Fig. 42. The circular copper disk (CSRR) is placed 0.78 mm below the radiating patch.

DOI 10.1007/978-981-4560-75-7_48-3

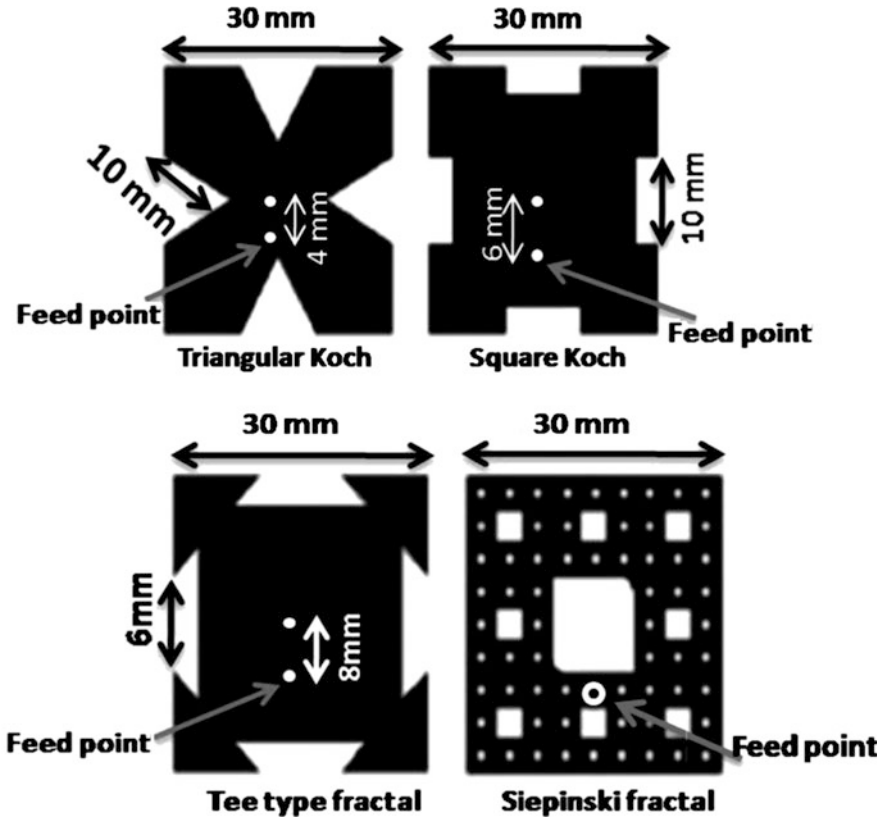


Fig. 40 Fractal shapes such as triangular Koch, tree type Fractal and Sierpinski Fractal applied to the edge of the square patch (© IEEE 2015)

Table 1 Comparison of characteristics of various Fractal shapes applied to patch antenna (© IEEE 2015)

Fractal	Resonance frequency (GHz)	Bandwidth (MHz)	Area (mm^2)	Perimeter (mm)
Peano	1.87	44	900	155.4
Triangular Koch	1.91	36	900	148.6
Square Koch	2.15	82	900	133.3
Tee type	2.14	40	900	134.6
Sierpinski	2.05	40	900	129

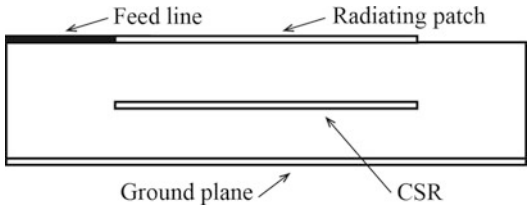


Fig. 41 Geometry of a complementary SRR (CSRR) integrated disk patch antenna

The radius of the patch is chosen for patch resonance at 2.45 GHz. Three patches are designed to have radii of 12, 8, and 6 mm, corresponding to size reduction values of 1/2, 1/3, and 1/4 compared to the traditional patch antenna, which does not use the CSRR and has radius of 23.1 mm, resonating at 2.45 GHz. The traditional antenna, which is used for comparison with the CSRR-loaded antenna, has a circular copper patch etched on top of a circular substrate ($\boldsymbol{\varepsilon_r} = 2.33$ and $\tan \boldsymbol{\delta} = 0.0012$) backed by a

DOI 10.1007/978-981-4560-75-7_48-3

Fig. 42 Geometry of the disk containing CSRR (© IEEE 2015)

Table 2 Dimensions of CSRR parameters for three different radii (© IEEE 2015)

Patch radius	N	R_1	R_2	W	G	S
12	1	23	7.1	1.5	1.15	-
8	1	15	6.2	1.6	1.75	-
6	3	11.5	9.9	1.65	1.9	1.05

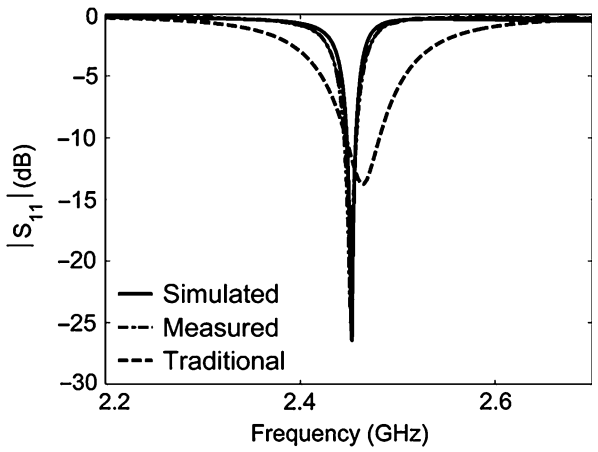


Fig. 43 Reflection coefficient S_{11} of the patch antenna with 1/4 size of the traditional patch © IEEE 2015

ground plane having the same diameter 46.2 mm with the substrate. The geometry of the CSRR within the cavity is optimized to produce an antenna resonance at a frequency well below that of the unloaded antenna. Alternatively, with the resonance frequency fixed, the loaded antenna can be small sized to a desired level of size reduction. Geometry of a proposed antenna, which is shown previously in Fig. 41, is determined in simulation by using in-house GA (genetic algorithm) so that the desired level of the size reduction is appropriate. Parameters of CSRR for each level of small sizing are given in Table 2. The parameters of CSRR are chosen to produce a reflection coefficient lower than 10 dB and the best identical

DOI 10.1007/978-981-4560-75-7_48-3

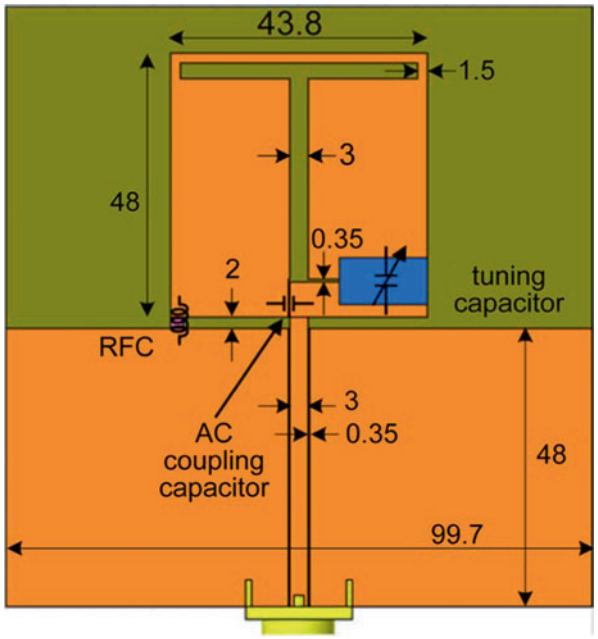


Fig. 44 An example antenna, on which a non-Foster matching network is embedded and the dimensions © IEEE 2015

cost values for each size reduction level. Reflection coefficient of the patch antenna with reduced size reduction of 1/4 (1/16 in terms of the area) is given in Fig. 43, where reflection coefficient of a traditional antenna is also shown as reference. From the figures, bandwidth increase in the CSRR antennas compared to the traditional antenna is observed. However, the radiation efficiency of the CSRR antenna is lower than that of the traditional antenna.

Impedance Matching

It is a common understanding that the minimum achievable Q of an antenna inversely relates to its physical size (Chu 1947). So as the size of an antenna tends smaller, antenna Q values tend higher, imposing a limitation in achievable matching bandwidth for matching of the antenna with lossless passive components. To overcome this problem, the use of active components or non-Foster reactance (−C or L) for the matching is considered as a useful means (Susman-Fort and Rudish 2009; Perry 1973).

Here, an antenna, to which non-Foster reactive elements are embedded inside the antenna (Fig. 44), instead of attached to the antenna terminals, is introduced (Mizaei and Eleftheriades 2013). A non-Foster reactive element does not obey Foster's reactance theorem (Foster 1924) for passive reactive elements. Non-Foster reactive elements can be implemented by using active networks such as NIC (negative impedance converter) or NII (negative impedance inverter) (Sussman-Fort 1998; Kolev et al. 2001). The underlying antenna used here is an ESA (electrically small antenna) with high Q and narrow bandwidth, and the intent is to achieve a wideband antenna by means of integration of NIC into the small antenna structure. An antenna chosen for embedding a non-Foster matching circuit is a frequency reconfigurable antenna, to which tunable reactance elements C_t and L_t are loaded. The tuning behavior of the antenna is given by a curve of $B_t = \boldsymbol{\omega} C_t$ or $X_t = \boldsymbol{\omega} L_t$ with respect to resonance frequency $\boldsymbol{\omega}$. The tuning behavior is simulated by using HFSS for some discrete values of the load capacitance C_t and shown in Fig. 45a for a large bandwidth and (b) tuning susceptance $B_t = \boldsymbol{\omega} C_t$ with respect to $\boldsymbol{\omega}$. The figure shows negative slope due to the non-Foster behavior. In the non-Foster version, C_t is replaced with a parallel combination of $-C_a$ and $-L_a$, while L_t is replaced with a series combination of these elements. The goal is to replace tuning element C_t or L_t with a network that satisfies the resonance condition. As the

DOI 10.1007/978-981-4560-75-7_48-3

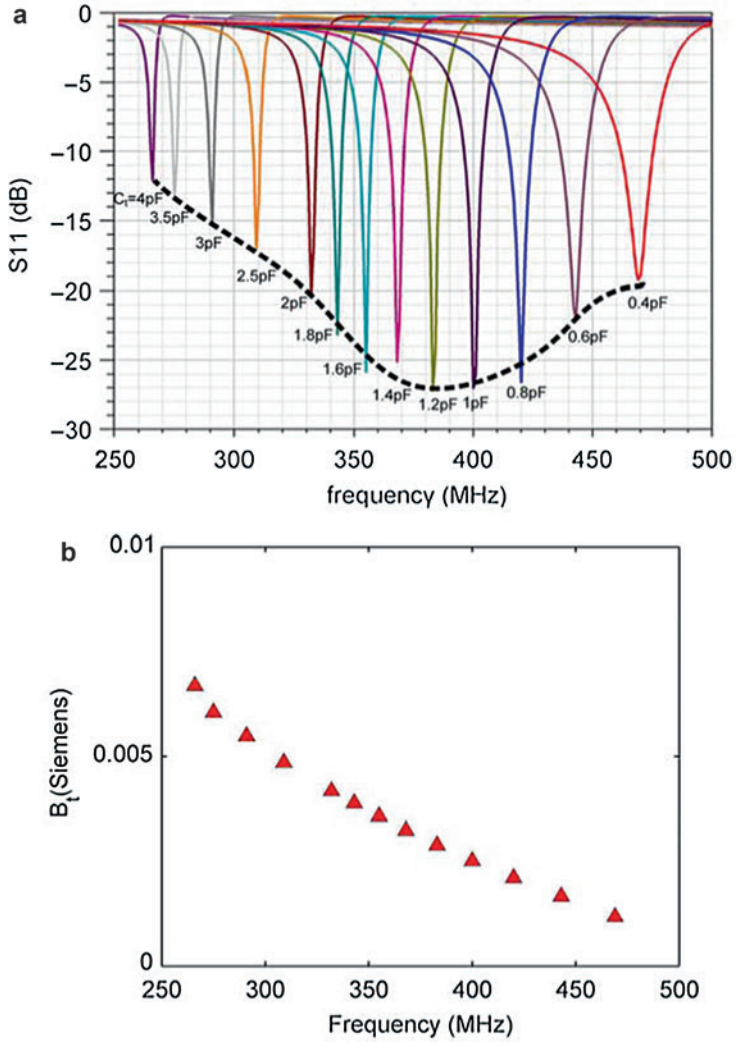


Fig. 45 Simulated tuning behavior of an antenna shown in fig.45; (a) variation with capacitor of S_{11} with C and (b) susceptance Bt vs frequency © IEEE 2015

network to fit the tuning behavior is not unique, different networks can be employed for non-Foster matching or network to be embedded on the antenna. For the antenna under consideration, three reactance functions are considered to fit the antenna tuning behavior shown in Fig. 45b. To select the best function, some factors are analyzed: complexity, values of network elements, and sensitivity. The appropriate function which fits to the tuning behavior should represent synthesizable reactance (susceptance function $B_t(\omega)$ or inductance function $X_t(\omega)$) using a combination of Foster or non-Foster elements. Two Tuning elements C_t or L_t with a network that satisfies the resonance condition in a wide frequency band. Such a network exhibiting behavior shown in Fig. 45b synthesizes a non-Foster reactance of its terminals and can be implemented by the addition of $-C$ or $-L$ elements to the set of the available reactive elements (capacitors or inductors). The antenna is designed in a single layer; however, the circuit uses both sides of the board for fitting the design into the available space. The bottom layer of the board is used mostly for running the DC bias lines all across the NIC circuit and for placing a trimmer capacitor for tuning the NIC circuit. Here the Linvill's NIC circuit scheme (Linvill 1953) is used for the NIC circuit.

Measured results of S_{11} for the fabricated antenna are shown in Fig. 46 for four trimmer states. Then, the results are compared with the tuning data of the passive antenna shown in Fig. 50b, and the results are

DOI 10.1007/978-981-4560-75-7_48-3

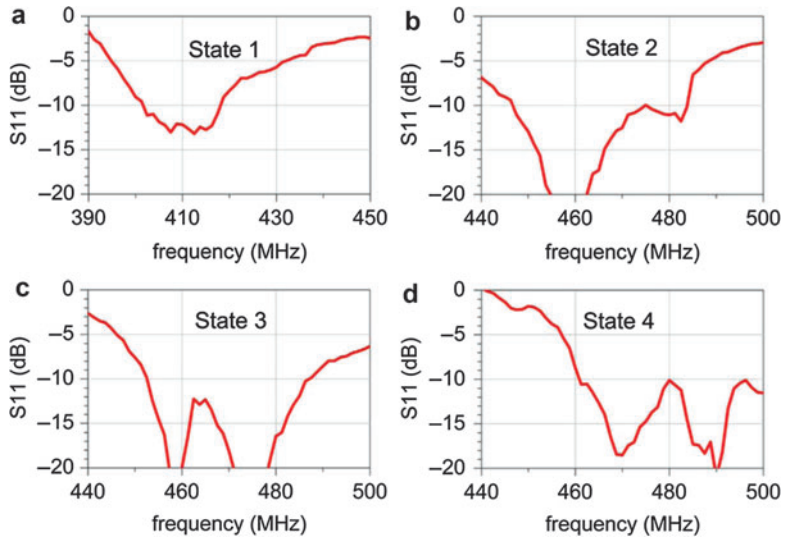


Fig. 46 Measured S_{11} for fabricated antenna for four trimmer states © IEEE 2015

Table 3 Comparison of measures S_{11} for the non-Foster circuit embedded antenna with simulated S_{11} of the associated passive antenna for the four different trimmer status (© IEEE 2015)

	−10 dB S_{11} fractional bandwidth	
Trimmer state	Passive antenna (%)	Non-Foster antenna (%)
State 1	2.0	4.0
State 2	2.9	8.1
State 3	2.9	7.4
State 4	3.0	8.3

provided in Table 3, where −10 dB S_{11} fractional bandwidths for four trimmer states are given. From the table, matching with a non-Foster network obtained a significant improvement in the bandwidth. This increase in the bandwidth is obtained by decreasing the reactive part of the antenna, instead of decreasing the antenna Q by loading with lossy elements. The stability of the embedded matching network is verified by the Nyquist test applied to the loop gain of the NIC circuit.

FSA

UWB (Ultra-wideband) Antennas

(a) Use of self-complementary structure

To obtain a wideband antenna, one of the most useful ways is to use self-complementary structure (SCS) (Mushiake 1996a), which is known to give frequency-independent performance (Rumsey 1957). By using SCS, infinite bandwidth will be realized theoretically if the ground plane is infinite. In practice, however, since the ground plane cannot be infinite, unlimited bandwidth cannot be obtained. Although infinite bandwidth cannot be realized, achieved bandwidths will be still fairly wide while using finite ground planes. How wide a bandwidth is achievable depends on the size of the ground plane when the ground plane is finite. This case is generally called "quasi-self-complementary (QSC)." An example is an

DOI 10.1007/978-981-4560-75-7_48-3

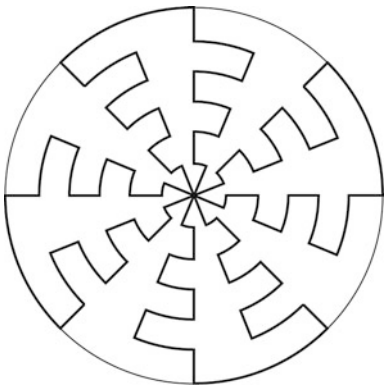


Fig. 47 Example of self-complementary antenna pattern

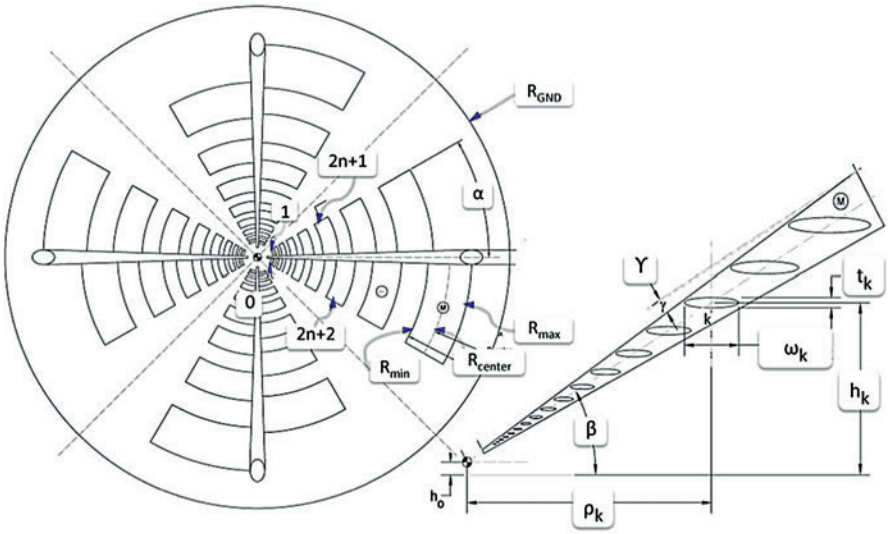


Fig. 48 Geometry of quasi-self-complementary (QSC) antenna and the side view of the one of the four arms © IEEE 2015

antenna pattern shown in Fig. 47, where a log-periodic element is used. The antenna element has log-periodic shape, while the space pattern is self-complementary to the antenna elements.

Here, a nonplanar QSC antenna is introduced (Cortes-Medellin 2011). The QSC antenna geometry is illustrated in Fig. 48, where the left side is top view, showing the projection of the antenna structure, and the right side is side view. As the figure shows, the antenna has four arms, having fourfold azimuth symmetry. The antenna is fed at the center of the structure, in a differential mode between two opposite arms, thus providing dual polarization. Each arm consists of a log-periodic (LP) SC element (it should be noted that LP structure does not always ensure frequency-independent characteristics as SC structure does (Mushiake 1996b)). This QSC antenna has a characteristic of differential VSWR with an input matching level better than a 1.7:1 bandwidth with a 270-ohm normalization, which corresponds to differential input match better than 10 dB from 1 to 20.0 GHz. The VSWR characteristics are shown in Fig. 49. The antenna has low profile, occupying a volume of a pillbox shape having diameter of $1.2\lambda_{min}$ and height of $0.25\lambda_{min}$ (λ_{min} is the wavelength of the lowest frequency for the −10 dB match).

Multiband Antennas

A compact, small-sized seven-band mobile handset antenna is described (Ban et al. 2013). Antenna Geometry and its detailed structure along with optimized dimensions are depicted in Fig. 50. where the

DOI 10.1007/978-981-4560-75-7_48-3

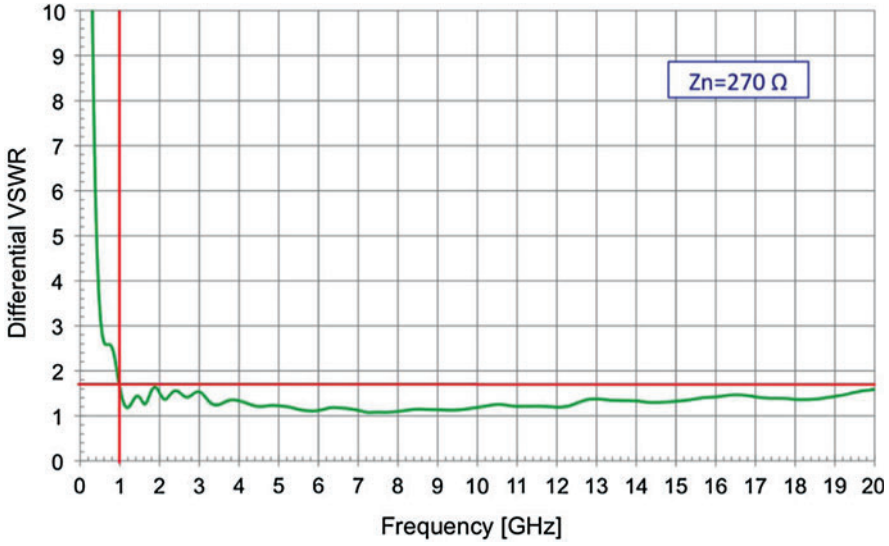


Fig. 49 Calculated differential VSWR, showing ultra-wideband (>10:1): red line indicates −10dB © IEEE 2015

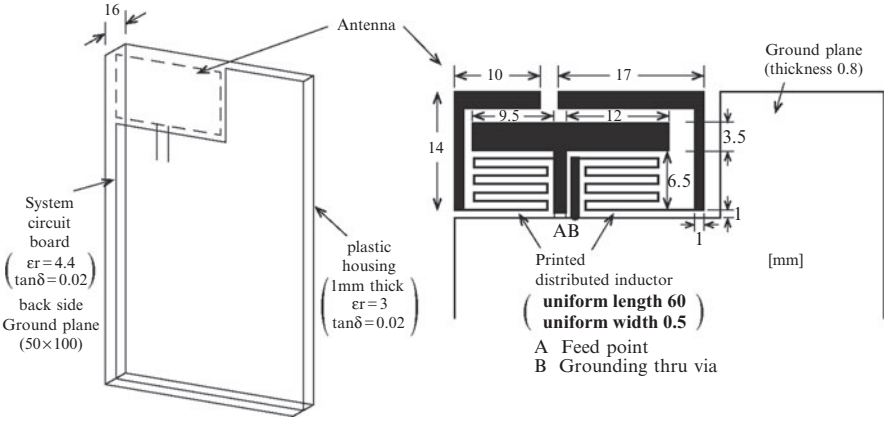


Fig. 50 Configuration of the proposed antenna: geometry of the 7-band antenna, mounted on the mobile phone (in the left) and the antenna structure (on the right)

left side shows the antenna mounted on the one corner of the bottom edge of the system circuit board and the right side expresses detailed dimensions of the metal pattern in the antenna area. The antenna consists of strips; one for feeding and other two (shorter and longer ones) for radiating, and symmetrically printed meandered inductive strips embedded in the shorter radiating strip and longer radiating strip, respectively, to generate dual resonance modes around 810 MHz and 962 MHz. The resonance for the lower band, covering GSM/850/900 (824–960MHz), is achieved by the longer radiating strip (0.25λ at 810 MHz) and the higher bands, covering GSM 850/900/1800/ 1900, UMTS 2100, and LTE 2300/2500 bands, is produced by the shorter radiating strip (0.25λ at 962 MHz). The length of the feeding strip is about 32 mm that supports 0.25λ at about 1800 MHz resonant mode to cover DCS 1800 operation. The shorter and longer radiating strips (total length 90 mm) generate a higher resonance mode at about 2450 MHz. By printing the antenna elements on one corner of the bottom edge of the system circuit board, the SAR (Specific Absorption Rate) values can be decreased. The circuit board uses FR4 substrate with ε_r of 4.4 and tan δ of 0.02, the size of 50 x 150 mm^2, and the thickness of 0.8 mm. There are two printed system ground planes: main ground plane of 50 x 100 mm^2 and a protruded one of 15 x 20 mm^2, leaving a non-ground region of 50 x 30 mm^2 at the top position of the circuit board where the proposed antenna shown in

DOI 10.1007/978-981-4560-75-7_48-3

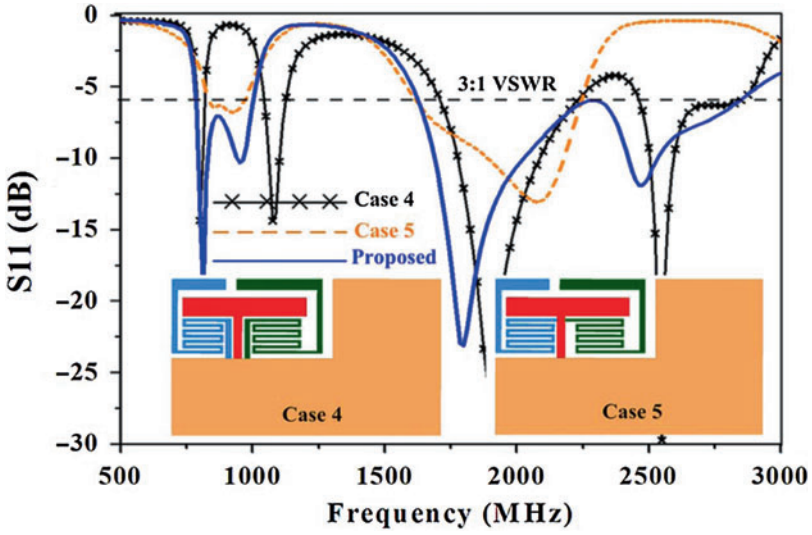


Fig. 51 Simulated S_{11} for the proposed antenna: case 4 and case 5, with the proposed antenna in the inset © IEEE 2015

Fig. 50 is placed. The system circuit board is enclosed by plastic housing with 1-mm-thick slab (ε_r of 4.4 and tan δ of 0.02 and total height of 10 mm) to simulate a practical mobile phone case.

Simulated S_{11} for proposed antennas are shown in Fig. 51, where the antennas of case 4 and case 5 are depicted in the inset. The terms *case 4* and *case 5* mean cases where the feeding method is different; the case 4 is when the shorter and longer strips are fed by coupling, while the case 5 is when those strips are fed directly. Remarkable effect due to the difference in the feeding method is observed in the figure.

By combining a feeding strip with two radiating strips and two symmetrical printed distributed inductors, the proposed antenna can exhibit two wide operating bands for lower frequency regions, 810 and 962 MHz, covering GSM 850/960 bands, and four wide bands for higher frequency regions, 1620–2830 MHz, covering DCS 1800/PCS 1900/UMTS 2100/LTE 2300/2500 operation.

Functional Antenna

A looped bow-tie antenna is designed for RFID tag which operates for a long-range and frequency reconfigurable performance (Lin et al. 2013). Configuration of the antenna is illustrated in Fig. 52, which gives top view with dimensional parameters, side view, and cross-sectional view, respectively. The antenna consists of two non-connected load bars and two bow-tie patches electrically connected to a conducting backplane by four pairs of vias, thus forming a looped bow-tie RFID tag antenna structure. By this structure, the tag antenna is made suitable for mounting on metallic objects. The two load bars are used as integral parts of a metal tag antenna by placing them near the edge so that the effect of the metal on the tag is reduced. The RFID chip is placed across a gap between the two patches and connected to them both. Each pair of vias is equivalent to a single loop, which can be viewed as a single inductor L.

In the RF tag design, four ways to adjust the self-resonance frequency are introduced to obtain the desired complex impedance of the antenna: two for inductances and another two for capacitances by varying dimensions of the antenna and number of vias. Multiple pairs of vias constitute multiple inductors L_i ($i = 1,2,..n$) in a parallel circuit, and a combination of them can yield an equivalent inductor L_a, which is shown in Fig. 53, which provides (a) general circuit model and (b) equivalent circuit. In the figure, C_a and L_a, respectively, are the equivalent capacitance and inductance of the antenna, which determine the antenna self-resonance frequency. Increasing the number of via pairs will lower the inductance, resulting in higher resonance frequency. The capacitances C_g and C_b, respectively, shown in Fig. 53, are capacitances due to the gap of the bow-tie patch elements and that contributed from the load bars with the length

DOI 10.1007/978-981-4560-75-7_48-3

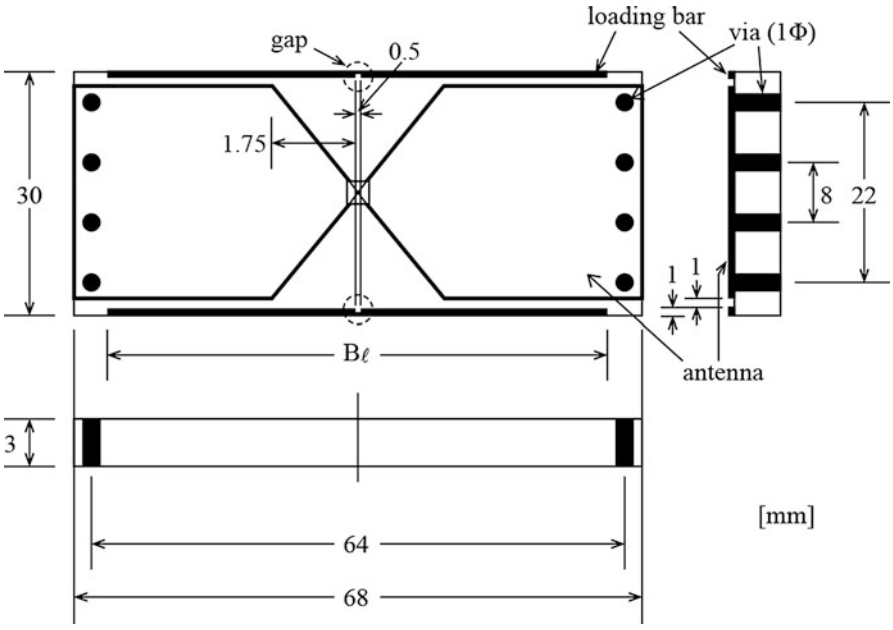


Fig. 52 Configuration of the proposed antenna; ((a) changed to) with dimensional parameters, ("and operating …2015" was changed to) side view (on the right) and cross sectional view (at the bottom) In the figure, dimensional parameters L, BW, Vr, Vd, V1, V2, and H are changed

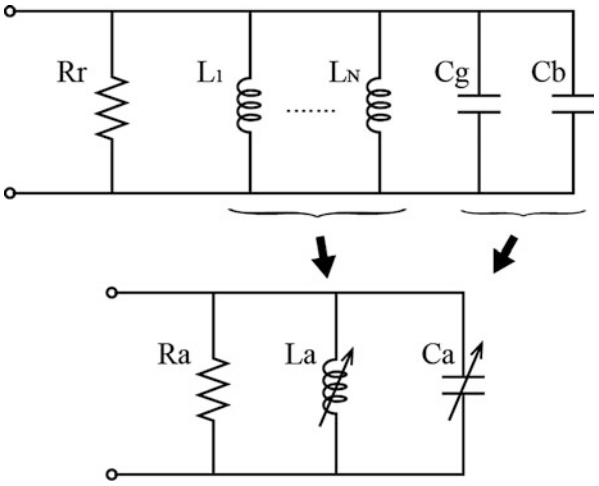


Fig. 53 Circuit model of the proposed antenna; general circuit (upper part) and (b) equivalent circuit (lower part)

of B_ℓ, respectively. Hence, by varying C_g and C_b, respectively, input impedance can be varied, and the resonance frequency can be controlled as well.

In addition, the looped bow tie itself serves as an inductor, which can be controlled by changing the dimensions V_1, V_2, and V_d, between the vias shown in Fig. 52a. By shortening V_d, inductance is decreased, which in turn shifts the resonance frequency higher. Variation of the resonance frequency depending on V_d is shown in Fig. 54. The load bars, each of which has a cutoff point, can be used for reconfiguring the operating frequency of the tag by exciting any one of three frequency modes; hence, the tag can be used for worldwide UHF RFID frequency bands. The experimental tests show that the

DOI 10.1007/978-981-4560-75-7_48-3

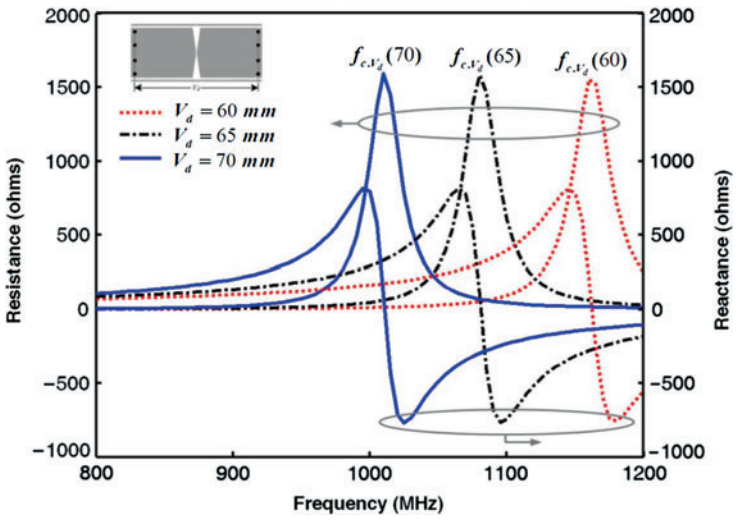


Fig. 54 Simulated input impedance of the proposed antenna for three different Vd's © IEEE 2015

maximum read range, located on a metallic object, was 3.0, 3.2, and 3.3 m for three different operating modes with only 0.4 W EIRP radiation power of the RFID reader.

The overall size of an implemented antenna is 68 x 30 x 3, all in mm.

PCSA

PEC Ground Plane

A printed wide-slot antenna with a parasitic patch in the center of the slot is proposed (Sung 2012). The antenna is designed for the purpose to enhance the bandwidth. Geometry of the proposed antenna is shown in Fig. 55, where a square of the central part shows the printed patch on the substrate and a square outside the central patch shows the etched slot. In the figure, dimensional parameters are also given, and the feed line on the opposite side is shown in dotted lines.

A fabricated antenna is printed on one side of the substrate FR4 ($\boldsymbol{\varepsilon_r} = 4.4$) with the thickness of 1.6 mm. The width of the feed line is 3 mm, by which the characteristic impedance is adjusted to 50 ohms. The distance between the center of the slot and the edge of the feed line is adjusted to obtain good matching.

Simulated and measured reflection coefficient of the proposed antenna is given in Fig. 56. Measured bandwidth (10 dB reflection coefficient) is as large as 3130 MHz (2225–5355 MHz) or 82 % with respect to the center frequency at 3790 MHz. The antenna can be a candidate for 2.4/5.2 GHz WLAN application.

Use of FSS Layers

An EBG (electromagnetic band gap) structure, which is composed by combining two complementary FSS (frequency-selective surface) layers (Munk 2000), is applied to compose a wideband Fabry-Perot resonator antenna (Wang et al. 2014). The EBG structure is illustrated in Fig. 57, which was changed to) antenna is fabricated on a substrate with thickness T and a dielectric constant εr, and the metal patches are on the top side, while the square apertures (metal grids) are on the bottom side. The aperture has complementary shape with the square patch. As the side view shows, the antenna is for the incidence of vertically polarized wave, each horizontal gap between two square patches on the top-side surface acts as a capacitor with capacitance C_p, and inductors are provided by the patch itself with a small inductance value L_p, which is shown in Part B in the equivalent circuit. The square apertures at the bottom shown in

DOI 10.1007/978-981-4560-75-7_48-3

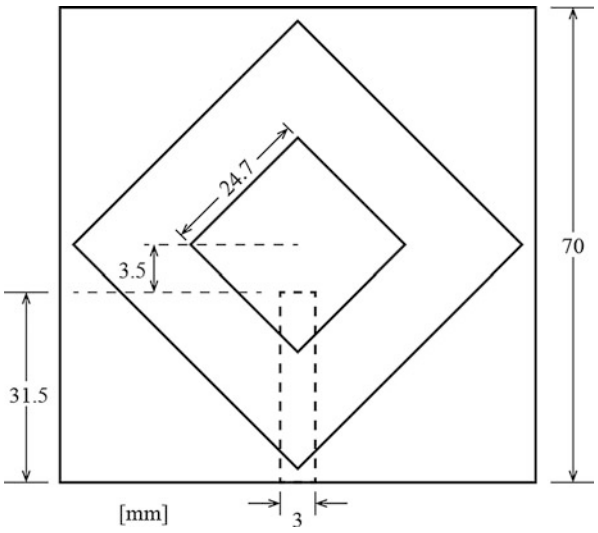


Fig. 55 Geometry of a proposed antenna

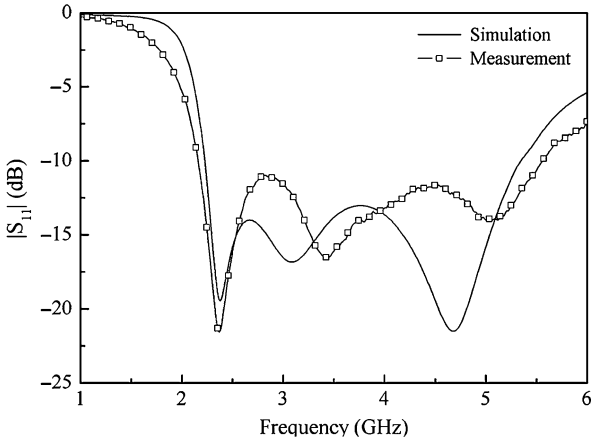


Fig. 56 Simulated reflection coefficient of proposed antenna © IEEE 2015

Fig. 57 are considered to compose long rods which act as inductors, and horizontal bars act as capacitors in parallel with the inductors as shown in Part D in the equivalent circuit given by Fig. 57. The dielectric substrate can be treated as a segment of transmission line with length L_{01} and a characteristic impedance of Z_{01} as shown in Part C in the equivalent circuit shown by Fig. 57. L_{01} is expressed by $T\sqrt{\varepsilon_r}$ and Z_{01} equals to $Z_{air}/\sqrt{\varepsilon_r}$ where Z_{air} is the wave impedance of air, about 377 ohms.

A fabricated Fabry-Perot resonance antenna designed by combining the suspended EBG layer and the feeding antenna is illustrated in Fig. 58, the feeding patch antenna is placed at the center of cavity. The patch antenna is coupled with the feed line through a slot in the ground plane and spaced from the ground plane by an air gap for suppressing the surface wave excitation. The EBG structure has the size 72 x 72 mm^2 and is comprised of 77 unit cells. The dimensions of the unit cell, the patch, and the aperture are designed to be 8 x 8 mm (less than l/3 λ at 10 GHz), 6.3 x 6.3 mm^2 and 5.6 x 5.6 mm^2, respectively.

Metamaterial Application

A compact patch antenna loaded with synthesized metamaterial (MM) is introduced (Dong et al. 2012). The MM is synthesized by using CSRR (complementary split-ring resonator) that is embedded on RIS

DOI 10.1007/978-981-4560-75-7_48-3

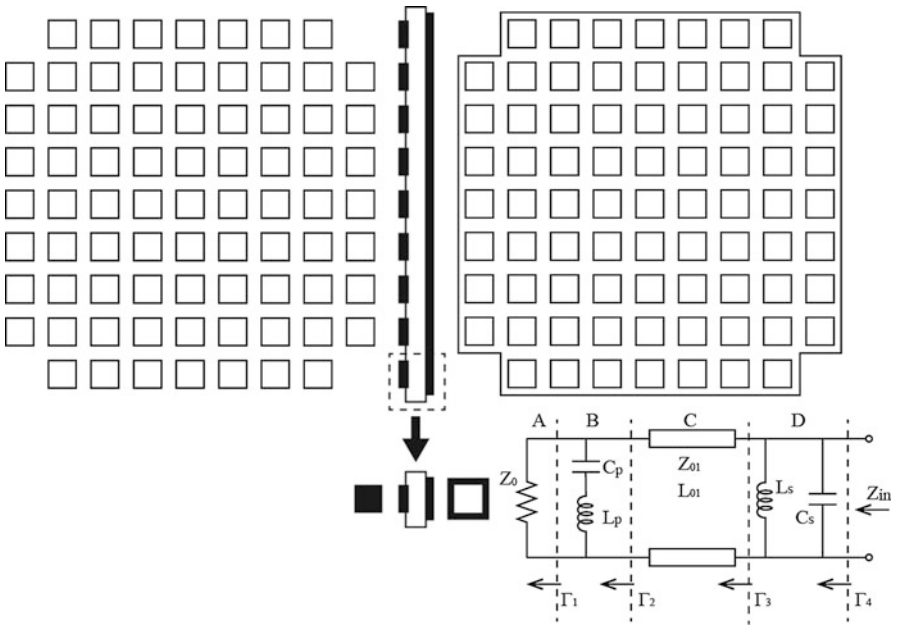


Fig. 57 EBG structure: top view (upper left), side view (upper middle), bottom view (upper right), unit cell (lower left) and equivalent circuit (lower right)

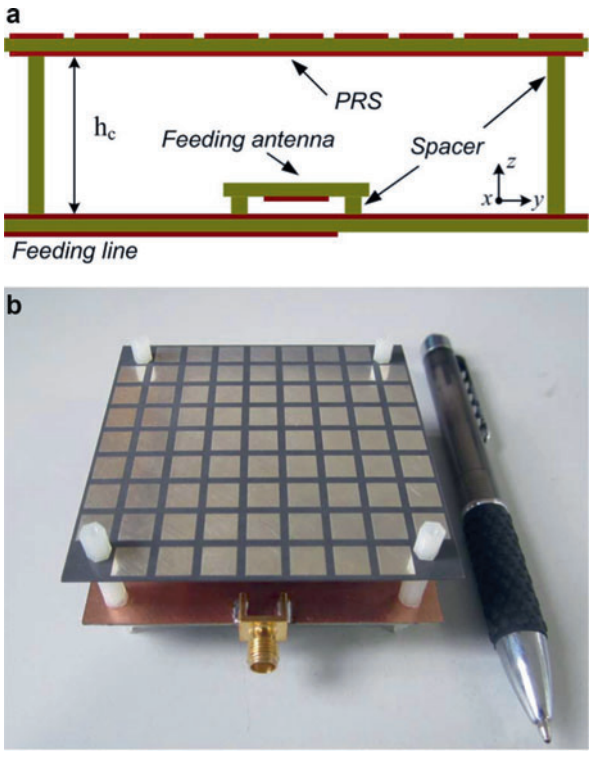


Fig. 58 (a) Side view of the FP resonator antenna and (b) a photo of the prototype antenna © IEEE 2015

(reactive impedance surface) and applied to make the antenna small sized. The CSRR is incorporated on the patch as a shunt LC resonator, which provides low resonance frequency, resulting in making the antenna small. The RIS is realized by using two-dimensional metallic patches printed on a metallic-grounded dielectric. The CSRR structure and its equivalent circuit are shown in Fig. 59. The CSRR is

DOI 10.1007/978-981-4560-75-7_48-3

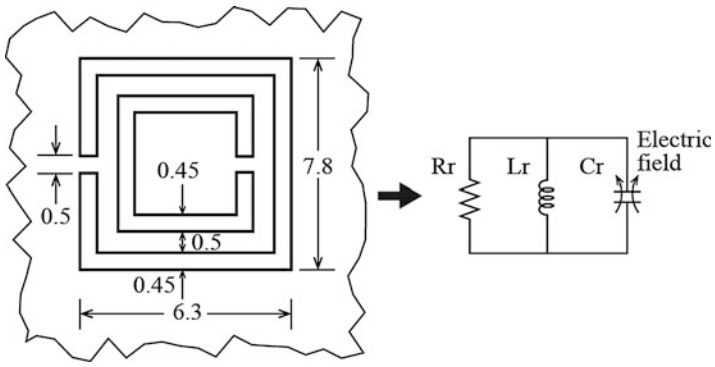


Fig. 59 CSRR structure and the equivalent circuit

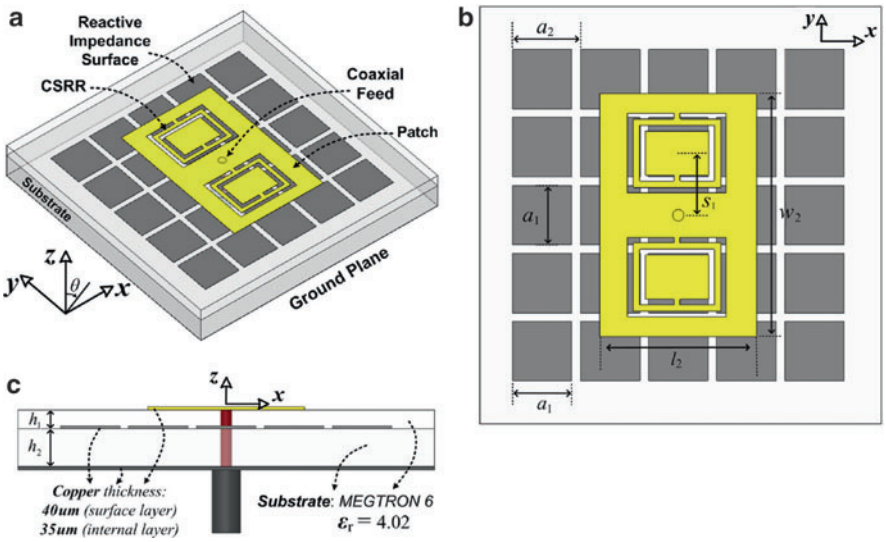


Fig. 60 Configuration of the proposed CSRR-loaded patch antenna placed on an HIS: (a) perspective view, (b) top view and (c) side view © IEEE 2015

modeled by a shunt LC resonator circuit and excited by the orthogonal electric field, so it is understood as an electric dipole placed along the ring axis. The RIS acts to further lower the resonance frequency and can improve the antenna radiation performance. Because it works as an inductor, it can store magnetic energy and increases the inductance value of the patch-type resonance. Then, the resonance frequency of the patch, which is inherently a parallel RLC resonator, is lowered, and the antenna is made small.

Configuration of a proposed antenna is shown in Fig. 60, which gives (a) perspective view, (b) top view, and (c) side view. The antenna consists of two CSRR elements face to back oriented with respect to the direction of the ring split. The antenna is fed by a coaxial probe in the center of the microstrip patch. The RIS is composed of a periodic array of 25 metallic square patches printed on a metal-backed dielectric substrate. As Fig. 60c illustrates, the antenna has three-layer structure, spacing with dielectric having $\varepsilon_r = 4.4$ and $\tan \delta = 0.009$ at 2.4 GHz. By illuminating a single cell by TEM plane wave, PEC (perfect electrically conducting) and PMC (perfect magnetically conducting) boundaries are established around the cell, which can be modeled as a parallel LC circuit, as shown in Fig. 61. The edge coupling of the square patch provides a shunt capacitor C, and the short-circuited dielectric-loaded transmission line can be modeled as a shunt inductor L. The ground size is 34 × 34 mm and the patch size is 12.4 × 19.2 mm.

DOI 10.1007/978-981-4560-75-7_48-3

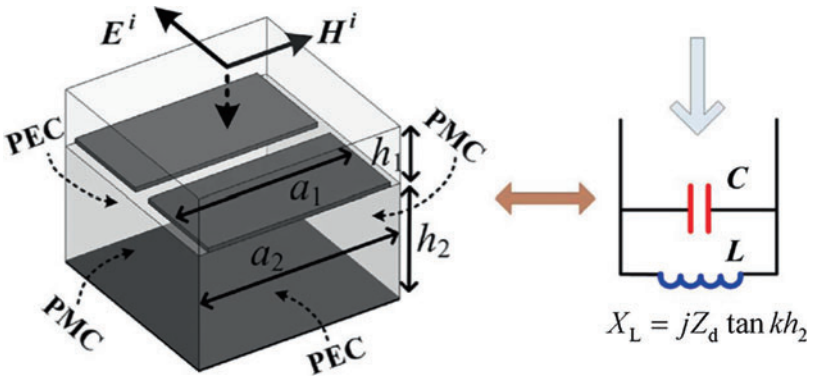


Fig. 61 Unit cell and the equivalent circuit of radiating patch over the RIS © IEEE 2015

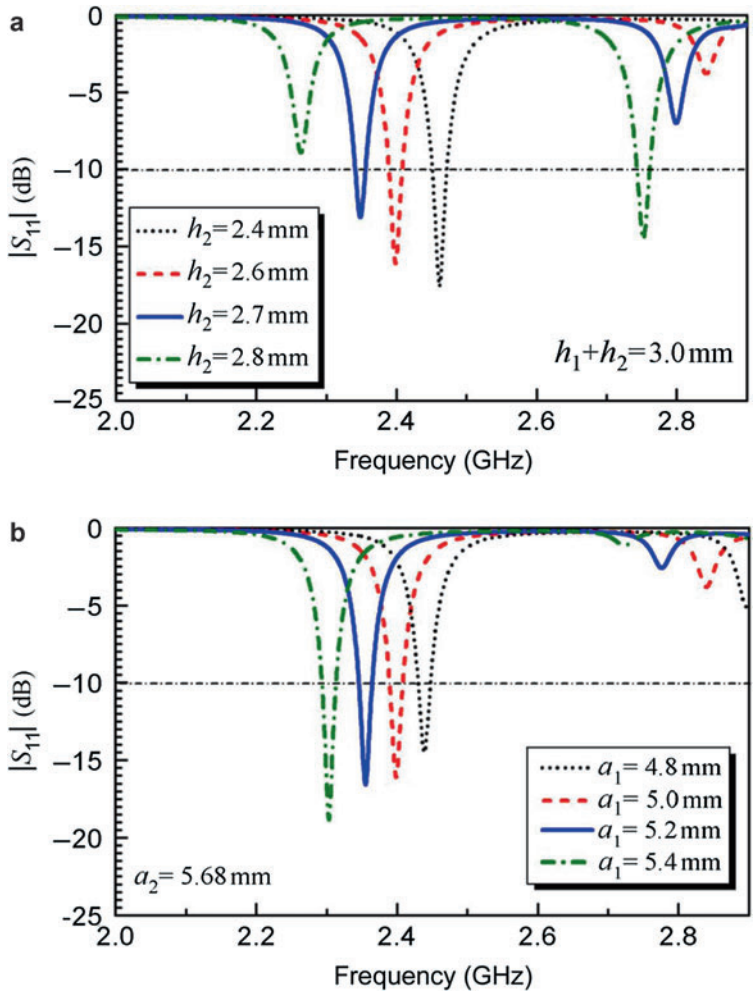


Fig. 62 Reflection coefficient S_{11}: (a) variation in h_2 with $h_1 + h_2$ fixed and (b) variation in a_1 with $a_2 = 5.68$ mm © IEEE 2015

DOI 10.1007/978-981-4560-75-7_48-3

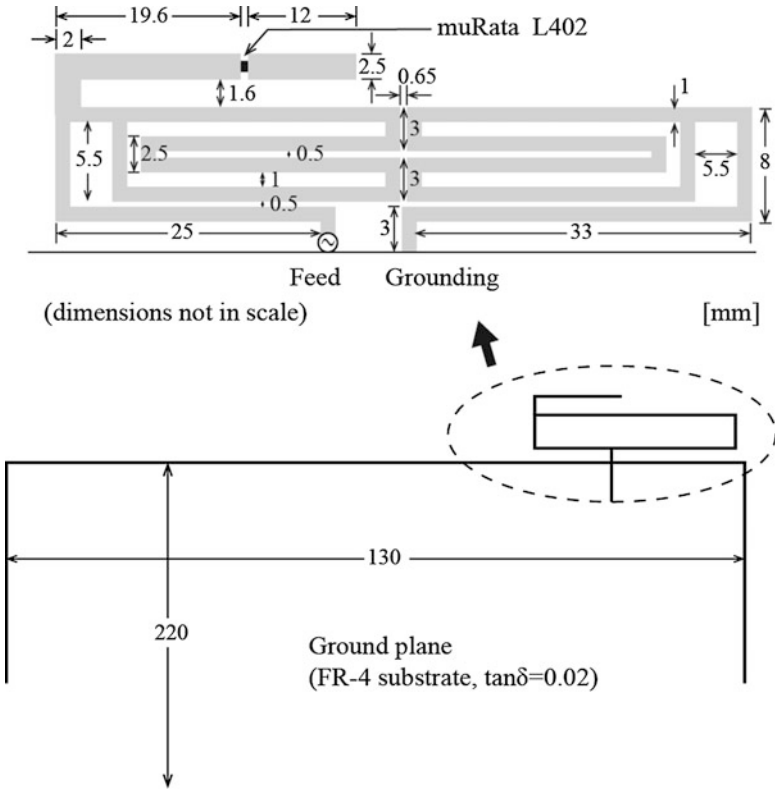


Fig. 63 Geometry of proposed antenna and parametric dimensions

An antenna, which is designed and optimized at 2.4 GHz, is analyzed parametrically. Reflection coefficients, for example, depending on h_2, keeping $h_1 + h_2 = 3$ mm, and on a_1 with $a_2 = 5.68$ mm, are shown in Fig. 62a, b, respectively.

The antenna constructed using results of the analysis is shown to perform specified functions such as circularly polarized radiation, dual- and triple-band operation with the same or different polarizations. The radiation characteristics can be controlled by changing the configurations of the CSRR. The antenna is compact, is simple in structure, and yet can be applied to versatile purposes.

PSA

Physically Small Antenna

(a.) Monopole antenna with closed-loop configuration

A PSA is designed to apply to a tablet and laptop computer (Chang and Liao 2012). It is a monopole type having closed-loop configuration as shown in Fig. 63 where its geometry and dimensions are given. The antenna was evolved from a simple loop to form a broadband antenna composed of two closed-loop elements broken from the simple loop and a folded metal strip line, on which an inductor is placed. The overall footprint of the antenna is 69 × 10 mm with the thickness of 1.6 mm of a dielectric substrate ($\varepsilon_r = 4.4$ and $\tan\delta = 0.02$). The antenna is fabricated by etching on 1.6-mm-thick fiberglass-reinforced epoxy (FR4) substrate. The overall antenna footprint is 0.16λ for

DOI 10.1007/978-981-4560-75-7_48-3

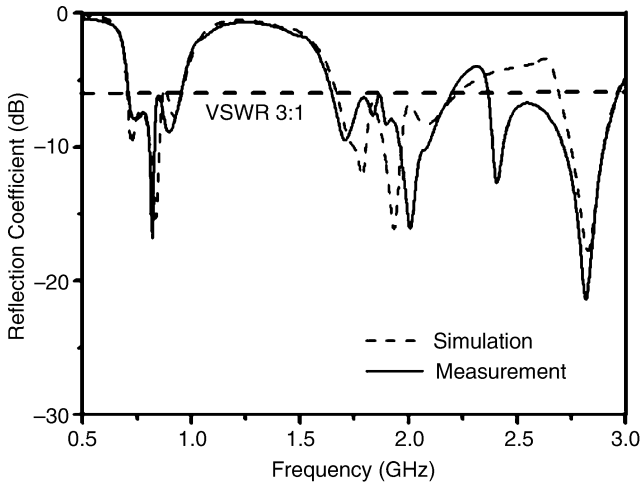


Fig. 64 Simulated and measured reflection coefficient of proposed antenna © IEEE 2015

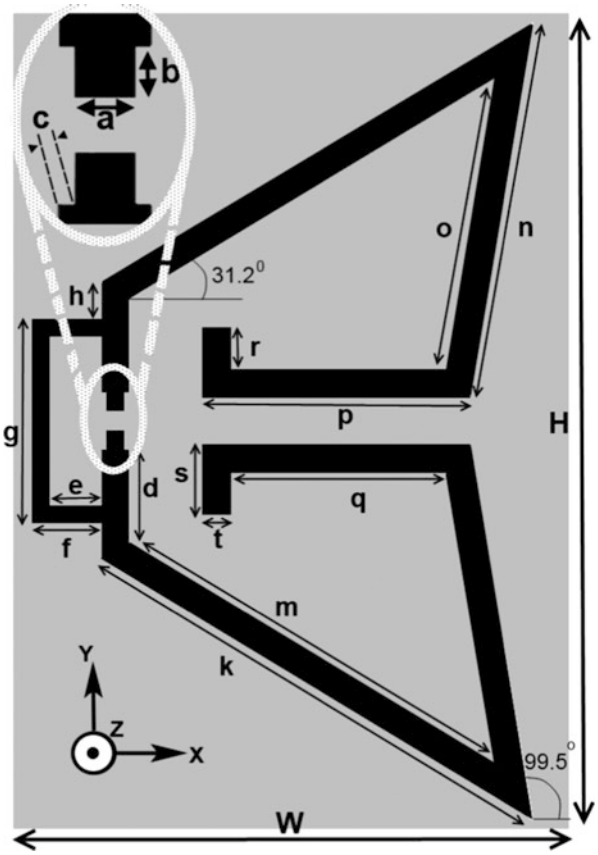


Fig. 65 Geometry of the developed antenna and the geometrical dimensions

the lowest 700-MHz bands. The fabricated antenna is applied to a tablet computer with the size of 220 x 130 mm^2 as shown in Fig. 63, mounted on the top right corner of 220 × 130 mm ground plane supporting an LCD panel of a 10-in. display. Figure 64 illustrates simulated and measured reflection coefficients of a prototype antenna mounted on a 130 × 220 mm test board. With this multiple

DOI 10.1007/978-981-4560-75-7_48-3

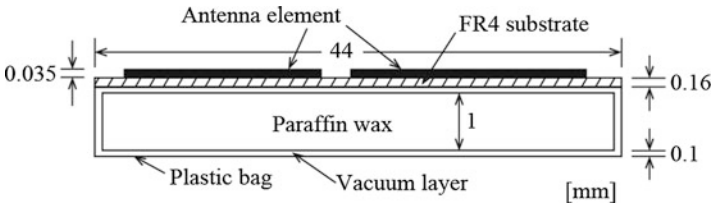


Fig. 66 Side view of the multilayer substrate of antenna shown in Fig. 65

Table 4 Geometrical parameters of the antenna (© IEEE 2015)

Line	Length (mm)	Width (mm)	Line	Length (mm)	Width (mm)
a	1		u	3	1.5
b	1		n	20.5	1.5
c	0.25		o	16	1.5
d	5	1.5	p	14.4	1.5
e	2.75	1	q	11.6	1.5
f	3.75	1	r	2.25	1.5
g	11	1	s	3.75	1.5
h	1.92	1.5	t	1.5	
k	27	1.5	H	44	
m	23	1.5	W	30	

operating capability with broadband in both lower- and higher-frequency bands, this antenna is a candidate as a portable terminal for LTE, both lower and higher band GSM and UMTS systems.

(b) Dipole structure with V shape

This is a UHF RFID tag for sensing applications (Babar et al. 2012). The antenna is a V-shaped dipole (Fig. 65) embedded on multilayer substrates, constituting 0.16-mm-thick FR4, a 0.1-mm-thick vacuumed plastic bag, and a layer of 1-mm-thick paraffin wax (Fig. 66). The paraffin wax acts as the main heat-sensitive layer of the substrates. A variation of the property and characteristics in the paraffin layer is used to detect any structural and physical changes of the substrate materials. Numerical parameters of the antenna geometry shown in Fig. 65 are given in Table 4. IC chips are laid on the terminals ***a*** and antenna feeding lines ***d*** are connected to the two arms of the V dipole. By using a V shape, the current path is made longer, and resonance frequency at the matched condition is lowered to 870 MHz with the help of a shorted line g which gives an additional inductance to make antenna size small along with a T-matching circuit. The overall antenna size is 44×30 mm^2.

The antenna is designed for a heat-sensing tag, detecting the exposure of various goods and products to high-heat environment by detuning the sensor tag. The maximum read range is shown in Fig. 67. The figure shows a frequency shift of about 30 MHz after heating. The paraffin wax acts as the heat-sensitive layer of the substrate, which changes the properties under the high-temperature environments. Change in the paraffin properties affects the dielectric properties of the whole multilayer substrate of the tag that influences the whole substrate, resulting in the change of the impedance and operating frequency of the antenna.

This type of heat-sensing tag is useful in many applications including supply chain operations and transportation of various heat-sensitive drugs and food items.

DOI 10.1007/978-981-4560-75-7_48-3

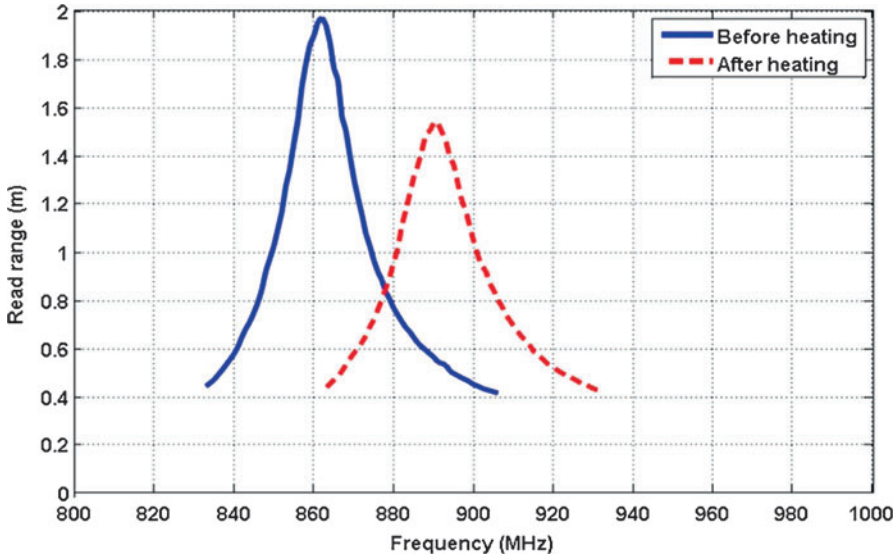


Fig. 67 The maximum read range calculated by using measured results; variation due to heat on an empty paper box on the +z axis © IEEE 2015

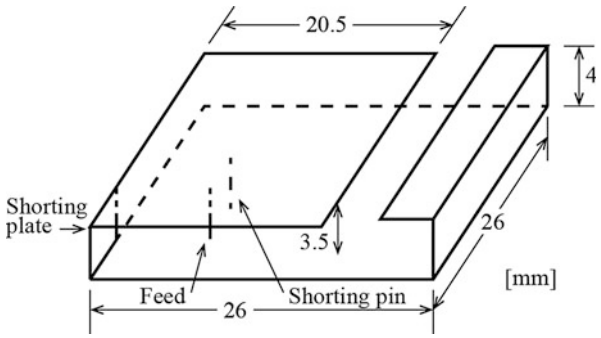


Fig. 68 The proposed antenna structure and its dimensions

(c) Modified PIFA

A compact PIFA operating in ISM band (2.45 GHz) is designed for on-body communication. The antenna (Lin et al. 2012) is composed of the rectangular planar horizontal element (patch) and folded end of the ground plane as the geometry with the dimensional parameters shown in Fig. 68. The horizontal patch (20.5 × 26 mm^2) is grounded by the small shorting plate, and the shorting pin is located near the feeding pin. The antenna size is 26 × 26 × 4 mm^3. The ground plane length ($L_1 + L_2 + L_3$) is 32 mm and the height h is 3.5 mm. With this antenna structure, the resonance frequency is 2.45 GHz. Three kinds of arm phantom are shown in Fig. 69a, while simulated reflection coefficients with those phantoms are given in Fig. 69b. These three arm phantoms are made in forms of rectangular solid, column, and rectangular solid with three layers. The phantoms are comprised of deionized water, agar salt, polyethylene powder, and TX-151. The proposed antenna is placed on a two-third muscle-equivalent phantom (dimensions of 50 × 50 × 400 mm^3, relative permittivity $\varepsilon_r = 35.2$, conductivity $\sigma = 1.16$ S/m). The antenna is put on the arm phantom 70 mm away from its end and takes 2-mm space between the antenna and the phantom. The three-layer phantom is comprised of equivalences of skin ($\varepsilon_r = 42.85$, $\sigma = 1.59$ S/m), fat ($\varepsilon_r = 5.28$, $\sigma = 0.10$ S/m), and

DOI 10.1007/978-981-4560-75-7_48-3

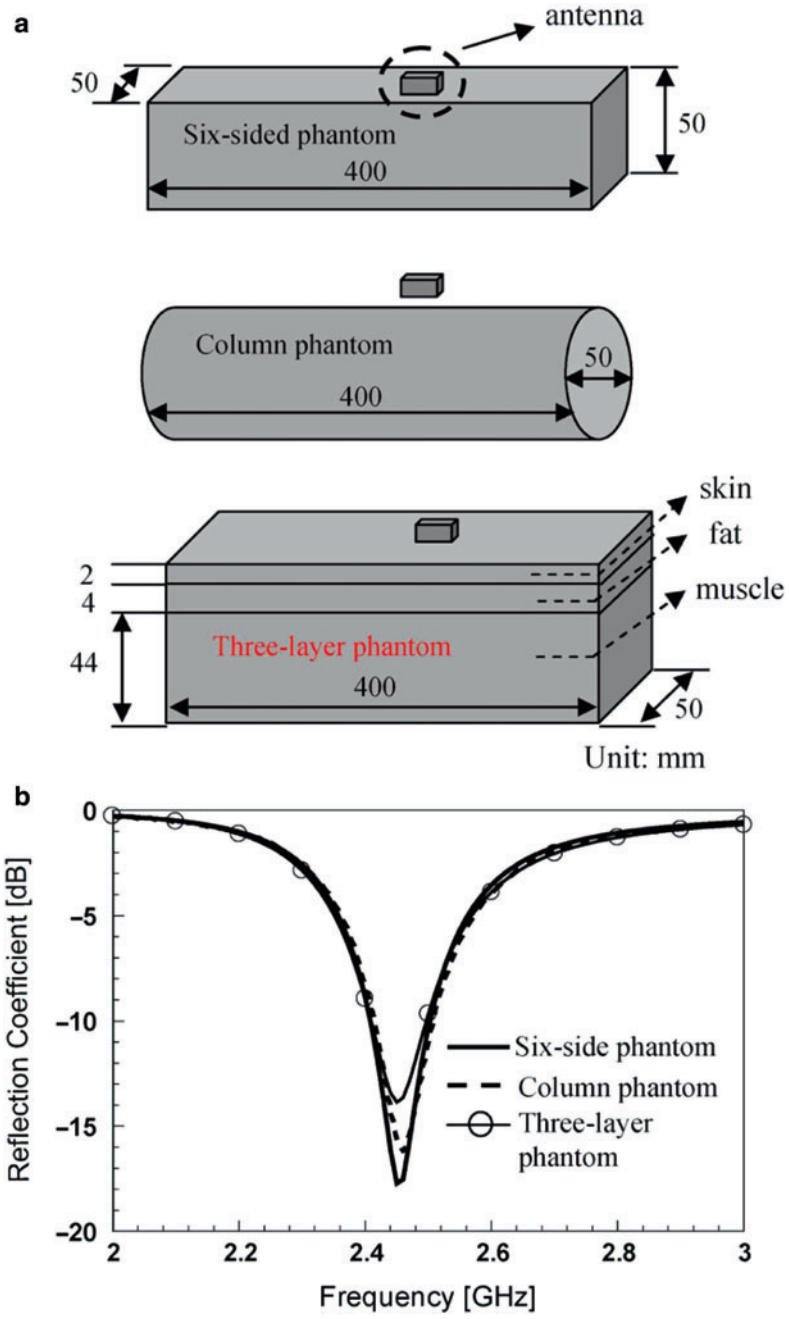


Fig. 69 (a) Three kinds of phantom and (b) simulated reflection coefficient © IEEE 2015

muscle ($\varepsilon_r = 53.59$, $\sigma = 1.81$ S/m). The resonance frequency of the antenna is affected very little when mounted on the human arm.

MW and Higher-Frequency-Range Antennas

(a) Wideband from about 5 to 25 GHz antenna

Hybrid monopole DRAs (dielectric resonator antenna) having hemispherical/conical structure illustrated in Fig. 70. (Guha et al. 2012) are described. The figure shows (a) top view of DRA either structure, (b) side view of monopole loaded with HDRDR (hemispherical dielectric ring resonator),

DOI 10.1007/978-981-4560-75-7_48-3

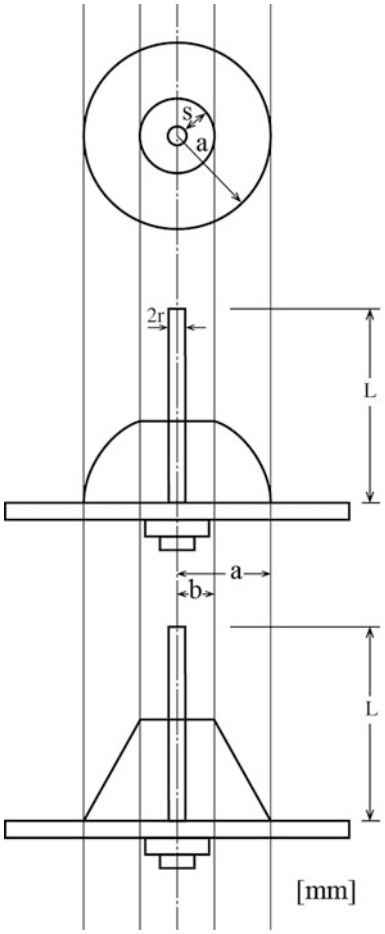


Fig. 70 Schematic diagram of the proposed antenna; from top to bottom; top view, cross sectional view, the HDRR-loaded and CoDRR-loaded monopole

Table 5 Parameters of HDRA/CoDRA (© IEEE 2015)

		Antenna parameters (mm)				
Design freq. f_L-f_H (GHz)	λ_l(cm)	*l*	*s*	*r*	*b*	*a*
#1: 4–16	6.00	15.0	1.35	0.65	2.0	6.8
#2: 5.6–22.4	4.29	10.7	1.3	0.65	1.95	4.8

and (c) side view of a monopole loaded with CoDRR (Conical Dielectric Ring Resonator). Dimensional parameters of the antenna are also given in the figure.

The antenna is designed to have ultra-wideband. The design procedure starts with specifying the operating frequency by using lowest and highest frequencies f_L and f_h, respectively, and they are set to $f_h = 4\ f_L$. Then the the design frequencies used are f_L is 4GHz in the design #1 and 5.6 GHz in the design #2. The parameters of the designed antenna are given in Table 5. In Table 5, parameters ε=10, h=a and d=0.6λ_1 are used, where d is the diameter of the ground plane and h is the height of both HDRR and CoDRR. Simulated S_{11} for HDRR and CoDRR is depicted in Fig. 71a and b, respectively, where the design is identified by #1 and #2.

DOI 10.1007/978-981-4560-75-7_48-3

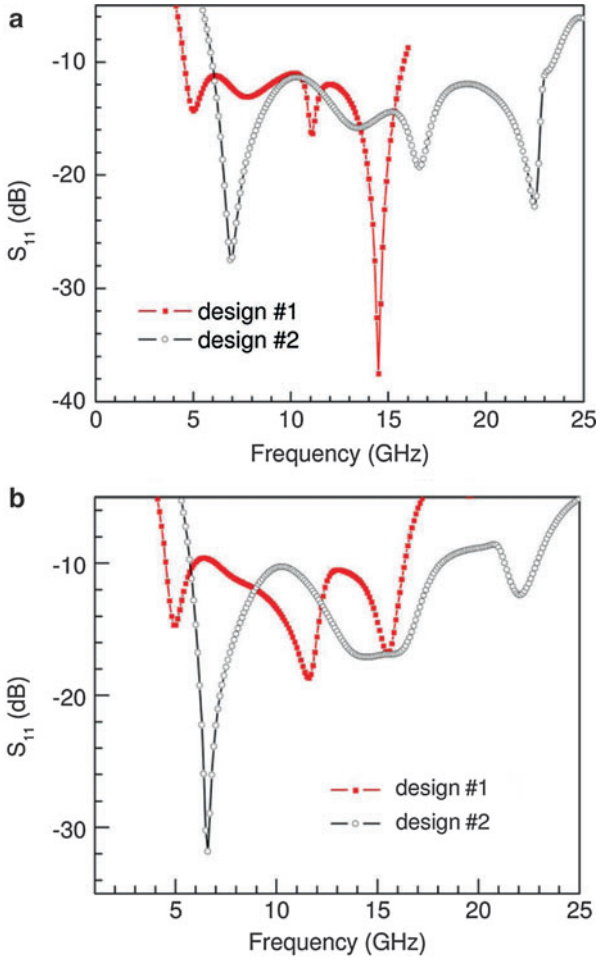


Fig. 71 Simulated S_{11} of hybrid monopole using (a) HDRR and (b) CoDRR © IEEE 2015

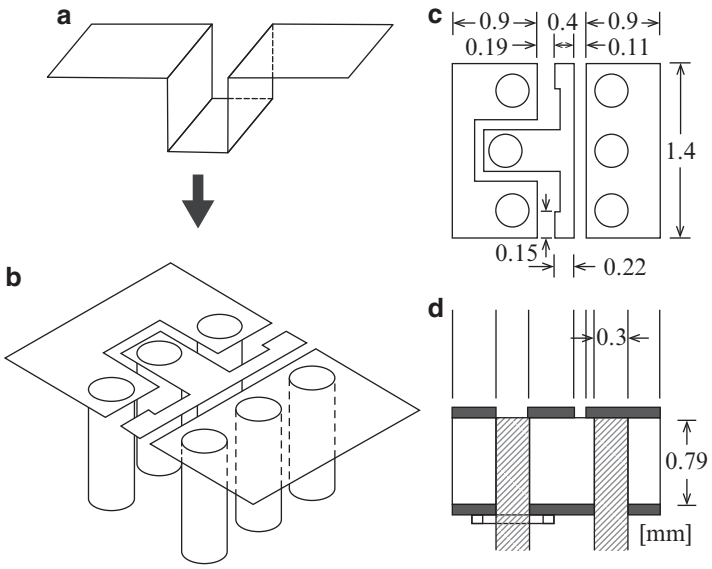


Fig. 72 Configuration of 60-GHz magneto-electric (ME) antenna; (a) principle of ME dipole, (b) 3D view and (c) side view

DOI 10.1007/978-981-4560-75-7_48-3

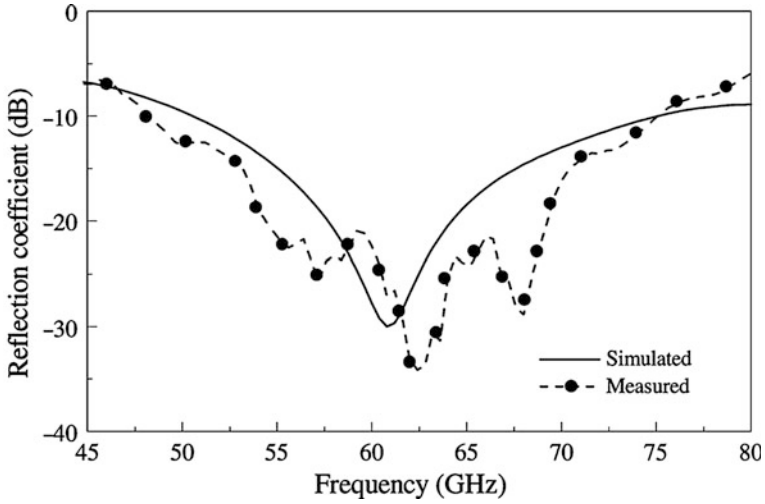


Fig. 73 Measured reflection coefficient for a plated through-hole printed ME dipole © IEEE 2015

Table 6 Dimensions of ME-dipole antenna (© IEEE 2015)

Parameters	W	$L1$	L_2	L_F	G_1	G_2	H	E_1	E_2	D
Unit (mm)	1.4	0.9	0.4	0.3	0.19	0.11	0.79	0.22	0.15	0.3
λg (guided wavelength)	0.42	0.26	0.12	0.09	0.06	0.03	0.23	0.07	0.04	0.09

The design procedure shown here is useful in producing practical antennas. A considerably compact monopole introduced here has nearly 126 % or 1 to 4 operating bandwidth with average 4 dBi peak gain. The antenna should be applicable in many areas from wide-range EM sensors to UWB communications.

(b) 60-GHz antenna

A planar quarter-wavelength (electric) dipole antenna combined with a quarter-wavelength patch (magnetic) antenna (Ng et al. 2012) is shown in Fig. 72a, while the 3D view and side view are illustrated in Fig. 72b, c, respectively. Practically, the planar dipole element is printed on a ROGERS laminate (RT/duroid 5880, $\varepsilon_r = 2.2$, thickness $h = 0.787$ mm). The quarter-wavelength patch antenna is formed by three metallic vias, fabricated by using plated-through-hole technology, which are located at the inner edge of each half planar dipole and shorted to the ground plane so they effectively form a vertically oriented antenna. Simulated and measured reflection coefficients for the prototype antenna are depicted in Fig. 73. It can be seen in the figure that the antenna has bandwidth as wide as 43 % for −10 dB S_{11} from 48 GHz to 75 GHz. Antenna gain is about 7.5 dB over 5070 GHz. Dimensions of the prototype antenna are given in Table 6.

(c) Antenna in THz and above frequency ranges

An antenna composed of bundled carbon nanotube (BCNT) (Choi and Sarabandi 2011), which is used as a conducting material to fabricate an antenna, is introduced, and its performances are evaluated in THz and above ranges. The performance is compared with the gold film, which is usually used for such antenna fabrication. The macroscopic behavior of BCNT is modeled by an isotropic resistive sheet model which is extracted from the discrete circuit model of a single-wall carbon nanotube (SWNT). By using the method of moment and mixed potential integral equation (MPIE), numerical simulations are performed to quantify radiation efficiency of resonant strip antennas composed of BCNT and gold film.

DOI 10.1007/978-981-4560-75-7_48-3

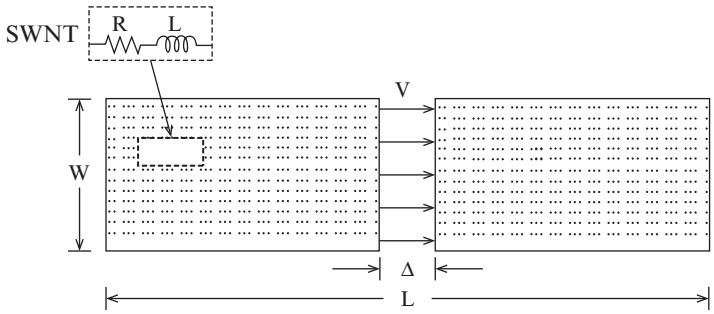


Fig. 74 Strip antenna geometry fed at the center using a thin voltage gap V; with equivalent circuit of SWNT unit

The SWNT has a cylindrical geometry with an extremely high aspect ratio and high resistance per unit length because of its very small radius in the order of nanometers. These properties limit its direct use for microwave application. To overcome this limitation, BCNT is developed to reduce the high intrinsic resistance of SWNT. Here, the performance of the BCNT is investigated as a material for fabricating antennas at THz frequencies and above. Figure 74 illustrates the proposed geometry of a strip dipole antenna constructed from strands of SWNT. In the figure, the macroscopic equivalent circuit model represented by the circuit elements of parallel-connected SWNTs in the bundle structure is given. These strands form an extremely thin conducting layer (about 10 nm), which can support the electric current along the antenna axis. In the bundle, each single nanotube, which is represented by the equivalent circuit model of the SWNT, is assumed to be connected in parallel, and electrons move only along the axis of SWNT; thus, only the electric-field component parallel to the strands can excite the surface currents.

In the THz or optical regions, the Drude model provides the conductivity of a good conductor (copper or gold) as a function of frequency. According to the Drude model, the negative real permittivity of gold at THz frequencies corresponds to an inductive reactance as opposed to the constant permittivity of free space assumed for good conductors at low frequencies.

To evaluate performance of an antenna composed of BCNT or thin gold film, the length L of the strip dipole antenna is varied to obtain fundamental resonance frequency over the range of 1–50 GHz. The antenna (Fig. 74) is excited by a thin voltage gap at the center. The antenna width W is chosen to be L/6 Antenna input impedance and reflection coefficient in terms of 50 Ω are studied to evaluate the performance of the BCNT and gold antennas operating at the same frequencies. Two strip BCNT antennas with density values of 10 and 50 CNTs/μm are considered. The antenna length is chosen to be 150 μm, which corresponds to an ideal dipole antenna operating at 1 THz. Although the density of such values can be realized by current nanofabrication facilities, antenna impedance is rather high in the range of kΩ, and the inductive reactance is also quite large, so matching to the 50-Ω load is difficult to perform. When the input impedance is varied by changing the BCNT density to above 10^3 CNTs/μm, the antenna can easily be matched without an external matching element. It is important to note that the antenna miniaturization effect is achieved by the BCNT antenna. The antenna miniaturization effect is defined by 2 L/λ, which is the normalized antenna length. As the density of BCNT increased from 10^3 to $5{\cdot}10^4$ [CNTs/μm], 2L/λ approaches unity, because the density of BCNT is inversely proportional to the inductance of the bundle and the lower density causes a lower resonance frequency, corresponding to a larger resonance wavelength. Antenna efficiency is compared between BCNT and gold film and given in Fig. 75, where (a) indicates for the densities of BCNT in the range of 10^2 to $5{\cdot}10^4$ [CNTs/μm] and (b) illustrates with magnified scale for the BCNT densities in the range of 10^4 to $5{\cdot}10^4$ [CNTs/μm].

DOI 10.1007/978-981-4560-75-7_48-3

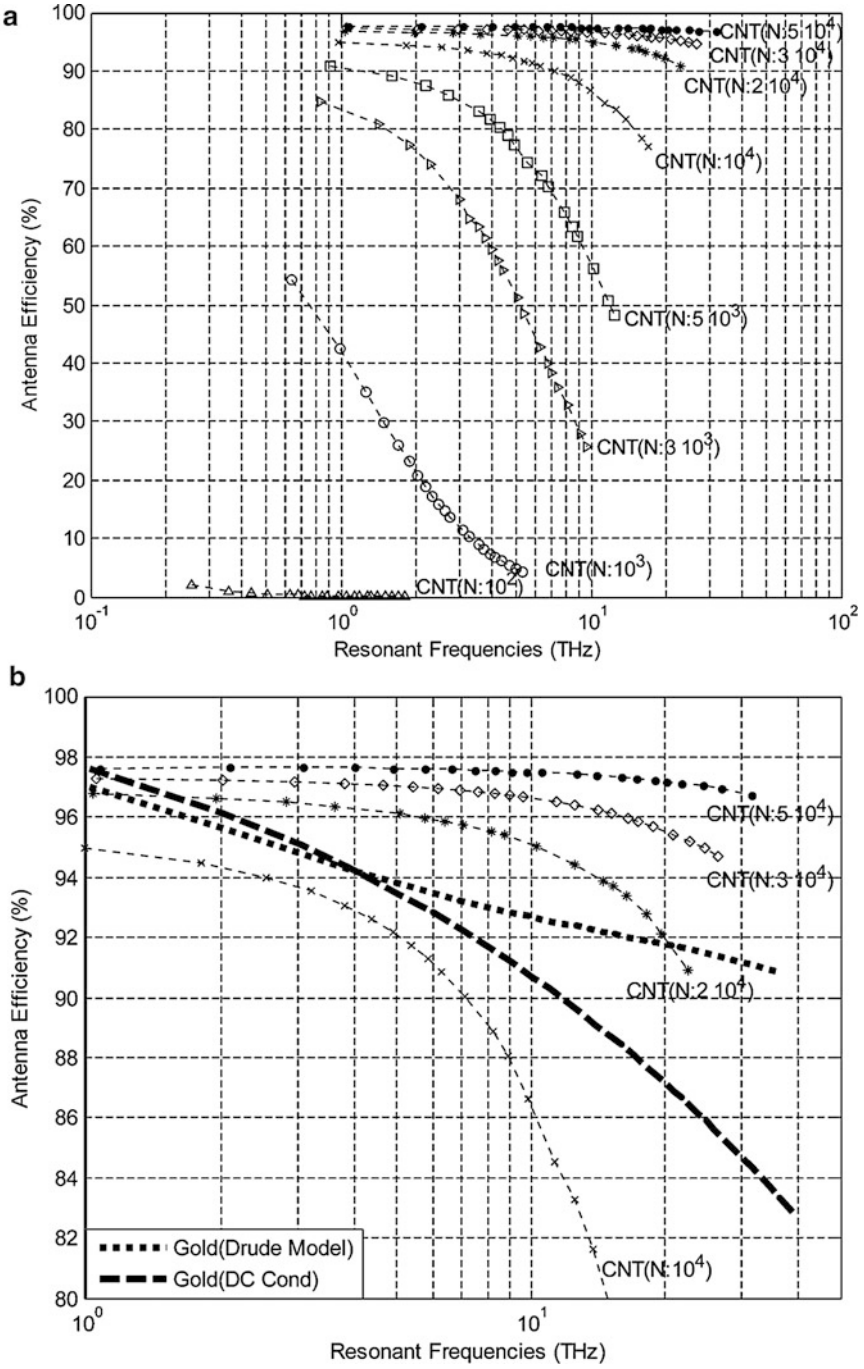


Fig. 75 Efficiency of strip antennas of BCNT's and thin gold film; (a) BCNT densities in the range of 10_3 to 5x 10_4[CNTs/ μm] and (b) a comparison of efficiency with that of thin gold film antenna © IEEE 2015

Future Directions and Open Problems

Small antennas are generally applied to numerous wireless systems and further extend to wide areas such as not only communications but also high-rate data and image transmission, radio control, and near-field systems, including RFID, wireless power transmission, body-centric communications, wearable systems, and so forth.

DOI 10.1007/978-981-4560-75-7_48-3

In each era, new systems appeared with advancement of technologies and social needs. Wireless systems were no exception, and neither were small antennas, as antenna technologies also made progress. In turn, it is true that wireless systems have also made progress along with small antennas. Impacts impressed upon wireless systems by small antennas are so great that systems to which small antennas are applied could be of little use without small antennas.

The trend is to extend further the application of small antennas to more sophisticated systems, more compact, small wireless terminals, and so forth. Antenna engineers should always encounter challenges to conquer difficulty in design and development of new small antennas that satisfy requirements of newly deployed systems, even though it might appear to exceed fundamental limitations imposed for small antennas. Integration techniques; application of electromagnetic materials, including metamaterials; and the cutting-edge design technologies expected along with the sophisticated theory will further enhance development and advancement of small antennas, which surely will give rise to a big impact on and contribute endlessly to the significantly wireless world.

SMALL ANTENNAS – BIG IMPACT!

References

Babar AA et al (2012) Passive UHF RFID tag for heat sensing applications. IEEE Trans Antennas Propag 60:4056–4064

Baena JD et al (2005) Equivalent-circuit models for split-ring resonator and complementary split-ring resonators coupled to planar transmission lines. IEEE Trans Microwave Theory Tech 53:1451–1461

Baliarda CP, Romeu J, Cardama A (2000) The Koch monopole: a small fractal antenna. IEEE Trans Antennas Propag 48:1773–1781

Ban Y-L et al (2013) Small size coupled-fed antenna with two printed distributed inductors for seven-band WWAN/LTE mobile handset. IEEE Trans Antennas Propag 61:5780–5784

Barlevy AS, Rahmat-Samii Y (2001) Characterization of electromagnetic bandgaps composed of multiple periodic tripoles with interconnecting vias: concept, analysis, and design. IEEE Trans Antennas Propag 49:343–353

Best SR (2003) A comparison of the resonant properties of small space-filling fractal antenna. IEEE Antennas Wireless Propag Lett 2:197–200

Best SR (2008) Small and fractal antennas. In: Balanis CA (ed) Modern antenna handbook. Wiley, New York, p 476

Bilotti F, Ali A, Vegni L (2008) Design of miniaturized metamaterial patch antenna with μ-negative loading. IEEE Trans Antennas Propag 56:1640–1647

Caloz C, Itoh T (2003) Novel microwave devices and structures based on the transmission line approach of meta-materials. IEEE MTT Int Symp Digest 1:195–198

Caloz C, Itoh T (2005) Electromagnetic metamaterials: transmission line theory and microwave applications. Wiley/IEEE Press, Hoboken

Chang S-H, Liao W-J (2012) A broadband LTE/ WWAN antenna design for tablet PC. IEEE Trans Antennas Propag 60:4354–4359

Chen X, Naeini, Liu Y (2003) A down-sized printed Hilbert antenna for UHF band. IEEE antennas and propagation society international symposium, pp 581–584

Choi S, Sarabandi K (2011) Performance assessment of bundled carbon nanotube for antenna application at terahertz frequencies and higher. IEEE Trans Antennas Propag 59:802–807

Chu LJ (1947) Physical limitations of omni-directional antennas. Research Laboratory of MIT, MIT tech report no 64

Cortes-Medellin G (2011) Non-Planar quasi-self complementary ultra-wideband feed antenna. IEEE Trans Antennas Propag 59:1935–1944
Dong Y, Toyao H, Itoh T (2012) Design and characterization of miniaturized patch antennas loaded with complementary split-ring resonators. IEEE Trans Antennas Propag 60:772–785
Dudley SM et al (2014) Practical issues for spectrum management with cognitive radio. IEEE Proc 102:249–250
Engheta N, Ziolkowsy RW (2006) Metamaterials: physics and engineering exploration. Wiley, Hoboken, pp 88–90
Foster RM (1968) A reactance theorem. Bell System Tech J 3:259267
Fujimoto K, Morishita H (2013) Modern small antennas. Cambridge University Press, Cambridge, UK, p 8
Fujimoto K (2005) Integrated antenna system. In: Chang K (ed) Encyclopedia of RF and microwave engineering. Wiley, New York, pp 2113–2147
Fujimoto K, Morishita H (2009) Modern small antennas. Cambridge University Press, Cambridge, UK, pp 64–65
Gianvittorio JP, Rahmat-Samii Y (2000) Fractal antenna: a novel antenna miniaturization technique and applications. IEEE Antennas Propag Mag 44:20–36
Guha D, Gupta B, Antar YMM (2012) Hybrid monopole-DR using hemispherical/conical-shaped dielectric ring resonators: improved ultrawideband designs. IEEE Trans Antennas Propag 60:393–398
Hansen RC, Collin RE (2009) A new formula for Q. Antennas Propag Mag 51:38–41
Johnk CTA (1975) Engineering electromagnetic field & waves. Wiley, New York, pp 16–17
Kildal P-S (1990) Artificial soft and hard surfaces in electromagnetics. IEEE Trans Antennas Propag 38:1537–1544
Kiourti A, Nikita K (2012) Miniature scalp-implantable antennas for telemetry in the MICS and ISM bands: design, safety considerations and link budget analysis. IEEE Trans Antennas Propag 60:3568–3575
Kolev S, Delacressonniere B, Gautier JL (2001) Using a negative capacitance to increase the tuning range of varactor diode. IEEE Trans Microwave Theory Tech 49:2425–2430
Lai A, Caloz C, Itoh T (2004) Composite right/left-handed transmission line metamaterials. IEEE Microw Mag 50:34–50
Lee Y, Tee S, Hao Y, Parini CG (2007) A compact microstrip antenna with improved bandwidth using complementary split-ring resonators (CSRR) loading. IEEE International symposium antennas and propagation and URSI radio science meeting digest, pp 1869–1872
Lee MWK, Leung KW, Chow YL (2014) IEEE Trans Antennas Propag 62:442–445
Lin C-H, Saito K, Takahashi M, Itoh K (2012) IEEE Trans Antennas Propag 60:4422–4426
Lin K-H, Chen S-L, Mittra R (2013) A looped-Bowtie RFID tag antenna design for metallic objects. IEEE Trans Antennas Propag 61:499–505
Linvill JG (1953) Transistor negative impedance converter. Proceedings of IRE. 41:725–729
McLean JS (1996) A re-examination of the fundamental limits on the radiation Q of electrically small antennas. IEEE Trans Antennas Propag 44:672–676
Mizaei H, Eleftheriades GV (2013) A resonant printed monopole antenna with an embedded non-Foster matching network. IEEE Trans Antennas Propag 61:5363–5371
Munk BA (2000) Frequency selective surface. Wiley, New York
Mushiake Y (1996a) Self-complementary antennas. Springer, London
Mushiake Y (1996b) Self-complementary antennas. Springer, London, pp 93–97

Ng KB, Wong H, So KK, Chan CH, Luk KM (2012) 60 GHz plated through hole printed magneto-electric dipole antenna. IEEE Trans Antennas Propag 60:3129–3136
Oraizi H, Hedayati S (2012) Miniaturization of microstrip antenna by the novel application of the Giuseppe Peano fractal geometries. IEEE Trans Antennas Propag 60:3559–3567
Ouedraogo RO, Rothwell EJ, Dias AR, Fuchi K, Temme A (2012) Miniaturization of patch antennas using a metamaterial-inspired technique. IEEE Trans Antennas Propag 60:2175–2181
Pendry JB, Holden AJ, Robbins DJ, Stewart WJ (1998) Low frequency plasmons in thin wire structures. J Phys Condens Matter 10:4785–4809
Pendry JB, Holden AJ, Robbins SWJ (1999) Magnetism from conduction and enhanced nonlinear phenomena. IEEE Trans Microwave Theory Tech 47:2075–2084
Perry AK (1973) Broadband antennas system realized by active circuit conjugate impedance matching. MS Thesis, Naval Postgraduate School, Monterey. Acc No. AD 769800
Rahmat-Samii Y, Mosallaei (2001) Electromagnetic band-gap structures; Classification, characterization, and application. Proceedings of 11th conference antennas and propagation, Manchester, pp 560–564
Rumsey VH (1957) Frequency independent antennas. IRE National Conv Rec. pt 1, 114–118
Rumsey VH (1981) Highlights of antenna history. IEEE Antennas and Propagation Society Newsletter, December p 8
Sagan H (1994) Space-filling curves. Springer, New York
Sievenpiper D et al (1989) High impedance electro-magnetic surfaces with a forbidden frequency band. IEEE Trans Microwave Theory Tech 47:2059–2074
Stuart HR, Pilwerbetsky A (2006) Electrically small antenna elements using negative permittivity resonator. IEEE Trans Antennas Propag 54:1644–1653
Sung Y (2012) Bandwidth enhancement of a microstrip line-fed printed wide-slot antenna with a parasitic center patch. IEEE Trans Antennas Propag 60:1712–1716
Susman-Fort SE (2006) Matching network design using non-Foster impedance. Int J RF Microwave Comput Aided Eng 16:135–142
Susman-Fort SE, Rudish RM (2009) Non-Foster impedance matching for electrically-small antennas. IEEE Trans Antennas Propag 57:2230–2241
Thal HL (2006) New radiation Q limits for spherical wire antennas. IEEE Trans Antennas Propag 54:2757–2763
Thal HL (2009) Gain and Q bounds for coupled TM-TE modes. IEEE Trans Antennas Propag 57:1879–1885
Volakis JL, Chen CC, Fujimoto K (2010a) Small antennas. McGraw-Hill, New York, pp 1–2
Volakis JL, Chen CC, Fujimoto K (2010b) Small antennas. McGraw-Hill, New York, p 103
Wang N, Liu Q, Wu C, Talbi L, Zeng Q, Xu J (2014) Wideband Fabry-Perot resonator antenna with two complementary FSS layers. IEEE Trans Antennas Propag 62:2463–2471
Werner DH, Gargiel S (2013) An overview of fractal antenna engineering research. IEEE Antennas Propag Mag 45:38–57
Wheeler HA (1947) Fundamental limitation of small antennas. Proceedings of IRE. 35:1479–1484
Wheeler HA (1958) The spherical coil as an inductor, shield, or antenna. Proceedings of IRE. 58:1595–1602
Yogo H, Kato K (1974) Circuit realization of negative impedance converter at UHF. Electron Lett 10:155–156
Zhu J et al (2004) Peano antennas. Antennas Wirel Propag Lett 3:71–74

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_49-1

Wideband Magneto-electric Dipole Antennas

Mingjian Li* and Kwai-Man Luk
Department of Electronic Engineering and State Key Laboratory of Millimeter Waves, City University of Hong Kong, Hong Kong SAR, People's Republic of China

Abstract

The magnetoelectric (ME) dipole antenna is a type of complementary antenna. The basic antenna geometry includes a planar electric dipole and a vertical shorted quarter-wave patch antenna. Traditionally, a proximity coupled feed is utilized to excite the antenna which performs as a combination of an electric dipole and a magnetic dipole. As a result, the antenna exhibits a wide impedance bandwidth, a stable gain, and a stable radiation pattern with low cross-polarization and back radiation levels over the operating frequencies. After the discourse of the fundamental magnetoelectric dipole design, several studies are devoted to antenna height reduction and alternative feed methods. And then some research works, focusing on designing magnetoelectric dipoles with different polarizations, are presented. The magnetoelectric dipole antenna is also modified for the fulfillments of various applications, such as UWB and 60-GHz wireless communications. Other than the magnetoelectric dipole, several other types of complementary antennas are also introduced.

Keywords

Magnetoelectric dipole antenna; Electric dipole; Shorted quarter-wave patch; Dual-polarized antenna; Circularly polarized antenna; Diversity antenna; Ultra-wideband antenna; 60-GHz antenna; Complementary antenna

Introduction

In 1954, Clavin proposed a new type of radiating element (Clavin 1954), as shown in Fig. 1. This antenna, consisting of an electric dipole and a coaxial line with an open end, is known as the complementary antenna. Practically, the complementary antenna is a kind of composite antenna, which comprises a pair of complementary sources. If the two sources of equal amplitude are combined in the proper phase, the resulting radiation pattern will be identical in the E- and H-planes, and furthermore, the back radiation will be suppressed. Arising from this concept, several antenna designs were proposed based on a dipole/slot (King and Owyang 1960) or a monopole/slot (Mayes et al. 1972) combination. The slot antenna was used to realize the magnetic dipole when the microstrip patch antenna was not yet developed. Excellent results in radiation patterns and bandwidth were achieved, but the antennas are not suitable for base stations in mobile communications mainly due to their bulky structures.

In 2006, Luk and Wong first utilized a dipole/shorted patch antenna combination to realize a wideband complementary antenna which was designated as the magnetoelectric (ME) dipole, as shown in Fig. 2. It was found that this antenna with excellent performance in all characteristics could be employed to develop wideband base station antennas for mobile communications. In comparison with the conventional

*Email: mingjiali2@um.cityu.edu.hk

DOI 10.1007/978-981-4560-75-7_49-1

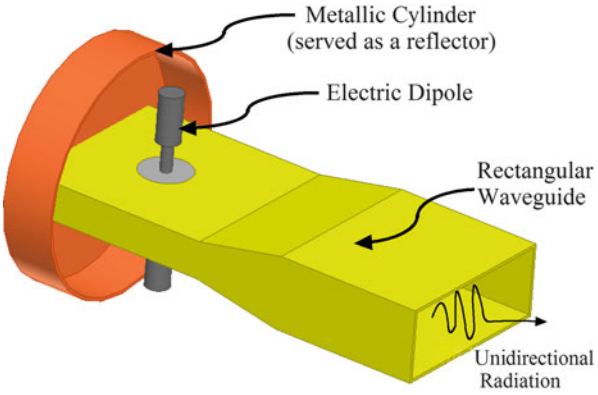


Fig. 1 Complementary antenna by Clavin (1954)

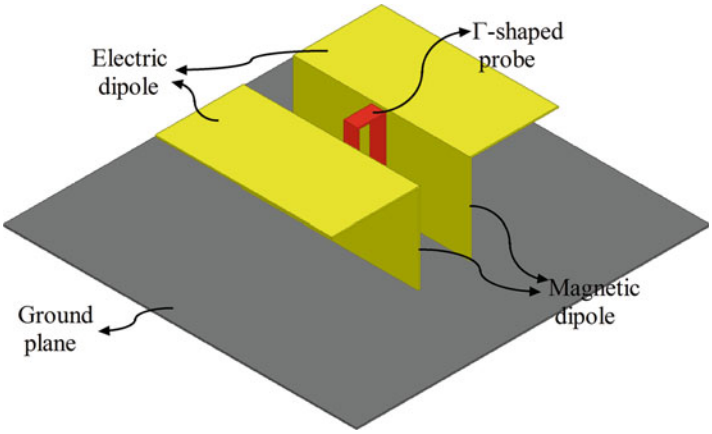


Fig. 2 Basic structure of the magnetoelectric dipole antenna

antennas such as the electric dipole antenna and microstrip antenna, the ME dipole can be used in array environment to produce wideband base station antenna with stable performance in gain, beamwidth, and radiation pattern over the operating frequencies. The advantages and disadvantages of the electric dipole, microstrip patch, and ME dipole are listed in Table 1.

Subsequently, the ME dipole antenna was developed into different forms with dual polarizations, circular polarization, or pattern and polarization diversities to fulfill various applications such as mobile, ultra-wideband and millimeter-wave communications.

In this chapter, the theory of complementary antenna is discussed in section "Theory of Complementary Antenna." The basic model of magnetoelectric dipole antenna, methods for height reduction, and other feed mechanisms are the topics of section "Basic Form of Magnetoelectric Dipole Antenna." In section "Magnetoelectric Dipole Antennas with Different Polarizations," examples with dual and circular polarizations are presented. And a magnetoelectric dipole antenna with pattern and polarization diversities is studied. Magnetoelectric dipole antennas for the ultra-wideband and millimeter-wave applications are discussed in sections "Designs for UWB Applications" and "Designs for 60-GHz Applications," respectively. Section "Other Complementary Antenna Designs" presents some other complementary designs. Concluding remarks are given in section "Conclusion."

DOI 10.1007/978-981-4560-75-7_49-1

Table 1 Comparisons between the electric dipole antenna, microstrip patch antenna, and ME dipole antenna

	Advantages	Disadvantages
Electric dipole	Wide impedance bandwidth	High profile Unstable gain over operating frequencies Unstable beamwidth over operating frequencies High cross-polarization and back radiation levels
Microstrip patch	Low profile Conformability to a shaped surface Ease of fabrication	Narrow bandwidth High cross-polarization and back radiation levels
Magnetoelectric dipole	Wide impedance bandwidth Stable gain over operating frequencies Stable beamwidth over operating frequencies Low cross-polarization and back radiation levels	High profile

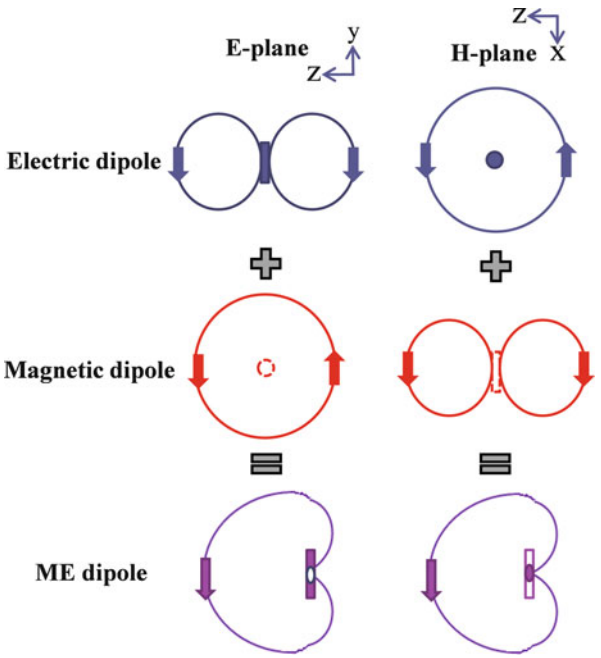


Fig. 3 Radiation mechanism of a complementary antenna

Theory of Complementary Antenna

Radiation Characteristic

Typically, a complementary antenna consists of an electric dipole and a magnetic dipole placed orthogonally and excited simultaneously. The radiation mechanism of a complementary antenna is illustrated in Fig. 3. The radiation pattern of an electric dipole looks like a figure-8 shape in the E-plane and a figure-O shape in the H-plane, whereas the radiation pattern of a magnetic dipole looks like a figure-O shape in the E-plane and a figure-8 shape in the H-plane. When the two sources of equal amplitude are combined in the proper phase, the resulting radiation pattern will have a cardiac shape with identical and symmetrical radiation patterns in the E-plane and the H-plane. More importantly, the back radiation will be zero which is desirable in cellular communications.

DOI 10.1007/978-981-4560-75-7_49-1

Analytically, an electric Hertzian dipole with a length of *dy*, placed along the *y*-axis, can be equivalent to an electric current $I = H_x dx$. Hence, the electric field in the far-field zone can be expressed as

$$\vec{E} = j\frac{Idy}{2\lambda r}\eta\left(e_\theta \cos\theta \quad \sin\varphi + e_\varphi \cos\varphi\right)e^{-jkr}. \tag{1}$$

While a magnetic Hertzian dipole with a length of *dx*, placed along the x-axis, is equivalent to a magnetic current, $I_m = E_y dy$. Hence, the electric field in the far-field zone can be expressed as

$$\vec{E} = \quad -j\frac{I_m dx}{2\lambda r}\left(e_\theta \sin\varphi + e_\varphi \cos\theta \quad \cos\varphi\right)e^{-jkr}. \tag{2}$$

A complementary antenna (or Huygens' source) can be realized by simultaneously exciting the electric Hertzian dipole and magnetic Hertzian dipole with the same magnitude in strength. Hence, the far-field electric field is obtained by adding Eqs. 1 and 2,

$$\vec{E} = j\frac{E_y dxdy}{2\lambda r}\left[e_\theta \sin\varphi(1 + \cos\theta) + e_\varphi \cos\varphi(1 + \cos\theta)\right]e^{-jkr}. \tag{3}$$

Consequently, the normalized radiation patterns in planes of any φ are the same and can be expressed as

$$F(\theta) = \frac{\sqrt{\sin^2\varphi(1 + \cos\theta)^2 + \cos^2\varphi(1 + \cos\theta)^2}}{2} = \frac{(1 + \cos\theta)}{2}. \tag{4}$$

When $\theta = 180°$, $F(\theta) = 0$, indicating that there is no back radiation. Hence, the radiation characteristic of the complementary antenna confirms that this concept can be applied for developing unidirectional antennas.

Impedance Characteristic

The impedance characteristic of complementary antenna is also investigated here. As is known to all, the microstrip patch antenna radiates as a magnetic current source. Hence, different from using the slot antenna in the classical complementary antenna, the patch antenna is employed to realize the magnetic dipole this time. An equivalent circuit of the complementary antenna is shown in Fig. 4. The patch antenna has a fundamental resonant mode which can be represented by a parallel resonant circuit (resistance R_p, capacitance C_p, and inductance L_p), whereas the electric dipole has a fundamental mode which can be represented by a series resonant circuit (resistance R_d, capacitance C_d, and inductance L_d). When the two circuits are connected in parallel, the input admittance is (Luk and Wu 2012)

$$\begin{aligned} Y_{in} &\approx \left[\frac{1}{R_d + j\left(\omega L_d - \frac{1}{\omega C_d}\right)}\right] + \left[\frac{1}{R_p} + j\left(\omega C_p - \frac{1}{\omega L_p}\right)\right] \\ &\approx \left[\frac{1}{R_d} - j\left(\omega L_d - \frac{1}{\omega C_d}\right)\frac{1}{R_d^2}\right] + \left[\frac{1}{R_p} + j\left(\omega C_p - \frac{1}{\omega L_p}\right)\right] \\ &= \left[\frac{1}{R_d} + \frac{1}{R_p}\right] - j\left[\left(\omega L_d - \frac{1}{\omega C_d}\right)\frac{1}{R_d^2} - \left(\omega C_p - \frac{1}{\omega L_p}\right)\right]. \end{aligned} \tag{5}$$

DOI 10.1007/978-981-4560-75-7_49-1

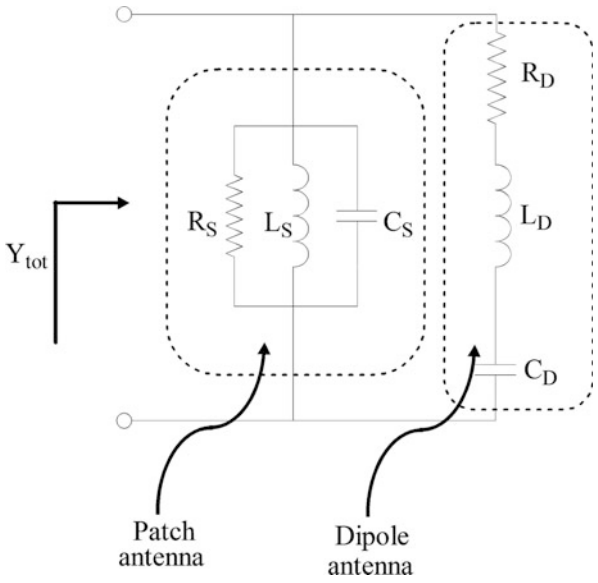


Fig. 4 Equivalent circuit of a complementary antenna

It can be observed that the imaginary part of the input admittance of the complementary antenna can be canceled out if

$$\begin{aligned} C_d L_d &= C_p L_p \\ R_d^2 &= L_d / C_p. \end{aligned} \tag{6}$$

The two equations can be satisfied simultaneously if the electric dipole and the patch antenna have the same resonant frequency, and the input resistance of the electric dipole is adjusted to a value related to the reactive components of the electric dipole and the patch antenna. Although this is not a rigorous proof, the result provides us with the insight that the complementary antenna can be wide in bandwidth if the dimensions of the antenna are selected appropriately.

Basic Form of Magnetoelectric Dipole Antenna

Antenna Configuration

According to the complementary antenna concept studied in the previous section, the magnetoelectric dipole antenna is developed by combining an electric dipole and a magnetic dipole. In the basic form, the electric dipole part is realized by a planar half-wavelength dipole as shown in Fig. 5a. The magnetic dipole part is implemented with the use of a shorted quarter-wavelength patch antenna as shown in Fig. 5b. From the classical cavity model theory, the radiation of a shorted quarter-wave patch antenna is mainly from the open end which radiates as a magnetic current. To combine the two antennas, the planar electric dipole, placed in front of a ground plane with a distance of quarter wavelength, has the inner edges shorted to the ground plane through two vertical metallic walls as shown in Fig. 5c. The metallic walls together with the ground plane between them form a vertical shorted quarter-wave patch antenna with a thickness of approximately 10 % of the operating wavelength. The feed of the antenna is a Γ-shaped probe, consisting of three portions. The first portion incorporated with its adjacent vertical metallic wall performs as an air microstrip line with 50-Ω characteristic impedance. This portion transmits the electric signal from the coaxial launcher to the second portion of the feed which is responsible to couple the electrical energy to

DOI 10.1007/978-981-4560-75-7_49-1

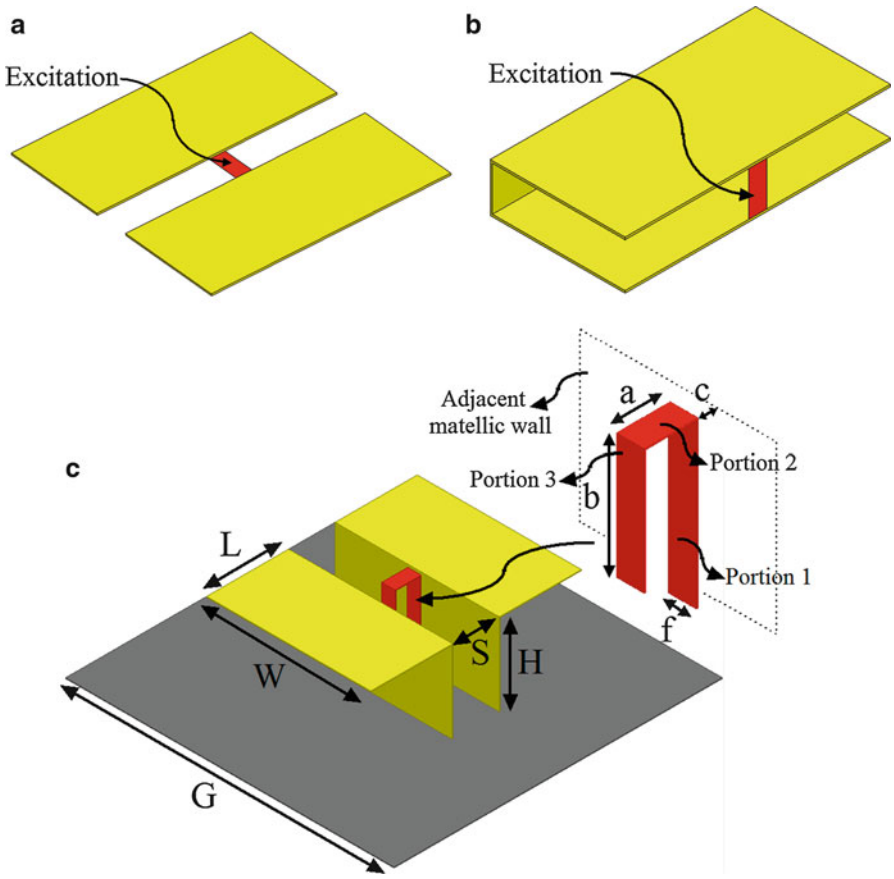


Fig. 5 (**a**) Planar electric dipole antenna, (**b**) shorted quarter-wave patch antenna, and (**c**) magnetoelectric dipole antenna

Table 2 Dimensions for the ME dipole antenna

Parameter	L	W	S	H	G
Value (mm)	30	60	17	30	160
	$0.25\lambda_0$	$0.50\lambda_0$	$0.14\lambda_0$	$0.25\lambda_0$	$1.33\lambda_0$
Parameter	a	b	c	f	
Value (mm)	9.5	22	1	4.91	
	$0.08\lambda_0$	$0.18\lambda_0$		$0.04\lambda_0$	

the magnetoelectric dipole. The third portion together with the vertical metallic wall acts as an open circuited microstrip stub which compensates the inductance introduced by the second portion.

Example of a Magnetoelectric Dipole Antenna

In the original study of Luk and Wong (2006), the performance of the antenna was demonstrated experimentally. The dimensions of the ME dipole operating at 2.5 GHz are shown in Table 2. Results for the impedance and radiation characteristics of one such antenna are shown in Figs. 6 and 7. The impedance bandwidth ($S_{11} \leq -10$ dB) is 49.1 % from 1.80 to 2.97 GHz. The antenna gain was approximately 7.9 dBi, varying from 7.54 to 8.14 dBi. The measured radiation patterns were stable across the band. The E-plane and H-plane beamwidths were generally equal. The cross-polarization and back radiation levels were very low across the entire operating bandwidth.

DOI 10.1007/978-981-4560-75-7_49-1

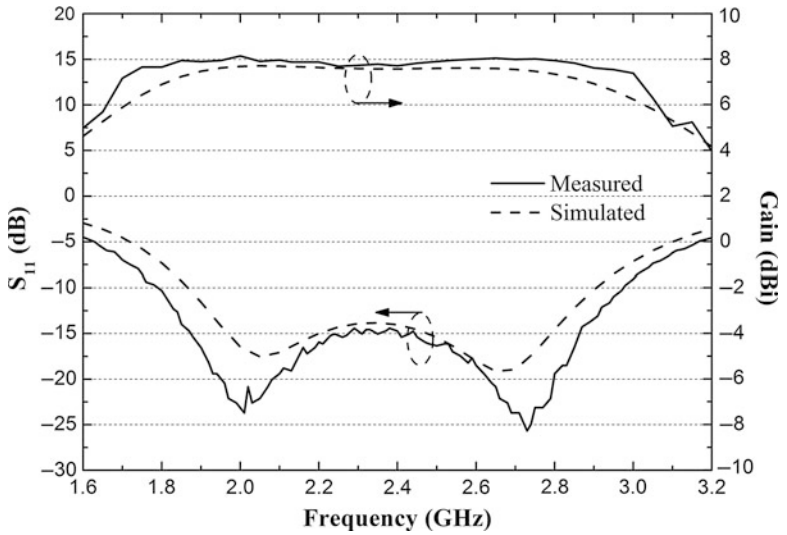


Fig. 6 Simulated and measured SWRs and gains

Magnetic Dipole

Geometrically, there is not much difference between this antenna and the common reflector-backed electric dipole excited by a conventional balun. However, the vertical metallic walls with the Γ-shaped probe make a substantial difference in the performance of the antenna. Three scenarios are studied here as shown in Fig. 8, i.e., ME dipole, planer electric dipole, and electric dipole. The simulations demonstrate that if the width of the vertical wall is reduced to $W_1 = 10$ mm, the beamwidth in H-plane becomes larger than that in E-plane and the back lobe is increased. The reason is that without the help of the magnetic dipole, the antenna works as a reflector-backed electric dipole.

The vertical shorted patch antenna has a height of quarter wavelength. Using the IE3D simulator, the SWR and gain are studied for various heights as shown in Fig. 9. It can be seen that if H is decreased such that the shorted quarter-wave patch antenna is reduced in resonant length, the second resonance is shifted upwards while the first resonance frequency is fixed at 2 GHz. This confirms that the second resonance is controlled by the size of the quarter-wave path antenna.

Electric Dipole

Typically, the planar electric dipole has a length of approximately half wavelength. The SWR and gain are studied for various dipole lengths as shown in Fig. 10. If L is increased such that the electric dipole is enlarged, the first resonance is shifted to a lower frequency, while the second resonance is fixed at 2.7 GHz. This confirms that the first resonance is mainly affected by the electric dipole.

Γ-Shaped Probe

Basically, the Γ-shaped probe performs as an air microstrip line connected to an L-shaped probe. It has a significant effect on the impedance matching of the antenna. The effect of the length of the Γ-shaped probe, **a**, is investigated as shown in Fig. 11. Reducing **a**, the first resonance moves downwards with a fixed second resonance and the in-band SWR is increased. The length of **b** of the Γ-shaped probe is also a crucial parameter on the impedance matching. The result is shown in Fig. 12. When **b** is decreased, the second resonance moves upwards with the first resonance fixed and the impedance matching at lower frequencies deteriorates.

DOI 10.1007/978-981-4560-75-7_49-1

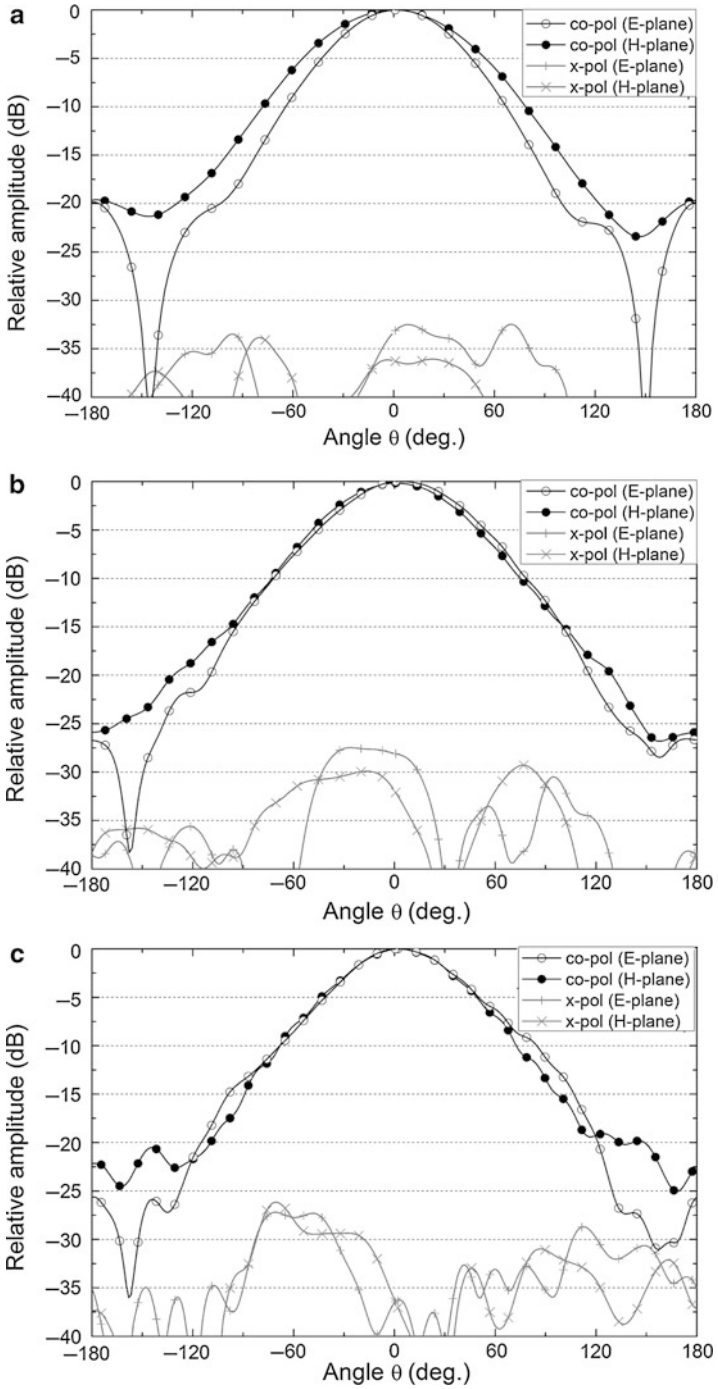


Fig. 7 Measured radiation pattern at 1.75, 2.5, and 3.0 GHz

DOI 10.1007/978-981-4560-75-7_49-1

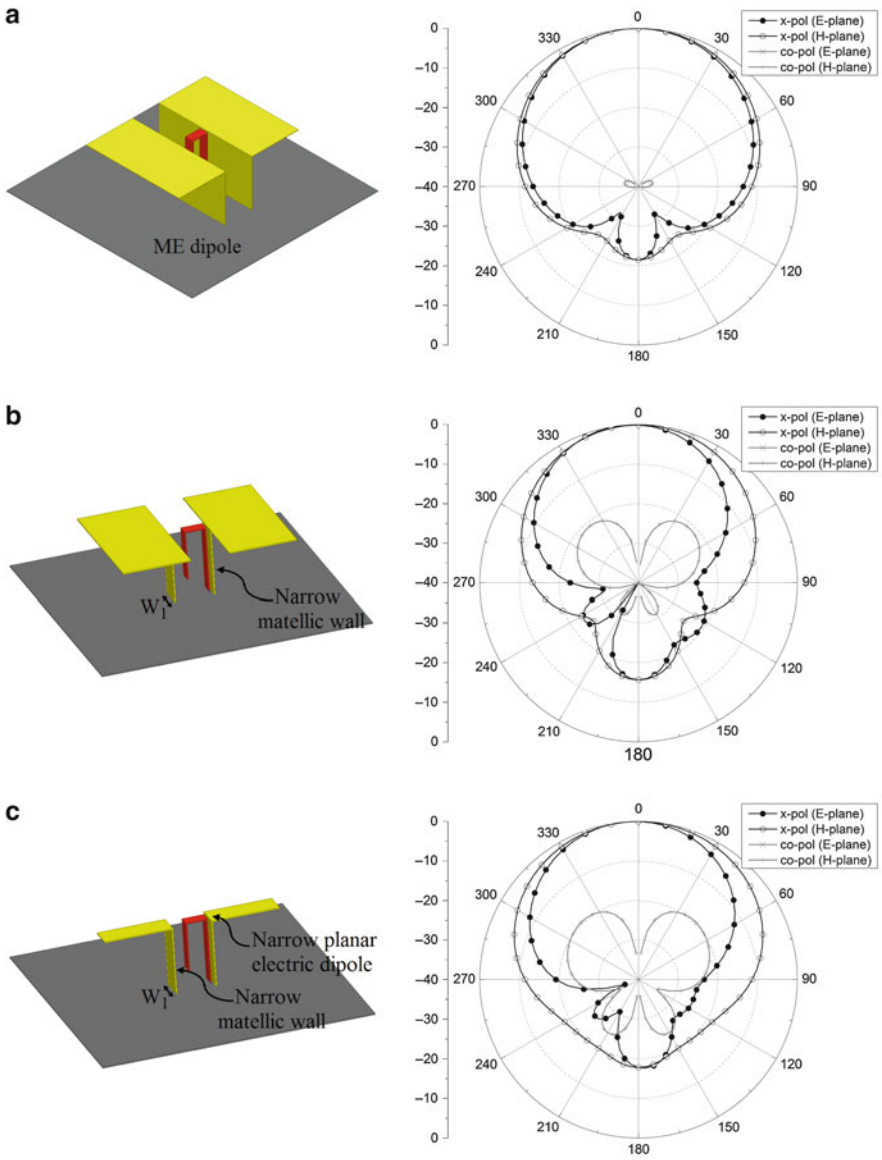


Fig. 8 (**a**) ME dipole and its radiation pattern, (**b**) planar electric dipole and its radiation pattern, and (**c**) thin wire electric dipole and its radiation pattern

Ground Plane

The ground plane acting as the reflector of the antenna has a significant effect on its performance, so the ground plane size effects have to be studied carefully. One extreme case is that the ME dipole does not have a ground plane. The radiation pattern at the center frequency is shown in Fig. 13a. As the aforementioned complementary antenna concept, the ideal complementary antenna has no back radiation. However, the ME dipole without a ground plane has a back lobe of approximately −7 dB. The reason is that when designing the ME dipole antenna, the antenna and the ground plane should be considered as a

DOI 10.1007/978-981-4560-75-7_49-1

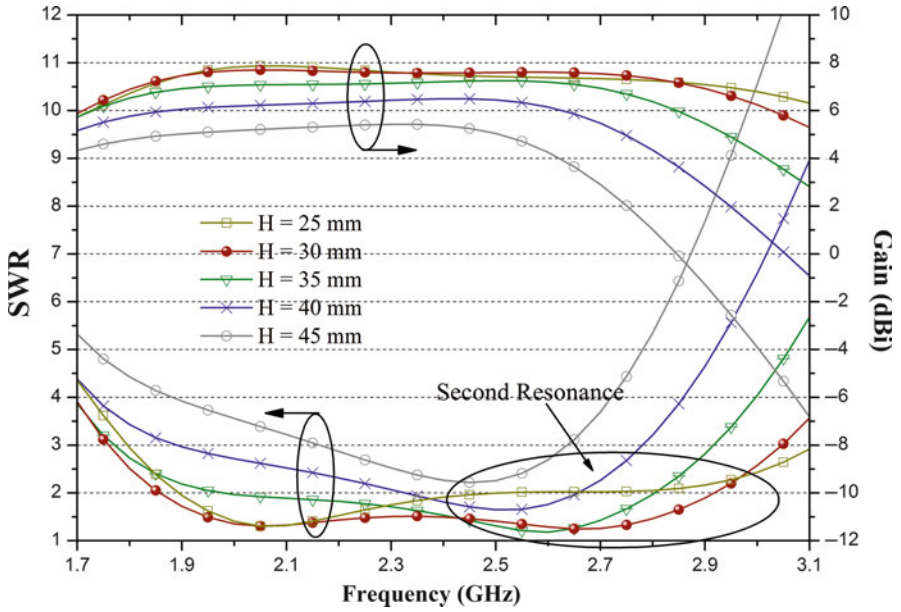


Fig. 9 Effect of the height of antenna H

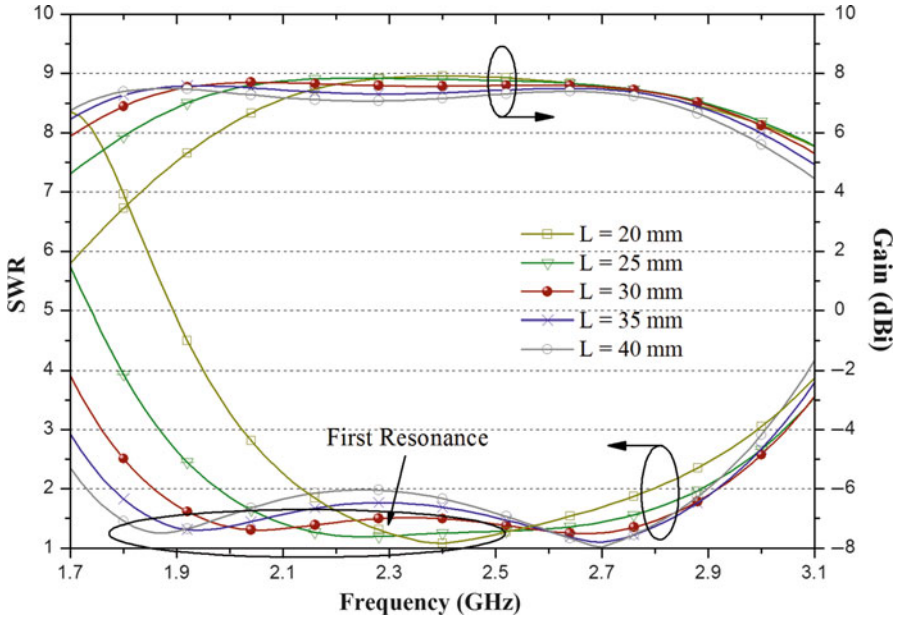


Fig. 10 Effect of the length of the planar electric dipole L

whole. In most cases, the ME dipole antenna cannot be mounted without a ground plane. It has to be noted that the ground plane between the vertical walls is a portion of the quarter-wave patch antenna and cannot be removed. As shown in Fig. 13b, c, and d, the back radiation increases with reducing G. The SWR and gain are also studied with different G, as shown in Fig. 14. The simulations demonstrate that as G is increased, the antenna gain is enhanced. But the ground plane size has no effect on the antenna impedance matching.

Design Guideline

Based on the above study, a design guide is recommended as follows:

DOI 10.1007/978-981-4560-75-7_49-1

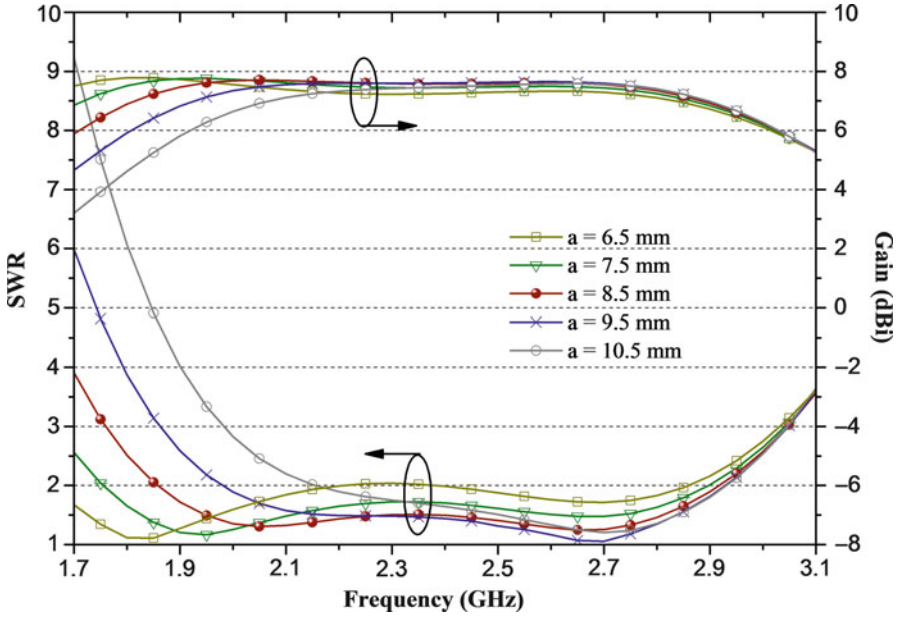


Fig. 11 Effect of the length of the second portion of the Γ-shaped probe

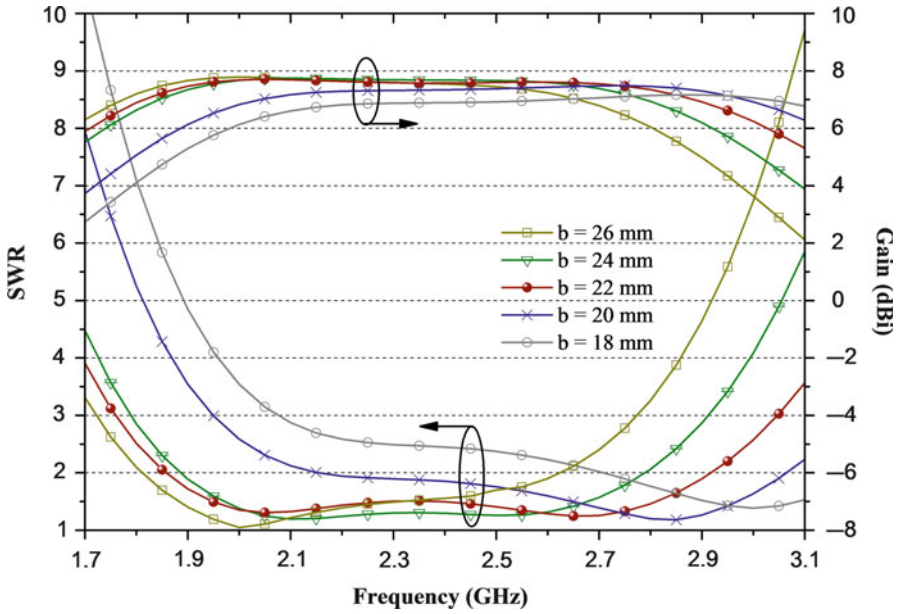


Fig. 12 Effect of the length of the third portion of the Γ-shaped probe

1. According to the required center frequency of the operating band (λ_0), determine the dimensions of the planar electric dipole antenna with $L = 0.25\lambda_0$ and $W = 0.5\lambda_0$.
2. Determine the dimensions of the shorted quarter-wave patch antenna with $H = 0.25\lambda_0$, $W = 0.5\lambda_0$, and $S \approx 0.14\lambda_0$.
3. Choose the ground plane with a proper size $G \geq \lambda_0$.
4. Design a Γ-shaped probe which has the air microstrip line portion with 50-Ω characteristic impedance. And finally tune **a** and **b** to achieve a good impedance matching.

DOI 10.1007/978-981-4560-75-7_49-1

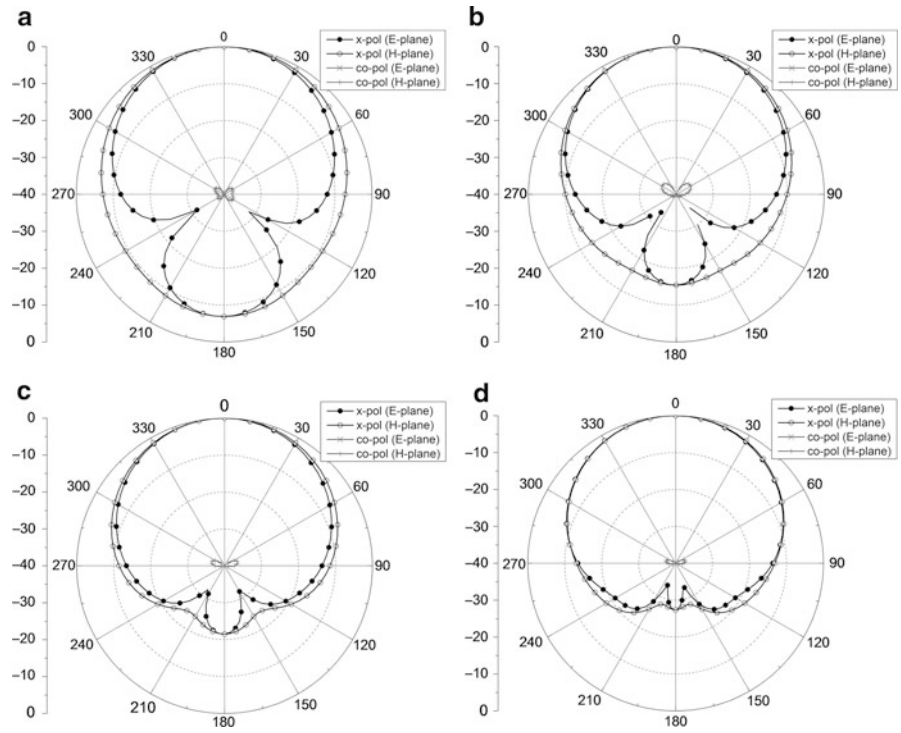


Fig. 13 Radiation patterns of ME dipole on ground plane with various length: (**a**) no ground plane, (**b**) G = 80 mm, (**c**) G = 120 mm, and (**d**) G = 160 mm

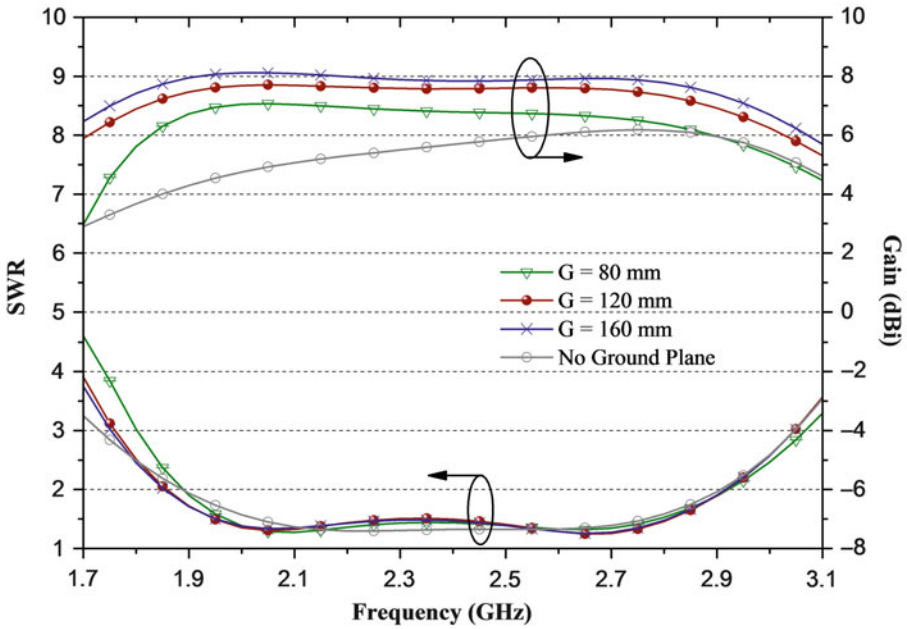


Fig. 14 Effect of the ground plane length G

DOI 10.1007/978-981-4560-75-7_49-1

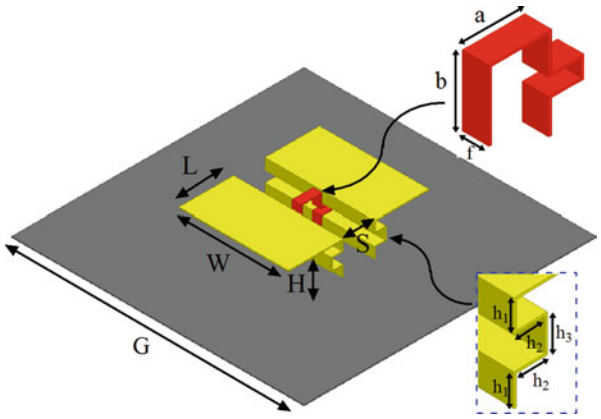


Fig. 15 Geometry of the folded ME dipole antenna

Table 3 Dimensions for the folded ME dipole antenna

Parameter	L	W	S	H	G	
Value (mm)	30	60	17	18	160	
	$0.29\lambda_0$	$0.59\lambda_0$	$0.17\lambda_0$	$0.17\lambda_0$	$1.5\lambda_0$	
Parameter	h_1	h_2	h_3	a	b	f
Value (mm)	6	6	6	11.5	13	4.91
	$0.06\lambda_0$	$0.06\lambda_0$	$0.06\lambda_0$	$0.11\lambda_0$	$0.13\lambda_0$	$0.05\lambda_0$

Methods of Reducing Height

Compared to microstrip patch antenna, ME dipole antenna has many good electrical characteristics. However, a notable drawback is that the ME dipole has a large height of approximately $0.25\lambda_0$, which may be undesirable for some practical applications. Hence, several methods to reduce the antenna height are introduced below.

Folded Magnetoelectric Dipole

A straightforward approach to the height reduction of ME dipole is to fold the vertical metallic walls as shown in Fig. 15. The folded ME dipole antenna consists of a planar electric dipole, a folded shorted quarter-wave patch antenna, and a folded Γ-shaped probe. The antenna height is $H = 2 \times h_1 + h_3$, whereas the electrical length of the shorted patch antenna is $2 \times h_1 + 2 \times h_2 + h_3$. Hence, this method confirms that the height of the antenna is reduced while keeping the physical length of the shorted patch antenna quarter wavelength. The original Γ-shaped probe is modified. The 50-Ω air microstrip line portion of the probe is folded and placed along the adjacent folded vertical metallic wall. The dimensions of the folded ME dipole operating at 2.95 GHz are shown in Table 3. Experimentally, the antenna with a height of $0.176\lambda_0$ (i.e., 18 mm, $h_1 = h_2 = h_3 = 6$ mm) was constructed and measured. Fig. 16 shows the simulated and measured SWRs and gains. It can be seen that the antenna exhibits a wide impedance bandwidth (SWR $\leq$2) of 46 % from 2.3 to 3.66 GHz and a gain of approximately 7.7 dBi over the operating frequency band. The radiation pattern at 2.8 GHz is shown in Fig. 17. It should be noted that the antenna height can be reduced further by using this method. With the decrease of the antenna height, the operating band is narrowed and shifted upwards. A design guide for this antenna is recommended as follows:

DOI 10.1007/978-981-4560-75-7_49-1

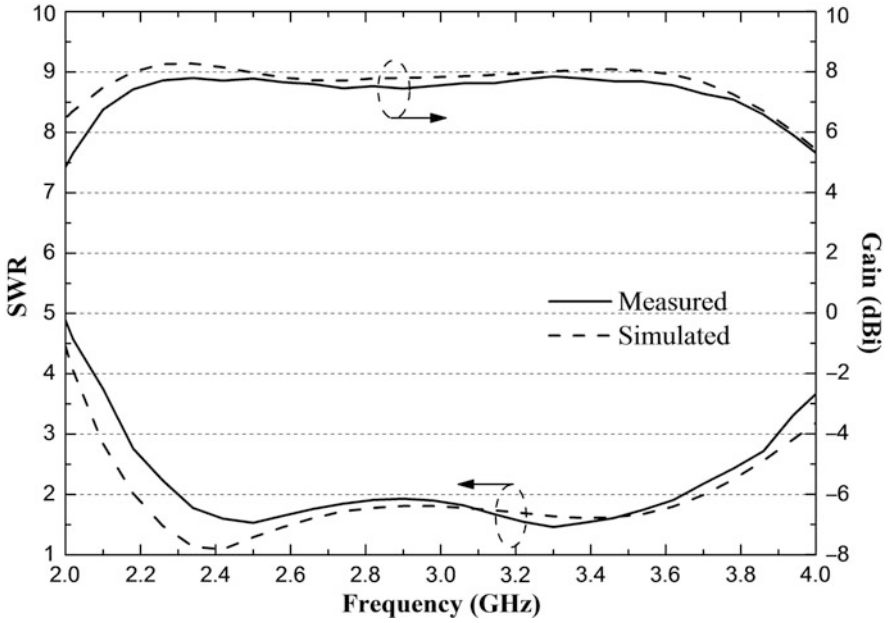


Fig. 16 Simulated and measured SWRs and gains

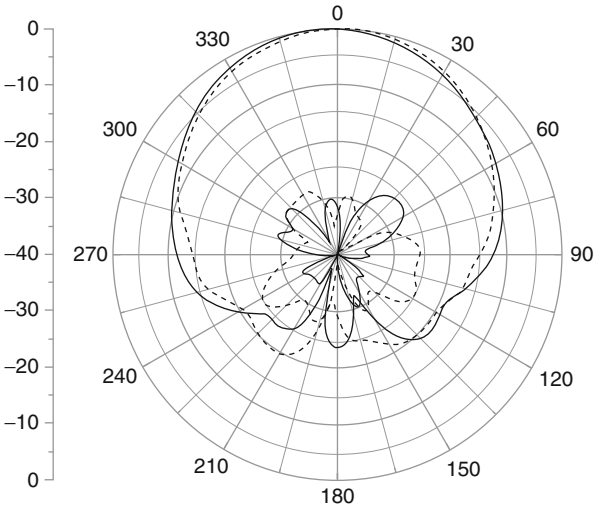


Fig. 17 Measured radiation pattern at 2.8 GHz (*black solid line* for copolarization in E-plane, *black dashed line* for copolarization in H-plane, *gray solid line* for cross-polarization in E-plane, and *gray dashed line* for cross-polarization in H-plane)

1. According to the required center frequency of the operating band (λ_0), determine the dimensions of the planar electric dipole antenna with $L = 0.29\lambda_0$ and $W = 0.59\lambda_0$.
2. According to the required antenna height, fold the shorted patch antenna with the physical length of the shorted patch antenna be approximately quarter wavelength.
3. Choose the ground plane with a proper size ($G \geq \lambda_0$).
4. Design a folded Γ-shaped probe. The air microstrip line portion with 50-Ω characteristic impedance is placed along a folded vertical metallic wall. And tune **a** and **b** to achieve a good impedance matching.

DOI 10.1007/978-981-4560-75-7_49-1

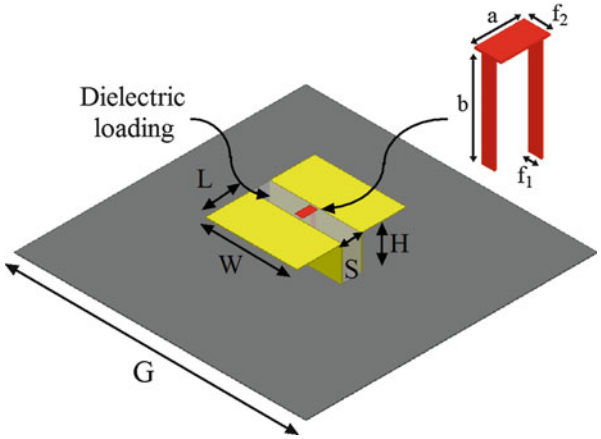


Fig. 18 Geometry of the ME dipole antenna with dielectric loading

Table 4 Dimensions for the ME dipole antenna with dielectric loading

Parameter	L	W	S	H	G
Value (mm)	23	47	10	18	150
	$0.20\lambda_0$	$0.42\lambda_0$	$0.09\lambda_0$	$0.16\lambda_0$	$1.34\lambda_0$
Parameter	a	b	f_1	f_1	
Value (mm)	7.7	15	4	2	
	$0.1\lambda_0$	$0.13\lambda_0$	$0.02\lambda_0$	$0.01\lambda_0$	

Magnetoelectric Dipole with Dielectric Loading

Another technique for reducing the ME dipole height is to use a higher-permittivity dielectric substrate instead of an air substrate in the shorted quarter-wave patch antenna. Siu et al. (2009) presented experimental study of an ME dipole with dielectric loading. The geometry of the antenna is depicted as shown in Fig. 18. It consists of a planar electric dipole, a dielectric-loaded shorted quarter-wave patch antenna, and a Γ-shaped probe. The region between the two vertical metallic walls is fully filled with the dielectric material with permittivity of 2.65; therefore, the height of the antenna is reduced to $0.16\lambda_0$, and the height can also be reduced further by using a substrate with higher permittivity. The dimensions of the ME dipole with dielectric loading operating at 2.67 GHz are shown in Table 4. As shown in Fig. 19, the antenna exhibits a wide impedance bandwidth ($S_{11} \leq -10$) of 49 % from 2.0 to 3.3 GHz and a gain of approximately 8.1 dBi. The radiation pattern at 2.5 GHz is illustrated in Fig. 20. The cross-polarization level is approximately −18 dB which is quite high in comparison with that of the original ME dipole design, which is probably due to the excitation of higher order modes in the shorted quarter-wave patch antenna.

Using Different Magnetic Dipoles

By using the aforementioned methods, the ME dipole height can be reduced effectively. However, it is obvious that the antenna structure becomes complex and the operating bandwidth is reduced. As a matter of fact, the height of ME dipole is mainly determined by the magnetic dipole which is a shorted quarter-wave patch antenna in the original ME dipole design. By utilizing other types of magnetic dipole, the antenna height can also be reduced while maintaining good electrical characteristics. In this section, two examples of ME dipole antennas employing a loop antenna as the magnetic dipole are presented.

DOI 10.1007/978-981-4560-75-7_49-1

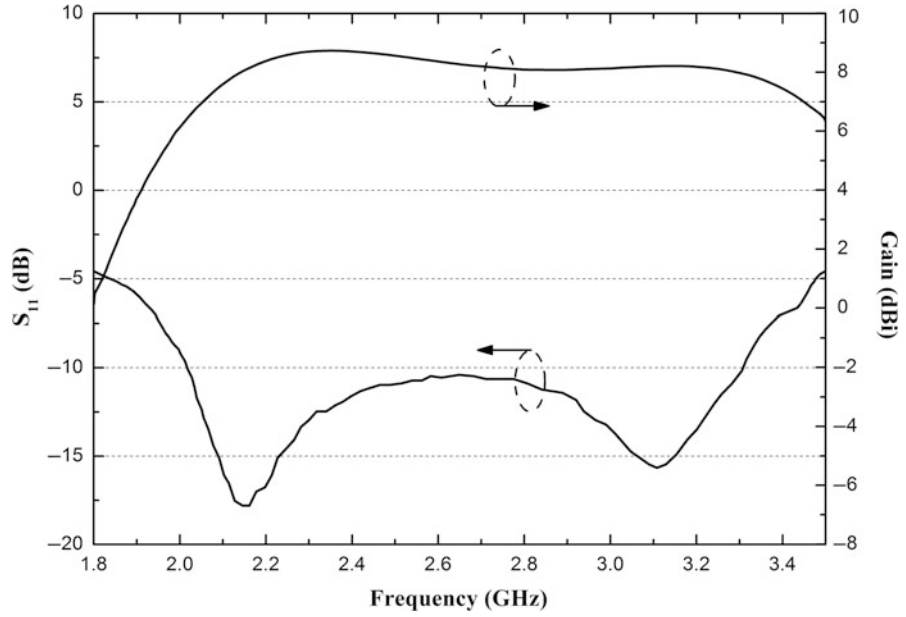


Fig. 19 Measured S_{11} and gain

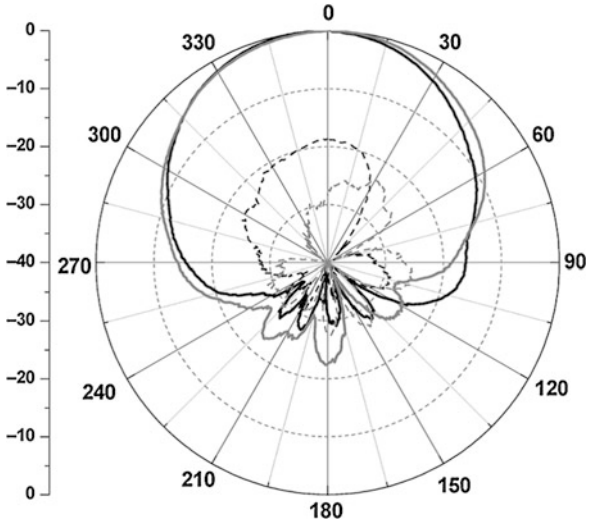


Fig. 20 Measured radiation pattern at 2.5 GHz (*black solid line* for copolarization in E-plane, *black dashed line* for cross-polarization in E-plane, *gray solid line* for copolarization in H-plane, and *gray dashed line* for cross-polarization in H-plane) (From (Siu et al.) © 2009 IEEE, Reprinted with permission)

Triangular Loop The triangular loop antenna was first used as a magnetic dipole by Luk and Wu (2012). As shown in Fig. 21, the antenna consists of a planar electric dipole, a triangular loop antenna, and a directly connected feed. The loop antenna is formed by changing the cross-sectional shape of the shorted quarter-wave patch antenna into a triangular shape. For excellent performance, the cross section should have a shape close to an equilateral triangle ($\theta = 60°$) with a total length of approximately $\lambda_0/2$. By using this magnetic dipole, the antenna height is reduced to $0.16\lambda_0$, and the gap between the planar dipole arms is only $0.02\lambda_0$ which allows the feasibility in designing the directly connected feed. This type of feed, including a segment of a coaxial cable and a small patch on the top corner of the triangular loop, will be

DOI 10.1007/978-981-4560-75-7_49-1

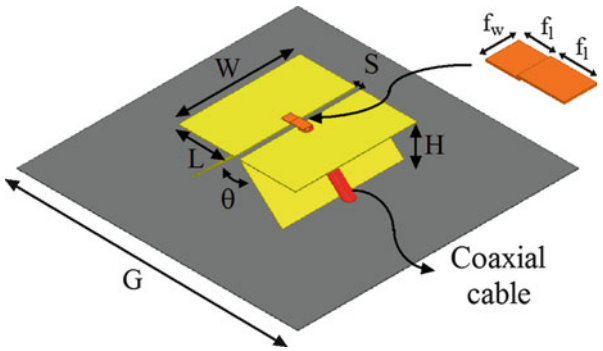


Fig. 21 Geometry of the triangular loop ME dipole antenna

Table 5 Dimensions for the triangular loop ME dipole antenna

Parameter	L	W	S	H
Value (mm)	30 $0.23\lambda_0$	65 $0.5\lambda_0$	2.6 $0.02\lambda_0$	21 $0.16\lambda_0$
Parameter	G	θ	f_1	
Value (mm)	150 $1\lambda_0$	60°	5 $0.03\lambda_0$	

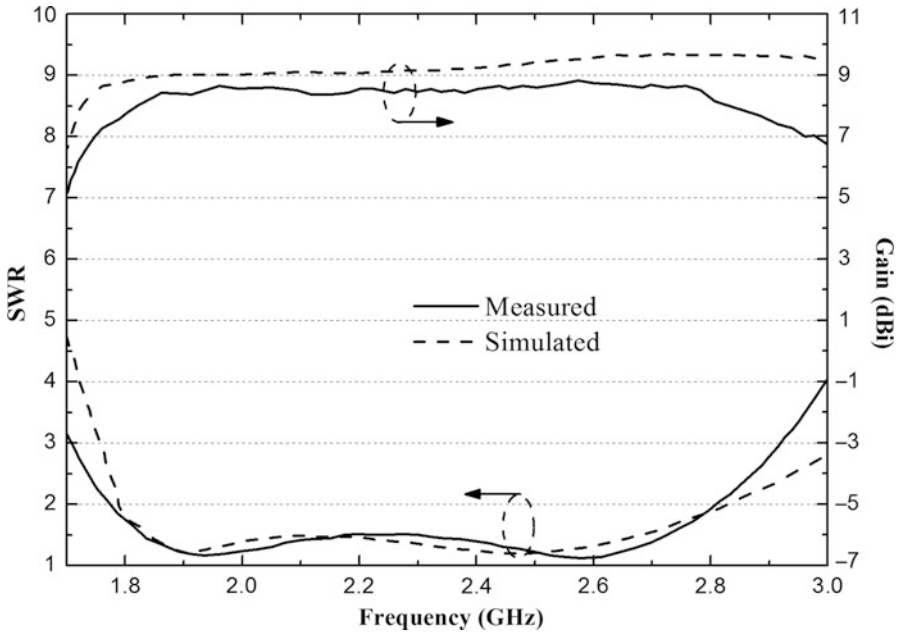


Fig. 22 Simulated and measured SWRs and gains (From (Luk and Wu) © 2012 IEEE, Reprinted with permission)

elaborated in section "Other Feed Mechanisms." The antenna prototype operating at 2.28 GHz was fabricated with dimensions shown in Table 5. Fig. 22 depicts the simulated and measured SWRs and antenna gains versus frequency. The simulated impedance bandwidth (SWR $\leq$1.5) is 39 % from 1.83 to 2.71 GHz, and the measured impedance bandwidth (SWR $\leq$1.5) is 40 % from 1.83 to 2.73 GHz. An agreement between simulated and measured results can be observed. Within the operating frequency range, the measured gain is 8.4 dBi with 0.2-dB variation, while the simulated gain is about 0.2 dB higher

DOI 10.1007/978-981-4560-75-7_49-1

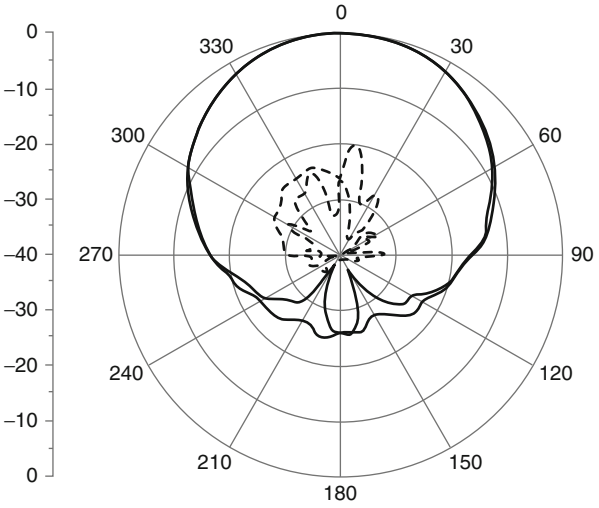


Fig. 23 Radiation pattern at 2.4 GHz (*black solid line* for copolarization in E-plane, *black dashed line* for cross-polarization in E-plane, *gray solid line* for copolarization in H-plane, and *gray dashed line* for cross-polarization in H-plane) (From (Luk and Wu) © 2012 IEEE, Reprinted with permission)

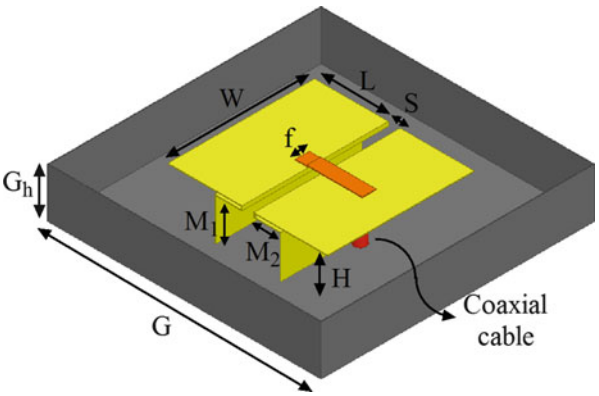


Fig. 24 Geometry of the rectangular loop ME dipole antenna

than the measured gain. The radiation pattern at 2.4 GHz is depicted in Fig. 23. A design guide for this antenna is recommended as follows:

1. According to the required center frequency of the operating band (λ_0), determine the dimensions of the planar electric dipole antenna with $L = 0.23\lambda_0$ and $W = 0.5\lambda_0$.
2. Determine the dimensions of the triangular loop antenna with $H = 0.16\lambda_0$ and $\theta = 60°$.
3. Choose the ground plane with a proper size ($G \geq \lambda_0$).
4. Design the directly connected feed.

Rectangular Loop Subsequently, Ge and Luk (2012a) utilized a rectangular loop antenna to perform as the magnetic dipole in the ME dipole antenna. Fig. 24 shows the geometry of the antenna. It consists of a planar electric dipole, a rectangular loop antenna, a directly connected feed, and a box-shaped reflector. By using the rectangular loop antenna, the antenna height can be reduced to $0.164\lambda_0$. A box-shaped reflector is employed to further reduce the back lobe and achieve a stable radiation pattern. The directly connected feed includes a segment of coaxial cable and an air microstrip line which extends along one arm

DOI 10.1007/978-981-4560-75-7_49-1

Table 6 Dimensions for the rectangular loop ME dipole antenna

Parameter	L	W	S	H	
Value (mm)	30	60	5.4	19	
	$0.26\lambda_0$	$0.52\lambda_0$	$0.05\lambda_0$	$0.16\lambda_0$	
Parameter	G	G_h	M_1	M_2	f
Value (mm)	112	20	17.2	11	5.7
	$1\lambda_0$	$0.17\lambda_0$	$0.15\lambda_0$	$0.1\lambda_0$	$0.05\lambda_0$

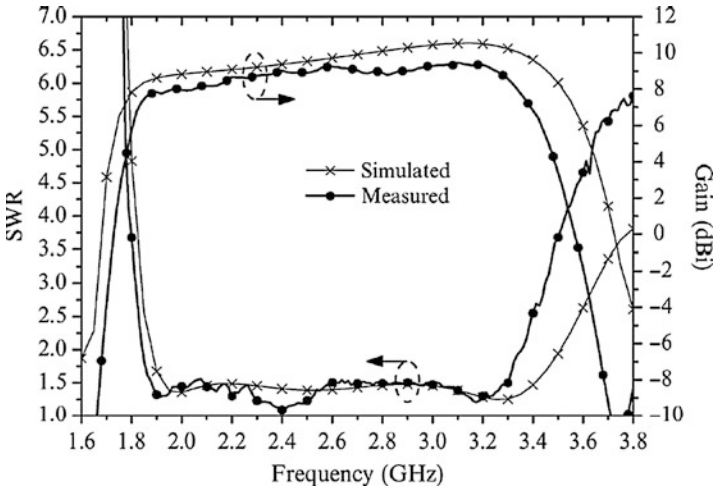


Fig. 25 Simulated and measured SWRs and gains (From (Ge and Luk) © 2012a IEEE, Reprinted with permission)

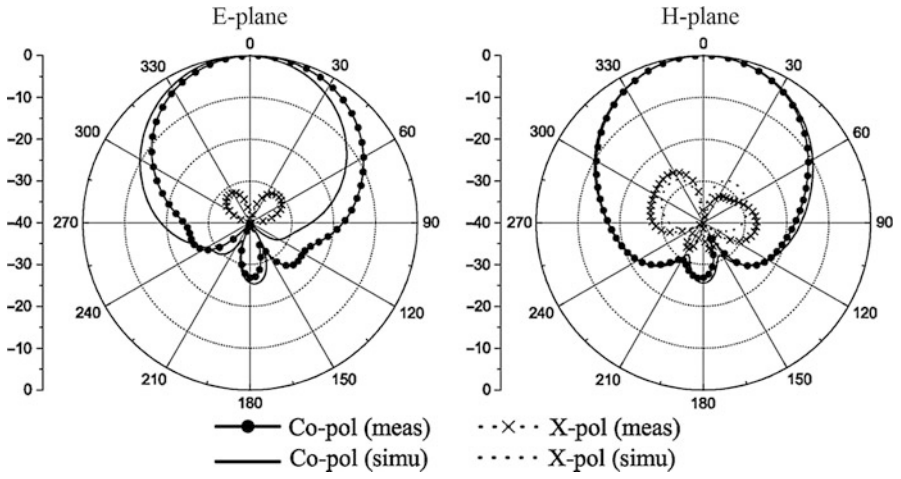


Fig. 26 Simulated and measured radiation patterns at 2.6 GHz (From (Ge and Luk) © 2012a IEEE, Reprinted with permission)

of the dipole, across the gap $S = 0.05\lambda_0$, and then directly connects to the other arm of the dipole. The antenna prototype operating at 2.59 GHz was fabricated with dimensions shown in Table 6. Fig. 25 shows the simulated and measured SWRs and gains. The measured impedance bandwidth is 54.8 % for SWR $\leq$ 1.5 from 1.88 to 3.3 GHz which is much wider than that of the triangular loop ME dipole antenna. The measured gain is approximately 8.6 with 0.8 dBi variation. The radiation pattern at 2.6 GHz is shown in Fig. 26. A design guide for this antenna is recommended as follows:

DOI 10.1007/978-981-4560-75-7_49-1

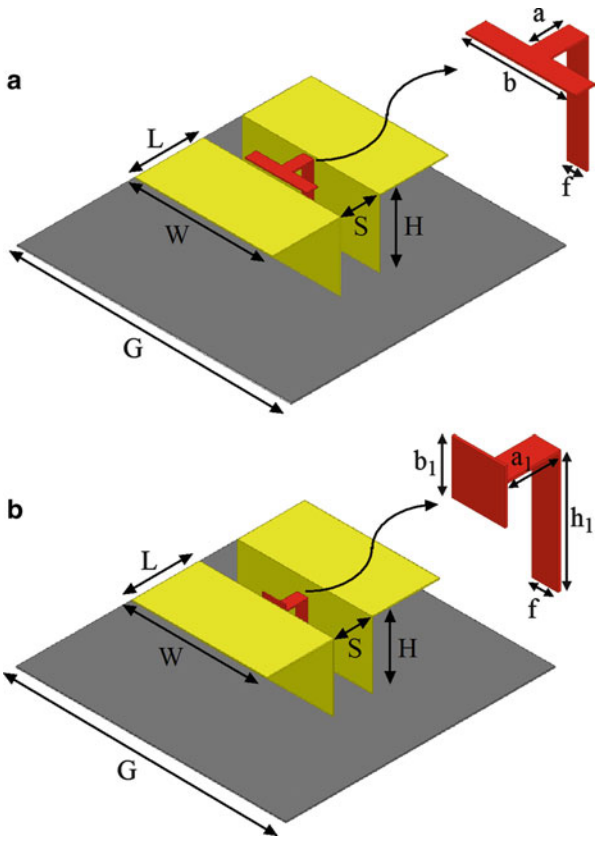


Fig. 27 Other types of ME dipole feeds. (**a**) T-shaped probe, (**b**) square-patch coupled probe

1. According to the required center frequency of the operating band (λ_0), determine the dimensions of the planar electric dipole antenna with $L = 0.259\lambda_0$ and $W = 0.518\lambda_0$.
2. According to the required antenna height, determine M_1 and M_2 by keeping $2 \times (M_1 + M_2) \approx 0.5\lambda_0$.
3. Choose the ground plane with a proper size ($G \geq \lambda_0$).
4. Design a directly connected feed and tune f to achieve a good impedance matching.

Other Feed Mechanisms

In the original ME dipole, a Γ-shaped probe is used to feed the antenna as illustrated in Fig. 5c. In practical use, the third portion of the probe can be deformed into different shapes to compensate the inductive reactance from the second portion. Two types of feeds are shown in Fig. 27, i.e., the T-shaped probe and the square-patch coupled probe. The T-shaped probe shown in Fig. 27a has the third portion of a horizontal patch. This confirms that the design in the millimeter-wave bands can be printed on a single-layer printed circuit board which will be discussed in section "Designs for 60-GHz Applications." The probe shown in Fig. 27b is shorter than the original Γ-shaped probe and has the third portion of a square patch. Although the antenna with either type of feed exhibits a wider impedance bandwidth compared to the case with original Γ-shaped probe, it suffers from higher back radiation and cross-polarization levels.

In addition, the aforementioned types of probes are all coupled proximity feeds and not direct-current (dc) grounded, which therefore cannot fulfill the requirement for lightning protection in outdoor applications. Hence, as studied in section "Methods of Reducing Height," the ME dipole antennas have loop antenna as the magnetic dipole (Luk and Wu 2012; Ge and Luk 2012a; Ge and Luk 2013a). The directly

DOI 10.1007/978-981-4560-75-7_49-1

connected feed is used to excite the antennas. This type of feed together with the loop antenna also functions as a folded balun and provides a wide operating bandwidth. However, because there is no portion compensating the inductance contributed by the portion of the feed across the separation between the dipole arms, it requires the separation to be very small, which definitely increases the complexity of the antenna structure. Moreover, due to the folded balun not working as well as the Γ-shaped probe, the maximum gain cannot occur exactly in the broadside direction at the higher frequencies.

Magnetoelectric Dipole Antennas with Different Polarizations

Dual-Polarized Antenna

General Remarks

The polarization diversity technique has been widely used in mobile communication base stations due to its attractive features and advantages. In real applications, a ± 45° dual-polarized antenna is used to receive and combine two uncorrelated signals to compensate for the signal degradation caused by the multipath-fading phenomena (Vaughan 1990), therefore increasing the signal reliability. Compared to the spatial diversity technique, the polarization diversity technique only requires one antenna for operation, thus reducing the installation cost and space (Turkmani et al. 1995).

In the literature, many dual-polarized antennas were presented based on the slot antenna (Paryani et al. 2010), dipole antenna (Cui et al. 2013), dielectric resonator antenna (Sun and Leung 2013), or microstrip patch antenna (Wong et al. 2004; Lai and Luk 2007; Caso et al. 2010). However, these designs have drawbacks, such as high cross-polarization, high back radiation, or low port isolation. With the rapid development of wireless communications, conventional antennas cannot satisfy the stringent requirements imposed on base station antennas.

Multiple Γ-Shaped Probes

Siu et al. (2009) first developed a dual-polarized ME dipole antenna, based on the linearly polarized design disclosed in the same paper. As shown in Fig. 28, the antenna consists of four square plates which are shorted to the ground plane by vertical metallic walls. The square plates act as planar electric dipoles, while the metallic walls form shorted patch antennas which are fully filled with dielectric substrate to

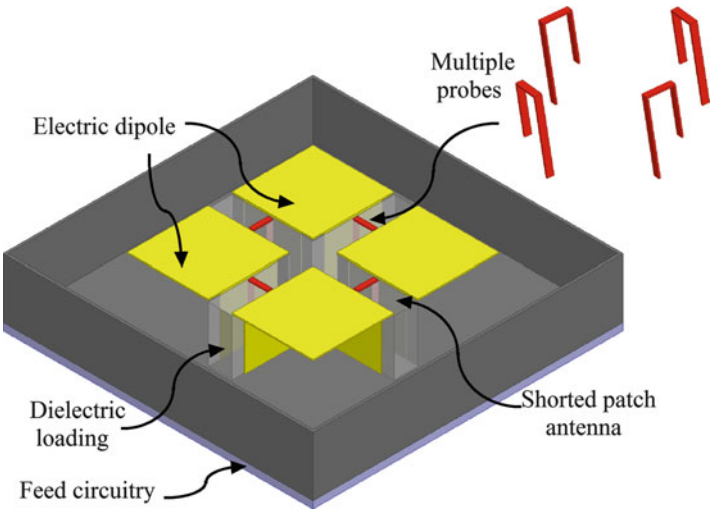


Fig. 28 Geometry of the dual-polarized ME dipole with multiple Γ-shaped probes

DOI 10.1007/978-981-4560-75-7_49-1

reduce the antenna height. Two pairs of Γ-shaped probes are placed between the vertical metallic walls to excite ± 45° polarizations of the antenna. A feed circuitry under the ground plane is needed for this antenna, therefore increasing the fabrication cost. The feed network simply consists of two power dividers. Each power divider splits and transmits the input electrical power to one pair of probes arranged side by side, further exciting one polarization of the antenna. Originally, the impedance bandwidth of the linearly polarized antenna discussed in section "Magnetoelectric Dipole with Dielectric Loading" is 48 %; however, the dual-polarized antenna only exhibits a narrow impedance bandwidth of 24 %, ranging from 1.65 to 2.12 GHz, over which the isolation between the two ports is larger than 29 dB. The average gains for both ports are about 8.2 dBi.

Orthogonal Γ-Shaped Probes

Shortly after the publication of this paper, another dual-polarized ME dipole antenna was introduced by Wu and Luk (2009a). As shown in Fig. 29, the design also consists of four square plates shorted to the ground by the vertical walls, which is very similar to the aforementioned antenna presented in Siu et al. (2009). But this time, only two single Γ-shaped probes are used to excite the ± 45° polarizations, therefore not requiring a feed circuitry. Each Γ-shaped probe consists of three portions. The first portion of the probe together with the adjacent vertical wall forms a tapered air microstrip line and transmits electrical energy from the coaxial cable to the second portion of the probe which comes out from the hole of one dipole arm, across the gap between the arms and enters into the hole of the other dipole arm. It couples the electrical energy to the antenna but introduces inductance which is compensated by the third portion of the probe. It should be noted that the probe is not electrically connected to the antenna.

When one polarization is excited as illustrated in Fig. 29, a pair of orthogonal opposite metallic plates performs as the electric dipole. Each metallic wall together with the adjacent wall and the ground plane between them performs as a shorted quarter-wave patch antenna; therefore, four shorted patch antennas can be formed. Each pair of the shorted patch antennas arranged side by side is equivalent to a magnetic dipole. Thus, the four shorted patch antennas can be equivalent to two magnetic dipoles which have a vector sum perpendicular to the electric dipole. Consequently, the magnetoelectric dipole antenna can be excited.

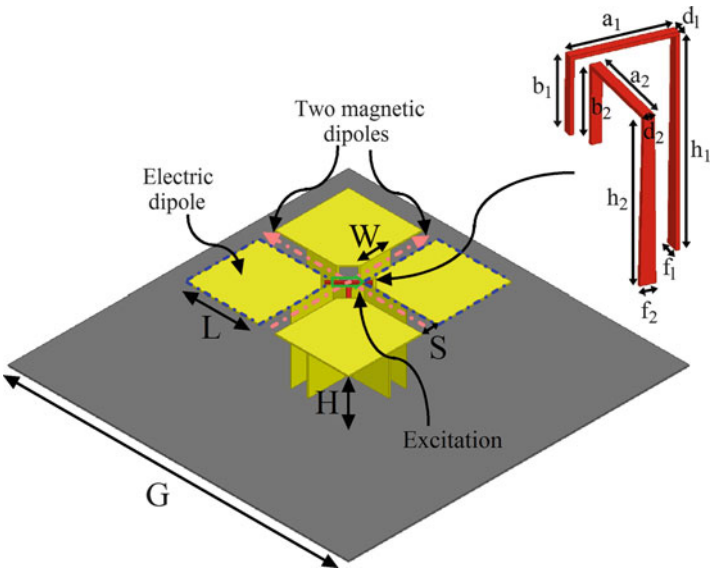


Fig. 29 Geometry of the dual-polarized ME dipole with orthogonal Γ-shaped probes

DOI 10.1007/978-981-4560-75-7_49-1

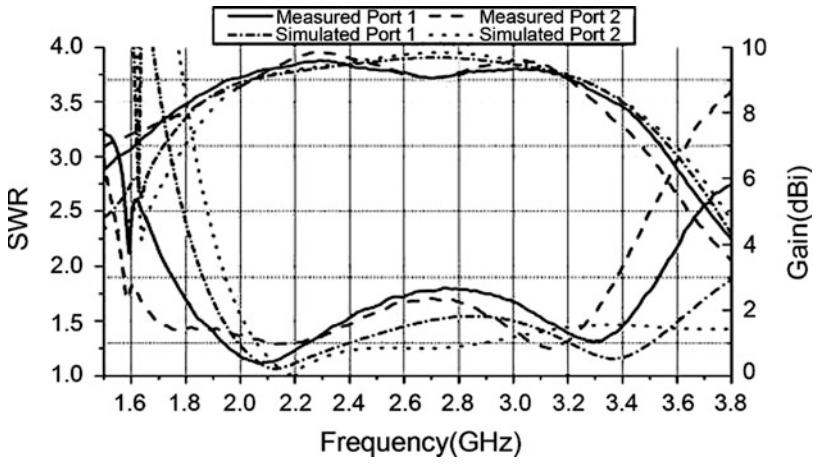


Fig. 30 Simulated and measured SWRs and gains (From (Wu and Luk) © 2009a IEEE, Reprinted with permission)

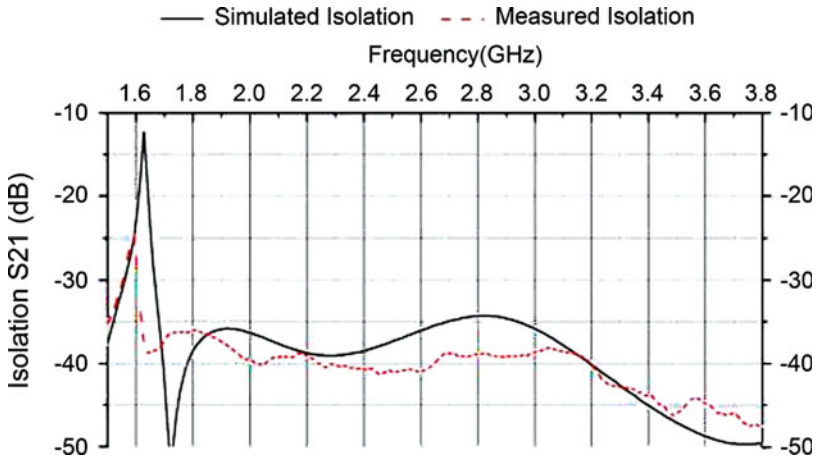


Fig. 31 Simulated and measured isolations (From (Wu and Luk) © 2009a IEEE, Reprinted with permission)

The dimensions of the antenna operating at 2.67 GHz are shown in Table 7. An antenna prototype was fabricated and tested. It can be seen from Fig. 30 that the measured impedance bandwidth (SWR $\leq$2) is 69.7 % and 74.6 % at port 1 and port 2, respectively. The common bandwidth of the two ports is 65.9 %, ranging from 1.7185 to 3.409 GHz, over which the isolation between both ports is more than 36 dB, as illustrated in Fig. 31. The antenna gain is approximately 9 dBi. Fig. 32 shows the measured radiation pattern at 2.701 GHz. It is obvious that, by using the orthogonal Γ-shaped probes, the antenna without the use of a feed circuitry achieves a wider operation bandwidth, compared with the design of Siu et al. (2009). However, the complex feed probes cause difficulties in fabrication and construction.

Simplification of the Orthogonal Γ-Shaped Probes

Subsequently, Li and Luk (2012) considered the simplification of the feeds by placing the two orthogonal Γ-shaped probes between the vertical metallic posts. More recently, Xue et al. (2013) modified the feeds to two pairs of differential feeds for the enhancement of the operating bandwidth and further suppression of the coupling between two polarizations. The antenna achieves a wide impedance bandwidth (SWR $\leq$2) of 68 %, from 0.95 to 1.92 GHz. The differential port-to-port isolation is better than 36 dB. The 3-dB gain bandwidth is 62 % ranging from 1.09 to 2.08 GHz. In what follows, the design with the simplified orthogonal Γ-shaped probes (Li and Luk 2012) is discussed in detail.

DOI 10.1007/978-981-4560-75-7_49-1

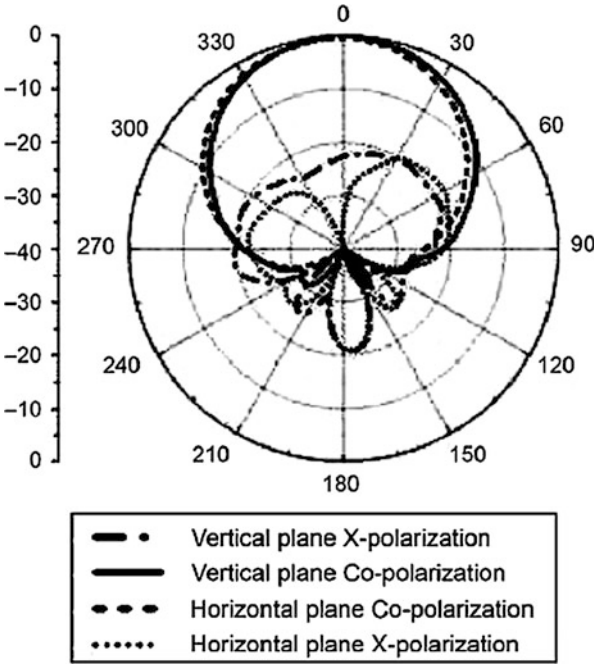


Fig. 32 Measured radiation pattern at 2.701 GHz (From (Wu and Luk) © 2009a IEEE, Reprinted with permission)

Table 7 Dimensions for the dual-polarized antenna (From (Wu and Luk) 2009a IEEE, Reprinted with permission)

Parameter	L	W	S	G	H	a_1	a_2	
Value(mm)	29.2 $0.23\lambda_0$	14 $0.10\lambda_0$	6.2 $0.05\lambda_0$	149.6 $1.28\lambda_0$	28 $0.23\lambda_0$	13.6 $0.12\lambda_0$	13.6 $0.12\lambda_0$	
Parameter	b_1	b_2	f_1	f_2	d_1	d_2	h_1	h_2
Value(mm)	4.5 $0.05\lambda_0$	4.0 $0.03\lambda_0$	2.0 $0.02\lambda_0$	2.0 $0.02\lambda_0$	1.18 $0.01\lambda_0$	1.31 $0.01\lambda_0$	27 $0.23\lambda_0$	21 $0.18\lambda_0$

As shown in Fig. 33, the antenna consists of four metallic plates, four rectangular metallic posts, two orthogonal Γ-shaped probes, and a box-shaped reflector. Each of metallic plate is shorted to the ground through a large metallic post. They are placed uniformly with separation of $S = 0.05\lambda_0$. Two Γ-shaped probes are placed orthogonally in the gaps between the metallic posts. The Γ-shaped probe consists of three portions. For high isolation, the two probes are required to have different heights. In operation, signals are launched to the orthogonal Γ-shaped probes through the two SMA connectors. For each probe, the copper strip, together with the two adjacent shorted metallic posts, forms a modified stripline structure. The surfaces of adjacent metallic posts perform as ground planes, and the middle copper strip transmits the signal. Since the metallic posts are located uniformly with two orthogonal gaps, the portions of horizontal strips in the intersection of two gaps are responsible to couple the electrical energy to the antenna element.

When one polarization is excited as indicated in Fig. 33, a pair of planar electric dipoles and a pair of shorted quarter-wave patch antennas are excited simultaneously. The four metallic plates perform as two planar electric dipoles. Each metallic post, together with an adjacent post and the ground between them, acts as a shorted quarter-wave patch antenna. Every two shorted patch antennas placed side by side can be equivalent to a magnetic dipole that is perpendicular to the electric dipoles. Consequently, the

DOI 10.1007/978-981-4560-75-7_49-1

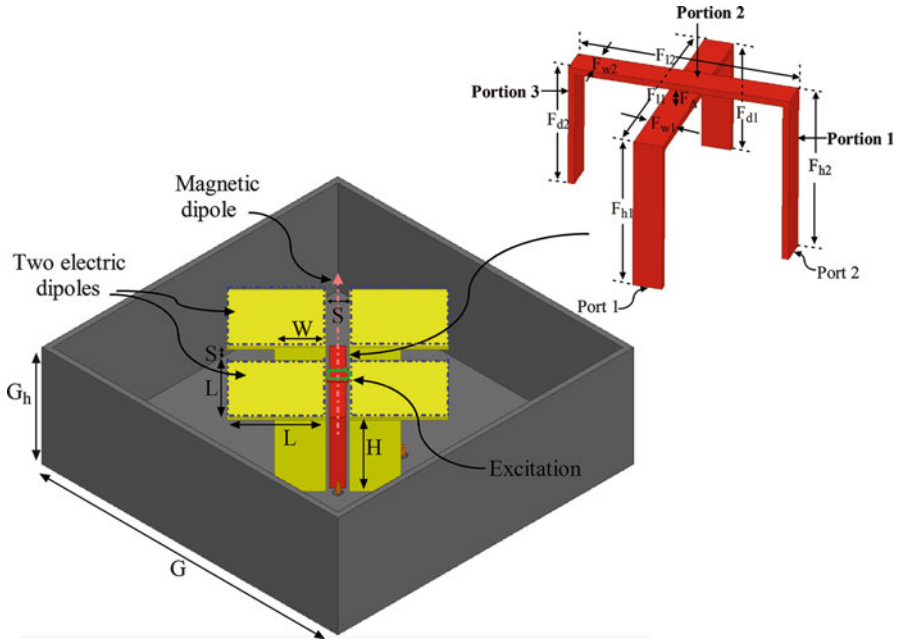


Fig. 33 Geometry of the dual-polarized ME dipole with simplified orthogonal Γ-shaped probes

magnetoelectric dipole antenna can be realized. It should be pointed out that this antenna has the different operation principle compared to the design described in the previous section (Wu and Luk 2009a).

Experimentally, a prototype at 2.2 GHz was fabricated. The antenna operates from 1.69 to 2.83 GHz with a bandwidth of 50 % (SWR $\leq$1.5) and from 1.68 to 2.76 GHz with a bandwidth of 49 % (SWR $\leq$1.5) for ports 1 and 2, respectively. The operating frequency ranges for the two ports are slightly different due to the unequal heights of the two orthogonal probes. The common bandwidth of the two ports is 48 %. Over the operating frequency range, the measured broadside gains for port 1 and port 2 are approximately 8.45 and 8.5 dBi, respectively. Nearly identical antenna gains are obtained for the two ports due to the symmetrical architecture of the proposed antenna design. The measured isolation between the two ports is larger than 30 dB, and the measured front-to-back ratios of the two ports are larger than 29 dB over the operating frequency range. A design guide for this antenna is suggested as follows:

1. According to the required center frequency of the operating band (λ_0), determine the dimensions of the planar electric dipole antenna with $L = 0.20\lambda_0$.
2. Determine the dimensions of the shorted quarter-wave patch antenna (metallic posts) with $H = 0.25\lambda_0$, $W = 0.10\lambda_0$, and $S = 0.05\lambda_0$.
3. Design a box-shaped reflector with dimensions of $G = 0.8\lambda_0$ and $G_h \approx 0.25\lambda_0$.
4. Design a pair of orthogonal Γ-shaped probes and tune f_Δ to achieve a high isolation and a relative acceptable overlapped impedance bandwidth.

The comparison between this antenna and the previous design in Wu and Luk (2009a) is summarized in Table 8. It can be found that the design in Li and Luk (2012) provides a wide operating bandwidth (SWR $\leq$1.5) with lower cross-polarization and back-lobe levels but suffers from an unstable radiation beamwidth and a lower isolation, when compared to the design in Wu and Luk (2009a).

DOI 10.1007/978-981-4560-75-7_49-1

Table 8 Comparison between this antenna and the antenna in Wu and Luk (2009a)

	(Li and Luk 2012)	(Wu and Luk 2009a)
Size (λ_0^3)	0.8 × 0.8 × 0.26	1.28 × 1.28 × 0.23
Impedance bandwidth	48 % (SWR ≤ 1.5)	65.9 % (SWR ≤ 2)
Isolation (dB)	>30	>36
Gain (dBi)	~8.5	~9
Cross-polarization level (dB)	< −21	< −18
Back-lobe level (dB)	< −29	< −18
Beamwidth (°)	57–79	58–64.5

Circularly Polarized Antenna

General Remarks

Circularly polarized (CP) radiation is useful for multipath interference suppression and polarization mismatch reduction. Owing to these features, circular polarization is highly desirable for many applications such as radar, satellite, and millimeter-wave systems. The basic operating principle of a circularly polarized antenna is to radiate two orthogonal field components with equal amplitude and in phase quadrature. Many omnidirectional CP antennas with good performances have been investigated in Pan et al. (2012) and Nakano et al. (2011). However, for some wireless communication systems, unidirectional CP antennas are required because security and efficiency issues should be considered by system designers.

Based on a dual-polarized antenna, a circularly polarized antenna can be realized by adding a feed circuitry. Many types of dual polarized antennas have been investigated in the literature, such as the dual-fed crossed-dipole antenna (Zhang et al. 2013), the quadrature-fed dielectric resonator antenna (Li et al. 2009), and the annular-ring microstrip antenna (Guo et al. 2009). The dual-polarized ME dipole antenna discussed in section "Simplification of the Orthogonal Γ-Shaped Probes" was also developed to a CP antenna (Li and Luk, 2013c). A feed circuitry, consisting of a broadband 90° phase shifter and a Wilkinson power divider, was used. Experimentally, the antenna prototype exhibited a wide impedance bandwidth (SWR ≤2) of 90 % from 1.23 to 3.23 GHz, which covered the whole 3-dB axial-ratio (AR) bandwidth of 82 % from 1.28 to 3.05 GHz. In this operating frequency band, the antenna had a broadside gain of larger than 5dBic above 1.45GHz. Considering the common overlapped bandwidth limited by the impedance, axial ratio, and gain, the antenna provided an effective bandwidth of 71 %.

In comparison with dual-fed CP antenna, single-fed CP antenna does not need feed circuitry and therefore has a simple structure. Thus, Li and Luk (2014) proposed a single-fed circularly polarized ME dipole antenna. In what follows, this design is presented in details.

Single-Fed Design

Fig. 34a shows the single-fed circularly polarized ME dipole antenna. Standing from the dual-polarized ME dipole antenna described in section "Simplification of the Orthogonal Γ-Shaped Probes," one pair of the diagonally opposite metallic plates is truncated at the corners in order to increase its resonant frequency. The other pair is added with hook-shaped strips in order to reduce its resonant frequency. And two of the four gaps are filled with small metallic blocks in order to increase the resonant frequency of the equivalent vertically quarter-wave shorted patch antenna. Thus, the proposed CP antenna consists of a pair of metallic plates with extended strips, a pair of metallic plates with corner truncated, a pair of U-shaped posts, an L-shaped probe feed, and a planar reflector. Fig. 34b depicts the top and side views of the proposed antenna. The metallic plates are direct-current grounded through the U-shaped posts. The

DOI 10.1007/978-981-4560-75-7_49-1

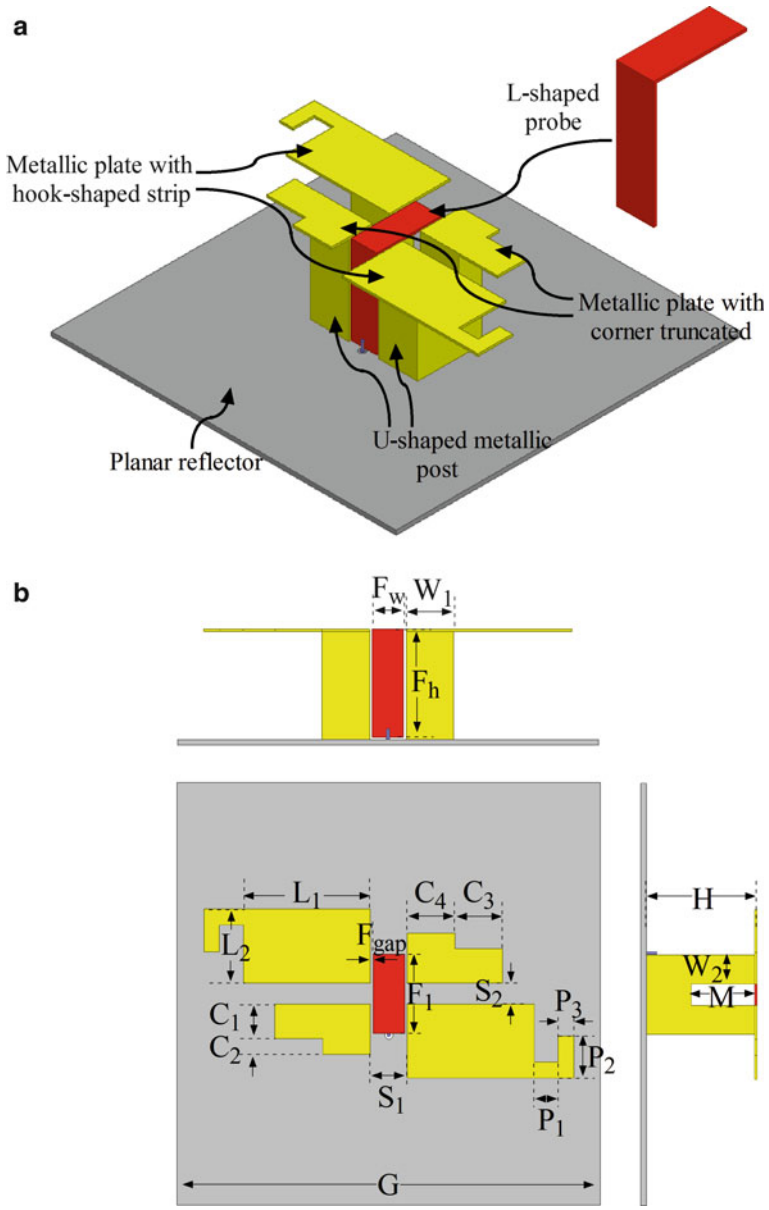


Fig. 34 Geometry of the single-fed circularly polarized ME dipole antenna. (**a**) Perspective view, (**b**) top and side views (From (Li and Luk) © 2014 IEEE, Reprinted with permission)

proposed antenna with the L-shaped probe is mounted on a planar reflector. A SMA connector is placed under the reflector for transmitting signal to the L-shaped probe. Detailed dimensions of the antenna are summarized in Table 9.

Fig. 35 shows simulated and measured SWRs and gains. The simulated impedance bandwidth is 52.3 % with SWR $\leq$1.5 from 1.58 to 2.7 GHz and 60.7 % with SWR $\leq$2 from 1.48 to 2.77 GHz. The measurements indicate that the antenna prototype can achieve a wide impedance bandwidth of 46.6 % (SWR $\leq$1.5) from 1.68 to 2.7 GHz and 73.3 % (SWR $\leq$2) from 1.32 to 2.85 GHz. Within the 1.5:1 SWR band, the antenna broadside gain varies in the range of 5 to 8.6 dBic. Fig. 36 plots the measured and simulated antenna axial ratios. It can be seen that the measured 3-dB AR band is 47.7 % from 1.69 GHz to 2.75 GHz, which is generally covered by the SWR $\leq$1.5 operating band. The measured and simulated radiation patterns at 2.1 GHz are depicted in Fig. 37.

DOI 10.1007/978-981-4560-75-7_49-1

Table 9 Dimensions for the single-fed circularly polarized ME dipole antenna (From (Li and Luk) © 2014 IEEE, Reprinted with permission)

Parameter	L_1	L_2	S_1	S_2	H	W_1	W_2
Value (mm)	48	28	14	8	42	18	11
	$0.33\lambda_0$	$0.19\lambda_0$	$0.10\lambda_0$	$0.06\lambda_0$	$0.29\lambda_0$	$0.13\lambda_0$	$0.08\lambda_0$
Parameter	P_1	P_2	P_3	C_1	C_2	C_3	C_4
Value (mm)	9	16	6	13	6	18	18
	$0.06\lambda_0$	$0.11\lambda_0$	$0.04\lambda_0$	$0.09\lambda_0$	$0.04\lambda_0$	$0.13\lambda_0$	$0.13\lambda_0$
Parameter	M	F_w	F_h	F_l	F_{gap}	G	
Value (mm)	24	12	40	30	1	160	
	$0.17\lambda_0$	$0.08\lambda_0$	$0.28\lambda_0$	$0.21\lambda_0$		$1.11\lambda_0$	

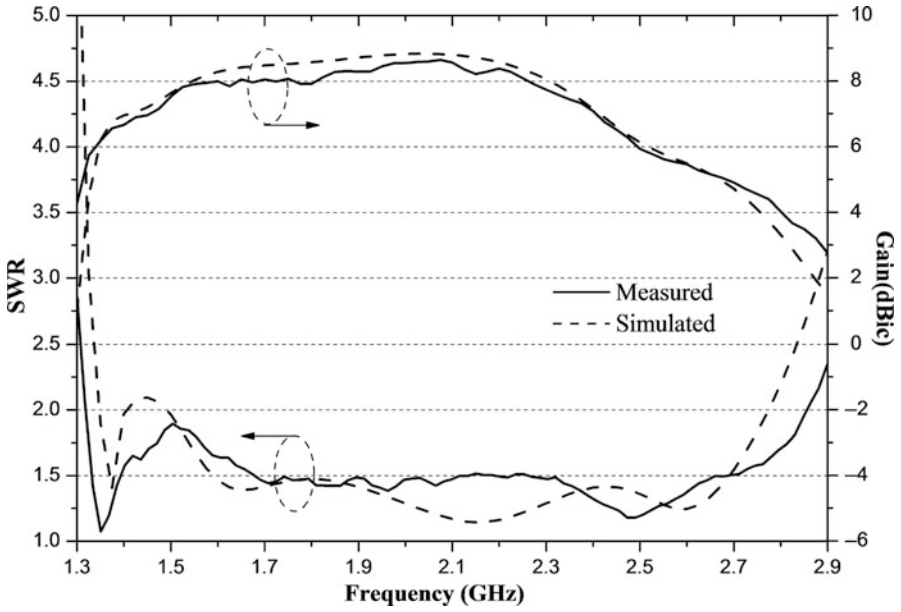


Fig. 35 Simulated and measured SWRs and gains of the single-fed CP antenna (From (Li and Luk) © 2014 IEEE, Reprinted with permission)

Antenna with Pattern and Polarization Diversities

General Remarks

Recently, diversity antenna has been considered to increase the communication system capability. There are different types of diversity antennas including spatial diversity antenna, frequency diversity antenna, polarization diversity antenna, and pattern diversity antenna (Yang and Luk 2006). As studied in section "Dual-Polarized Antenna," the dual-polarized antennas are two-port antennas with polarization diversity. To further solve multipath-fading effects in complex environments and enhance the performance of a wireless system, a four-port antenna with polarization and pattern diversities is recommended. This type of diversity antenna is capable of radiating and receiving signals through different radiation modes simultaneously and therefore has the advantage of performing higher effective gain while maintaining

DOI 10.1007/978-981-4560-75-7_49-1

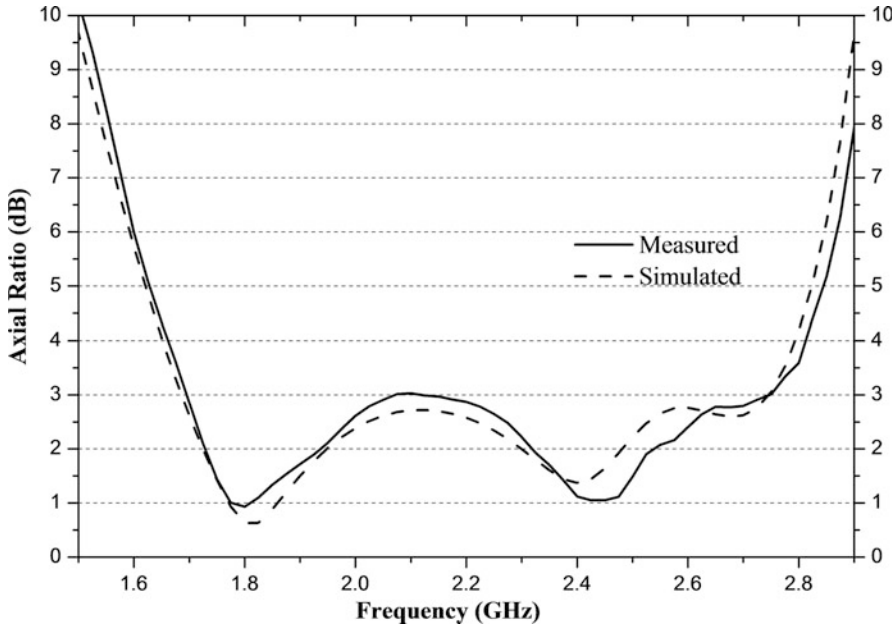


Fig. 36 Simulated and measured ARs of the single-fed CP antenna (From (Li and Luk) © 2014 IEEE, Reprinted with permission)

the same installation space. Yang et al. (2008) first demonstrated that this type of antenna could be realized based on microstrip patch antenna. Unfortunately, the antenna has a drawback of the narrow overlapped bandwidth of the excitation ports. Subsequently, the ME dipole was employed by Wu and Luk (2011a) to develop this kind of diversity antenna with wideband performance. Note that the preliminary work on the design of a magnetic monopole was presented in Wu and Luk (2009c).

Diversity Antenna Design

The geometry of the diversity antenna is shown in Fig. 38. It consists of four ME dipoles arranged in a ring configuration and a wire electric monopole. Each ME dipole is realized by deforming the original ME dipole discussed in section "Antenna Configuration." A pair of sectorial-shape horizontal plates has the inner radius R_1 and outer radius R_2, acting as an electric dipole. The magnetic dipole is performed by a shorted quarter-wave patch antenna realized by two vertical metallic walls and the portion of the ground plane between them. A Γ-shaped probe is employed to excite each ME dipole. Four identical ME dipoles are arranged in a ring configuration and are located at 0°, 90°, 180°, and 270°.

The wire electric monopole is located at the center of the four ME dipoles, radiating a conical beam with vertical polarization (i.e., electric monopole conical mode). As shown in Fig. 39, the feed network, mounted under the ground plane, consists of two hybrid rings and a Wilkinson power divider. Two stages of excitation are used. In the first stage, the output ports of the two hybrid rings are connected to the Γ-shaped probes of the ME dipoles. Two isolated input ports of hybrid rings are used to excite two modes of broadside radiation with horizontal polarization (i.e., ±45° broadside modes). In the second stage, the other input port of each hybrid ring is connected to an output port of a Wilkinson power divider, the input port of which is used to radiate a conical beam with horizontal polarization (i.e., magnetic monopole conical mode). For further understanding the antenna operation, the current distributions of the antenna at $t = 0$ and T/4 are depicted in Fig. 40.

DOI 10.1007/978-981-4560-75-7_49-1

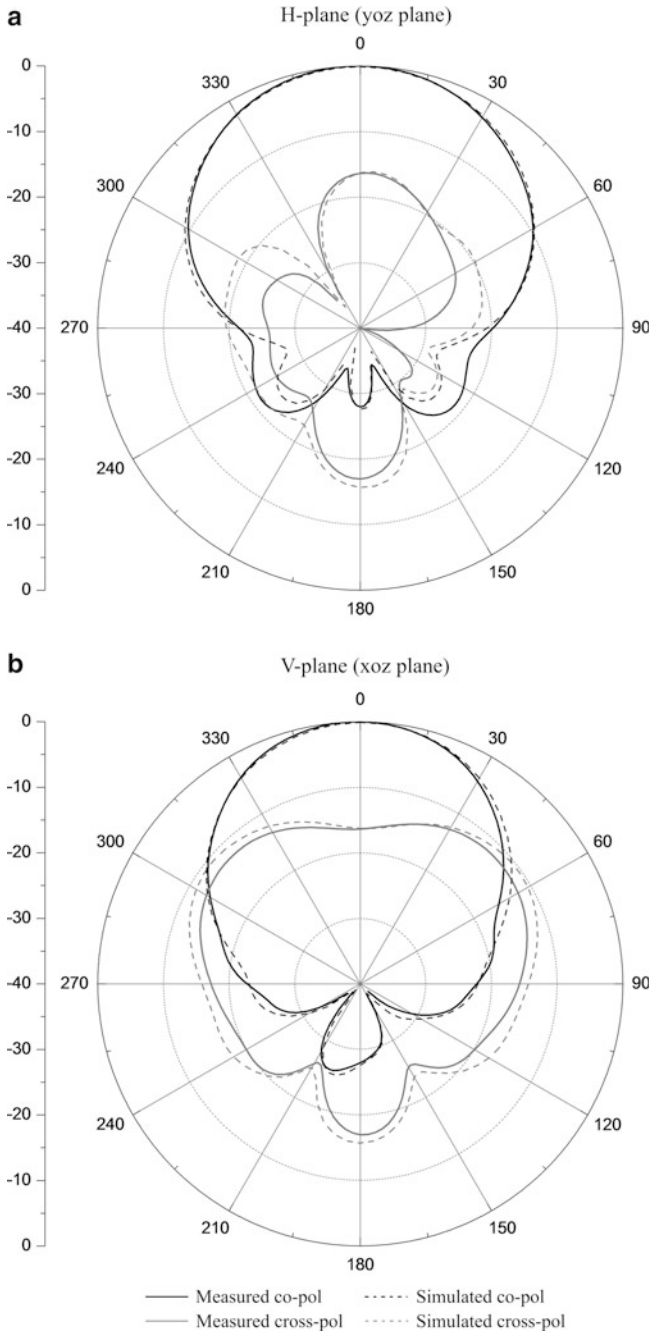


Fig. 37 Simulated and measured radiation pattern at 2.1 GHz (From (Li and Luk) © 2014 IEEE, Reprinted with permission)

Detailed dimensions of the antenna are summarized in Table 10. As shown in Fig. 41, the measured impedance bandwidths (SWR $\leq$2) for electric monopole mode, magnetic monopole mode, +45° broadside mode, and −45° broadside mode are 38.6 % (1.90–2.81 GHz), 22.2 % (2.12–2.66 GHz), 43.1 % (1.87–2.90 GHz), and 39.2 % (1.95–2.90 GHz), respectively. The measured antenna gains for electric and

DOI 10.1007/978-981-4560-75-7_49-1

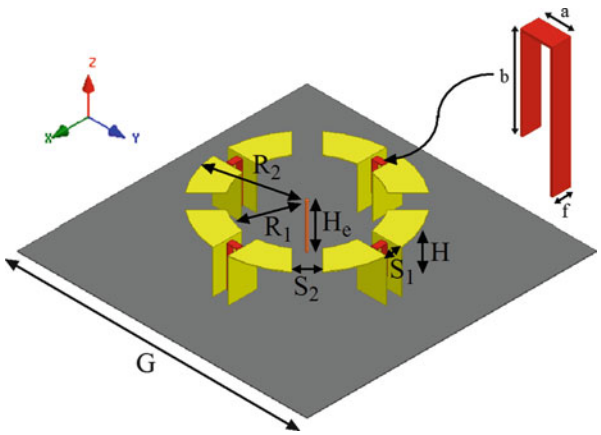


Fig. 38 Geometry of the antenna with pattern and polarization diversities

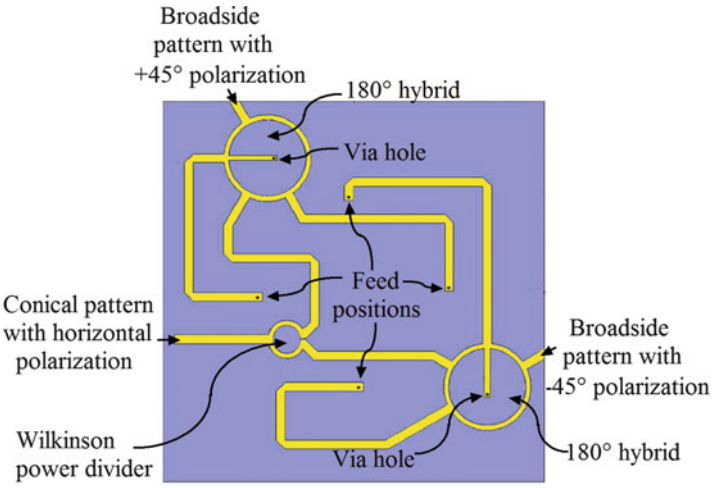


Fig. 39 Feed network of the antenna

magnetic monopole modes are approximately 6 dBi and for ±45° broadside modes are approximately 11 dBi. Thus, the overlapped impedance bandwidth is 22.2 % ranging from 2.12 to 2.66 GHz, over which the measured isolation is larger than 26 dB. The measured radiation patterns at 2.4 GHz for different modes are depicted in Fig. 42.

DOI 10.1007/978-981-4560-75-7_49-1

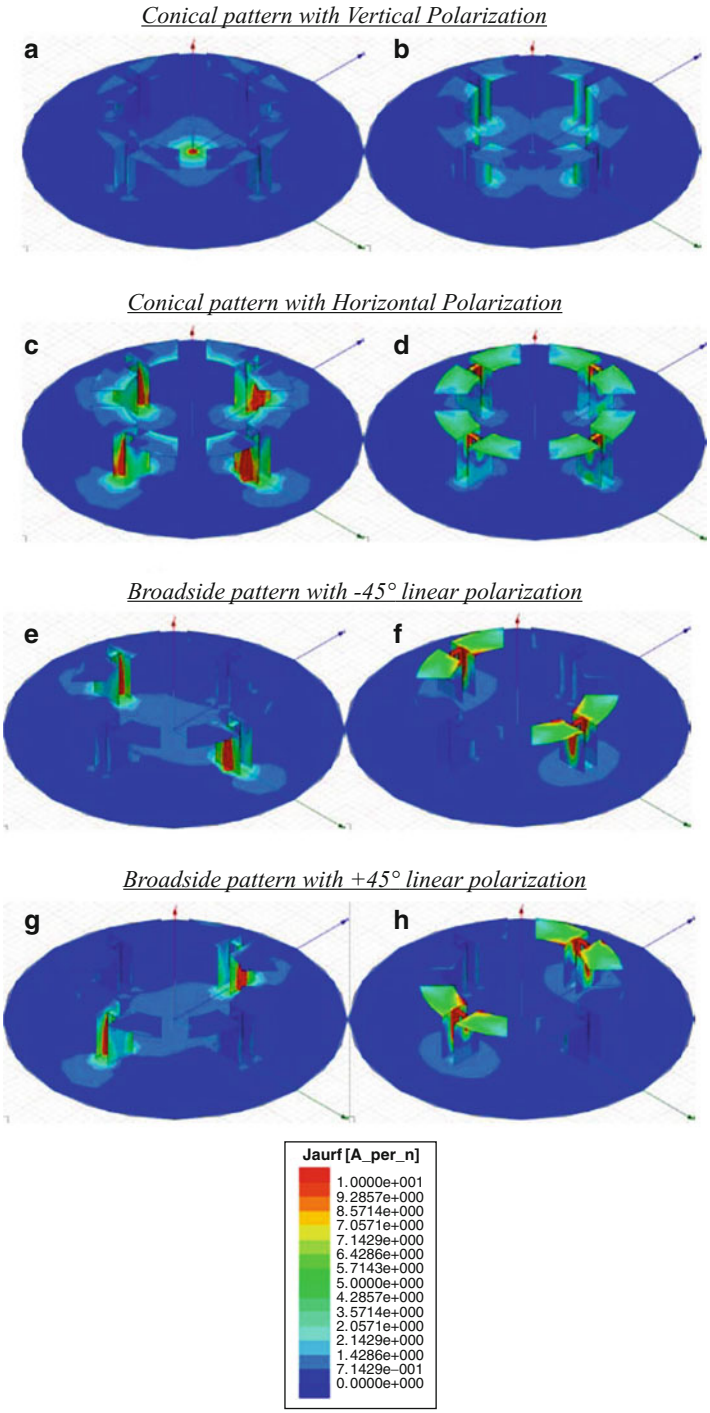


Fig. 40 Current distribution of four radiation modes: (**a**) t = 0, (**b**) t = T/4, (**c**) t = 0, (**d**) t = T/4, (**e**) t = 0, (**f**) t = T/4, (**g**) t = 0, (**h**) t = T/4 (From Wu and Luk) © 2011a IEEE, Reprinted with permission)

DOI 10.1007/978-981-4560-75-7_49-1

Table 10 Dimensions for the single-fed circularly polarized ME dipole antenna

Parameter	R_1	R_2	S_1	S_2	H
Value (mm)	40	60	10	15	29
	$0.32\lambda_0$	$0.48\lambda_0$	$0.08\lambda_0$	$0.12\lambda_0$	$0.23\lambda_0$
Parameter	H_e	f	a	b	G
Value (mm)	30	4	7	20	200
	$0.24\lambda_0$	$0.03\lambda_0$	$0.06\lambda_0$	$0.16\lambda_0$	$1.6\lambda_0$

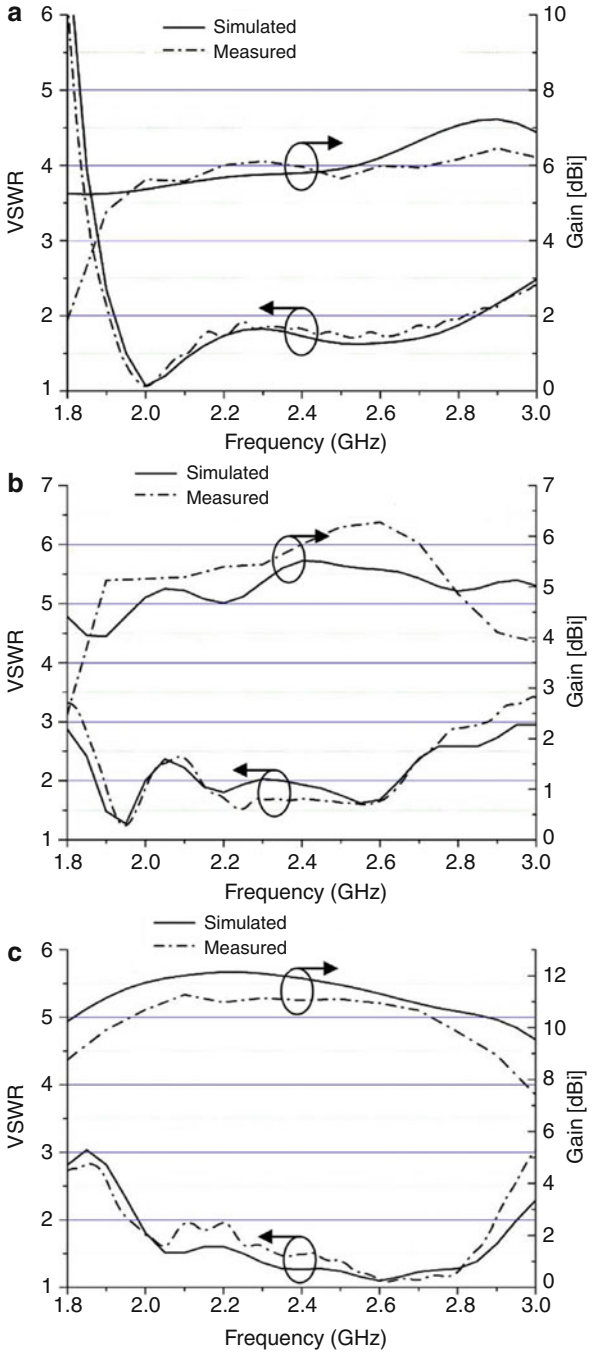


Fig. 41 (continued)

DOI 10.1007/978-981-4560-75-7_49-1

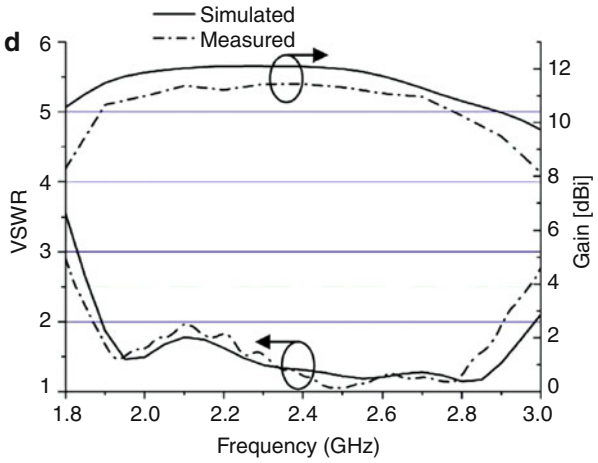


Fig. 41 Simulated and measured SWRs and gains. (**a**) Electric monopole conical mode, (**b**) magnetic monopole conical mode, (**c**) +45° broadside mode, and (**d**) −45° broadside mode (From (Wu and Luk) © 2011a IEEE, Reprinted with permission)

Designs for UWB Applications

Since the Federal Communications Commission (FCC) allocated the 7.5GHz (3.1GHz–10.6GHz) band for unlicensed use, ultra-wideband (UWB) wireless communication technology has attracted much interest. Many applications of UWB technology have been developed including medical imaging systems, ground penetrating radars, and high data rate communication systems.

Typically, the bandwidth of ME dipole antenna is limited to 40–50 % which cannot meet the bandwidth requirement of UWB communication. By deforming antenna structure, the bandwidth can be broadened. Wu and Luk (2009b) first attempted to develop a UWB antenna based on the ME dipole, as shown in Fig. 43. It consists of a pair of trapezoidal plates shorted to the ground plane by right-angle corner walls and a modified box-shaped reflector. A Γ-shaped probe, similar to the probe shown in Fig. 28, is used to excite the antenna. The antenna exhibits a wide impedance bandwidth (SWR $\leq$2) of 91.6 % ranging from 3.05 to 8.2 GHz. The measured average gain is about 9.3 dBi. Unfortunately, the operating band of this antenna cannot cover the whole UWB unlicensed spectrum. Later, a dual-polarized UWB antenna was presented by Wu and Luk (2011b). They utilized a bow-tie antenna and loop antenna as the electric and magnetic dipoles, respectively. The operating bandwidth of this antenna is from 2.4 GHz to 10.6 GHz, which covered the whole UWB frequency band.

After the publication of these papers, another UWB ME dipole antenna was demonstrated by Ge and Luk (2012b). The geometry of the antenna is shown in Fig. 44. It consists of a bow-tie-shaped planar electric dipole and a folded shorted quarter-wave patch antenna. The impedance bandwidth of the antenna (SWR $\leq$2) is 95.2 % from 1.65 to 4.65 GHz. Within the operating frequency band, the antenna gain is about 7.9 dBi with 0.9 dBi variation. Subsequently, this antenna was modified to achieve a wider bandwidth for the coverage of the whole UWB unlicensed band (Ge and Luk 2013b). The modified antenna, as illustrated in Fig. 45, is excited by a printed balun instead of the folded Γ-shaped probe. And a pair of metallic cylindrical posts performs as the magnetic dipole in the new design. The measured impedance bandwidth (SWR $\leq$2) is about 110 % from 3.08 to 10.6 GHz, over which the measured antenna gain varies between 6.8 and 10.6 dBi with 3.8 dBi variation.

Another method for broadening the bandwidth of the ME dipole is to utilize a pair of differentially feeds to excite the antenna. Experimental and parametric studies were presented by Li and Luk (2013a). Fig. 46 shows the geometry of the antenna. Similar to the original ME dipole shown in Fig. 5, this antenna is composed of a planar electric dipole and a shorted patch antenna. Two slots in the electric dipole are

DOI 10.1007/978-981-4560-75-7_49-1

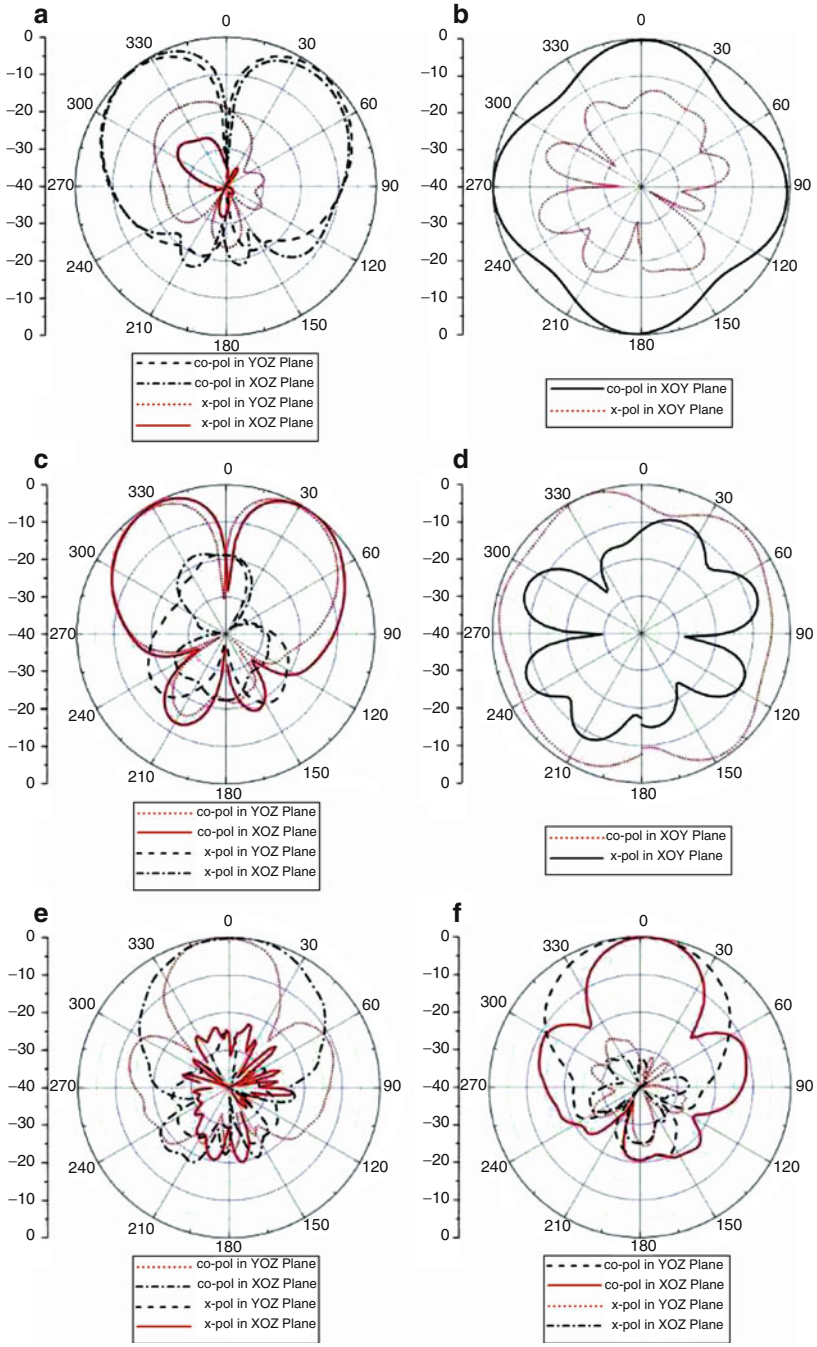


Fig. 42 Measured radiation patterns at 2.4 GHz for various modes. (**a**) Elevation plane for electric monopole conical mode, (**b**) azimuth plane for electric monopole conical mode, (**c**) elevation plane for magnetic monopole conical mode, (**d**) azimuth plane for magnetic monopole conical mode, (**e**) elevation plane for +45° broadside mode, and (**f**) elevation plane for −45° broadside mode (From (Wu and Luk) © 2011a IEEE, Reprinted with permission)

introduced to suppress the high-order mode radiation, further reducing the side-lobe level at the high frequencies. Instead of a planar reflector, the ME dipole is mounted on a rectangular box-shaped reflector which can reduce back radiation and stabilize antenna gain. The antenna provides an impedance bandwidth of 114 % from 2.95 GHz to 10.73 GHz, covering the whole UWB bands. More importantly, the antenna broadside gain is 8.25 dBi with a small variation of ± 1.05 dBi, which is much more stable than other unidirectional UWB antennas.

DOI 10.1007/978-981-4560-75-7_49-1

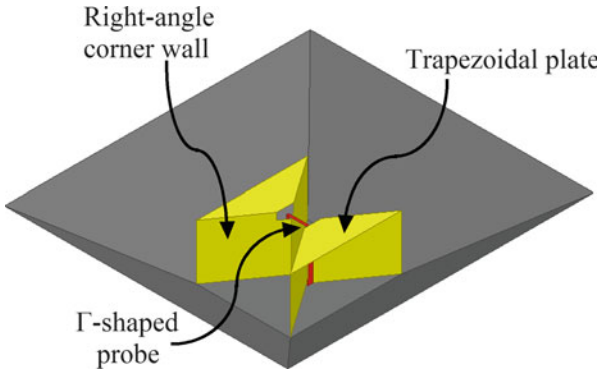


Fig. 43 Geometry of the UWB trapezoidal ME dipole antenna

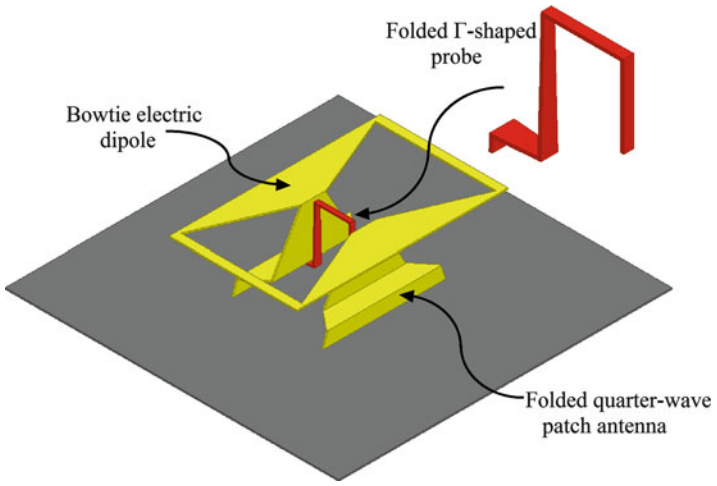


Fig. 44 Geometry of the UWB bow-tie ME dipole antenna

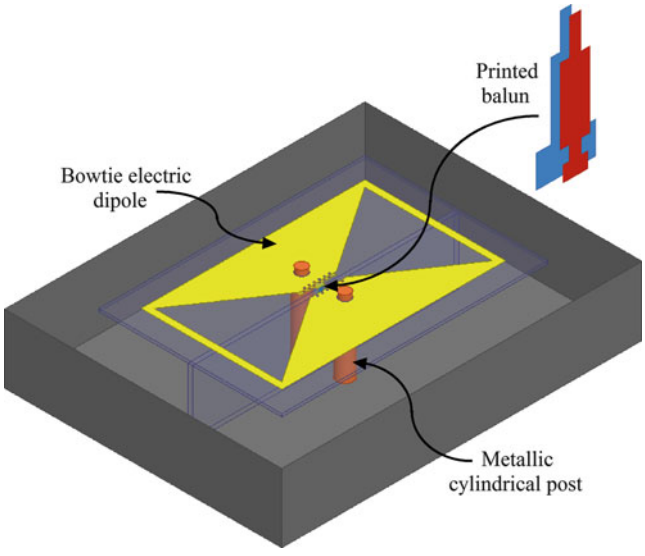


Fig. 45 Geometry of the UWB-modified bow-tie ME dipole antenna

DOI 10.1007/978-981-4560-75-7_49-1

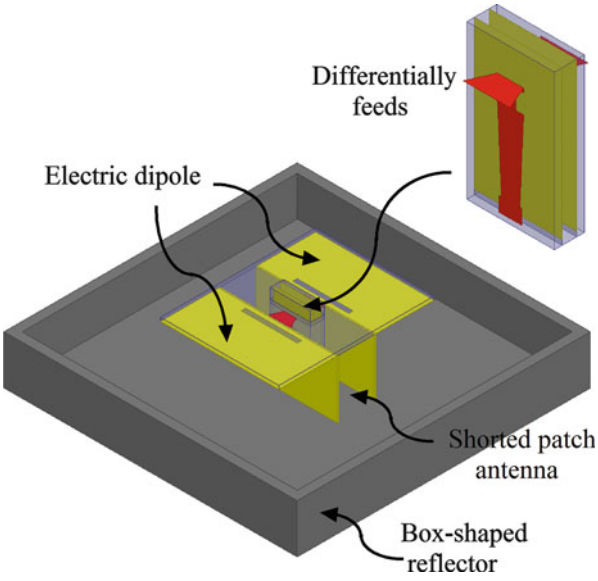


Fig. 46 Geometry of the differentially fed UWB ME dipole antenna

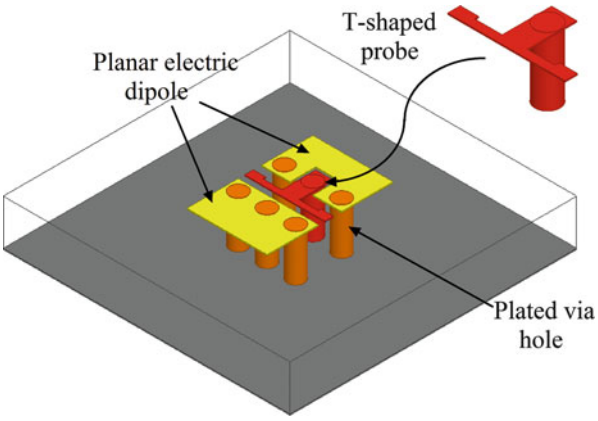


Fig. 47 Geometry of the 60-GHz ME dipole antenna with T-shaped probe

Designs for 60-GHz Applications

Within the millimeter-wave spectrum, the 60-GHz band is useful for short-range, high-frequency reuse and secure communications because of an oxygen absorption peak occurring in this band. High data rates of multi-gigabits are supported by 60-GHz wireless communications, allowing for many exciting applications, such as a high-definition multimedia interface, uncompressed video streaming, mobile distributed computing, rapid large file transfer, high-speed Internet, sensing, and radar. As a result, the 60-GHz radio technology, representing a valuable opportunity for circuit and antenna engineers, has attracted considerable interest, especially after a global consensus on the 60-GHz band allocation was reached by federal agencies worldwide.

Ng et al. (2012) first proposed a 60-GHz ME dipole antenna on a single-layer printed circuit board. Duroid 5880 substrate with thickness of 0.787 mm, metal thickness of 1/2 oz, $\varepsilon_r = 2.2$, and $\tan\delta = 0.004$ at 60 GHz was chosen for its low loss tangent and approximately $\lambda_g/4$ thickness. The geometry of the antenna is shown in Fig. 47. It can be seen that a planar electric dipole is shorted to the ground by five

DOI 10.1007/978-981-4560-75-7_49-1

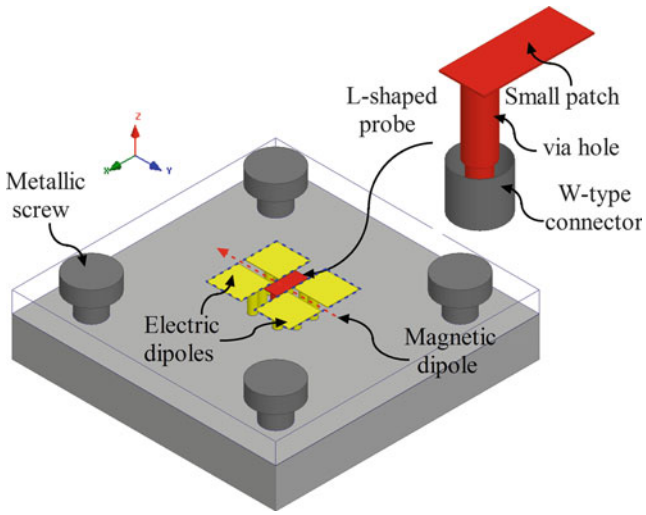


Fig. 48 Geometry of the 60-GHz ME dipole antenna with L-shaped probe

plated via holes. These via holes together with the ground plane between them form a shorted quarter-wave patch antenna. Instead of the Γ-shaped probe of the original model shown in Fig. 5, a T-shaped probe is applied because it can be easily printed on a PCB. As discussed in section "Other Feed Mechanisms," the ME dipole with T-shaped probe suffers from higher back radiation and cross-polarization levels. This antenna exhibits a wide impedance bandwidth ($S_{11} \leq -10$) of 43 % from 48 to 75 GHz, over which the average gain is approximately 7.5 dBi. The radiation pattern is stable and equal in the E- and H-planes. However, the cross-polarization level of the antenna is not as low as that of the original ME dipole, varying between −15 dB and −21 dB across the operating bandwidth from 50 to 70 GHz.

Subsequently, another 60-GHz ME dipole antenna was designed by Luk and Li (2013). This antenna used the same Duroid 5880 substrate of the previous design. Fig. 48 shows the geometry of the antenna which consists of four rectangular metallic patches, four sets of via holes, and an L-shaped probe. The metallic patches perform as two planar electric dipoles. Each set of via holes, consisting of three metallic plated via holes, is located close to the inner corner of a rectangular metallic patch. The via holes together with the ground plane between them form a shorted quarter-wave patch antenna. The L-shaped probe is composed of a plated via hole and a small patch. The via hole, together with the adjacent via holes, accomplishes a transmission line to transmit signals to the small patch. This patch is placed between two planar dipoles which act as its electrical ground. This combination works as a CPW line, the portion of which across the separation between two dipole arms is responsible to couple the electrical energy to the antenna element. Thus, the L-shaped probe can excite the planar dipoles and shorted patch antenna simultaneously. A W-type connector (Anritsu: W1-103 F) is launched underneath the antenna fixture for transmitting signal to the L-shaped probe. Four metallic screws were used to fix the substrate on the antenna fixture. The impedance bandwidth of this antenna is 50.5 % with SWR ≤2 from 41.5 to 69.5 GHz. The antenna gain is about 8 dBi with a variation of about 1.1dBi. Over the entire operating band, the cross-polarization level is less than −20 dB, which is lower than the design that is excited by a T-shaped probe.

DOI 10.1007/978-981-4560-75-7_49-1

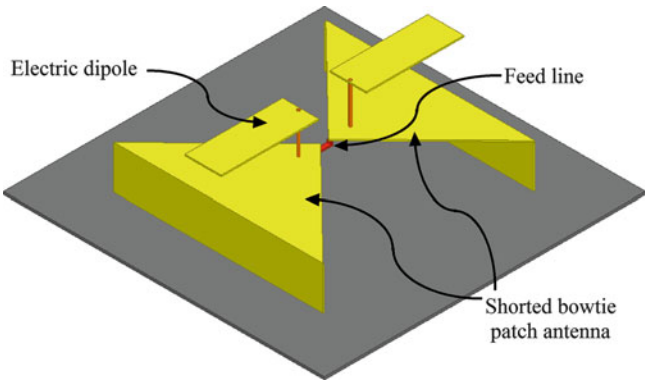


Fig. 49 Geometry of the shorted bow-tie patch antenna with electric dipole

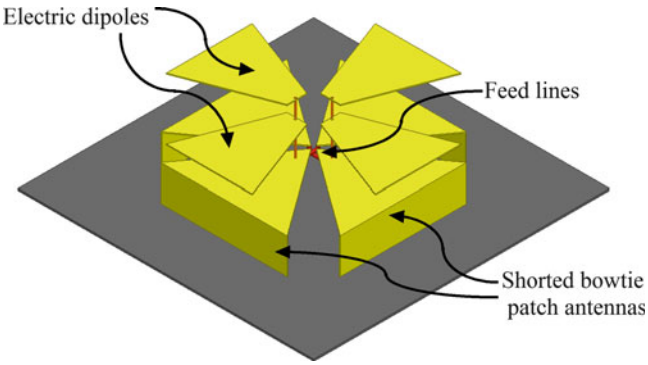


Fig. 50 Geometry of the dual-polarized shorted bow-tie patch antenna with electric dipole

Other Complementary Antenna Designs

Apart from ME dipole antenna, many other complementary antennas have been demonstrated in the literature. The combination of an electric dipole and a shorted bow-tie patch antenna resulted in a wideband complementary antenna, as shown in Fig. 49 (Wong et al. 2008). This antenna exhibits an impedance bandwidth of 63 %. It should be pointed out that without the electric dipole, the antenna can also achieve an impedance bandwidth of 62 %. Thus, the addition of an electric dipole can only enhance the gain, suppress the back radiation, and stabilize the frequency response of gain and radiation patterns. Based on this design, a dual-polarized antenna was developed by Mak et al. (2007). The geometry of the antenna is depicted in Fig. 50. This antenna, operated in 1.71–2.17 GHz, has a stable gain of 6.6 dBi, a stable beamwidth of 80°, a large front-to-back ratio of over 19 dB, and a high isolation of over 28 dB. Later, the attempt to reduce the height of this type of antenna was made by Li and Luk (2013b), as shown in Fig. 51. By placing two pairs of electric dipoles next to the shorted bow-tie patch antenna, the antenna height can be reduced to less than $0.1\lambda_0$ which is close to the height level of a wideband microstrip patch antenna. Meanwhile, it maintains all good electrical characteristics of a complementary antenna.

Another kind of the complementary antenna is composed of a printed electric dipole and a printed wire loop. An antenna of this type was investigated by Chan et al. (2007). It radiates unidirectionally without the use of a reflector but suffers from a narrow impedance bandwidth of 7.4 %. When a semiring is placed on a ground plane to perform as the magnetic dipole, instead of the loop antenna, the antenna bandwidth can be broadened (Tang and Xue 2013; Wang et al. 2014).

DOI 10.1007/978-981-4560-75-7_49-1

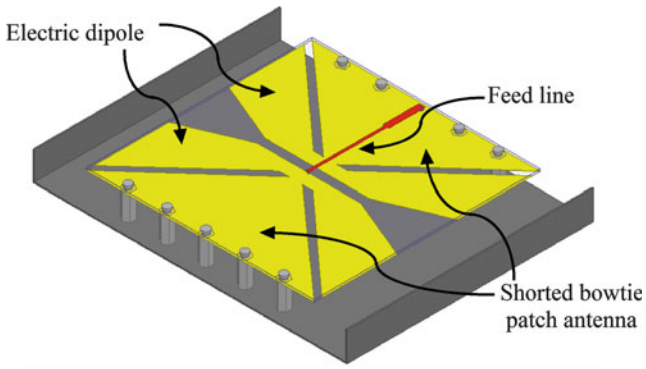


Fig. 51 Geometry of the low-profile shorted bow-tie patch antenna with electric dipole

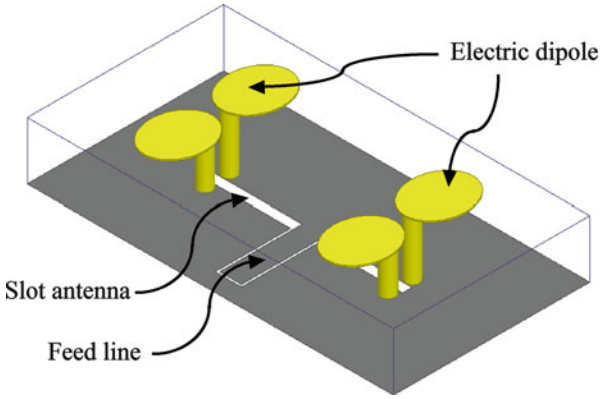


Fig. 52 Geometry of the dipole/slot antenna combination at 60 GHz

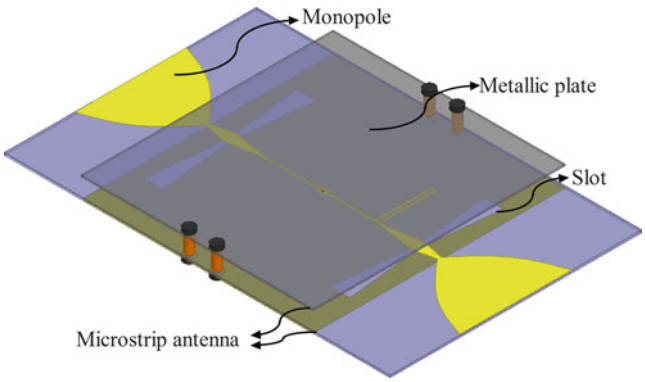


Fig. 53 Geometry of the low-profile monopole/slot/microstrip antenna combination

Subsequently, a 60-GHz complementary antenna was realized by the use of slot antenna and parasitic dipole (Ng et al. 2014). The geometry of the antenna is depicted in Fig. 52. It can be seen that the traditional CPW-fed slot antenna acts as a magnetic source, and a parasitic dipole, located across the slot, acts as an electric source. More recently, a combination of monopole/slot/microstrip antenna was demonstrated by Li and Luk (2015). This design, as shown in Fig. 53, has a low profile of $0.035\lambda_0$ and exhibits a wide impedance bandwidth of over 20 %, a stable broadside gain of 6.11 dBi, and a

DOI 10.1007/978-981-4560-75-7_49-1

unidirectional radiation pattern with low back radiation (< -16 dB) and low cross-polarization (< -25 dB).

Conclusion

This chapter began with discussions on the theory of the complementary antenna, the basic configuration of the magnetoelectric dipole antenna, methods of height reduction, and feed mechanisms. And then some magnetoelectric dipole antennas with dual polarizations, circular polarization, and pattern polarization diversities have been studied in this chapter. These designs can be used to solve multipath-fading effects and enhance the performance of wireless system. The chapter then proceeded to present several antenna solutions for UWB and 60-GHz wireless communications followed by some other types of complementary antennas.

Since the magnetoelectric dipole antenna was created in 2006, there have been many studies in the literature, focusing on the development of the ME dipole antenna to have different polarizations and fulfill various applications. However, a number of challenges remain. For example, most of the designs provide wide bandwidth, while dual-band or multiband ME dipole designs have not been adequately considered. As we know, the mobile base station usually operates in multibands with dual polarizations nowadays, but the design of dual-band or tri-band dual-polarized ME dipole antenna is a big challenge to us. In addition, the research on wideband and low-profile ME dipole antenna is still an active research area. Furthermore, the realization of ME dipoles at higher frequencies, such as 120 GHz or even Terahertz, is an unknown area till now.

Acknowledgment

This work was supported in part by the Research Grants Council of the Hong Kong SAR (Project No. CityU 119511 (9041677)).

Cross-References

▶ Antenna Systems for Cellular Base Stations
▶ Circularly Polarized Antennas
▶ Microstrip Patch Antennas
▶ Millimeter-Wave Antennas and Arrays
▶ Ultra-wideband Antennas

References

Caso R, Serra AA, Pino M, Nepa P, Manara G (2010) A wideband slot-coupled stacked-patch array for wireless communications. IEEE Antennas Wirel Propag Lett 9:986–989

Chan PW, Wong H, Yung EKN (2007) Unidirectional antenna composed of dipole and loop. Electron Lett 43(22):1176–1178

Clavin A (1954) A new antenna feed having equal E-and H-plane patterns. IRE Trans Anntenas Propag 2(3):113–119

DOI 10.1007/978-981-4560-75-7_49-1

Cui YH, Li RL, Wang P (2013) A novel broadband planar antenna for 2G/3G/LTE base stations. IEEE Trans Antennas Propag 61(5):2767–2774

Ge L, Luk KM (2012a) A low-profile magneto-electric dipole antenna. IEEE Trans Antennas Propag 60(4):1684–1689

Ge L, Luk KM (2012b) A wideband magneto-electric dipole antenna. IEEE Trans Antennas Propag 60(11):4987–4991

Ge L, Luk KM (2013a) A magneto-electric dipole antenna with low-profile and simple structure. IEEE Antennas Wirel Propag Lett 12:140–142

Ge L, Luk KM (2013b) A magneto-electric dipole for unidirectional UWB communications. IEEE Trans Antennas Propag 61(11):5762–5765

Guo YX, Bian L, Shi XQ (2009) Broadband circularly polarized annular-ring microstrip antenna. IEEE Trans Antennas Propag 57(8):2474–2477

King RWP, Owyang GH (1960) The slot antenna with coupled dipoles. IRE Trans Anntenas Propag 8(2):136–143

Lai HW, Luk KM (2007) Dual polarized patch antenna fed by meandering probes. IEEE Trans Antennas Propag 55(9):2625–2627

Li MJ, Luk KM (2012) A wideband dual-polarized antenna with very low back radiation. In: 2012 Asia-Pacific microwave conference proceedings (APMC), IEEE, Kaohsiung, pp 61–63

Li MJ, Luk KM (2013a) A differential-fed magneto-electric dipole antenna for UWB applications. IEEE Trans Antennas Propag 61(1):92–99

Li MJ, Luk KM (2013b) A low-profile wideband planar antenna. IEEE Trans Antennas Propag 61(9):4411–4418

Li MJ, Luk KM (2013c) A wideband dual-fed circularly polarized antenna. In: 2013 international workshop on electromagnetics proceedings (iWEM), IEEE, Hong Kong, pp 112–114

Li MJ, Luk KM (2014) A wideband circularly polarized antenna for microwave and millimeter-wave applications. IEEE Trans Antennas Propag 62(4):1872–1879

Li MJ, Luk KM (2015) A low-profile, low-backlobe, and wideband complementary antenna for wireless application. IEEE Trans Antennas Propag 63(1):7–14

Li B, Hao CX, Sheng XQ (2009) A dual-mode quadrature-fed wideband circularly polarized dielectric resonator antenna. IEEE Antennas Wirel Propag Lett 8:1036–1038

Luk KM, Li MJ (2013) Magneto-electric dipole antennas for millimeter-wave applications. In: Asia-Pacific microwave conference proceedings (APMC), IEEE, Seoul, pp 304–306

Luk KM, Wong H (2006) A new wideband unidirectional antenna element. Int J Microgr Opt Technol 1(1):35–44

Luk KM, Wu B (2012) The magnetoelectric dipole – a wideband antenna for base stations in mobile communications. Proc IEEE 100(7):2297–2307

Mak KM, Wong H, Luk KM (2007) A shorted bowtie patch antenna with a cross dipole for dual polarization. IEEE Antennas Wirel Propag Lett 6:126–129

Mayes PE, Warren W, Wiesenmeyer F (1972) The monopole slot: a small broad-band unidirectional antenna. IEEE Trans Antennas Propag 20(4):489–493

Nakano H, Oyanagi H, Yamauchi J (2011) A wideband circularly polarized conical beam from a two-arm spiral antenna excited in phase. IEEE Trans Antennas Propag 59(10):3518–3525

Ng KB, Wong H, So KK, Chan CH, Luk KM (2012) 60 GHz plated through hole printed magneto-electric dipole antenna. IEEE Trans Antennas Propag 60(7):3129–3136

Ng KB, Chan CH, Zhang H, Zeng G (2014) Bandwidth enhancement of planar slot antennas using complementary sources technique in millimeter wave applications. IEEE Trans Antennas Propag 62(9):4452–4458

Pan YM, Leung KW, Lu K (2012) Omnidirectional linearly and circularly polarized rectangular dielectric resonator antennas. IEEE Trans Antennas Propag 60(2):751–759

Paryani RC, Wahid PF, Behdad N (2010) A wideband, dual-polarized, substrate-integrated cavity-backed slot antenna. IEEE Antennas Wirel Propag Lett 9:645–648

Siu L, Wong H, Luk KM (2009) A dual-polarized magneto-electric dipole with dielectric loading. IEEE Trans Antennas Propag 57(3):616–623

Sun YX, Leung KW (2013) Dual-band and wideband dual-polarized cylindrical dielectric resonator antennas. IEEE Antennas Wirel Propag Lett 12:384–387

Tang C, Xue Q (2013) Vertical planar printed unidirectional antenna. IEEE Antennas Wirel Propag Lett 12:368–371

Turkmani AMD, Arowojolu A, Jefford PA, Kellett CJ (1995) An experimental evaluation of the performance of two-branch space and polarization diversity schemes at 1800 MHz. IEEE Trans Veh Technol 44(2):318–326

Vaughan RG (1990) Polarization diversity in mobile communications. IEEE Trans Veh Technol 39(3):177–186

Wang H, Liu SF, Chen L, Li WT, Shi XW (2014) Gain enhancement for broadband vertical planar printed antenna with H-shaped resonator structures. IEEE Trans Antennas Propag 62(8):4411–4415

Wong H, Lau KL, Luk KM (2004) Design of dual-polarized L-probe patch antenna arrays with high isolation. IEEE Trans Antennas Propag 52(1):45–52

Wong H, Mak KM, Luk KM (2008) Wideband shorted bowtie patch antenna with electric dipole. IEEE Trans Antennas Propag 56(7):2098–2101

Wu BQ, Luk KM (2009a) A broadband dual-polarized magneto-electric dipole antenna with simple feeds. IEEE Antennas Wirel Propag Lett 8:60–63

Wu BQ, Luk KM (2009b) A magneto-electric dipole with a modified ground plane. IEEE Antennas Wirel Propag Lett 8:627–629

Wu BQ, Luk KM (2009c) A wideband, low-profile, conical-beam antenna with horizontal polarization for indoor wireless communications. IEEE Antennas Wirel Propag Lett 8:634–636

Wu BQ, Luk KM (2011a) A 4-port diversity antenna with high isolation for mobile communications. IEEE Trans Antennas Propag 59(5):1660–1667

Wu BQ, Luk KM (2011b) A UWB unidirectional antenna with dual-polarization. IEEE Trans Antennas Propag 59(11):4033–4040

Xue Q, Liao SW, Xu JH (2013) A differentially-driven dual-polarized magneto-electric dipole antenna. IEEE Trans Antennas Propag 61(1):425–430

Yang SLS, Luk KM (2006) Design of a wide-band L-probe patch antenna for pattern reconfiguration or diversity applications. IEEE Trans Antennas Propag 54(2):433–438

Yang SLS, Luk KM, Lai HW, Kishk AA, Lee KF (2008) A dual-polarized antenna with pattern diversity. IEEE Antennas Propag Mag 50(6):71–79

Zhang ZY, Liu NW, Zhao JY, Fu G (2013) Wideband circularly polarized antenna with gain improvement. IEEE Antennas Wirel Propag Lett 12:456–459

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_50-2

Reduced Surface Wave Microstrip Antennas

David R. Jackson*
Department of ECE, Cullen College of Engineering, University of Houston, Houston, TX, USA

Abstract

Reduced surface wave (RSW) microstrip antennas are a class of microstrip antennas that excite much less surface wave fields than do conventional microstrip antennas. Furthermore, they have a greatly reduced amount of lateral (horizontal) radiation that propagates outward from the antenna along the ground plane. This results in less edge diffraction from the edges of the ground plane or supporting structure, which in turn results in smoother front-side patterns and less back radiation. The RSW antennas also exhibit less mutual coupling when the antennas are widely spaced, due to the reduced surface wave field and lateral radiation. This chapter reviews RSW antenna design and the different methods for realizing such antennas, and illustrates their main performance features.

Keywords

Microstrip antenna; Reduced surface wave antenna; Surface waves; Lateral wave; Mutual coupling; Edge diffraction; GPS

Introduction

A common property of most microstrip antennas is that the antenna excites the dominant TM_0 surface wave mode of the grounded substrate, in addition to the fields radiated into space (see Fig. 1). The surface wave is a guided mode with a wavenumber β_{TM0} that spreads out radially in the horizontal plane, decaying slowly as $1/\rho^{1/2}$, where ρ is the distance from the antenna in cylindrical coordinates. Physically, for large distances ρ, the surface wave behaves as a plane wave that bounces up and down inside the substrate beyond the critical angle, decaying exponentially in the air region above (symbolized in Fig. 1 by the horizontal lines of decreasing size), since $\beta_{TM0} > k_0$ (Pozar 1998; Harrington 2001). Because of the slow decay of the surface wave with distance ρ, it can cause coupling and other undesirable effects at a considerable distance from the antenna, if the substrate extends out to large distances. The power launched into the surface wave is power that will eventually be lost, at least for the case of an infinite substrate. For finite-size substrates, the surface wave field will reflect and diffract from the edges of the substrate, resulting in a disturbance of the radiation pattern. Furthermore, the excitation of the TM_0 surface wave also results in increased mutual coupling between distant antenna elements, since the surface wave fields decay more slowly with radial distance than does the space-wave field. For these reasons, the excitation of the surface wave from a microstrip antenna is undesirable. Since the dominant TM_0 surface wave of a grounded dielectric layer has a zero cutoff frequency, a microstrip antenna will usually always excite some surface wave power unless some mitigating technique is used. A reduced surface wave (RSW) microstrip antenna is one that is designed to have a minimum amount of surface wave excitation. This chapter reviews the design and performance of RSW microstrip antennas.

*Email: djackson@uh.edu

DOI 10.1007/978-981-4560-75-7_50-2

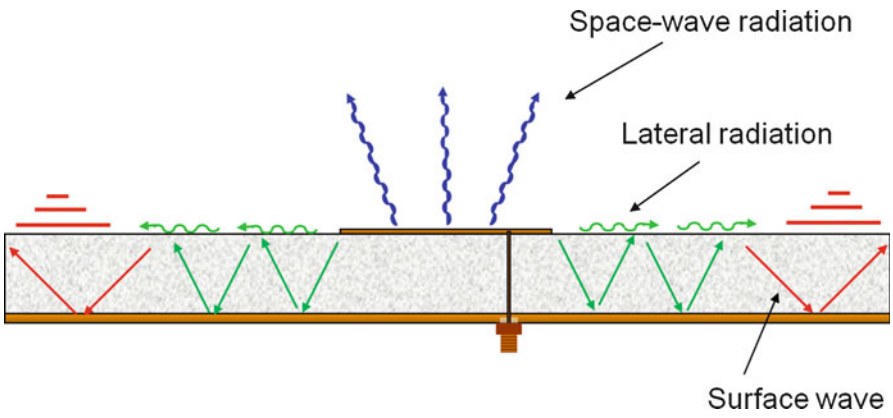


Fig. 1 A sketch showing the different types of waves radiated by a microstrip antenna

In addition to surface wave excitation, microstrip antennas produce lateral radiation, i.e., radiation that propagates outward horizontally (laterally) from the antenna. The lateral radiation exists at the dielectric interface and in the vicinity of the interface in the air region above, with the wavenumber of free space (see Fig. 1). Inside the substrate, for large distances ρ, this field has the character of a plane wave field that bounces between the interface and the ground plane at the critical angle. The lateral radiation initially decays as $1/\rho$, as it would without the substrate. At extreme distances, the lateral radiation field evolves into a "lateral wave" type of field that decays as $1/\rho^2$ due to the interface effect, assuming that the substrate extends out to such extreme distances (King 1992). The lateral-radiation field is often more significant than the surface wave field for small and moderate distances from the antenna, depending on the substrate permittivity and thickness. For example, for a typical microstrip substrate, the lateral-radiation field might be the most significant contributor to the field on the interface out to a distance of about 10 wavelengths, after which the surface wave field becomes dominant. While the space-wave radiation away from the antenna is desirable, as it forms the radiation pattern of the antenna, the lateral radiation along the substrate and ground plane is undesirable. The lateral radiation will diffract from the edges of the ground plane or supporting structure, and this edge diffraction will often cause significant distortion of the radiation pattern, including rippling of the pattern in the forward direction and significant back radiation behind the ground plane. The lateral radiation will also result in significant mutual coupling between antennas and between the antenna and other components.

The same design principles that are used to design RSW antennas can be used (with slight modification) to design reduced lateral wave (RLW) antennas as well. (The term "lateral wave" is used here to denote lateral radiation in general, regardless of the distance from the antenna or whether the substrate is present.) Fortunately, in most practical cases, the wavenumber β_{TM0} of the TM_0 surface wave is fairly close to that of the free-space wavenumber k_0 (unless the substrate is electrically thick). This means that an RSW antenna, designed to greatly reduce the excitation of the TM_0 surface wave, will automatically have a greatly reduced lateral radiation as well.

This chapter will first review the basic design concept of the RSW antenna, based on the "shorted annular ring" (SAR) concept, which is perhaps the simplest way to fabricate an RSW antenna (Jackson et al. 1993). The SAR-RSW antenna design is based on a circular patch of radius b that has a ring of shorting vias located at a radius a in the interior, forming an active annular cavity region $a < \rho < b$. The inner cavity region ($\rho < a$) is a dead region, which can be removed to form an annular ring microstrip antenna structure that is shorted at the inner boundary (and hence the name SAR). Other design variations use a circular patch that has a cylindrical region in the middle of the substrate cored out and replaced by a lower permittivity material, or a circular patch that has the entire substrate inside the cavity region

DOI 10.1007/978-981-4560-75-7_50-2

replaced by an artificial substrate with a lower permittivity than that of the surrounding substrate outside the patch. Results will be presented to demonstrate the improved radiation patterns of the SAR-RSW antennas and the reduced mutual coupling between pairs of such antennas, relative to conventional microstrip antennas on the same substrate (Khayat et al. 2000).

By feeding the SAR-RSW antenna with two 90° feed ports, a circularly polarized (CP) RSW antenna is obtained. One application for CP RSW antennas is for applications that require low levels of lateral radiation and corresponding edge diffraction that would otherwise result in undesirable multipath reception from signals arriving from the back region or at low angles. Conventional antenna systems in applications such as GPS use choke rings to suppress the lateral radiation and reduce the edge diffraction. The RSW antenna is an alternative, being planar and significantly smaller in size (Basilio et al. 2005).

Another design variation of the SAR antenna, called the ISAR ("inverted shorted annular ring") antenna, is also examined. In this design variation, an annular ring microstrip antenna is shorted at the outer boundary, and the inner edge of the annular ring becomes the radiating edge (Basilio et al. 2008). The interior of the annular ring is open space, and other components may be inserted in this region. By inserting an SAR-RSW antenna in this region, a composite SAR/ISAR RSW antenna is obtained. By designing the resonances of the SAR and ISAR antennas to be at different frequencies, a dual-band RSW antenna can be obtained (Basilio et al. 2007). Results will be presented for such a structure.

A general "RSW theorem" is then discussed, which gives a general method for constructing RSW antennas, which do not have to be based on a circular shape as the SAR and ISAR designs are (Komanduri et al. 2013). The theorem states that a microstrip antenna of arbitrary shape will become an RSW antenna if the cavity filling material is chosen to have an effective relative permittivity that is the same as the square of the normalized wavenumber of the TM_0 surface wave, i.e., $\varepsilon_r^{eff} = (\beta_{TM0}/k_0)^2$. This generalizes the design of the RSW antenna to other shapes than those based on a circular geometry.

Although the SAR-RSW antenna has good performance, its main limitation is that its size is larger than that of a convention microstrip antenna. Furthermore, the size of the RSW antenna does not scale with the substrate permittivity, as it does for conventional microstrip antennas. The diameter of the RSW antenna is roughly 0.59 λ_0, independent of the substrate permittivity. Furthermore, the mutual coupling between RSW antennas becomes less than that for conventional microstrip antennas only when the spacing between the antenna centers is larger than about 0.75 λ_0. This means that the RSW antenna is not suitable for phased-array applications that require a large scan range. However, it may be suitable for arrays with limited scan ranges, and the performance of RSW antennas in such arrays is examined here (Chen et al. 2005).

RSW Design Principle

The RSW antennas discussed in the next section are all based on the same principle, namely that a ring of magnetic current inside of a grounded substrate that varies as $\cos\phi$ will not excite the TM_0 surface wave if the radius of the ring is chosen properly. This principle is explained here.

The designs introduced in the next section are all based on a modification of a circular patch, which has a circular radiating edge. If we consider a circular patch antenna operating in the TM_{mn} mode with $m = 1$, the electric field at the edge of the patch varies as $\cos\phi$. The electric field of the resonant mode at the patch edge is given by $E_z = \cos\phi$. In the magnetic current model of the microstrip antenna (Jackson 2007), which is based on the equivalence principle (Harrington 2001), this corresponds to a magnetic surface current at the edge of the patch that is

$$M_{s\phi} = \cos\phi. \tag{1}$$

DOI 10.1007/978-981-4560-75-7_50-2

The surface wave excitation from this sheet of magnetic surface current may be understood by examining the canonical problem of a ring of magnetic current of radius b having a current $K = \cos\phi$ at some height z_0 above the ground plane as shown in Fig. 2, as the sheet of magnetic surface current $M_{s\phi}$ can be thought of as a superposition of such rings. This current ring model is only an approximate model of a circular patch antenna because the aperture is not a perfect magnetic conductor in reality. Furthermore, the simple current function of Eq. 1 neglects all higher-order TM_{mn} modes with $m \neq 1$ that are excited by the feed probe. The approximation of the patch by the magnetic current in Eq. 1 becomes more accurate as the quality factor Q of the patch antenna increases, which means the substrate becomes thinner.

Consider first the radiation from a single infinitesimal (Hertzian) dipole of magnetic current oriented in the x direction, at a height z_0 above the ground plane along the z axis. This infinitesimal dipole will excite a TM_0 surface wave field of the form

$$\psi^{sw} = A(z, z_0)\, H_1^{(2)}(\beta_{TM0}\rho)\, \sin\phi, \tag{2}$$

where ψ denotes the field component of interest, e.g., the field E_z along the ground plane. The wavenumber of the TM_0 surface wave β_{TM0} can be determined numerically once the substrate thickness, permittivity, and operating frequency are specified (for example, using a bisection or secant method search on the real axis between k_0 and k_1). The amplitude factor $A(z,z_0)$ depends on the height of the source and the observation point above the ground plane. This factor can be determined from the residue of the surface wave pole in a spectral-domain solution (Pozar 1982), though the value is not needed for the present discussion. Far away from the dipole, the surface wave field has the approximate form

$$\psi^{sw} = A(z, z_0)\sqrt{\frac{2}{\pi}}\, e^{j3\pi/4}\, \frac{e^{-j\beta_{TM0}\rho}}{\sqrt{\beta_{TM0}\rho}}\, \sin\phi. \tag{3}$$

Next, consider the surface wave excitation from the ring of magnetic current $K = \cos\phi$ shown in Fig. 2. The surface wave field from the ring may be found from the field of the Hertzian dipole, integrating over the ring current. If the observation point is sufficiently far from the ring, a far-field approximation for the phase of the surface wave radiated from each point on the ring may be used. The phase term is then written as

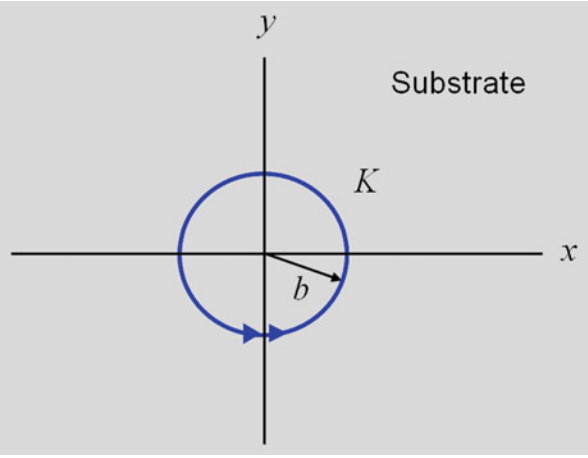


Fig. 2 A ring of magnetic current $K = \cos\phi$ inside of a grounded substrate (*top view*)

DOI 10.1007/978-981-4560-75-7_50-2

$$e^{-j\beta_{\mathrm{TM0}}|\underline{r}-\underline{r}'|} \approx e^{-j\beta_{\mathrm{TM0}}\rho} e^{+j\beta_{\mathrm{TM0}}(x'\cos\phi + y'\sin\phi)} = e^{-j\beta_{\mathrm{TM0}}\rho} e^{+j\beta_{\mathrm{TM0}} b\cos(\phi-\phi')}, \tag{4}$$

where ϕ is the observation angle and ϕ' is the angle of the source point on the ring. The surface wave field excited from the ring is then

$$\begin{aligned}\psi^{sw} = A(z,z_0)\sqrt{\frac{2}{\pi}}\, e^{j3\pi/4}\, \frac{e^{-j\beta_{\mathrm{TM0}}\rho}}{\sqrt{\beta_{\mathrm{TM0}}\rho}} \\ \cdot \int_0^{2\pi} \cos\phi' \sin(\phi - (\phi' + \pi/2)) e^{+j\beta_{\mathrm{TM0}} b\cos(\phi-\phi')} b\, d\phi'.\end{aligned} \tag{5}$$

Using $\alpha = \phi - \phi'$ and shifting the range of integration back to $(0, 2\pi)$, the integration may be written as

$$\begin{aligned}\psi^{sw} = -A(z,z_0)\sqrt{\frac{2}{\pi}}\, e^{j3\pi/4}\, \frac{e^{-j\beta_{\mathrm{TM0}}\rho}}{\sqrt{\beta_{\mathrm{TM0}}\rho}} \\ \cdot \int_0^{2\pi} \left(\cos\phi\cos^2\alpha + \sin\phi\sin\alpha\cos\alpha\right) e^{+j\beta_{\mathrm{TM0}} b\cos\alpha} b\, d\alpha.\end{aligned} \tag{6}$$

To evaluate the integral in Eq. 6, the Jacobi–Anger expansion is used (Abramowitz and Stegun 1970), which is

$$e^{j\beta_{\mathrm{TM0}} b\cos\alpha} = J_0(\beta_{\mathrm{TM0}} b) + 2\sum_{n=1}^{\infty} j^n J_n(\beta_{\mathrm{TM0}} b)\cos(n\alpha). \tag{7}$$

The trigonometric identities

$$\cos^2\alpha = \frac{1+\cos(2\alpha)}{2}, \quad \sin\alpha\cos\alpha = \frac{1}{2}\sin(2\alpha) \tag{8}$$

are then used, and orthogonality then allows us to keep only the $n = 0$ and $n = 2$ terms in Eq. 7. Using the identity

$$J_1'(x) = \frac{1}{2}(J_0(x) - J_2(x)) \tag{9}$$

then allows Eq. 6 to be written as

$$\psi^{sw} = -A(z,z_0)(2\pi b)\sqrt{\frac{2}{\pi}}\, e^{j3\pi/4}\, \frac{e^{-j\beta_{\mathrm{TM0}}\rho}}{\sqrt{\beta_{\mathrm{TM0}}\rho}} \cos\phi J_1'(\beta_{\mathrm{TM0}} b). \tag{10}$$

Equation 10 implies that a ring of magnetic current will not excite the TM_0 surface wave, provided the radius is chosen to satisfy

$$J_1'(\beta_{\mathrm{TM0}} b) = 0, \tag{11}$$

DOI 10.1007/978-981-4560-75-7_50-2

which implies that

$$\beta_{TM0} b = x'_{11} = 1.8412, \tag{12}$$

where x'_{11} denotes the first (smallest) root of the $J_1'(x)$ Bessel function. Any other root x'_{1n} can also be used in Eq. 12, but to have the smallest possible ring, the value $x'_{11} = 1.8412$ is chosen. This corresponds to an SAR-RSW antenna with the smallest possible radius.

A circular patch having a radius chosen according to Eq. 12 will not in general be of resonant size. According to the cavity model, a resonant circular patch operating in the TM_{11} mode will have a radius determined from

$$k_1 b = x'_{11}, \tag{13}$$

where k_1 is the wavenumber of the substrate, which is the same as the wavenumber inside the patch cavity for a conventional circular patch. Since $k_0 < \beta_{TM0} < k_1$, the radius of a circular patch cannot satisfy both Eqs. 12 and 13. A circular patch with a radius chosen from Eq. 12 will be too large to be resonant at the operating frequency (at which β_{TM0} is calculated). Hence, the patch must be modified to make it resonant, while keeping the radius of the radiating edge fixed according to Eq. 12. There are various ways of modifying the resonance frequency of a circular patch, and this leads to several different RSW designs discussed below.

RSW Designs

Shorted Annular Ring (SAR)

The SAR-RSW antenna is shown in Fig. 3. It consists of a circular patch of radius b, with a ring of shorting vias located at radius a, and is fed by a probe at radius ρ_0. The radius b of the antenna is chosen according to Eq. 12. The inner radius a of the shorting vias is then chosen to make the antenna resonant at the operating frequency (the frequency at which β_{TM0} is calculated). The region $\rho < a$ inside the patch cavity is a "dead" region where the fields are zero. Hence, the patch metal can be removed from this circular region, creating an annular ring that is shorted on the inner boundary at $\rho = a$ (and hence the name "shorted annular ring"). The term "shorted annular ring" is used to denote the structure of Fig. 3, regardless of whether or not the interior metal region is present.

A simple cavity model analysis can be used to determine the radius a of the via ring in Fig. 3 to make the SAR-RSW structure resonant. In this approximate analysis, a perfect magnetic conductor (PMC) wall is assumed at the outer radius b, and a perfect short circuit is assumed at the inner radius a. The electric field inside the cavity is assumed to have the form

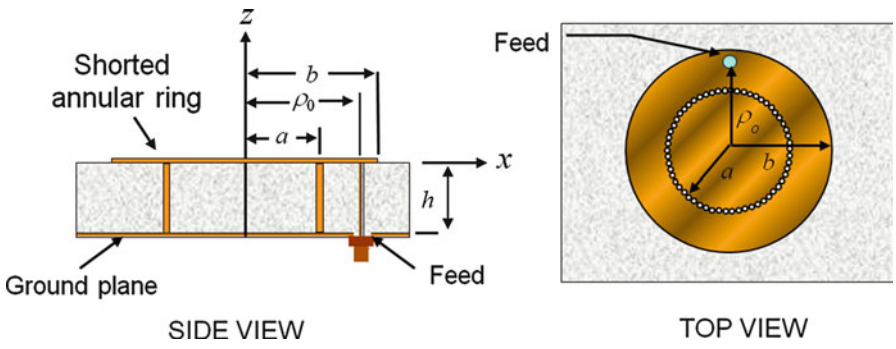


Fig. 3 Geometry of the shorted annular ring reduced surface wave antenna (SAR-RSW)

DOI 10.1007/978-981-4560-75-7_50-2

$$E_z(\rho,\phi) = \cos\phi(AJ_1(k_1\rho) + BY_1(k_1\rho)). \tag{14}$$

Enforcing the boundary conditions at the inner and outer radii, the electric field vanishes at the inner boundary and its normal derivative vanishes at the outer boundary. This yields a transcendental equation for the inner radius a as

$$\frac{J_1(k_1a)}{Y_1(k_1a)} = \frac{J_1'\left(x_{11}'\dfrac{k_1}{\beta_{\mathrm{TM0}}}\right)}{Y_1'\left(x_{11}'\dfrac{k_1}{\beta_{\mathrm{TM0}}}\right)}, \tag{15}$$

which is easily solved numerically. Equation 15 always has one solution for the desired radius a. If the permittivity of the substrate is sufficiently large, a second solution (with a smaller value of a) will also exist as well. This will occur when k_1 is larger than a certain value that corresponds to the solution $a = 0$, so that the argument of the J_1' term in Eq. 15 is at a root of the J_1' function, namely the root x_{12}'. This yields the condition for the second solution to exist as

$$k_1 > \beta_{\mathrm{TM0}}\left(\frac{x_{12}'}{x_{11}'}\right). \tag{16}$$

If the wavenumber of the TM_0 surface wave is approximated as k_0 (this approximation becoming more accurate for thin substrates), then this condition reduces to $\varepsilon_r > 8.38$.

The design procedure for the SAR-RSW antenna is first to calculate the wavenumber β_{TM0} of the TM_0 surface wave from the given substrate thickness, permittivity, and operating frequency chosen. Next, the outer radius b is determined from Eq. 12 and the inner radius a is determined from Eq. 15.

If desired, one can design the SAR structure to be a "reduced lateral wave" (RLW) antenna instead of a reduced surface wave antenna, reducing the lateral radiation that propagates outward horizontally from the antenna with the wavenumber of free space k_0. The design principle of the RLW antenna is very similar to that of the RSW antenna, with k_0 replacing β_{TM0}. To see this, first consider the space-wave field produced by an x-directed magnetic dipole along the z axis at a height z_0 above the ground plane, with no substrate present. The space-wave field has the form

$$\psi^{sp} = B(z,z_0)\frac{e^{-jk_0\rho}}{\rho}\sin\phi, \tag{17}$$

with some amplitude coefficient $B(z, z_0)$. Comparing Eq. 17 with Eq. 3, we see that the forms are similar, with β_{TM0} replaced by k_0. Repeating the same type of deviation above, we find that the ring of radius b in Fig. 2 radiates a space-wave field along the ground plane given by

$$\psi^{sp} = -B(z,z_0)(2\pi b)\frac{e^{-jk_0\rho}}{\rho}\cos\phi J_1'(k_0b). \tag{18}$$

Hence, the outer radius of the SAR-RLW antenna would be chosen as

$$k_0b = x_{11}' = 1.8412. \tag{19}$$

DOI 10.1007/978-981-4560-75-7_50-2

The antenna would then not produce any space-wave field along the ground plane, at least to leading order (a higher-order space-wave field would exist that decays faster than $1/\rho$).

If an infinite substrate were present, the dipole would produce a lateral radiation field where the $1/\rho$ term in Eq. 17 is replaced by a function $F(\rho)$, giving the radial dependence of the lateral radiation. For small and moderate values of ρ, this function would vary as $1/\rho$, and for extreme distances, it would vary as $1/\rho^2$. The lateral radiation field would then be given as

$$\psi^{lw} = -B(z, z_0)(2\pi b)e^{-jk_0\rho}F(\rho)\cos\phi J_1'(k_0 b). \tag{20}$$

In view of Eq. 19, the antenna would not produce any lateral radiation (to leading order) when a substrate is present. The same design, based on Eq. 19, would thus minimize space-wave radiation along the ground plane when no substrate is present and would minimize lateral radiation inside and near the substrate when an infinite substrate layer is present. The inner radius a for the RLW antenna would be chosen according to

$$\frac{J_1(k_1 a)}{Y_1(k_1 a)} = \frac{J_1'\left(x_{11}'\dfrac{k_1}{k_0}\right)}{Y_1'\left(x_{11}'\dfrac{k_1}{k_0}\right)}, \tag{21}$$

which comes from Eq. 15, replacing β_{TM0} with k_0.

Fortunately, for most substrates, the wavenumber β_{TM0} of the TM_0 surface wave is fairly close to that of the free-space wavenumber k_0. Hence, the SAR-RSW and the SAR-RLW designs would be very close to each other. In practice, the SAR-RSW design has both reduced surface wave excitation and a reduced amount of lateral radiation.

Cored Substrate Design

In this design, a cylindrical region of radius a is cored out of the substrate cavity and replaced by a material having a lower relative permittivity ε_{rc}, as shown in Fig. 4 (Jackson et al. 1993). The value of ε_{rc} is chosen to be less than that of the substrate permittivity ε_r, but is otherwise somewhat arbitrary. The core region could be chosen as an air region, in which case $\varepsilon_{rc} = 1$. The lower permittivity of the core region could be obtained by drilling holes in the substrate to create an artificial dielectric in the core region with a lower permittivity than the surrounding substrate (Gauthier et al. 1997; Papapolymerou et al. 1998). The cavity region below the circular patch thus consists of a core region of radius $\rho < a$ and a substrate region with ε_r in the region $a < \rho < b$. The outer radius b is chosen from the basic RSW design Eq. 12, as it is in the SAR design. The inner radius a of the core region is chosen to make the patch resonant at the operating frequency.

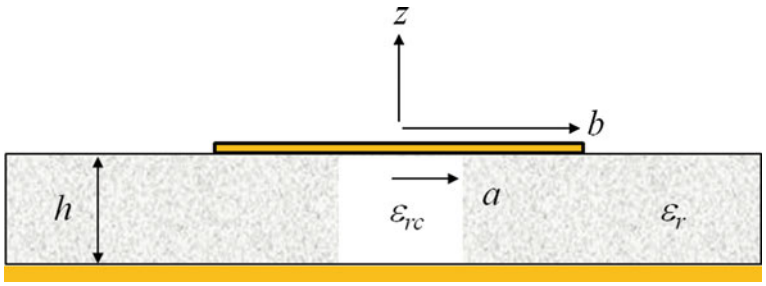


Fig. 4 Geometry of the cored-substrate reduced surface wave antenna

DOI 10.1007/978-981-4560-75-7_50-2

A cavity-model analysis can be used to predict the necessary inner radius a. In the core region, the field is assumed to be

$$E_{z1}(\rho,\phi) = A\cos\phi J_1(k_c\rho), \tag{22}$$

while the field in the substrate region below the patch is assumed to be

$$E_{z2}(\rho,\phi) = \cos\phi(BJ_1(k_1\rho) + CY_1(k_1\rho)). \tag{23}$$

The wavenumbers k_c and k_1 correspond to the core and substrate regions, respectively. Enforcement of the boundary conditions at $\rho = a$ and $\rho = b$ results in the following transcendental equation for determining the inner radius a:

$$\begin{aligned}\frac{\eta_1}{\eta_c}J_1'(k_ca)\left[J_1(k_1a)Y_1'(k_1b) - J_1'(k_1b)Y_1(k_1a)\right] \\ = J_1(k_ca)\left[J_1'(k_1a)Y_1'(k_1b) - J_1'(k_1b)Y_1'(k_1a)\right],\end{aligned} \tag{24}$$

where η_c and η_1 are the intrinsic impedances of the core region and substrate region, respectively. (Alternatively, if the radius a is specified, this becomes a transcendental equation for determining the permittivity ε_{rc} of the core region.)

Patch with Artificial Substrate

This design is similar to that of the cored design in that an artificial substrate is used inside the patch cavity. In this design, the entire substrate within the patch cavity is replaced by an artificial substrate having a relative permittivity of ε_{rp} under the patch, inside the cavity region. The artificial substrate now occupies the entire cavity region $\rho < b$ as shown in Fig. 5. The patch will be resonant when

$$k_pb = x_{11}', \tag{25}$$

with k_p the wavenumber in the patch cavity region. Since we wish for the patch to be an RSW antenna, Eq. 12 is also enforced. Both equations will be satisfied when

$$k_p = \beta_{\mathrm{TM0}}, \tag{26}$$

implying that the filling material of the cavity region must have the same wavenumber as that of the TM_0 surface wave. This means that

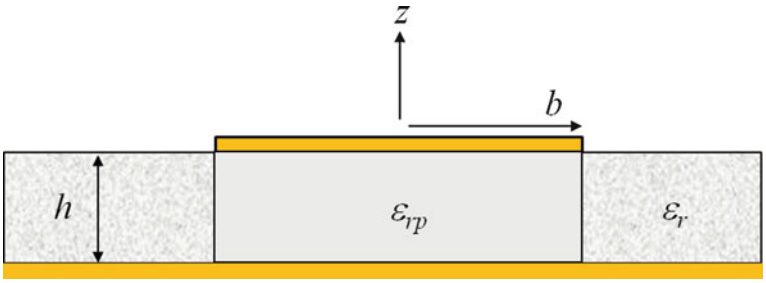


Fig. 5 Geometry of the circular patch with an artificial substrate inside

DOI 10.1007/978-981-4560-75-7_50-2

$$\varepsilon_{rp} = (\beta_{\mathrm{TM0}}/k_0)^2. \tag{27}$$

Since $\beta_{\mathrm{TM0}} < k_1$, it follows that $\varepsilon_{rp} < \varepsilon_r$. This lower permittivity can be realized by drilling holes in the substrate within the patch cavity to lower the permittivity (Gauthier et al. 1997; Papapolymerou et al. 1998), or by adding a dense array of vertical vias inside the patch cavity to create an artificial substrate having a lower effective permittivity (Nikolic et al. 2005).

Inverted Shorted Annular Ring (ISAR)

The previous RSW designs were all based on a circular structure that radiates from the outer edge, with the radius of the outer radiating edge chosen according to Eq. 12. A variation of the design uses a circular structure that radiates from the *inner edge*. The inverted shorted annular ring (ISAR) RSW antenna is based on this design variation (Basilio et al. 2008). The ISAR antenna is shown in Fig. 6 below. It consists of an annular ring structure that is shorted on the outer edge (using shorting vias), with the inner edge being the radiating edge whose radius is chosen according to Eq. 12, with b replaced with a. The outer radius is chosen to make the antenna resonant at the operating frequency. This yields a transcendental design equation for the outer radius b as

$$\frac{J_1(k_1 b)}{Y_1(k_1 b)} = \frac{J_1'(k_1 a)}{Y_1'(k_1 a)}. \tag{28}$$

The ISAR patch is larger than the SAR patch, so for most applications the SAR patch would probably be preferable. However, one advantage of the ISAR patch is that it has a large interior region $\rho < a$ that is empty, into which another RSW antenna can be placed, e.g., an SAR-RSW antenna. The interior SAR-RSW antenna would have to be resonant at a higher frequency than the ISAR antenna, so that the radius of its outer radiating edge is smaller than the inner radius of the ISAR patch. In this way, a dual-band RSW antenna can be realized that is planar. (Stacking RSW antennas of different size would be another way to obtain a dual-band structure.)

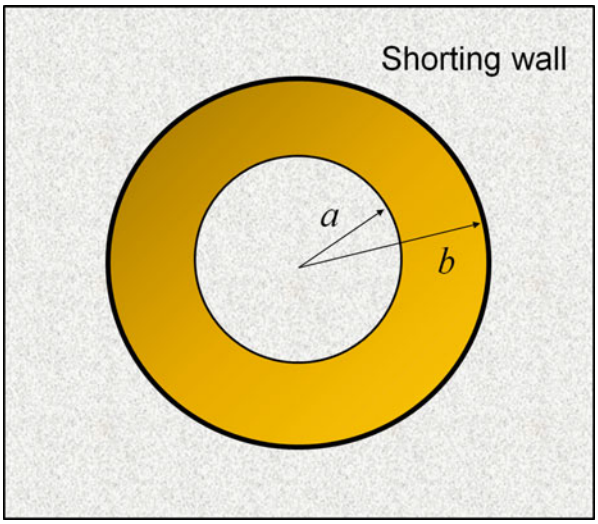


Fig. 6 Geometry of the inverted shorted annular ring (ISAR) patch

DOI 10.1007/978-981-4560-75-7_50-2

Annular Ring

The previous designs were all variations of an annular ring that had one of its edges shorted, either the inner or outer edge. Another design concept is to use an annular ring microstrip antenna with both edges open, and thus radiating from both edges (Hagerty and Popovic 2002). The geometry is the same as that of the ISAR shown in Fig. 6, without the short circuit on the outer boundary. Instead of having a single radiating edge located at a critical radius that does not excite the TM_0 surface wave, the two edges of the annular ring are positioned so that the surface wave field excited by the two edges cancels out. The TM_{12} mode of the annular ring was used in (Hagerty and Popovic 2002). This mode has a $\cos\phi$ variation and an electric field in opposing directions at the two edges, resulting in magnetic currents that are in the same direction (and hence a broadside pattern as for a conventional circular patch).

In view of Eq. 10, the excitation of the TM_0 surface wave field from the two edges can be found from the amplitudes of the magnetic currents at the edges, given by

$$K(\phi) = \pm E_z(\phi)h, \tag{29}$$

where E_z is the electric field at the corresponding edge, with the "+" sign being for the outer edge and the "−" sign being for the inner edge. The surface wave field is then given by

$$\psi^{sw} = -A(z,z_0)(2\pi h)\sqrt{\frac{2}{\pi}}e^{j3\pi/4}\frac{e^{-j\beta_{TM0}\rho}}{\sqrt{\beta_{TM0}\rho}}\cos\phi\left(-aE_z(a)J_1'(\beta_{TM0}a) + bE_z(b)J_1'(\beta_{TM0}b)\right). \tag{30}$$

The electric field inside the patch is given by Eq. 14. Enforcing a zero radial derivative at the two edges (assuming an approximate open circuit or PMC boundary), the field inside the patch is described by

$$E_z = \cos\phi\left(J_1(k_1\rho)Y_1'(k_1a) - Y_1(k_1\rho)J_1'(k_1a)\right), \tag{31}$$

where

$$J_1'(k_1b)Y_1'(k_1a) - Y_1'(k_1b)J_1'(k_1a) = 0. \tag{32}$$

At a given frequency, if the inner radius a is chosen, Eq. 32 gives the corresponding outer radius b for the ring to be resonant. In order for the ring to not excite the TM_0 surface wave, Eq. 30 implies that the inner radius must be chosen to satisfy

$$-aE_z(a)J_1'(\beta_{TM0}a) + bE_z(b)J_1'(\beta_{TM0}b) = 0. \tag{33}$$

Hence, substituting from Eq. 31, it follows that

$$\begin{aligned} -a\left(J_1(k_1a)Y_1'(k_1a) - Y_1(k_1a)J_1'(k_1a)\right)J_1'(\beta_{TM0}a) \\ + b\left(J_1(k_1b)Y_1'(k_1a) - Y_1(k_1b)J_1'(k_1a)\right)J_1'(\beta_{TM0}b) = 0. \end{aligned} \tag{34}$$

For a given value of the inner radius a, the outer radius b is found from Eq. 32, so that b is a function of a. Hence, Eq. 34 may be viewed as a transcendental equation for the unknown inner radius a.

Note that if the inner or outer radius of the ring is short-circuited, Eq. 33 reduces to the design equation of the SAR or ISAR design equation for the radius of the radiating edge.

Because the interior of the annular ring is open substrate space, components can be placed there. In (Hagerty and Popovic 2002) an X-band oscillator using a class-E amplifier was inserted inside the ring to

DOI 10.1007/978-981-4560-75-7_50-2

make a radiating element for spatial power combining. Similarly, components can be placed inside the ISAR antenna as well. (Components can also be placed inside the SAR antenna, though the inner radius is short-circuited so care must be taken if the components are to be connected to the antenna.)

Non-circular Designs

Other designs of RSW antennas, based on geometries that are not circular in shape, have also been explored in addition to the ones reviewed above that were based on circular geometries. In Hickey et al. (2001), a twin-arc slot design was introduced. This design uses two opposing slots that lie along arcs of a circle, with the slots lying in the ground plane of a grounded substrate. The diameter of the circle and the slot lengths are allowed to vary independently in order to minimize surface wave excitation while maintaining resonance. Although the surface wave excitation is not completely eliminated, it is greatly reduced. The slots were fed by coplanar waveguide and the substrate was electrical thick, about a quarter-wavelength in the dielectric of the substrate. The design greatly reduced the surface wave excitation, which would have otherwise been quite serious for such a thick substrate. The twin-arc slot design is a useful alternative to the patch antenna when it is desired to implement a slot antenna.

In Amendola et al. (2005), a modification of the SAR was introduced, in which the inner and outer edges of the annular ring have elliptical boundaries instead of circular boundaries. The antenna is called the shorted annular elliptical patch (SAEP). The use of elliptical boundaries allows for dual-band operation. As the ellipticity increases, the separation between the two bands increases. The surface wave excitation cannot be eliminated exactly at both frequency bands, so a compromise must be made between the band separation and the amount of surface wave reduction. As long as the band separation is not large, the surface wave excitation is minimal at both bands.

It is also possible to design a rectangular patch that has reduced surface wave excitation. If the two radiating edges of the rectangular patch are spaced apart by a distance L (the resonant length of the patch) then the surface wave excitation will be eliminated in the E-plane provided $\beta_{\mathrm{TM0}} L = \pi$, so that the surface wave excitation from the two edges cancels in the E-plane direction. The surface wave excitation will also be zero in the H plane, since the TM-surface wave field pattern from the magnetic currents on the radiating edges will have a null in this direction. The surface wave field excitation pattern thus has the shape of a four-leaf clover, being zero in the E and H planes and maximum approximately in the four diagonal-plane directions. Although the surface wave excitation is not eliminated, it is greatly reduced. The patch will not be resonant (unless the substrate permittivity satisfies Eq. 27) so the interior of the patch must be modified to make the patch resonant, and this can be done in a variety of ways. In Papapolymerou et al. (1995), a cored substrate inside the rectangular patch was used. In Boccia et al. (2005), shorting pins were used to create an inner rectangular-shaped shorted cavity region, similar to the circular-shorted cavity region ($\rho < a$) of the SAR design.

RSW Theorem

The previous RSW antennas were all circular in shape and were based on the basic design condition of Eq. 12. It is possible to design more general RSW antennas of arbitrary shape, provided the cavity region inside the patch antenna is chosen properly. An "RSW theorem" was introduced in (Komanduri et al. 2013) that allows for such designs. The RSW theorem applies to a planar microstrip patch antenna of arbitrary shape, whose cavity region below the patch has been replaced by an artificial material having a relative permittivity ε_{rp} that is given by Eq. 27. The general patch configuration is shown in Fig. 7.

DOI 10.1007/978-981-4560-75-7_50-2

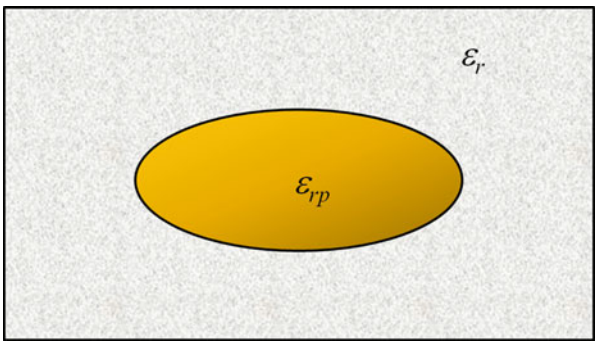


Fig. 7 Geometry of an arbitrary patch having a cavity that is filled with a material having a relative permittivity ε_{rp}

Note that this design generalizes the circular patch filled with an artificial material, shown in Fig. 5. The RSW theorem indicates that any patch shape can be used, and the patch will not excite the TM_0 surface wave when the corresponding cavity is filled with the relative permittivity specified by Eq. 27.

A proof of the RSW theorem is based on the fact that the radiation from the patch can be found from the magnetic current model, modeling the patch by a sheet of magnetic surface current at the boundary of the patch. For simplicity, a single ring of magnetic current varying as $K(\phi')$ inside the substrate, flowing around the boundary C of the patch, is considered. The surface wave field at some z coordinate (e.g., at $z = 0$) launched by such a ring has an amplitude given by

$$\psi^{sw} = A_1 \left(\frac{e^{-j\beta_{\mathrm{TM0}}\rho}}{\sqrt{\rho}} \right) I, \tag{35}$$

where

$$I = \int_C K(\ell')\, e^{j\beta_{\mathrm{TM0}}\left(\underline{\rho}'\cdot\underline{\hat{\rho}}\right)} \sin\left(\phi - \alpha(l')\right) dl', \tag{36}$$

with α being the local tilt angle of the magnetic current in the ring with respect to the x axis, and A_1 being a constant. We wish to prove that the integral I in Eq. 36 is zero when condition Eq. 27 is satisfied. To do this, we consider a scenario when the entire half-space space above the ground plane (both inside the patch cavity and outside of it) is filled with a material having an effective relative permittivity ε_{rp}. According to the cavity model, the patch in this scenario has (approximately) the same fields inside the cavity and the same resonance frequency as the actual patch structure does (which has the material ε_{rp} inside the patch cavity and a substrate with ε_r and thickness h outside the patch cavity). The space wave field ($1/\rho$ field along the ground plane) produced by the same magnetic current ring (now modeling the patch in this new half-space scenario) has the form

$$\psi^{sp} = A_2 \left(\frac{e^{-j\beta_{\mathrm{TM0}}\rho}}{\rho} \right) I, \tag{37}$$

where A_2 is another constant.

The space wave field from the patch in the new scenario can also be calculated using the electric current model, in which a planar surface current radiates at a height h above a ground plane, when the entire medium above the ground plane is a homogeneous one having $\varepsilon_r = \varepsilon_{rp}$. It can be shown by image theory that a horizontal electric current radiating above an infinite ground plane in a homogenous medium will

DOI 10.1007/978-981-4560-75-7_50-2

have a space wave field along the ground plane that falls off with distance faster than $1/\rho$. Therefore, the integral I appearing in Eq. 37 must be zero. Since this is the same I term appearing in Eq. 36, then it must be true that I in Eq. 36 is also zero, and hence that ψ^{sw} is zero. This completes the proof.

As a variation of the RSW theorem, an "RLW theorem" can also be stated. For a patch of arbitrary shape, if the patch cavity is filled with air, then the patch will operate as a reduced lateral wave antenna, with minimal lateral radiation along the substrate and ground plane. The proof of this parallels the preceding proof of the RSW theorem. The lateral-radiation field ψ^{lw} along the substrate or ground plane has the form

$$\psi^{lw} = B_1 e^{-jk_0\rho} F(\rho) I \tag{38}$$

where

$$I = \int_C K(\ell')\, e^{jk_0\left(\underline{\rho}'\cdot\underline{\hat{\rho}}\right)} \sin\left(\phi - \alpha(l')\right) dl'. \tag{39}$$

The term $F(\rho)$ transitions from a $1/\rho$ behavior to a $1/\rho^2$ behavior at extreme distances (if the substrate is infinite). In the proof, we now consider a new scenario where there is free space above the ground plane, both inside and outside the patch, instead of a half-space of material. The space wave radiation along the ground plane in this new scenario is predicted by

$$\psi^{sp} = B_2 \left(\frac{e^{-jk_0\rho}}{\rho}\right) I. \tag{40}$$

The space wave radiation in this scenario falls off with distance faster than $1/\rho$ because of the homogenous (air) region above the ground plane, so the integral term I in Eq. 40 must be zero. The same integral term in Eq. 38 is therefore zero. Hence, the original patch filled with an air cavity will have a lateral radiation field along the substrate or ground plane that is zero to leading order, meaning that the lateral radiation field is a weaker, higher-order field compared to that produced by a conventional microstrip antenna.

The fact that a patch of arbitrary shape filled with an air cavity will not produce lateral radiation along the ground plane was first observed in Nikolic et al. (2005), where the authors demonstrated reduced mutual coupling from the corresponding RLW antennas. For a thin substrate, an RLW antenna should also be an RSW antenna, since β_{TM0} is close to k_0. For thicker and higher-permittivity substrates where β_{TM0} is not very close to k_0, it is worthwhile to distinguish between the two designs. In this case, it may be desirable to design the antenna as an RSW antenna if the goal is to reduce mutual coupling for widely separated antennas, since the surface wave may be more dominant than the lateral radiation. This would depend on the separation between the antennas, however.

Results for RSW Antennas

In this section results are presented for different types of RSW antennas, both for the single element and for the mutual coupling between a pair of elements. Results are also shown for an array of RSW antennas.

DOI 10.1007/978-981-4560-75-7_50-2

Patterns of SAR-RSW Antenna

Figures 8 and 9 compare the E-plane patterns of two microstrip patch antennas that are designed for 2.0 GHz: a conventional circular patch antenna (Fig. 8) and the SAR-RSW antenna (Fig. 9) (Khayat et al. 2000). The patches are on a substrate of thickness 0.1524 cm with a relative permittivity of 2.94 and a loss tangent of 0.0012. The conductivity of the copper patch and ground plane was chosen as 3.0×10^7 S/m in the simulations. The conventional patch has a radius of 2.66 cm. The SAR-RSW patch has an inner radius of 2.42 cm and an outer radius of 4.39 cm. Both antennas are fed by a standard 50 Ω SMA connector on the *x* axis. For each patch, the feed was positioned to give a 50 Ω match, giving a radial feed location of 0.61 cm for the conventional patch and 2.95 cm for the SAR-RSW antenna. The patterns were measured on a 1-m diameter circular ground plane. The measured patterns are also compared with the theoretical patterns that assume an infinite ground plane.

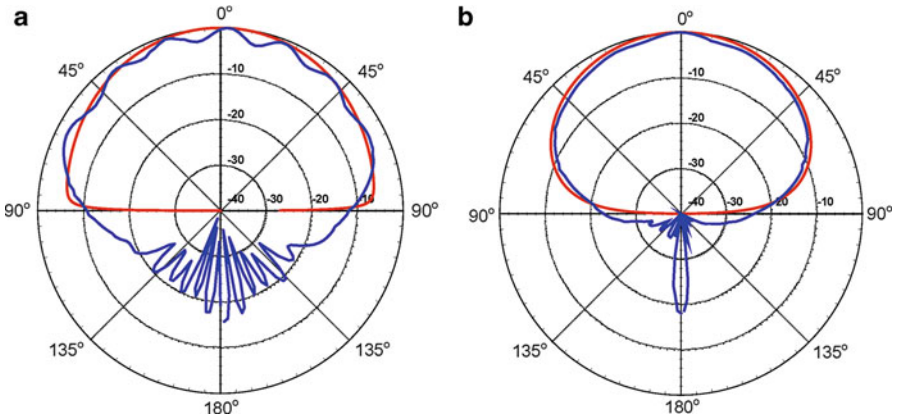


Fig. 8 (**a**) Comparison of the theoretical E-plane radiation pattern for a conventional circular patch on an infinite ground plane (*red curve*) and the measured E-plane pattern for a conventional circular patch on a 1-m diameter ground plane at 2.0 GHz (*blue line*). (**b**) Comparison of the theoretical H-plane radiation pattern for a conventional circular patch on an infinite ground plane (*red curve*) and the measured H-plane pattern for a conventional circular patch for a 1-m diameter ground plane at 2.0 GHz (*blue line*)

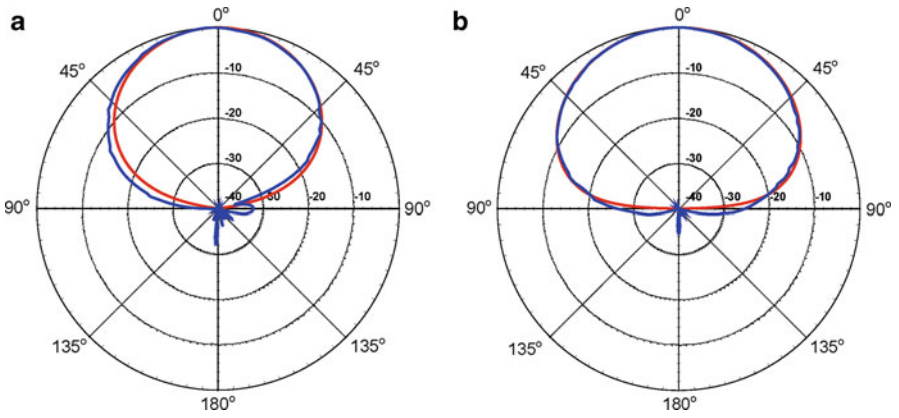


Fig. 9 (**a**) Comparison of the theoretical E-plane radiation pattern for a SAR-RSW patch on an infinite ground plane (*red curve*) and the measured E-plane pattern for a SAR-RSW patch on a 1-m diameter ground plane at 2.0 GHz (*blue line*). (**b**) Comparison of the theoretical H-plane radiation pattern for a SAR-RSW patch on an infinite ground plane (*red curve*) and the measured H-plane pattern for a SAR-RSW patch on a 1-m diameter ground plane at 2.0 GHz (*blue line*)

DOI 10.1007/978-981-4560-75-7_50-2

The E-plane pattern of the conventional patch in Fig. 8a shows significant back radiation and scalloping (rippling) of the pattern in the forward direction. Both of these effects are due to edge diffraction from the space wave radiated by the patch when it hits the edge of the ground plane. According to the geometrical theory of diffraction (GTD) (Mcnamara 1990; Balanis 2005), the E-plane pattern consists of the pattern of the antenna on an infinite ground plane together with the pattern due to two diffraction points at the E-plane edges of the ground plane (i.e., at $\phi = 0°$ and 180°, assuming a probe feed along the x axis). These two diffraction points act as sources that radiate, and this radiation produces the back radiation as well as the radiation that interferes with the main pattern of the antenna (the pattern of the antenna on an infinite ground plane) in the forward region.

It is seen from Fig. 9a that the E-plane pattern of the SAR-RSW antenna has much less back radiation as well as a greatly reduced scalloping of the forward pattern. Because of the greatly reduced edge diffraction, the measured pattern also agrees much better with the theoretical pattern than for the conventional patch. The E-plane beamwidth of the SAR-RSW antenna is smaller than that of the conventional patch. This is due to the larger size of the SAR-RSW patch, which introduces an array-factor effect. This array factor almost has a null at the horizon ($\theta = 90°$) since the outer patch radius approximately satisfies Eq. 19, since $\beta_{TM0} \approx k_0$. The SAR-RSW antenna therefore also has a larger directivity than a conventional circular patch, being about 10 dBi for the SAR-RSW antenna compared to about 7 dBi for the conventional circular patch. The radiation efficiency was found to be very similar to that of the conventional circular patch, around 75 %. (Although the SAR antenna produces only a very small amount of surface wave power, the overall efficiency is dominated by the conductor and dielectric loss, which is about the same as for the conventional patch.) The gain of the SAR-RSW antenna (about 9 dBi) was therefore about 3 dB higher than that of the conventional circular patch (about 6 dBi). The bandwidth was also found to be similar between the two antennas (about 1 %, based on an SWR < 2 definition).

Figures 8b and 9b show the H-plane patterns of the same two patches. According to GTD, the two diffraction points that determine the H-plane pattern are located at the edge of the ground plane along the H plane. Since the far-field radiation pattern component E_θ varies as $\cos\phi$, this component has a null in the H plane. Therefore, there is minimal diffraction in the H plane, even for the conventional patch. Hence, the H plane pattern of the conventional patch (Fig. 8b) shows little back radiation and forward scalloping. Two exceptional points are at $\theta = 0°$ and 180°; in these directions the E and H planes coincide. A fairly significant back radiation occurs at $\theta = 180°$, while a bump is observed at $\theta = 0°$. These effects are removed in the H-plane pattern of the SAR-RSW patch (Fig. 9b).

One interesting observation is that the E and H plane patterns of the SAR-RSW patch are nearly equal, with the E-plane pattern being slightly narrower. The SAR-RSW patch has nearly the same H plane pattern as does the conventional patch, while the E plane pattern is narrower, making the two planes of the SAR-RSW patch more nearly equal than for the conventional patch. This makes it possible to obtain a circularly polarized patch with a good axial ratio over a fairly wide angle range, as will be shown below.

Mutual Coupling Between SAR-RSW Antennas

Figure 10 shows the geometry for the mutual coupling between a pair of antennas. For E-plane coupling, $\Delta y = 0$, while for H-plane coupling $\Delta x = 0$.

Figure 11 shows the mutual coupling parameter S_{21} in the E plane between a pair of conventional circular patches and a pair of SAR-RSW antennas, with the antennas being the same as in Figs. 8 and 9. The coupling is shown versus the normalized center-to-center separation $\Delta x/\lambda_0$. Measured and calculated values are shown, and the measured results are seen to agree well with the calculated values (Khayat et al. 2000). The E-plane coupling between conventional patches falls off with separation approximately as $1/\rho$, while for SAR-RSW patches the E-plane coupling falls off as $1/\rho^3$, so it becomes very weak as the

DOI 10.1007/978-981-4560-75-7_50-2

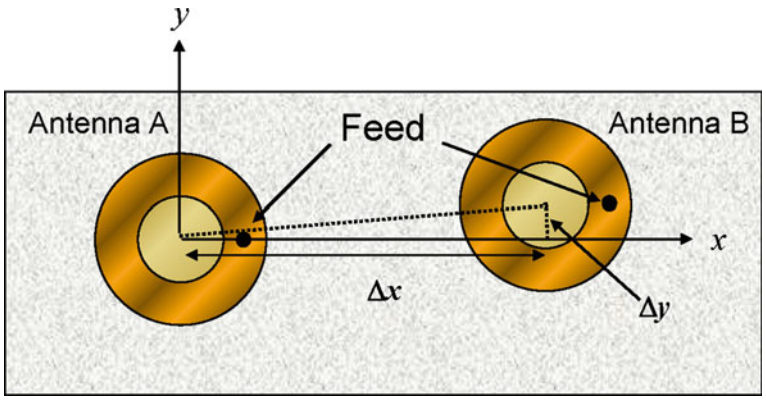


Fig. 10 Geometry setup for mutual coupling between either a pair of conventional or SAR-RSW antennas (SAR-RSW antennas are shown here)

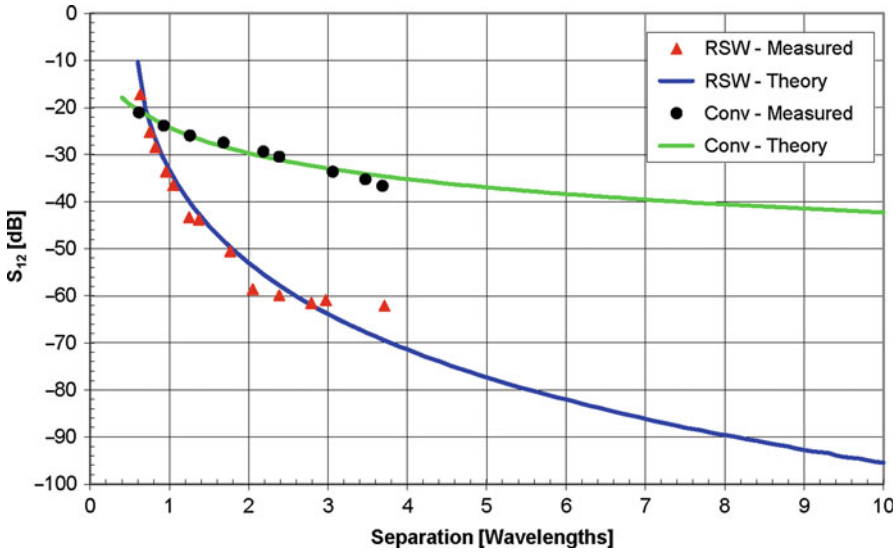


Fig. 11 Comparison of E-plane mutual coupling between a pair of conventional circular patches and a pair of SAR-RSW patches at 2.0 GHz. The patches are the same as in Figs. 8 and 9

separation increases (Khayat et al. 2000). The mutual coupling between the SAR-RSW antennas becomes less than that between conventional circular patches when the separation between the centers exceeds roughly 0.75 λ_0.

Figure 12 shows the calculated mutual coupling between conventional and SAR-RSW patches in both the E and H planes. Figure 12 shows that the mutual coupling between conventional patches falls off much slower in the E plane than the H plane as expected (Pozar 1982), since the E_θ component of the radiation field varies as $\cos\phi$ along the ground plane far away from each antenna. The E-plane coupling between conventional patches therefore falls off as $1/\rho$, while the H-plane coupling falls off as $1/\rho^2$. The H-plane coupling between the SAR-RSW antennas is only slightly lower than that between the conventional antennas, and also falls off as $1/\rho^2$. The E-plane coupling between the SAR-RSW antennas is the lowest of all the couplings for large separations, falling off as $1/\rho^3$.

Figure 13 shows the calculated E-plane mutual coupling between the pair of conventional circular patch antennas decomposed into its constituent components, including the surface wave component, the lateral-radiation component (called the “lateral wave”) and the total (the complex sum of the two)

DOI 10.1007/978-981-4560-75-7_50-2

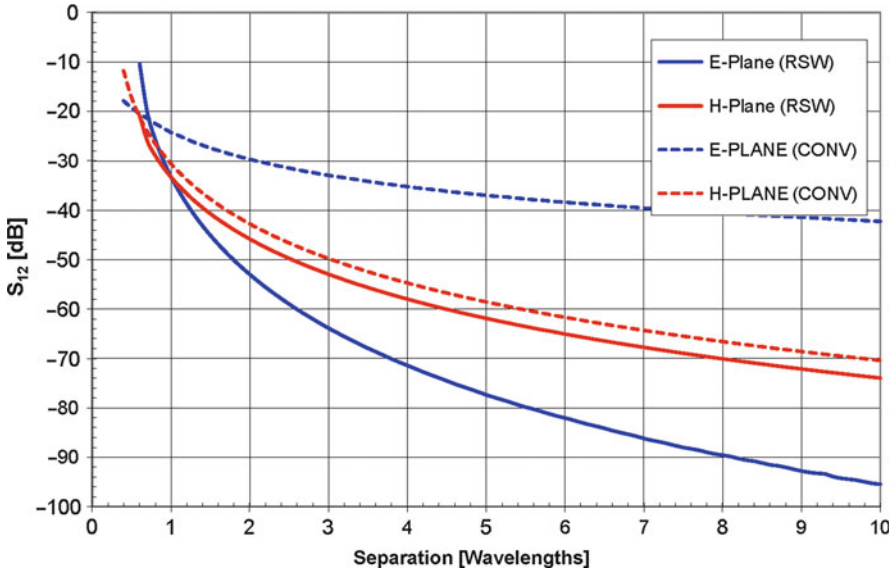


Fig. 12 Comparison of mutual coupling between a pair of conventional circular patches and a pair of SAR-RSW patches at 2.0 GHz, in different planes. The patches are the same as in Figs. 8 and 9

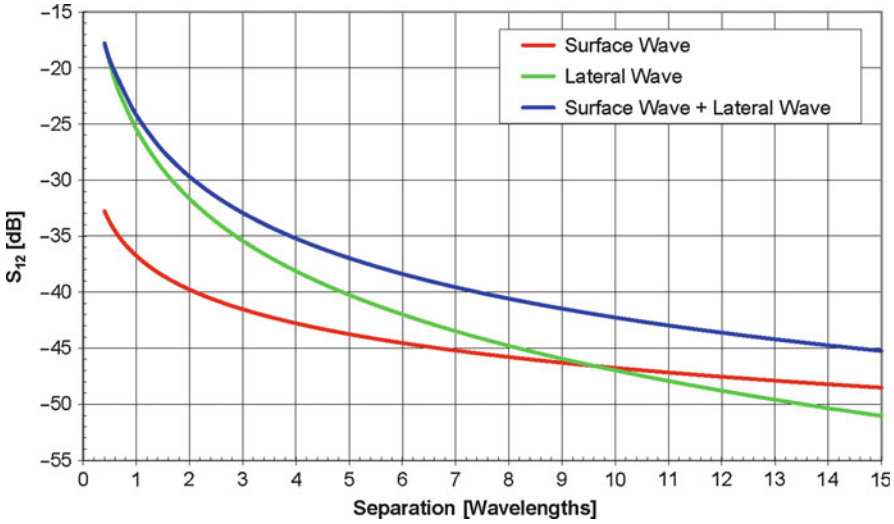


Fig. 13 Constituent components of the E-plane mutual coupling between a pair of conventional circular patches at 2.0 GHz. The patches are the same as the conventional patches in Figs. 8 and 9

(Khayat et al. 2000). It is seen that for this typical substrate ($\varepsilon_r = 2.94$, $h/\lambda_0 = 0.010$), the lateral-radiation field is the dominant coupling mechanism out to a separation of about 9.5 λ_0. Beyond this separation, the surface wave field becomes the dominant coupling mechanism. This confirms that for separations that are not too large, it is more important to reduce the lateral radiation than the surface wave.

Circularly Polarized SAR-RSW Antenna

The SAR-RSW antenna can be made into a circularly polarized (CP) antenna by feeding at two orthogonal feed ports spaced 90° apart, with a 90° phase difference between the two feed ports. A RHCP version of the SAR-RSW antenna was designed for the GPS L1 frequency of 1.575 GHz using a substrate of

DOI 10.1007/978-981-4560-75-7_50-2

thickness 0.1524 cm with a relative permittivity of 2.94 and a loss tangent of 0.0012. The SAR-RSW patch has an inner radius of 3.03 cm and an outer radius of 5.58 cm. The short-circuit inner boundary of the SAR-RSW was realized using vias that were spaced 10 mils (0.0254 cm) apart (from edge to edge) and were each 25 mils (0.0635 cm) in diameter. This is sufficient to ensure a good short circuit boundary at the inner radius. Both antennas are fed by a standard 50 Ω SMA connector at each feed port. For each antenna, the feed was positioned to give a 50 Ω match, corresponding to a radial feed location of 3.70 cm from the center of the patch. A standard 90° quadrature hybrid was used to feed the patch at the two orthogonal feed points.

Figure 14 shows the RHCP and LHCP far-field radiation components of the antenna, measured on a 1 m diameter circular ground plane. There is little back radiation, and the LHCP cross-pol level is lower than −30 dB. The low level of back radiation and the null in the pattern at the horizon reduce the susceptibility of the antenna to low-angle multipath and multipath arising from reflections from the ground.

Because of the low level of lateral radiation, the ground plane size can be reduced relative to that of a conventional microstrip antenna, while maintaining an acceptable pattern. Figure 15 shows the patterns for the same antenna in Fig. 14 when mounted on a smaller ground plane. In Fig. 15a, the antenna is mounted on a circular ground plane that is 35.5 cm (14 in.) in diameter. In Fig. 15b, the antenna is mounted on a square ground plane that is 12.7 cm (5 in.) on a side. In both cases, the pattern degradation due to the smaller ground plane is minimal, though in Fig. 15b the back radiation has now exceeded −20 dB.

Dual-Band SAR/ISAR-RSW Antenna

As mentioned in the Introduction, the SAR and ISAR designs can be combined to form a dual-band RSW antenna, with the SAR inside of the ISAR. The inner SAR antenna resonates at the lower frequency, while the outer ISAR antenna resonates at the upper frequency. The design procedures for the SAR and ISAR are given in sections "Shorted Annular Ring (SAR)" and "Inverted Shorted Annular Ring (ISAR)", respectively. As was discussed in the previous section, a quadrature hybrid can also be used together with two feed ports on each patch to obtain circular polarization. The design is shown in Fig. 16.

As an application, a dual-band RHCP antenna was designed for the L1 (1.575) and L2 (1.227 GHz) bands. The substrate had a thickness of 0.1524 cm with a relative permittivity of 2.94 and a loss tangent of

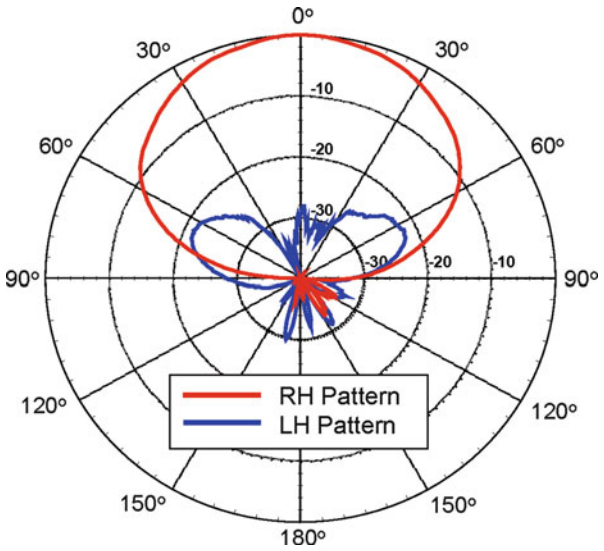


Fig. 14 RHCP (*red*) and LHCP (*blue*) patterns from the CP SAR-RSW antenna that was designed to radiate RHCP at 1.575 GHz. Measurements were taken on a 1 m diameter ground plane

DOI 10.1007/978-981-4560-75-7_50-2

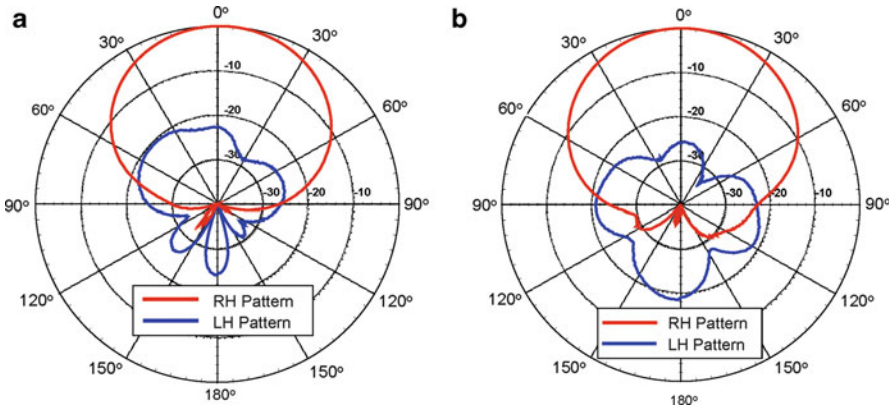


Fig. 15 RHCP (*red*) and LHCP (*blue*) patterns from the CP SAR-RSW antenna that was designed to radiate RHCP at 1.575 GHz. Measurements were taken on smaller ground planes. (**a**) A circular ground plane that is 35.5 cm (14 in.) in diameter. (**b**) A square ground plane that is 12.7 cm (5 in.) on a side

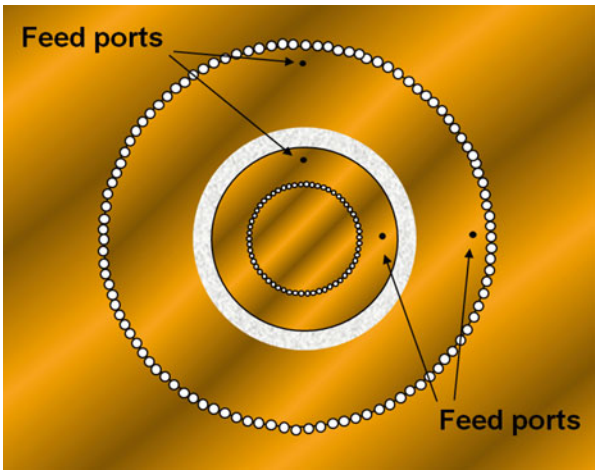


Fig. 16 Top view of the dual-band CP SAR/ISAR RSW antenna, showing the feed ports

0.0012. The SAR antenna had the same dimensions as in the previous section, while the ISAR had an inner radius of 5.577 cm and an outer radius of 8.727 cm. The inner radius of the SAR and the outer radius of the ISAR were short-circuited using vias that were spaced 10 mils (0.0254 cm) apart (from edge to edge) and were each 25 mils (0.0635 cm) in diameter.

Figure 17 shows the measured patterns at the L1 and L2 frequencies. As seen, the patterns are good at both frequencies, almost as good as those for the single CP SAR antenna of the previous section.

Antennas Based on the RSW Theorem

The RSW theorem of section "RSW Theorem" indicates that a general method for constructing an RSW antenna is to fill the cavity region of the patch with an artificial material having an effective permittivity that is chosen according to Eq. 27. The patch itself may have an arbitrary shape. To test this, both a rectangular patch and a circular patch were designed using a three-layer structure as shown in Fig. 18a (Komanduri et al. 2013). Note that the upper layer is an electrically thick ($h_3 = 0.062\ \lambda_0$) high-permittivity layer ($\varepsilon_{r3} = 9.8$), which would tend to create a significant surface wave excitation when excited by a microstrip source. However, a bottom layer is introduced and a second air spacer layer is placed on top of the bottom layer, below the patch. The thickness of the air spacer layer is chosen so that the effective

DOI 10.1007/978-981-4560-75-7_50-2

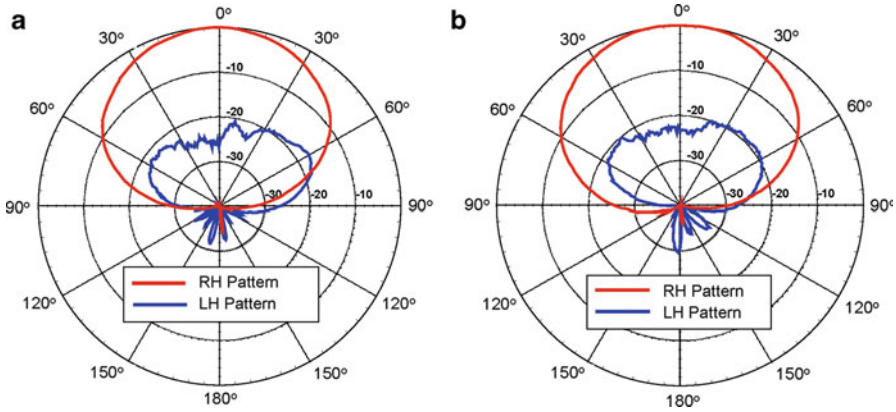


Fig. 17 RHCP and LHCP patterns of the dual-band CP SAR/ISAR RSW antenna. (**a**) Patterns at 1.575 GHz. (**b**) Patterns at 1.227 GHz. Patterns are measured on a 1 m diameter circular ground plane

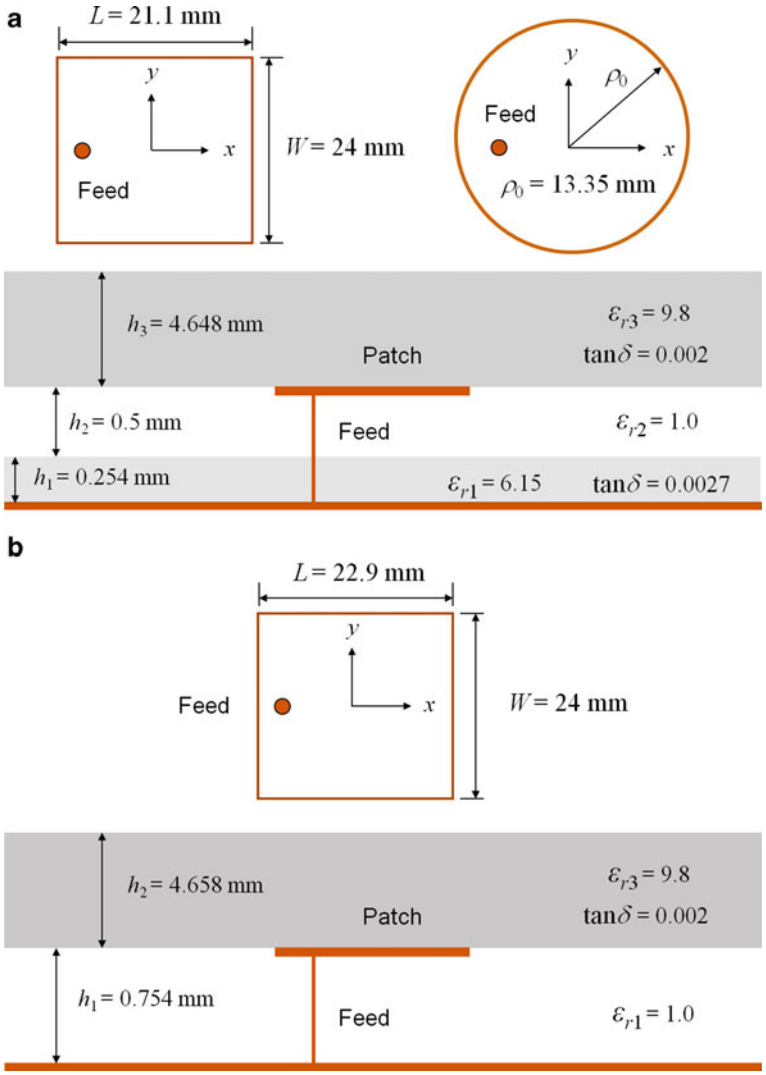


Fig. 18 Geometry of multilayer antenna for testing the RSW theorem. (**a**) Three-layer structure designed to be an RSW antenna. Both rectangular and circular patches are shown. (**b**) Two-layer structure that has the same *upper* superstrate layer as in part (**a**), but is not designed to be an RSW antenna. The patches are rectangular

DOI 10.1007/978-981-4560-75-7_50-2

permittivity inside the patch cavity (based on a simple series capacitor type of calculation) satisfies Eq. 27. The wavenumber of the TM_0 surface wave on the three-layer structure is $\beta_{TM0} = 1.180\, k_0$, found from a numerical solution of the appropriate transcendental equation for the wavenumber on the three-layer structure (details omitted here). This corresponds to an effective relative permittivity inside the patch cavity (the region below the patch) of $\varepsilon_{rp} = 1.393$. For comparison purposes, Fig. 18b shows a patch design that has the same upper superstrate and maintains the same composite thickness below the patch. However, the lower substrate has been replaced with air. This patch design is referred to as the "non-RSW" design. In both cases the patches are resonant at 4 GHz and are fed at the 50 Ω match points.

Figure 19a shows the E-plane coupling between the two RSW patches (rectangular and circular) and the non-RSW rectangular patches. It is seen that the E-plane coupling for both of the RSW antennas is much less than that of the non-RSW rectangular patches for larger separations, verifying that the RSW patches are operating with reduced surface wave excitation, and that the RSW property is independent of the patch shape. Figure 19b shows that the H-plane coupling is about the same for the RSW and non-RSW antennas, which is consistent with the trends observed with the SAR-RSW antenna, as seen in Fig. 12.

If a patch antenna is printed on a single substrate layer, as opposed to using multiple layers, then an artificial substrate must be used in order to fill the patch cavity with a material satisfying Eq. 27, where the filling material has a lower permittivity than the surrounding substrate. One way to lower the effective permittivity of a substrate region is to drill holes in it (Gauthier et al. 1997; Papapolymerou et al. 1998). Another way is to add shorting vias inside the patch cavity. Shorting vias behave as inductors at

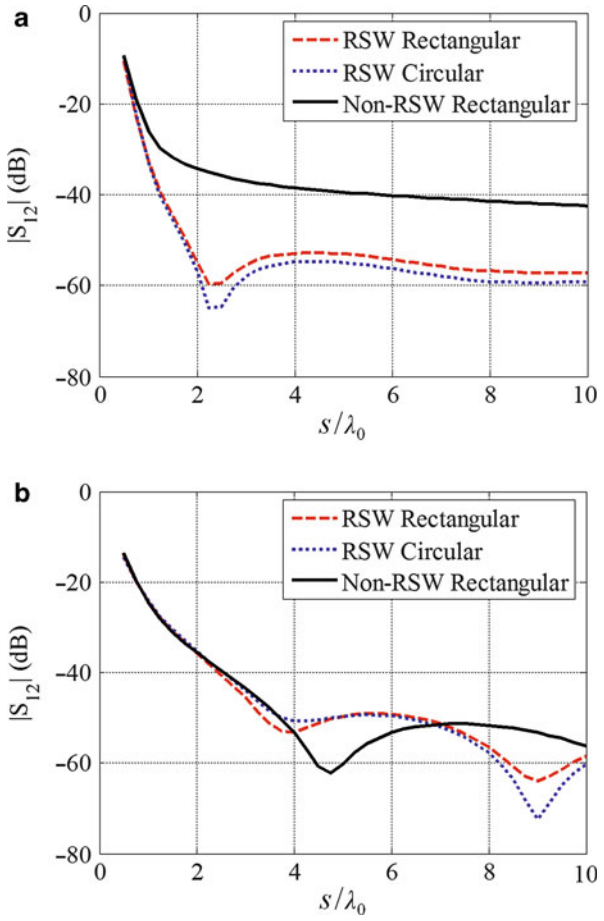


Fig. 19 Mutual coupling versus the normalized separation s/λ_0 between the patch centers for the antennas shown in Fig. 18. (**a**) E-plane coupling. (**b**) H-plane coupling

DOI 10.1007/978-981-4560-75-7_50-2

microwave frequencies, so that when they are added inside of a parallel capacitor region they lower the effective capacitance of the capacitor. (If the vias become very close together, they can also be used to create a short-circuit wall, as was done for the SAR and ISAR designs.) From another point of view, the inductive current on the vias partially cancels the capacitive polarization currents set up by the electric field inside the dielectric material (Nikolic et al. 2005). If the vias are chosen so that the artificial substrate below the patch has a permittivity satisfying Eq. 27, the patch will be an RSW antenna. If the artificial substrate has a free-space permittivity, the patch will be an RLW antenna.

Figure 20 shows a rectangular RSW antenna that has an array of vias (Komanduri et al. 2013). The patch is on an electrically thick substrate with $\varepsilon_r = 9.2$, and $\tan\delta = 0.002$, and a thickness $h = 3.25$ mm. The substrate has an electrical thickness (in terms of wavelength in the dielectric) of $h/\lambda_d = 0.073$ at a resonance frequency of 2.14 GHz. The dimensions are $r_1 = 0.3$ mm, $s_1 = 3$ mm, $s_2 = 6$ mm, $s_3 = 5.5$ mm, $d_1 = 22$ mm, $d_2 = 27$ mm, and $d3 = 32$ mm. The radius of the shorting vias was $r_1 = 0.3$ mm. The radius of the two SMA feed probes was 0.635 mm. Because of the electrically thick substrate, the SMA feed probe radiation is significant, so a balanced two-probe feed system was used to minimize probe radiation, consisting of two SMA feeds 180° out of phase, symmetrically located 5.5 mm from the center of the patch.

In the design process, the wavenumber β_{TM0} was first calculated, and from this the effective relative permittivity of the patch cavity was calculated using Eq. 27. Based on this permittivity, the length of the patch was calculated to make the patch resonant at the design frequency. Vias were then added into the patch until the simulated resonance frequency was the desired design frequency. The vias were kept away from the feed probes to avoid unwanted interaction with the probes. The length of the patch (resonant dimension) was 63 mm and the width was 45 mm.

For comparison, conventional rectangular patches on the same substrate were simulated as well, having a length (resonant dimension) of 22.4 mm and a width of 13.75 mm. The conventional antenna was also designed to resonate at 2.14 GHz and was also fed with a balanced SMA probe feed.

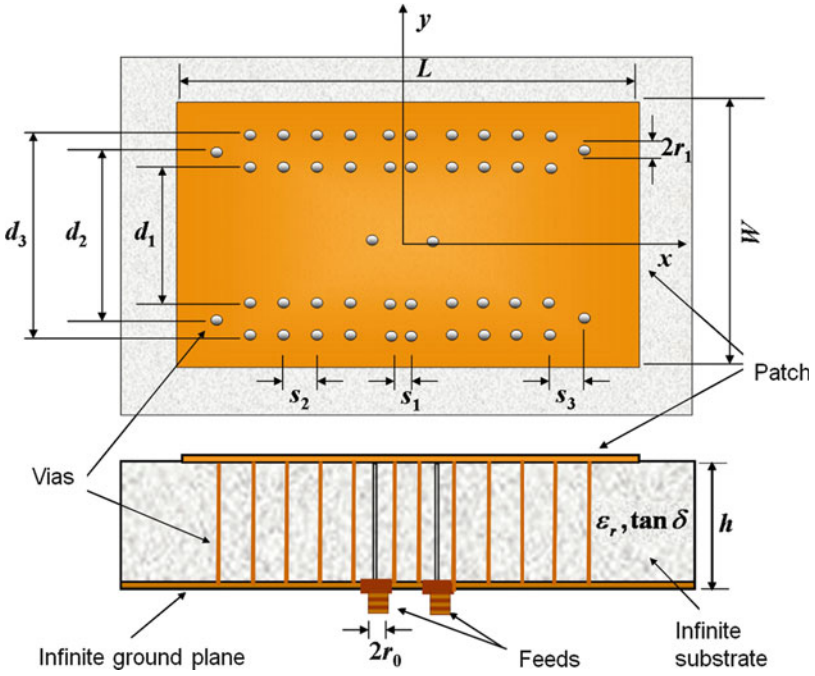


Fig. 20 Rectangular patch loaded with vias. The patch is on an electrically thick substrate with vias added. A balanced two-probe feed system (using two probes that are 180° out of phase) is used to improve symmetry

DOI 10.1007/978-981-4560-75-7_50-2

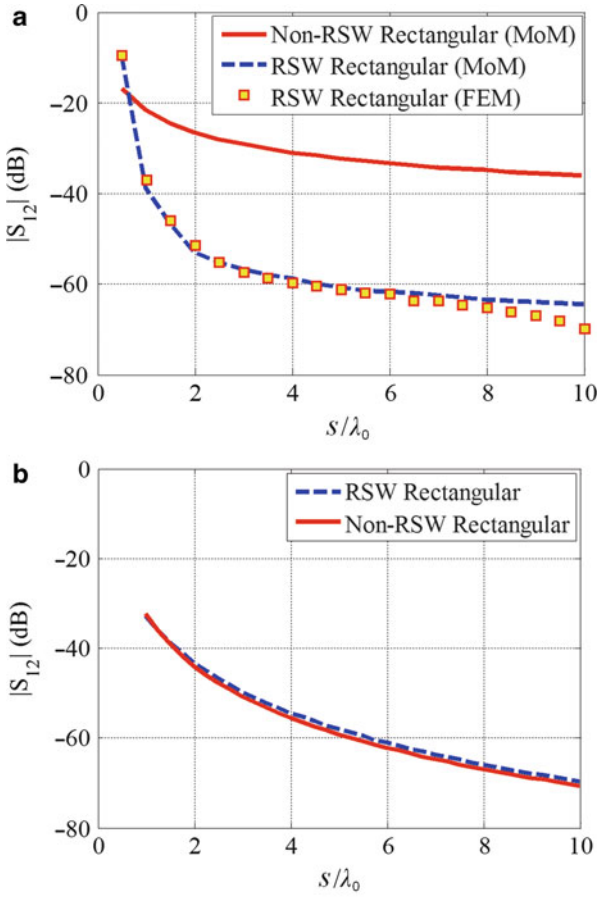


Fig. 21 Mutual coupling between two via-loaded rectangular patches and between two conventional rectangular patches versus the separation between centers. (**a**) E-plane. (**b**) H-plane

Figure 21 shows the simulated mutual coupling between a pair of the via-loaded RSW antennas and a pair of the conventional rectangular patch antennas, versus the separation between the patch centers. The simulation is done using a method of moments (MoM) simulation, using Ansys Designer. For the E-plane coupling (Fig. 21a), a finite element method (FEM) simulation was also used (Ansys HFSS) for validation. The E-plane coupling is much lower with the RSW antennas than with the conventional antennas as the separation increases. The H-plane coupling was approximately the same for both antennas, consistent with the mutual coupling results seen for the SAR-RSW antenna (Fig. 12).

Array of RSW Antennas

The mutual coupling between RSW antennas becomes small when the separation between the antennas increases, as seen in previous results. This is because of the greatly reduced surface wave excitation as well as the reduced lateral radiation along the ground plane. Because of this, there may be an advantage to using an array of RSW antenna to reduce scan blindness, which occurs due to the coherent buildup of the surface wave field from microstrip antennas in an array, adding constructively from one element to the next (Pozar and Schaubert 1984). The scan blindness is most serious in the E plane of an array, since the surface wave is excited the strongest in this direction. However, one limitation of using RSW antennas in an array is the larger size relative to a conventional microstrip antenna. The diameter of the SAR-RSW antenna is about 0.59 λ_0, which is the same size as a conventional circular patch on an air substrate. The interelement spacing between SAR-RSW elements in an array must therefore be larger than 0.59 λ_0, and in

DOI 10.1007/978-981-4560-75-7_50-2

practice larger than about 0.75 λ_0 in order to avoid having too strong of a coupling between nearest neighbors. This large separation will set a limit to the scan range of an array before grating lobes occur (Balanis 2005). For a 0.75 λ_0 separation, the scan range in the E plane will be limited to about 20°, assuming a square lattice element layout. Results are presented below to compare an array of SAR-RSW antennas with an array of conventional circular patches (Deshpande and Bailey 1989).

Figure 22 shows the broadside scan impedance of the center element of a two-dimensional array of antenna elements. The center element is assumed to be matched to the feed line (zero reflection coefficient) when the center element is isolated, with no other elements present. All antennas are fed equally and in phase. The total number of elements in the array is varied (an odd number since the elements are symmetric about the center) with the separation between the centers of the elements being 0.75 λ_0. Results are shown for SAR-RSW elements and conventional circular microstrip antennas. The substrate has a relative permittivity of $\varepsilon_r = 2.94$ and a loss tangent $\tan\delta = 0.0012$, and a thickness of $h = 0.1524$ cm.

For the SAR-RSW antennas, the inner radius is 2.39 cm and the outer radius is 4.38 cm. Each antenna is fed at a radius $\rho_0 = 2.91$ cm, which gives a 50 Ω match when the antenna is isolated. The SAR-RSW antennas resonate at 1.99 GHz. The conventional circular patches have a radius of 2.64 cm and are fed at a radius of $\rho_0 = 0.63$ cm, which gives a 50 Ω match when the antenna is isolated. The conventional circular patch antennas resonate at 1.88 GHz. All antennas are constructed from copper, with a conductivity of $\sigma = 3.0 \times 10^7$ S/m assumed in the simulations. In both cases, the interelement spacing is 0.75 λ_0 at the frequency of operation.

From Fig. 22 it is seen that for the array of SAR-RSW antennas the scan impedance stabilizes quickly, after the array size is only about three elements on a side (nine total elements). This is because the mutual coupling falls off fast enough that it has a negligible effect for elements that are more than one unit cell away. The converged value for the infinite array is about $40 + j10$ Ω, which is not greatly different from the

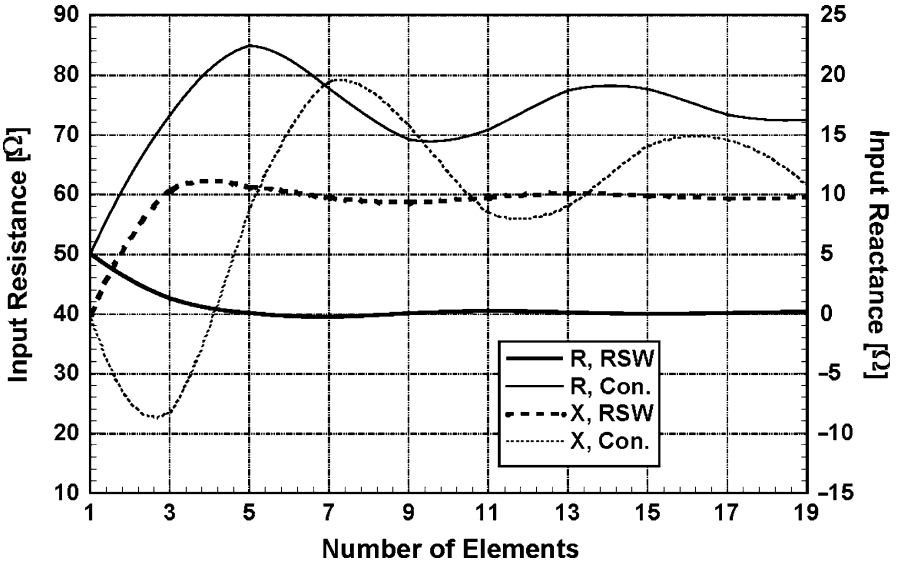


Fig. 22 Broadside scan impedance versus the number of elements N on each side of a two-dimensional $N \times N$ array with an element separation of 0.75 λ_0. The real and imaginary parts of the scan impedance are shown for an array of conventional circular patches and an array of SAR-RSW antennas

DOI 10.1007/978-981-4560-75-7_50-2

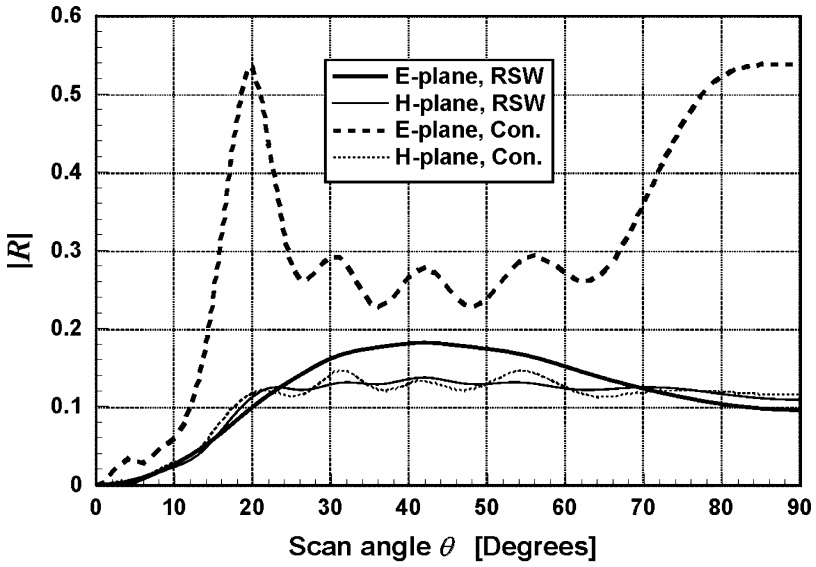


Fig. 23 Magnitude of the reflection coefficient versus the scan angle for a two-dimensional 19 × 19 array with an element separation of 0.75 λ_0. The reflection coefficient is shown for an array of conventional circular patches and an array of SAR-RSW antennas

isolated value of $50 + j0\ \Omega$ for the single patch. The array of conventional circular patches shows a very large variation in the scan impedance as the number of elements increases, and even for an array of 19 × 19 elements, convergence has not yet been obtained. The converged value appears to be around $75 + j12\ \Omega$, which is significantly different from the isolated input impedance of the single patch, namely $50 + j0\ \Omega$.

Figure 23 shows the magnitude of the reflection coefficient for the same array with 19 × 19 elements, as the scan angle is varied by changing the interelement phase shift in either the E plane or the H plane. Because of the large separation between elements, a scan blindness occurs in the E plane for the array of conventional patches, at about 20°. At this scan angle, the magnitude of the reflection coefficient rises to slightly more than 0.5, due to the TM_0 surface wave fields of the patches adding coherently in phase with the $n = -1$ space harmonic (Floquet wave) of the guided wave on the periodic structure. As the size of the patch array increases, the magnitude of the reflection coefficient would rise even higher (Pozar 1986). For the array of SAR-RSW antennas, no visible scan blindness occurs, due to the very small amount of surface wave field being excited by each antenna element. The H-plane for the array of conventional circular patches does not show a scan blindness, since the TM_0 surface wave is not excited in the H plane direction for a ring of magnetic current varying as $\cos\phi$, far away from the current. This can be seen from Eq. 10. The H-plane scan impedances of the SAR-RSW and conventional antennas are almost identical, consistent with the fact that the H-plane mutual coupling between conventional and RSW antennas due to lateral radiation both fall off with distance y in the H plane as $1/y^2$ (Khayat et al. 2000).

Of course, the scan blindness problem with the array of conventional patches can be avoided by simply choosing a smaller element spacing, such as 0.5 λ_0. One cannot use this smaller spacing for the array of SAR-RSW antennas, due to their larger size. However, if one wishes to have a larger array spacing such as 0.75 λ_0, e.g., to keep the number of elements in the array to a minimum, the use of SAR-RSW elements offers an opportunity to avoid the scan blindness problem. In this case, the array scan angle is not limited by the scan blindness effect, but by the onset of grating lobes. For an array spacing of 0.75 λ_0, grating

DOI 10.1007/978-981-4560-75-7_50-2

lobes will appear when the array is scanned in either the E plane or the H plane beyond about 19.5°. Therefore, the use of SAR-RSW antennas in a phased array is limited to applications requiring only a small scan range. The use of a hexagonal lattice will allow the scan range to increase to approximately 30° before the onset of grating lobes, for this same element spacing.

Conclusion

A reduced surface wave (RSW) microstrip antenna is a type of microstrip antenna that excites much less surface wave field than does a conventional microstrip antenna on the same substrate. The antenna is designed so that the TM_0 surface wave field is not excited by the magnetic current that models the dominant patch cavity mode, and hence the amount of surface wave excitation is very small and is mainly due to higher-order effects such as the higher-order modes of the patch. The same concepts used to design an RSW antenna can also be used to design a reduced lateral wave (RLW) antenna, in which case the lateral radiation (horizontally propagating radiation along the substrate and ground plane) is reduced instead of the TM_0 surface wave. In most practical cases where the substrate is relatively thin compared to a wavelength, the wavenumber β_{TM0} of the TM_0 surface wave is fairly close to the wavenumber k_0 of free space. This implies that an antenna designed to reduce the TM_0 surface will also greatly reduce the lateral radiation.

One class of RSW antennas is based on the concept that a ring of magnetic current inside of a grounded substrate that varies as $\cos\phi$ will not excite the TM_0 surface wave if the radius b of the ring is chosen correctly, according to $\beta_{TM0}b = x_{11}{}' = 1.8411$, where $x_{11}{}'$ is the first root of the $J_1{}'$ Bessel function. Various designs can be implemented based on this concept, which are all variations of a circular patch and are circularly symmetric with a dominant mode field that varies as $\cos\phi$. The designs all have a circular radiating edge that has the appropriate critical radius b.

One design is the shorted annular-ring RSW (SAR-RSW) antenna. This design consists of an annular-ring microstrip antenna that is shorted-circuited at the inner radius $\rho = a$, using shorting vias. The outer radius $\rho = b$ is selected to reduce the surface wave excitation, while the inner radius is chosen to make the antenna resonant at the operating frequency. Another design variation starts with a circular patch having the critical radius b and cores out a circular region inside the patch cavity of radius a and replaces it with material having a lower permittivity, whose value is chosen according to the radius a. Another design variation starts with a circular patch having the critical radius b and fills the patch cavity with an effective permittivity selected so that $\varepsilon_r^{eff} = (\beta_{TM0}/k_0)^2$. Another design variation, called the inverted shorted annular ring RSW (ISAR-RSW) antenna, uses a shorted annular ring that is short circuited at the outer radius $\rho = b$ with radiation coming from the inner edge at $\rho = a$. The inner radius now satisfies $\beta_{TM0}a = x_{11}{}' = 1.8411$ and the outer edge at $\rho = b$ is selected to make the antenna resonant. A dual-band RSW antenna can be constructed by placing an SAR-RSW antenna inside of an ISAR-RSW antenna, with the SAR-RSW antenna operating at the lower frequency.

Another class of RSW antennas is based on an "RSW theorem" that states that a resonant planar microstrip antenna of arbitrary shape will not excite the TM_0 surface wave if the patch cavity is filled with a material having an effective permittivity chosen according to $\varepsilon_r^{eff} = (\beta_{TM0}/k_0)^2$. (Alternatively, the antenna will minimize lateral radiation if the filling material is selected to be air.) The circular patch filled with such an effective permittivity is thus a special case of this more general class of RSW antennas.

The advantages of RSW antennas include improved radiation patterns due to less edge diffraction from the supporting ground plane and less mutual coupling between antenna elements (or between the antenna and other circuit components) that are widely separated. The radiation patterns of the RSW antenna are also slightly more directive than for conventional microstrip antennas (due to the larger size). The E and

DOI 10.1007/978-981-4560-75-7_50-2

H plane patterns are more similar for the RSW antenna than for conventional microstrip antennas (which have a broader E plane pattern). This allows for a good axial ratio to be obtained over a fairly large range of angles when using a circularly polarized version of the RSW antenna (using two feeds 90° out of phase).

The main disadvantage of the RSW antennas is that they are larger than conventional microstrip antennas. The SAR-RSW antenna has a diameter of about 0.59 λ_0 and the size cannot be made smaller by increasing the permittivity of the substrate, as it can for a conventional microstrip antenna.

Results show that the SAR-RSW antenna has an E-plane radiation pattern that has much less scalloping on the front side and much less total radiation on the back side, due to less diffraction from the edges of the supporting ground plane. The E-plane mutual coupling between SAR-RSW antennas is also much less than for conventional microstrip antennas when the antennas are widely separated. This is because of the reduced surface wave excitation, and also because of the reduced lateral radiation, since the lateral-radiation field falls off with separation distance x in the E plane as $1/x^3$ for the SAR-RSW antenna, while only falling off as $1/x$ for the conventional microstrip antenna. The H plane lateral-radiation field falls off as $1/y^2$ for both the SAR-RSW and the conventional microstrip antennas. The mutual coupling between SAR-RSW antennas starts to become less than that for conventional microstrip antennas when the element spacing becomes larger than about 0.75 λ_0.

Because the RSW antenna excites very little surface wave field, an array of such antennas will not exhibit scan blindness, which is an effect that arises due to surface wave excitation. However, due to the larger size of the RSW antennas, the interelement spacing in the array also needs to be larger than what is allowable for conventional microstrip antennas. This limits the scan range before grating lobes occur. For example, with an interelement spacing of 0.75 λ_0, the scan range is limited to about 20° from broadside when using a square lattice of elements. For such limited scan applications, RSW antennas may be beneficial.

Cross-References

- ▶ Circularly Polarized Antennas
- ▶ Conformal Array Antennas
- ▶ Low-Profile Antennas
- ▶ Microstrip Patch Antennas
- ▶ Phased Arrays

References

Abramowitz M, Stegun IE (1970) Handbook of mathematical functions. Dover, New York

Amendola G, Boccia L, Di Massa G (2005) Shorted elliptical patch antennas with reduced surface waves on two frequency bands. IEEE Trans Antennas Propag 53:1946–1956

Balanis CA (2005) Antenna theory: analysis and design. Wiley-Interscience, Hoboken

Basilio LI, Williams JT, Jackson DR, Khayat MA (2005) A comparative study of a new GPS reduced-surface-wave antenna. IEEE Antennas Wirel Propag Lett 4:233–236

Basilio LI, Chen RL, Williams JT, Jackson DR (2007) A new planar dual-band GPS antenna designed for reduced susceptibility to low-angle multipath. IEEE Trans Antennas Propag 55:2358–2366

Basilio LI, Williams JT, Jackson DR, Chen RL (2008) Characteristics of an inverted shorted annular-ring-reduced surface-wave antenna. IEEE Antennas Wirel Propag Lett 7:123–126

DOI 10.1007/978-981-4560-75-7_50-2

Boccia L, Amendola G, Di Massa G (2005) Rectangular patch antennas with reduced surface wave excitation. In: IEEE international symposium antennas and propagation, Washington, DC

Chen RL, Jackson DR, Williams JT, Long SA (2005) Scan impedance of RSW microstrip antennas in a finite array. IEEE Trans Antennas Propag 53:1098–1104

Deshpande MD, Bailey MC (1989) Analysis of finite phased arrays of circular microstrip patches. IEEE Trans Antennas Propag 37:1355–1360

Gauthier GP, Courtay A, Rebeiz GH (1997) Microstrip antennas on synthesized low dielectric-constant substrate. IEEE Trans Antennas Propag 45:1310–1314

Hagerty JA, Popovic Z (2002) A 10 GHz integrated class-E oscillating annular ring element for high-efficiency transmitting arrays. In: IEEE international microwave symposium, Seattle

Harrington RF (2001) Time harmonic electromagnetic fields. Wiley-IEEE Press, Hoboken

Hickey MA, Qiu M, Eleftheriades GV (2001) A reduced surface-wave twin arc-slot antenna for millimeter-wave applications. IEEE Microwave and Wireless Components Letters 11:459–461

Jackson DR (2007) Microstrip antennas, Chapter 7. In: Volakis JL (ed) Antenna engineering handbook. McGraw Hill, New York

Jackson DR, Williams JT, Bhattacharyya AK, Smith RL, Buchheit SJ, Long SA (1993) Microstrip patch designs that do not excite surface waves. IEEE Trans Antennas Propag 41:1026–1037

Khayat MA, Williams JT, Jackson DR, Long SA (2000) Mutual coupling between reduced surface-wave microstrip antennas. IEEE Trans Antennas Propag 48:1581–1593

King RWP (1992) Electromagnetic field of dipole and patch antennas on microstrip. Radio Sci 77:71–78

Komanduri VR, Jackson DR, Williams JT, Mehrotra A (2013) A general method for designing reduced surface wave microstrip antennas. IEEE Trans Antennas Propag 61:2887–2894

Mcnamara DA (1990) Introduction to the uniform geometrical theory of diffraction. Artech House, Norwood

Nikolic MM, Djordjevic AR, Nehorai A (2005) Microstrip antennas with suppressed radiation in horizontal directions and reduced coupling. IEEE Trans Antennas Propag 53:3469–3476

Papapolymerou I, Drayton RF, Katehi LPB (1995) Surface wave mode reduction for rectangular microstrip antennas. In: IEEE international symposium antennas and propagation, Newport Beach

Papapolymerou I, Drayton RF, Katehi LPB (1998) Micromachined patch antennas. IEEE Trans Antennas Propag 46:275–283

Pozar DM (1982) Input impedance and mutual coupling of rectangular microstrip antennas. IEEE Trans Antennas Propag 30:1191–1196

Pozar DM (1986) Finite phased arrays of rectangular microstrip patches. IEEE Trans Antennas Propag 34:658–665

Pozar DM (1998) Microwave engineering. John Wiley & Sons, Hoboken

Pozar DM, Schaubert DH (1984) Scan blindness in infinite phased arrays of printed dipoles. IEEE Trans Antennas Propag 32:602–610

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_51-1

Waveguide Slot Array Antennas

Jiro Hirokawa* and Miao Zhang
Tokyo Institute of Technology, Tokyo, Japan

Abstract

At first, the history on studies on waveguide slot antennas and arrays is described briefly. Then the components for the waveguide slot antennas are explained. Various configurations of the waveguide slot antenna are described, and the equivalent circuits for some configurations are given. The procedure to design a waveguide slot array antenna is explained, and also the analysis model to design the element is discussed as well as the scattering matrix in connecting the equivalent circuit with a transmission line. Finally, the design examples for one-dimensional and two-dimensional arrays of the waveguide slot antennas are explained.

Keywords

Slot antenna; Waveguide; Array; Equivalent circuit; Longitudinal slot antenna; Inclined slot antenna; Edge-slot antenna; Reflection-canceling slot antenna; Standing-wave excitation; Traveling-wave excitation; Series feed

Introduction

A waveguide slot antenna is a radiating slot antenna cut on a waveguide. This chapter focuses on slot antennas and their arrays on rectangular waveguide(s) of metal. Typically, the waveguide is hollow but sometimes it is filled with dielectric to reduce the guided wavelength. The waveguide slot antenna has features of low loss and high-power capability. It is mainly used for communication and radar systems in the microwave band and higher.

The theories for the analyses and the designs on the waveguide slot antennas and their arrays have been developed well for a long time by using knowledge on mathematics. Recently, electromagnetic wave simulators have also been developed well, and they are used to analyze and design the waveguide slot antennas and their arrays. This chapter pays attention to the analysis and the design of using them.

History on Waveguide Slot Antennas and Arrays

Watson investigated experimentally at first (Watson 1946, 1947). A slot cut on a waveguide wall interrupts the current flow and couples a power from the waveguide modal field into free space. Stevenson formulated firstly the theoretical analysis as a boundary value problem (Stevenson 1948). He found that an aperture electric field of a resonant slot had sinusoidal distribution in the first approximation. Expressions for the normalized conductance and resistance were derived for various resonant slots. Oliner derived an equivalent impedance of a slot on a transmission line model by the variational method (Oliner 1957). His

*Email: jiro@antenna.ee.titech.ac.jp

DOI 10.1007/978-981-4560-75-7_51-1

expression for the conductance was depended on both the offset and length of the slot, but the susceptance was not depended the offset. He also included the effects of the waveguide wall thickness to verify the theoretical results with experimental ones. Oliner's variational solution was extended to include the effects of the offset by Yee (1974). Stegen measured carefully a set of data and facilitated the slot array design by making plots for the conductance and susceptance of longitudinal slot antennas cut on the broad wall of a rectangular waveguide as a function of slot length normalized to its resonant length (Stegen 1971). Khac and Carson solved the slot analysis problem numerically at first, using moment method (Khac and Carson 1973). They used pulse basis functions and point matching to solve for the unknown electric fields on the slot apertures. Lyon and Sangster used entire basis sinusoidal functions and concluded two bases functions are sufficient to get accurate solutions (Lyon and Sangster 1981). Stern and Elliott used piecewise sinusoidal functions in Galerkin's method (Stern and Elliott 1985). Seki and Goto analyzed a circularly polarized slot pair by taking into the mutual coupling between the two slots on the narrow wall of a rectangular waveguide together with the wall thickness effect (Seki and Goto 1981). Josefsson and Rangarajan analyzed accurately longitudinal and compounded slots on the broad wall of a rectangular waveguide, respectively (Josefsson 1987; Rangarajan 1989). Elliott presented a design theory of linear or planar arrays of waveguide-fed slots, including the effects of external mutual coupling (Elliott and Kurtz 1978; Elliott 1983).

Components

Slot Antenna

Figure 1 shows the typical structure of a slot antenna. It is a narrow straight rectangular aperture on a flat metal plate. The length is about a half wavelength for the slot antenna to resonate. The width is much smaller than the length and the wavelength.

The slot antenna is excited by an electric current crossing with it (equivalently a magnetic field along it). The excitation is controlled by the intensity of the current. It is also controlled by the crossing angle of the current. When the current is perpendicular to the slot antenna (the magnetic field is parallel to the slot antenna), the excitation is the maximum. When the current is parallel to the slot antenna (the magnetic field is perpendicular to the slot antenna), the excitation is zero. Also, the length is smaller than a half wavelength; the excitation becomes weaker.

The polarization of the slot antenna (the electric field on the slot antenna) is oriented along the width. The amplitude of the electric field is sinusoidal along the length and zero at the edges. It is diverged at the edges along the width. The radiation pattern in the E-plane is uniform. It is a figure of eight in the H-plane. The radiation is the maximum in the boresight and null along the length.

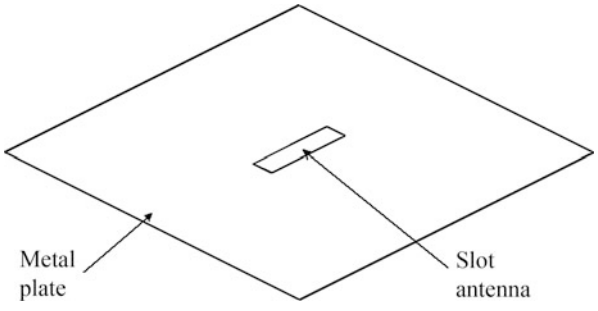


Fig. 1 Slot antenna

DOI 10.1007/978-981-4560-75-7_51-1

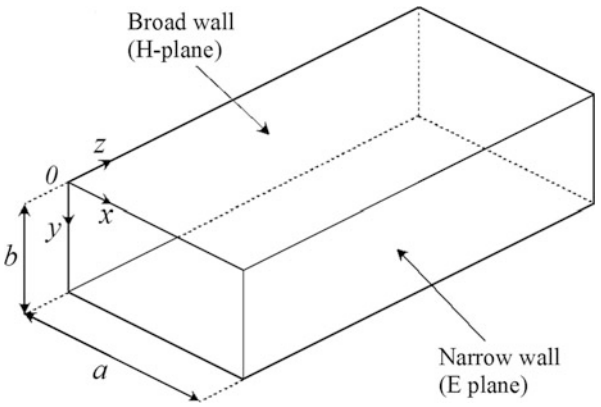


Fig. 2 Rectangular waveguide

Rectangular Waveguide

Figure 2 shows the structure of a rectangular waveguide. The broad-wall width a should be in a range from a half wavelength to one wavelength, and the narrow-wall width b should be less than a half wavelength to propagate only a dominant TE_{10} mode. The broad wall is called as the H-plane because the magnetic field of the TE_{10} mode is parallel to the broad wall. Similarly, the narrow wall is called as the E-plane because the electric field of the TE_{10} mode is parallel to the narrow wall. The field components of the $+z$-directed TE_{10} mode are given by

$$H_z = A\cos\frac{\pi x}{a}\exp(-j\beta z)$$

$$H_x = \frac{j\beta}{k_c}A\sin\frac{\pi x}{a}\exp(-j\beta z)$$

$$E_y = -jAZ_h\frac{\beta}{k_c}\sin\frac{\pi x}{a}\exp(-j\beta z)$$

$$E_z = E_x = H_y = 0$$

where k_c, β, and Z_h are the cutoff wave number, the propagation constant, and the characteristic impedance, respectively. They are given by

$$k_c = \frac{\pi}{a}$$

$$\beta = \sqrt{k_0^2 - \left(\frac{\pi}{a}\right)^2}$$

$$Z_h = -\frac{E_y}{H_x} = \frac{k_0}{\beta}Z_0$$

where k_0 and Z_0 are the wave number and the wave impedance in the free space, respectively. They are given by

DOI 10.1007/978-981-4560-75-7_51-1

$$k_0 = \frac{2\pi}{\lambda_0}$$

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} = 120\pi$$

where λ_0, ε_0, and μ_0 are the wavelength, the permittivity, and the permeability in the free space, respectively.

The transverse component of the electric current (the longitudinal component H_z of the magnetic field) on the broad wall is sinusoidal, which is the maximum at the edges and zero at the center along the broad-wall width. The longitudinal component of the electric current (the transverse component H_x of the magnetic field) on the broad wall is sinusoidal, which is the maximum at the center and zero at the edges along the broad-wall width. The electric current on the narrow walls has the component along height (the magnetic field has only the longitudinal component). The components of the fields and the current are uniform along the narrow wall.

The guided wavelength λ_g of a hollow rectangular waveguide is given as follows.

$$\lambda_g = \frac{2\pi}{\beta} = \frac{\lambda_0}{\sqrt{1 - \left(\frac{\lambda_0}{2a}\right)^2}}$$

It is longer than the free-space wavelength. When slot antennas are arrayed with a spacing of one guided wavelength for inphase excitation on a hollow rectangular waveguide, undesired grating lobes are observed. To suppress the grating lobes, the spacing should be a half guided wavelength which is shorter than the free-space wavelength. Or slow-wave structure such as filling dielectric should be placed in the waveguide.

Elements

Longitudinal Slot Antenna

Figure 3 shows the structure of a longitudinal slot antenna. It is cut in parallel to the waveguide with an offset d from the center on the broad wall. The polarization is perpendicular to the waveguide. The

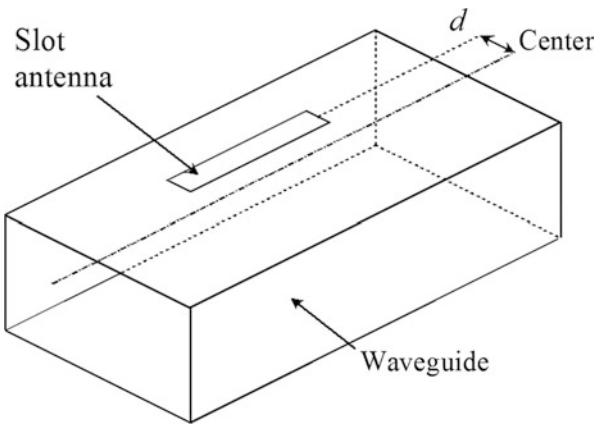


Fig. 3 Longitudinal slot antenna

DOI 10.1007/978-981-4560-75-7_51-1

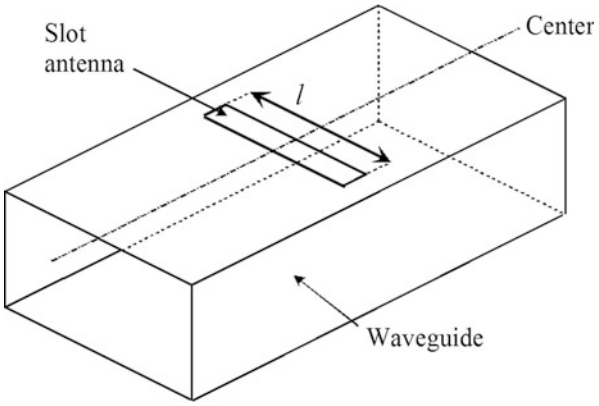


Fig. 4 Transverse slot antenna

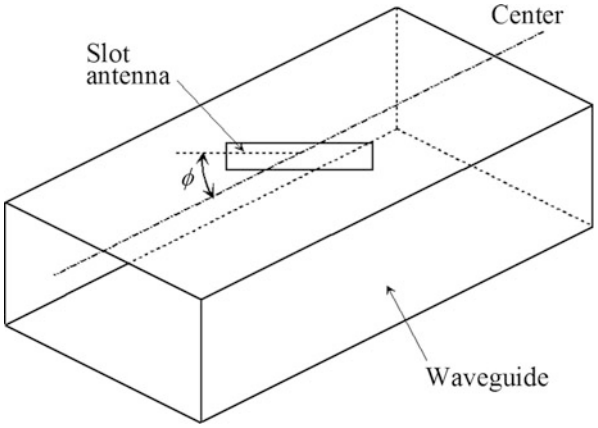


Fig. 5 Inclined slot antenna

excitation is controlled mainly by the offset. It is the maximum at the edges and zero at the center. The equivalent circuit on a transmission line is a parallel admittance because the longitudinal slot antenna is excited by the transverse component of the current flowing between the upper and the bottom broad walls through the narrow wall.

When the longitudinal slot antenna is resonant, the equivalent circuit for the design becomes a parallel conductance. In order to radiate in the boresight, the longitudinal slot antennas are arrayed by a spacing of a half guided wavelength, and the offset direction is opposite among the adjacent slots.

Transverse Slot Antenna

Figure 4 shows the structure of a transverse slot antenna. It is cut perpendicularly to the waveguide on the broad wall. It is placed at the center. The polarization is parallel to the waveguide. The excitation is mainly controlled by the length l. In order to radiate in the boresight, the transverse slots should be arrayed by a spacing of one guided wavelength, and the slow-wave structure such as filling dielectric should be placed in the waveguide.

Inclined Slot Antenna

Figure 5 shows the structure of an inclined slot antenna. It is cut with a tilting angle from the center axis on the broad wall. The excitation is controlled by the tilting angle ϕ. It becomes stronger for a larger tilting angle, and it is zero for 0°. The equivalent circuit on a transmission line is a series impedance because the

DOI 10.1007/978-981-4560-75-7_51-1

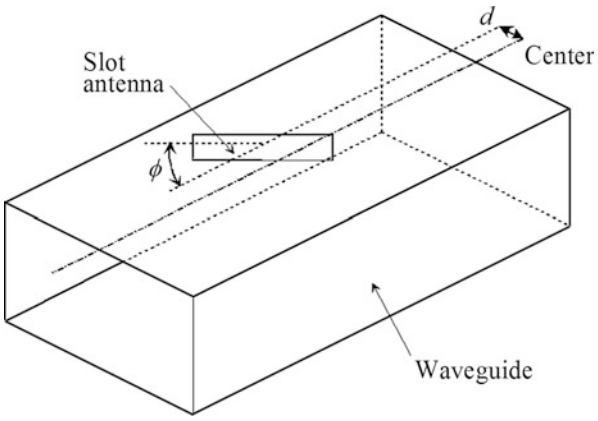


Fig. 6 Compound slot antenna

inclined slot antenna is excited by the longitudinal component of the current flowing along the center of the broad wall. When the inclined slot antenna is resonant, the equivalent circuit becomes a series resistance.

In order to radiate in the boresight, the inclined slot antennas are arrayed with a spacing of a half guided wavelength, and the inclining direction is opposite among the adjacent slots. The main polarization is perpendicular to the waveguide axis. Undesired cross polarization is observed in tilted directions with the angle θ from the boresight given by $\sin^{-1}\frac{\lambda_0}{\lambda_g}$. Its level is higher for a larger tilting angle of the slot.

The inclined slots are often used to feed radiating waveguides in a two-dimensional waveguide slot array antenna.

Compound Slot Antenna

Figure 6 shows the structure of a compound slot antenna (Maxum 1960). The compound slot antenna has both the offset and the tilting angle from the waveguide center axis on the broad wall. For a resonant length, it can arbitrarily control the phase of the excitation, while the longitudinal or the inclined slot antennas have the fixed phase of the excitation. The inclination makes the cross polarization.

Edge-Slot Antenna

Figure 7 shows the structure of an edge-slot antenna. The edge-slot antenna is cut with a tilting angle from the waveguide center axis on the narrow wall. In order to get a resonant length of a half wavelength, the edges are often cut on the broad wall. The excitation is controlled by the tilting angle ϕ. It becomes smaller for a larger tilting angle, and it is zero for 0°. The equivalent circuit on a transmission line is a parallel admittance because the edge-slot antenna is excited by the transverse component of the current flowing on the narrow wall between its top and its bottom. When the edge-slot antenna is resonant, the equivalent circuit becomes a parallel conductance.

In order to radiate in the boresight, the edge-slot antennas are arrayed with a spacing of a half guided wavelength, and the tilting direction is opposite among the adjacent slots. The main polarization is parallel to the waveguide axis. Undesired cross polarization is observed in tilted directions with the angle θ from the boresight given by $\sin^{-1}\frac{\lambda_0}{\lambda_g}$. Its level is higher for a small tilting angle of the slot.

Reflection-Canceling Slot Antenna

Each of the abovementioned slot antennas gives reflection. In an array of these slot antennas, the reflections can make the operation and the design complicate. In order to avoid the complexity, the

DOI 10.1007/978-981-4560-75-7_51-1

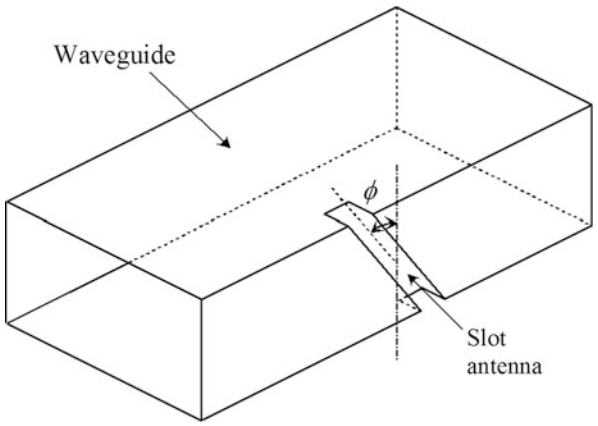


Fig. 7 Edge-slot antenna

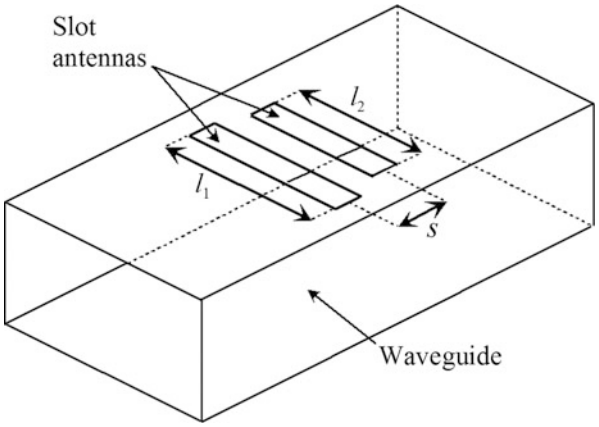


Fig. 8 Transverse slot antenna pair

reflection-canceling slot antennas were introduced. Figure 8 shows the structure of a transverse slot antenna pair on the broad wall (Sakakibara et al. 1994). The two slot antennas in a pair have a spacing of about a quarter guided wavelength. The spacing is smaller than a quarter guided wavelength due to the mutual coupling between the two slot antennas. The reflected waves from the two slot antennas have an effective path length difference of a half guided wavelength so that they are canceled out. The length of one slot antenna is used to control the amplitude of the excitation. The length of the other slot antenna and the spacing between the two slot antennas are used to suppress the reflection. Another example of the reflection-canceling slot antenna is the configuration consisting of a slot antenna with a post or a wall (Park et al. 2003). In an array of the reflection-canceling slot antennas, the phase of the excitation among the adjacent elements can be easily controlled by the spacing of the elements.

A circularly polarized element on a rectangular waveguide has generally small reflection. It consists of orthogonal two slot antennas excited by 90-degree phase difference. The reflected waves from the two slot antennas have 180-degree phase difference so that they are canceled out.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_51-1

Design Procedure for a Waveguide Slot Array Antenna

The design procedure for a waveguide slot array antenna is as follows.

1. Design of the excitation distribution of the elements
2. Design of the elements for their desired excitations
3. Design modification of the elements by analyzing the overall array antenna

For a two-dimensional slot array antenna, the array of slot antennas on the radiating waveguide and that on the feed waveguide should be designed separately. Each step in the above procedure is explained below.

Design of the Excitation Distribution of the Elements

The excitation distribution of the elements in the array should be designed to have a desired radiation pattern using pattern synthesis technique with the consideration of the element pattern.

Design of the Elements for Their Desired Excitations

The dimensions in each element should be designed to have the desired excitation. In other words, they should be designed to have the desired value of the parallel admittance or the series impedance in the equivalent circuit or the desired coupling in the reflection-canceling element. Figure 9 shows the analysis model of a longitudinal slot antenna as an example. A ground plane is typically embedded in the external region. Its size is infinite or finite depending on the analysis method for an isolated slot antenna. In many cases, such as using an electromagnetic wave simulator, the ground plane is finite, and the external region is truncated as a radiation box. The truncation of the ground plane affects the radiation pattern and the coupling. The faces of the box are assumed to be radiation boundaries except the ground plane. For an example, the size of the radiation box is chosen as $1\lambda_0(x) \times 1\lambda_0(y) \times 0.5\lambda_0(z)$ not to affect the coupling very much. The consideration of the mutual coupling in the infinite array with uniform excitation could give better initial values of the dimensions of the slot antenna. For a one-dimensional array, two faces parallel to the *xy*-plane in the radiation box should be replaced with periodic boundary walls. The spacing of the two faces in the *z*-direction should be equal to the element spacing. For a two-dimensional array, not only two faces parallel to the *xy*-plane but also those parallel to the *yz*-plane should be replaced with

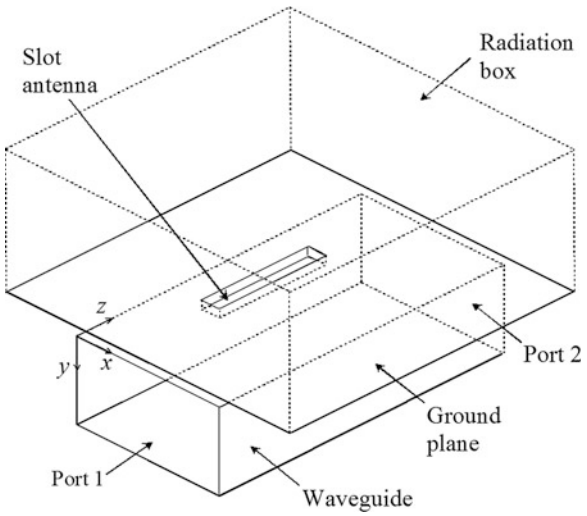


Fig. 9 Analysis model of a longitudinal slot antenna

DOI 10.1007/978-981-4560-75-7_51-1

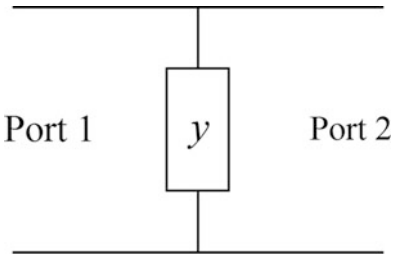


Fig. 10 Parallel admittance

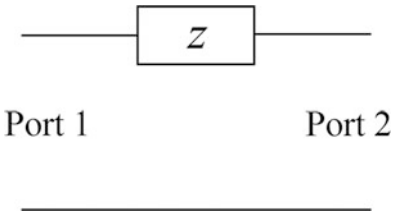


Fig. 11 Series impedance

periodic boundary walls. The spacing of the two faces in the *x*-direction should be equal to the broad-wall width plus the wall thickness of the waveguide. The wall thickness as well as the shape (sometimes rounded from the fabrication point of view) of the edges in the slot antenna should be included because it could affect the frequency dependence of the excitation.

Ports should be placed at both ends of the waveguide. The distance between the port and the center of the slot is large enough to attenuate high modes. For example, it is chosen to be a half guided wavelength. Once the scattering matrix of a waveguide slot antenna is obtained by analyzing it, the parameter of the equivalent circuit can be derived. When the equivalent circuit is a parallel admittance as shown in Fig. 10, the value y which normalized the characteristic admittance of the transmission line is given by the components of the 2-port scattering matrix as follows.

$$y = -2\frac{S_{11}}{1+S_{11}} = 2\frac{1-S_{21}}{S_{21}}$$

Similarly, when the equivalent circuit is a series impedance as shown in Fig. 11, the value z which normalized the characteristic impedance of the transmission line is given as follows.

$$z = 2\frac{S_{11}}{1-S_{11}} = 2\frac{1-S_{21}}{S_{21}}$$

Note that $S_{12} = S_{21}$ is satisfied because of the reciprocity, and $S_{22} = S_{11}$ is satisfied because of the symmetry of the circuit.

Design Modification of the Elements by Analyzing the Overall Array Antenna

The dimensions in each element should be modified by analyzing the overall structure of the waveguide slot antenna to have the desired excitation.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_51-1

One-Dimensional Slot Array Antennas

Standing-Wave Excitation

Figure 12 shows the structure of a one-dimensional array of resonant longitudinal slot antennas with standing-wave excitation. The number of the slot antennas is N. The array antenna is fed with the end feed. It is fed from one end of the waveguide and is terminated by a short circuit at the other end. The spacing between the short circuit and the last slot antenna is a quarter guided wavelength at the design frequency. The short circuit makes a standing wave in the waveguide, which excites the slot antennas. The longitudinal slot antennas are spaced with a half guided wavelength at the design frequency with the offsets opposite among the adjacent elements. The main beam is oriented to the broadside.

Figure 13 shows the equivalent circuit of the array antenna with the standing-wave excitation. The designed excitation amplitude of the n-th slot is given as α_n. The normalized conductance to be determined of the n-th slot is g_n. The n-th slot is excited by a constant voltage V. The radiation power of the n-th slot is $g_n V^2$, which is proportional to a_n^2. Therefore, $g_n = Ca_n^2$ where C is a constant. For others to match at the feed, $\sum_{n=1}^{N} g_n = C\sum_{n=1}^{N} a_n^2 = 1$ should be satisfied. From this equation, C is determined, and then g_n is specified. The parameters for the n-th slot are determined for g_n.

Another way to feed the one-dimensional array is the center feed where it is fed from the center of the waveguide and is terminated by short circuits at both ends. In this way, the matching condition at the feed becomes $\sum_{n=1}^{N} g_n = C\sum_{n=1}^{N} a_n = 2$.

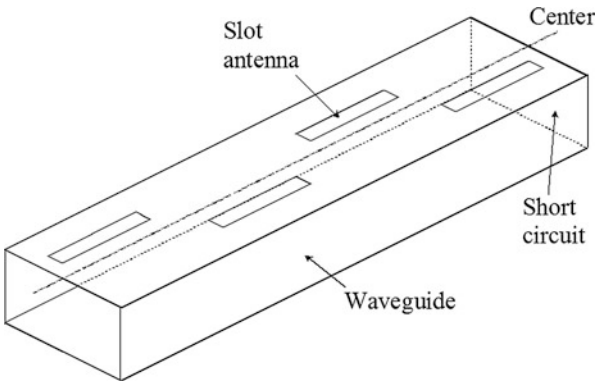


Fig. 12 One-dimensional array of resonant longitudinal slot antennas with standing-wave excitation

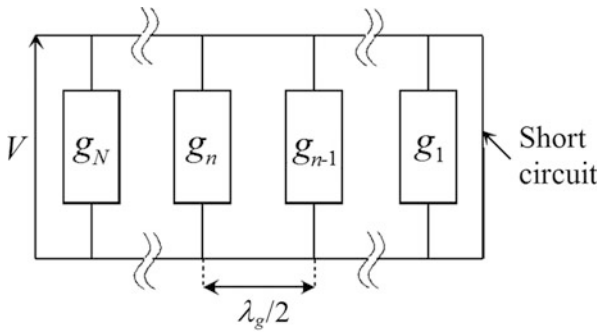


Fig. 13 Equivalent circuit of the resonant longitudinal slot array antenna with the standing-wave excitation

DOI 10.1007/978-981-4560-75-7_51-1

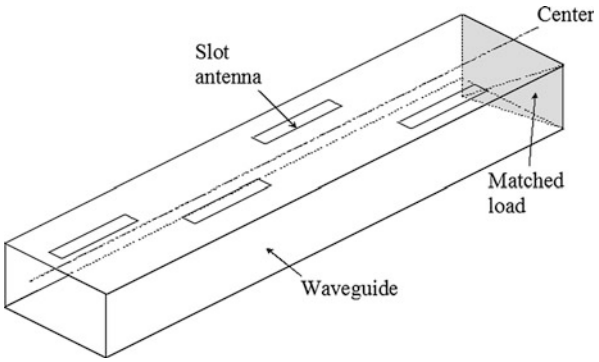


Fig. 14 One-dimensional array of resonant longitudinal slot antennas with traveling-wave excitation

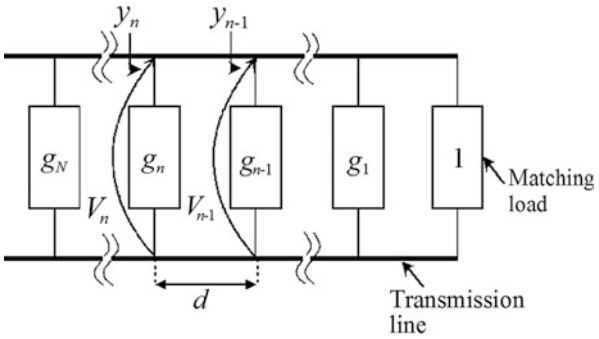


Fig. 15 Equivalent circuit of the resonant longitudinal slot array antenna with the traveling-wave excitation

Traveling-Wave Excitation

Figure 14 shows the structure of a one-dimensional array of resonant longitudinal slot antennas with traveling-wave excitation. The number of the slot antennas is N. The array antenna is fed with the end feed. It is terminated by a matching load. The matching load could be replaced with a matching element that radiates all the residual power at the end. No reflected wave from the end makes a traveling wave in the waveguide, which excites the slot antennas. The longitudinal slot antennas are spaced with other than a half guided wavelength at the design frequency with the offsets opposite among the adjacent elements. The main beam is tilted from the broadside.

Figure 15 shows the equivalent circuit of the array antenna with the traveling-wave excitation. The slot antennas are numbered from the end to the feed. The designed excitation amplitude of the n-th slot is given as a_n. The normalized conductance to be determined of the n-th slot is g_n. The spacing is set to be constant d, and the propagation constant in the waveguide is β. At the n-th slot, the total admittance looking to the matched load and the voltage are y_n and V_n. They satisfy the following recurrence relations.

$$y_n = g_n + \frac{y_{n-1}\cos\beta d + j\sin\beta d}{\cos\beta d + jy_{n-1}\sin\beta d}$$

$$V_n = V_{n-1}(\cos\beta d + jy_{n-1}\sin\beta d)$$

The radiating power of the n-th slot is $g_n|V_n|^2$, which is proportional to a_n^2. Therefore, $g_n|V_n|^2 = Ca_n^2$ where C is a constant. According to the above recurrence relations, g_n is specified.

DOI 10.1007/978-981-4560-75-7_51-1

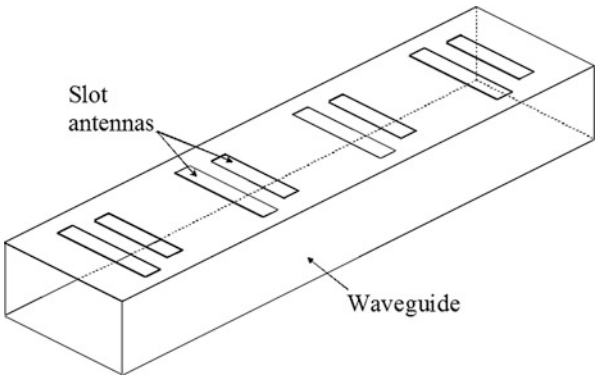


Fig. 16 One-dimensional array of transverse slot antenna pairs

When the number of the slot antennas is large, the reflections among them are negligible, and all the slots are excited uniformly; the design can be simplified. The constant spacing d is determined to suppress the overall reflection at the feed.

$$d = \left(\frac{1}{2} \pm \frac{1}{2N}\right)\lambda_g$$

The beam-tilting angle θ toward the end is given as

$$\sin\theta = \lambda_0\left(\frac{1}{\lambda_g} - \frac{1}{2d}\right)$$

where λ_0 is the free-space wavelength, and λ_g is the guided wavelength. The parameters for the n-th slot are determined for the conductance g_n given similarly to the standing-wave excitation.

Reflection-Canceling Slot Array Antennas

Figure 16 shows the structure of a one-dimensional array of transverse slot antenna pairs. Each transverse slot antenna pair is designed to suppress the reflection. The number of the slot antenna pairs is N. The array antenna is fed with the end feed. It is terminated by a matching element that radiates all the residual power at the end.

The slot antenna pairs are numbered from the end to the feed. The designed excitation amplitude and phase in degrees of the n-th slot antenna pair are given as a_n and P_n, respectively. The coupling c_n of the n-th slot antenna pair is determined as

$$c_n = \frac{a_n^2}{\sum_{m=1}^{n} a_m^2}$$

where the coupling is defined as the ratio of the radiating power over the incident power. The dimensions or the n-th slot are determined for c_n. The transmission phase p_{21n} and the radiation phase p_{31n} in degrees are also estimated for c_n. Figure 17 shows the phase relation in the array antenna. The element spacing d_n between the elements $n + 1$ and n is derived by the following equation.

DOI 10.1007/978-981-4560-75-7_51-1

$$p_{n+1} - p_{31n+1} + p_{21n+1} - 360\frac{d_n}{\lambda_g} + p_{31n} = p_n - 360$$

Two-Dimensional Slot Array Antennas

Structure

Figure 18 shows a typical structure of a two-dimensional array of longitudinal slot antennas. Radiating waveguides are arrayed on the upper layer. A feed waveguide is placed at the lower layer. The feed waveguide is typically located at the center of the radiating waveguides. It is fed in phase toward both directions by a longitudinal feed slot antenna at its center. Inclined coupling slot antennas are cut to couple between the radiating and the feed waveguides as shown. They are spaced with a half guided wavelength in the feed waveguide and have opposite orientation among the adjacent slot antennas to excite them in phase at the design frequency. Figure 19 shows the structure of the inclined coupling slot antenna. The radiating and the feed waveguides are designed separately. Mutual coupling among the radiating slot antennas in the external region should be included in the design, while mutual coupling among the coupling slot antennas outside the feed waveguide does not need to be included in the design.

Series Feed and Corporate Feed

In the series-feed array antennas excited in phase at the design frequency, when the frequency is away from the design one, the slot antennas are not excited in phase anymore due to the change of the guided

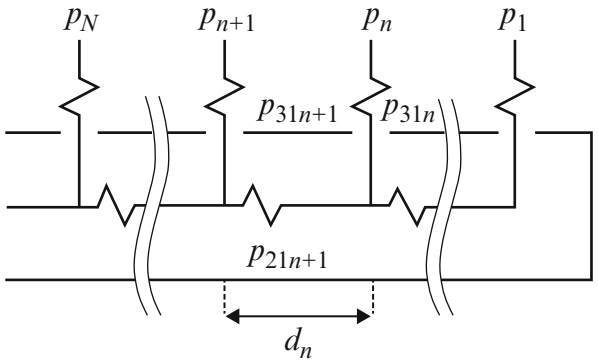


Fig. 17 Phase relation in the transverse slot antenna pair array

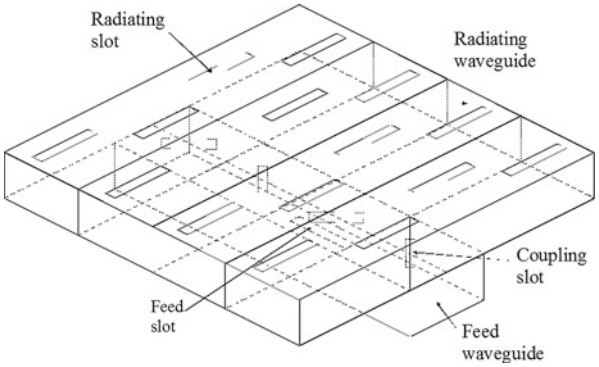


Fig. 18 Two-dimensional array of longitudinal slot antennas

DOI 10.1007/978-981-4560-75-7_51-1

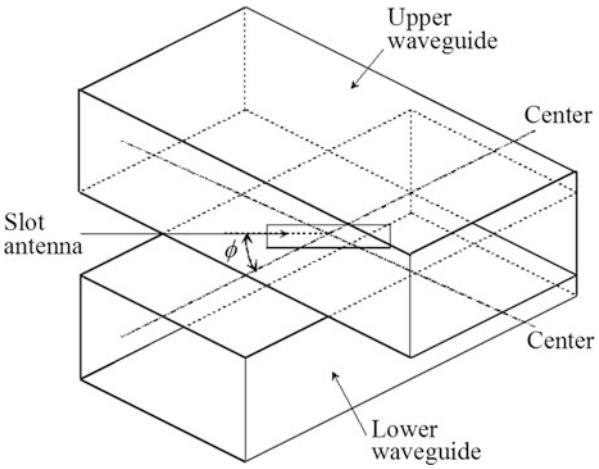


Fig. 19 Inclined coupling slot antenna

wavelength, and the directivity is degraded. On the other hand, in the corporate-feed array antennas, even when the frequency is away from the design one and the guided wavelength is changed, the slot antennas are kept to excite in phase, and the directivity is not degraded. The bandwidth of the corporate-feed array antenna is dependent on those of the elements in the radiating and the feed waveguides.

Design Examples

Alternating-Phase Fed Single-Layer 18 × 18-Element Waveguide Slot Array

An alternating-phase fed single-layer (Sakakibara et al. 1997) 18 × 18-element array with uniform distribution is designed at 25.3 GHz. There are 18 radiating waveguides connected with a feeding circuit at one end. Eighteen slots cut in each radiating waveguide are fed in series. The dimensions of the radiating waveguide are 7.9 mm in width and 3.0 mm in height. On the other hand, the thickness of waveguide sidewall and that of slot thickness are 1.5 and 0.2 mm, respectively. In order to suppress the reflection from the slots as well as to decrease the slot spacing below a half guided wavelength, backward beam-tilting technique is adopted in the design of the slot array. The slot spacing calculated from the equation $d = (1/2 + m/18)\lambda_g$ is fixed at 7.96 mm in common for $m = -1$, and the main beam is tilted at -4.7° from boresight to the feed.

Step 1: To realize a uniform aperture distribution as well as a maximum antenna gain, the targeted values of slot admittances are determined from the equivalent circuit model first. A matching slot that radiates all the residual power is adopted at the end. The normalized conductance and susceptance in a one-dimensional array are summarized in Fig. 20. Since an array of resonant slots is assumed to be realized, the susceptance will be zero for each element. Additionally, the total reflection at the waveguide input is estimated at −19.9 dB from its equivalent circuit mode as well.

Step 2: To obtain the design chart as well as the initial slot parameters, a single resonant slot located in a uniformly excited infinite two-dimensional array is simulated. Here, one pair of periodic boundary walls in the longitudinal direction and the other pair of PEC (perfect electric conductor) walls in the transverse direction are assumed in the external region of slot array as illustrated in Fig. 21. These PEC walls will be replaced by periodic boundary walls for an inphase fed array instead of an alternating-phase fed one. The slot length and offset for a resonant slot as a function of normalized conductance are

DOI 10.1007/978-981-4560-75-7_51-1

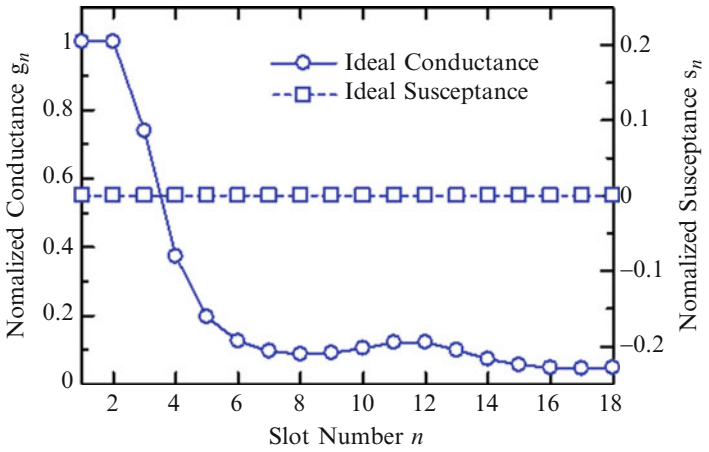


Fig. 20 Normalized conductance and susceptance determined in an equivalent circuit 2-D array

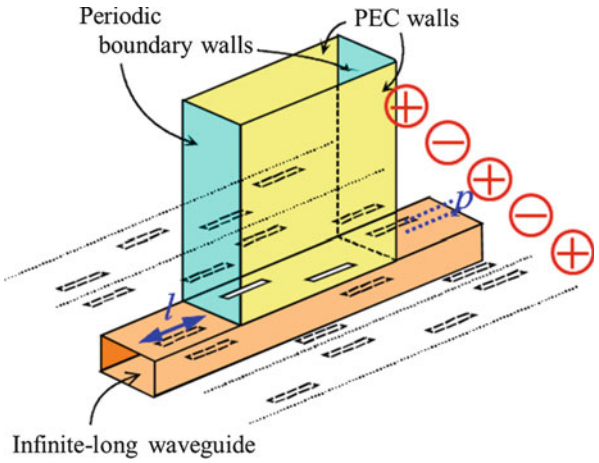


Fig. 21 Single-slot analysis model with periodic boundary walls in the external region

summarized in Fig. 22 which is used in the design chart. The initial slot parameters shown in Fig. 23 are determined to realize the slot admittances in Fig. 20.

Step 3: The one-dimensional array with slot parameters designed above is analyzed with the effects of mutual coupling taken into account. Here, either a method of moment (MoM) analysis (Sakakibara et al. 1996) or an electromagnetic (EM) field simulator adopting periodic boundary walls in the transverse direction can help us to evaluate the external mutual coupling. It is a good approximation for an actual two-dimensional 18 × 18-element array and has the feature of short computation time. The slot parameters will be iteratively designed for further improvement, and the optimized slot parameters are also summarized in Fig. 23 for comparison.

Step 4: After the design of slots, a feed circuit consisting of a series of T-junctions will be designed to excite the radiating waveguides in identical amplitude and alternating phase with a difference of 180°. The full-structure analysis can be conducted by applying a full-wave MoM code (Zhang et al. 2011) or by just an EM field simulator to confirm the antenna performances. The excitation coefficients obtained from the MoM analysis are shown in Fig. 24. The nonuniformities in amplitude and phase are 5 dB and ± 25°, respectively. The simulated antenna gain is 32.3 dBi with an aperture efficiency of 79 %.

DOI 10.1007/978-981-4560-75-7_51-1

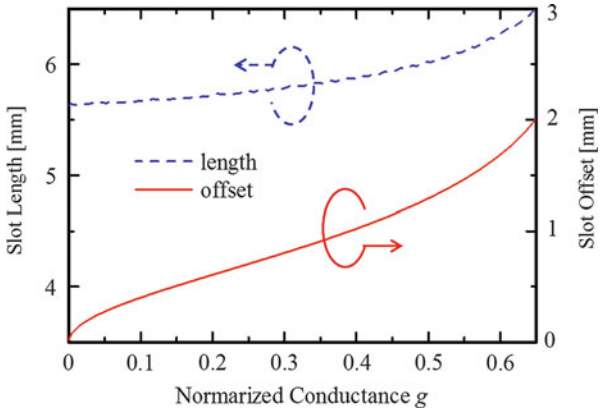


Fig. 22 Slot length and offset as a function normalized conductance obtained in an infinite ground plane

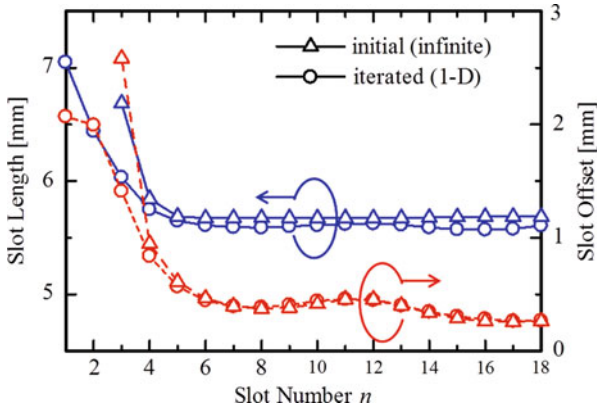


Fig. 23 Initial and iterated values of slot length and offset

By the way, an advanced design technique has been developed (Zhang et al. 2011) to improve the reflection as well as uniformity furthermore. Firstly, with the explicit restraint of input matching, the admittance and element spacing are determined in the equivalent circuit from antenna input toward the termination. At the same time, a significant deviation in the excitation phase as shown in the y direction of Fig. 24b can be compensated by changing the slot spacing d_n or by tuning the admittance y_n off resonance. Secondly, to realize the admittance specified above, the slot length and offset are determined by iterative use of the full-wave MoM analysis of the finite two-dimensional waveguide slot array. Hence, the uniformity in both x and y directions as well as the aperture efficiency can be further improved.

Partially Corporate-Feed Double-Layer 20 × 20-Element Waveguide Slot Array

A 20 × 20-element double-layer antenna with uniform excitation is to be designed at 38.7 GHz. The widths of the radiating and feeding waveguides are 5.0 and 5.4 mm, respectively. The waveguide heights are 2.0 mm in common, and the thickness of sidewalls are 1.0 mm. The widths and thicknesses of both the radiating and coupling slots are 1.0 and 0.4 mm, respectively. By introducing a partially corporate-feed structure as illustrated in Fig. 25, the whole antenna is decomposed into 2 × 2 sub-arrays. Those four sub-arrays are fed in same amplitude and phase without frequency dependence in theory. In addition, the inphase fed array adopted in this paper has regular slot arrangement and the advantage of suppressing the second-order beam (Kurtz and Yee 1957). That is, all radiating waveguides are fed in phase through slot

DOI 10.1007/978-981-4560-75-7_51-1

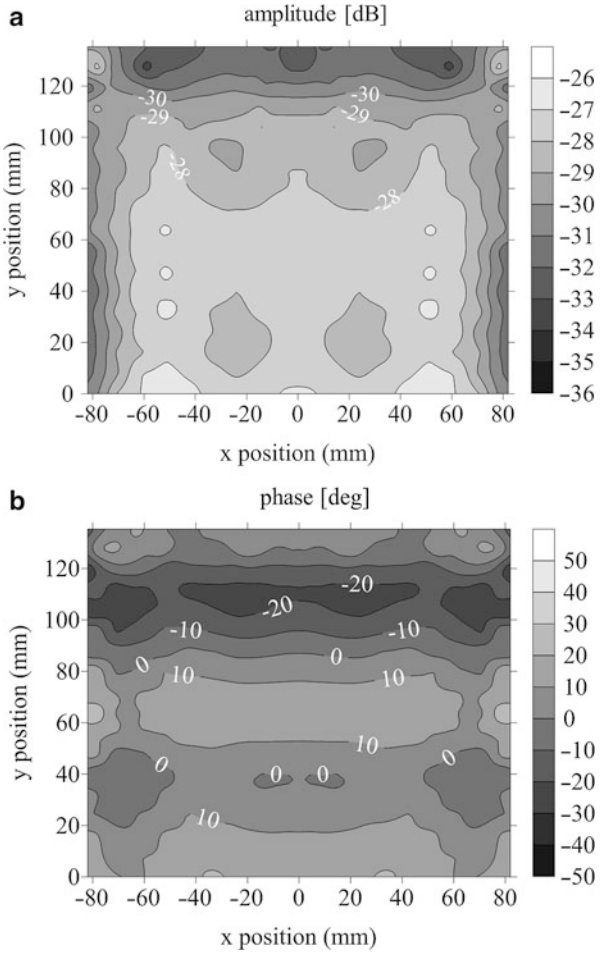


Fig. 24 Excitation coefficients of the designed 18 × 18-element array

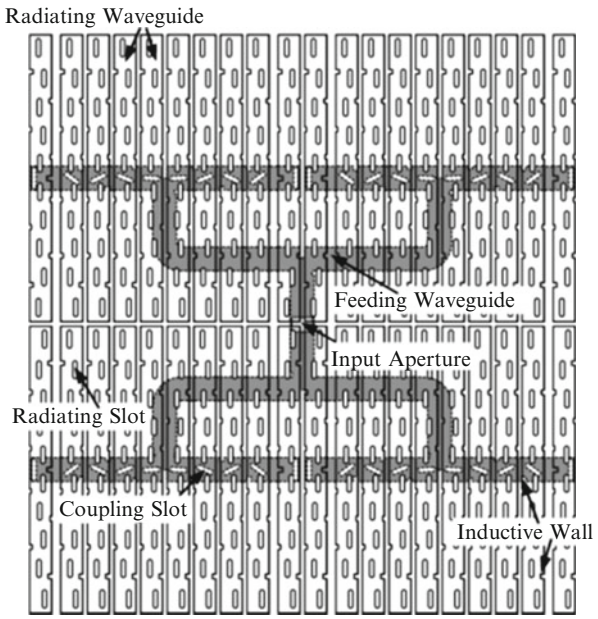


Fig. 25 Structure of a double-layer 20 × 20-element waveguide slot array

DOI 10.1007/978-981-4560-75-7_51-1

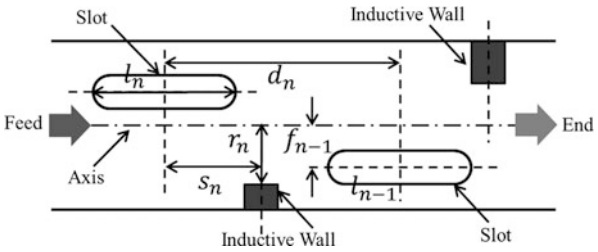


Fig. 26 Design parameters for the slot with an inductive wall

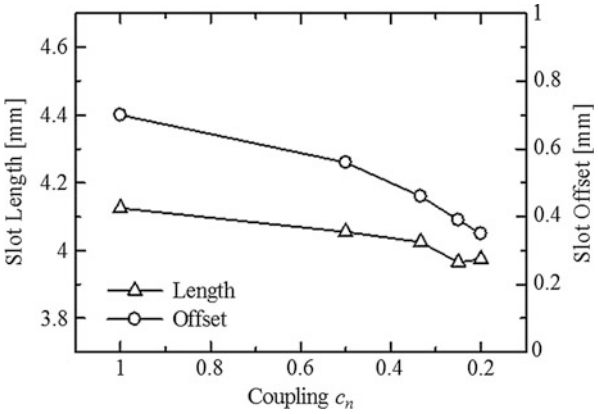


Fig. 27 Designed values of slot length and offset as a function of coupling

couplers. Generally, it is enough to design only five elements in both the feeding and radiating parts, because the 20 × 20-element array is simplified into 2 × 2 sub-arrays according to the partially corporate feed.

In the center-feed and partially corporate-feed arrays, it is necessary to maintain the main beam radiation of the sub-array at the boresight. In addition, the number of elements in series is only five. Instead of beam-tilting technique, inductive walls are introduced to all the radiating slots to realize reflection suppression as well as traveling-wave excitation. The main design parameters for slot and wall are illustrated in Fig. 26. It should be noted that the slot shape treated in this design has round edges at both ends. Similar to the procedure illustrated in the previous design example, a MoM analysis or an EM field simulator is applied in the element design and the one-dimensional array analysis. Since the number of elements is only five, it is not necessary to make a design chart. Each element can be designed individually with a coupling $c_n = 1/n$, where n is the slot number. The slot parameter length and offset are summarized in Fig. 27 as a function of coupling c_n. The positions of inductive walls as illustrated in Fig. 26 are optimized to suppress the reflection locally, and their parameters are summarized in Fig. 28. By the way, the first slot called matching slot can be matched by itself without the assistance from an inductive wall. Since an array of resonant slots is adopted here, the radiation phase p_{31n} is simply assumed to have identical value for all n. The slot spacing d_n is increased from a half guided wavelength to compensate the phase enhancement p_{21n} due to the introduction of inductive wall. The values of spacing and phase are summarized in Fig. 29. The frequency characteristics of reflection for each element are summarized in Fig. 30. It is easily observed that in an array with traveling-wave excitation, the slot closer to the end has stronger coupling and shows narrower bandwidth in reflection. In addition, the reflection of the five-element sub-array is also included in Fig. 30 and is well suppressed below −30 dB at the center frequency. The first sidelobe level is −12.8 dB, and a relative good uniformity in excitation is also predicted.

DOI 10.1007/978-981-4560-75-7_51-1

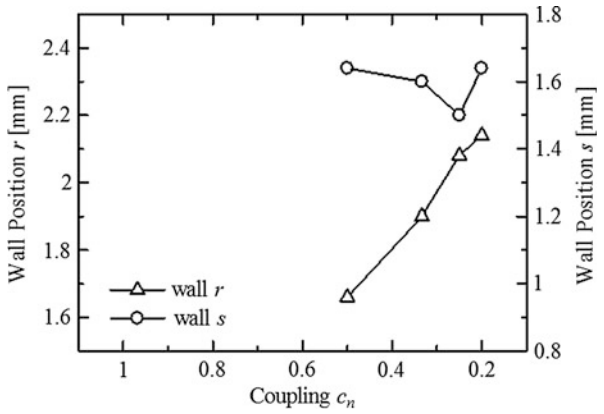


Fig. 28 Designed values of wall positions as a function of coupling

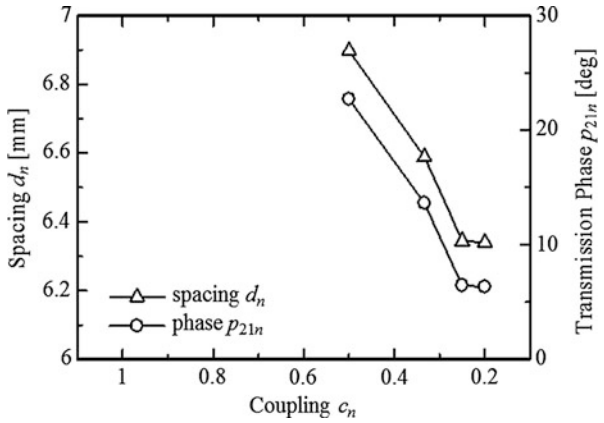


Fig. 29 Slot spacing and transmission phase as a function of coupling

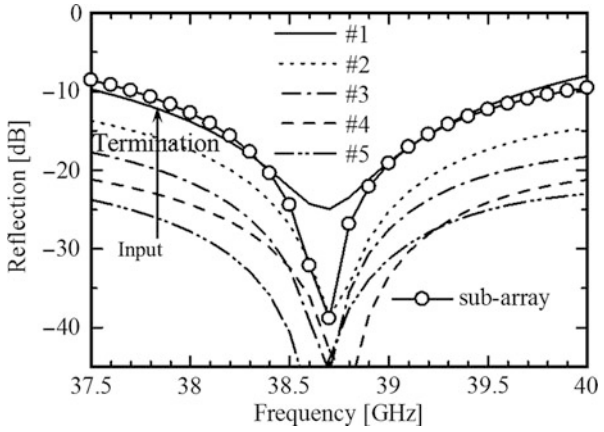


Fig. 30 Reflection of the radiating slots with inductive walls

Summary

This chapter discusses on slot antennas and their arrays on rectangular waveguide(s). The elements and their equivalent circuits are summarized. The design procedures for one- and two-dimensional arrays of the waveguide slot antennas are explained together with their examples.

DOI 10.1007/978-981-4560-75-7_51-1

Cross-References

- ▶ GAP Waveguides
- ▶ Radial-Line Slot Antennas
- ▶ Substrate-Integrated Waveguide Antennas

References

Elliott RS (1983) An improved design procedure for small arrays of shunt slots. IEEE Trans Antennas Propag 31:48–53

Elliott RS, Kurtz LA (1978) The design of small slot arrays. IEEE Trans Antennas Propag 26:214–219

Josefsson LG (1987) Analysis of longitudinal slots in rectangular waveguide. IEEE Trans Antennas Propag 35:1351–1357

Khac TV, Carson CT (1973) Impedance properties of a longitudinal slot antenna in the broad face of a rectangular waveguide. IEEE Trans Antennas Propag 21:708–710

Kurtz LA, Yee JS (1957) Second-order beams of two-dimensional slot arrays. IRE Trans Antennas Propag 5:356–362

Lyon RW, Sangster AJ (1981) Efficient moment method analysis of radiating slots in a thick-walled rectangular waveguide. IEE Proc 128(pt.H):197–205

Maxum BJ (1960) Resonant slots with independent control of amplitude and phase. IRE Trans Antennas Propag 8:383–388

Oliner AA (1957) The impedance properties of narrow radiating slots in the broad wall of rectangular waveguide. IRE Trans Antennas Propag 5:1–20

Park SH, Hirokawa J, Ando M (2003) Simple analysis of a slot with a reflection-canceling post in a rectangular waveguide using only the axial uniform currents on the post surface. IEICE Trans Commun 86(8):2482–2487

Rangarajan SR (1989) Compound radiating slots in a broad wall of a rectangular waveguide. IEEE Trans Antennas Propag 37:1116–1123

Sakakibara K, Hirokawa J, Ando M, Goto N (1994) A linearly-polarized slotted waveguide array using reflection-canceling slot pairs. IEICE Trans Commun 77:511–518

Sakakibara K, Hirokawa J, Ando M, Goto N (1996) Periodic boundary condition for evaluation of external mutual coupling in a slotted waveguide arrays. IEICE Trans Commun 79:1156–1164

Sakakibara K, Kimura Y, Akiyama A, Hirokawa J, Ando M, Goto N (1997) Alternating phase-fed waveguide slot arrays with a single-layer multiple-way power divider. IEE Proc Microw Antennas Propag 144:425–430

Seki H, Goto N (1981) Synthesis of circular polarization with non resonant slots in the narrow wall of a rectangular waveguide. Trans IECE 64(pt.B):1000–1007

Stegen RJ (1971) Longitudinal shunt slot characteristics, Technical report 261, Hughes technical memorandum

Stern GJ, Elliott RS (1985) Resonant length of longitudinal slots and validity of circuit representation: theory and experiment. IEEE Trans Antennas Propag 33:1264–1271

Stevenson AF (1948) Theory of slots in rectangular waveguides. J Appl Phys 19:24–38

Watson WH (1946) Resonant slots. IEE J 93(pt.3A):747–777

DOI 10.1007/978-981-4560-75-7_51-1

Watson WH (1947) The physical principles of wave guide transmission and antenna system. Oxford University Press

Yee HY (1974) Impedance of a narrow longitudinal shunt slot in a slotted waveguide array. IEEE Trans Antennas Propag 22:589–592

Zhang M, Hirokawa J, Ando M (2011) Full-wave design considering slot admittance in 2-D waveguide slot arrays with perfect input matching. IEICE Trans Commun 94:725–734

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_52-1

Omnidirectional Antennas

Xianming Qing[a]* and Zhi Ning Chen[a,b]
[a]Institute for Infocomm Research (I^2R), Singapore, Singapore
[b]National University of Singapore, Singapore, Singapore

Abstract

An omnidirectional antenna which radiates electromagnetic wave uniformly in a specific plane (often in the azimuth plane) is one of the most popularly used antennas in wireless applications. This chapter illustrates the basic principles and recent development of the omnidirectional antennas. The discussion is carried out based on a category of polarization of the omnidirectional antennas. The polarizations of the omnidirectional antennas include linear polarizations of vertical polarization and horizontal polarization, dual linear polarization, and circular polarization. A brief literature review about omnidirectional antennas is also presented along with state-of-the-art designs.

Keywords

Omnidirectional antenna; Vertically polarized omnidirectional antenna; Horizontally polarized omnidirectional antenna; Dual polarized omnidirectional antenna; Circularly polarized omnidirectional antenna; Dipole antenna; Monopole antenna; Loop antenna; Slot antenna; Dielectric resonator antenna

Introduction

Antennas are essential components of all wireless systems. An antenna consisting of an arrangement of metallic conductors (elements) is used with either a wireless transmitter or a wireless receiver. An oscillating current of electrons forced through the antenna by a transmitter will create an oscillating magnetic field around the antenna elements, while the charge of the electrons also creates an oscillating electric field along the elements. These time-varying fields radiate away from the antenna into space as a moving transverse electromagnetic field wave. Conversely, during reception, the oscillating electric and magnetic fields of an incoming electromagnetic wave exert the force on the electrons in the antenna elements, causing them to move back and forth, creating oscillating currents in the antenna.

The tremendous numbers of antennas have been designed and applied in various systems since 1888 when Heinrich Hertz carried out his pioneering experiments to prove the existence of electromagnetic waves. Generally, the antennas can be classified on the basis of:

- Frequency: very-low-frequency (VLF) antenna, low-frequency (LF) antenna, high-frequency (HF) antenna, very-high-frequency (VHF) antenna, ultrahigh-frequency (UHF) antenna, microwave antenna, millimeter/submillimeter (mmW) antenna, Terahertz antenna, and optical antenna.
- Aperture: wire antenna, reflector antenna, microstrip patch antenna, slot antenna, and so on.
- Polarization: linearly (vertically/horizontally) polarized antenna, dual linearly polarized antenna, and circularly polarized antenna.

*Email: qingxm@i2r.a-star.edu.sg

DOI 10.1007/978-981-4560-75-7_52-1

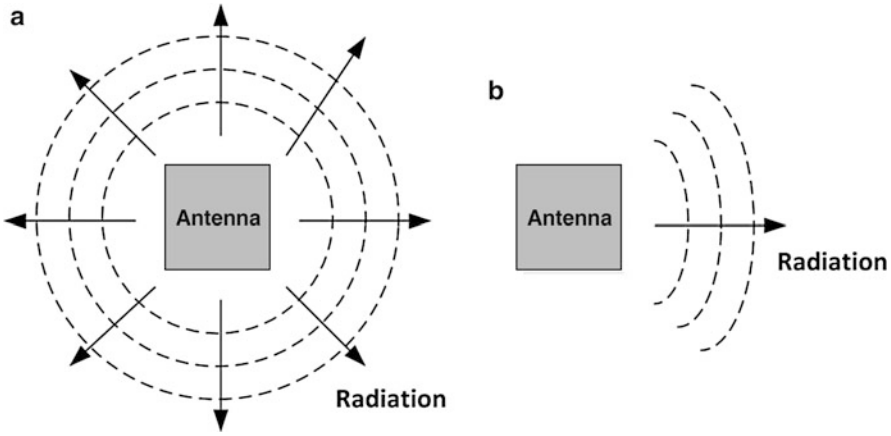


Fig. 1 Omnidirectional antenna versus directional antenna; (**a**) radiation from omnidirectional antenna and (**b**) radiation from directional antenna

- Radiation: isotropic antenna, omnidirectional antenna, and directional antenna.

As shown in Fig. 1, an omnidirectional antenna is able to radiate radio wave uniformly in all directions and offers a 360° coverage in a specific plane (normally in Azimuth plane), which make it very suitable to communicate multiple users. Omnidirectional antennas are widely used for radio broadcasting, satellite, global positioning system (GPS), access point of wireless local area network (WALN), and mobile devices such as cell phone, cordless phone, FM radio, laptop, and so on (Riblet 1947; Croswell and Cockrell 1969).

The first omnidirectional antenna dates back to very early days, a simple dipole or monopole antenna generates vertically polarized omnidirectional radiation in a specific plane. Generally, the omnidirectional radiation is achieved either by the circularly symmetrical current distribution on the radiating element such as monopole antennas and dipole antennas or by combining the radiation of cylindrically positioned multiple directional radiating elements such as microstrip patches and slot antennas. This chapter is organized to concisely summarize the fundamentals of omnidirectional antennas first. Various types of omnidirectional antennas are then reviewed; some of recently developed technologies such as metamaterial-based omnidirectional antennas are also discussed.

Omnidirectional Radiation Pattern

Based on the IEEE Standard Definitions of Terms for Antennas (IEEE Standard 145–1983), the relevant definitions regarding the antenna radiation are as follows:

Isotropic antenna: a hypothetical, lossless antenna having equal radiation intensity in all directions
Omnidirectional antenna: an antenna having an essentially nondirectional pattern in a given plane of the antenna and a directional pattern in any orthogonal plane
Directional antenna: an antenna having the property of radiating or receiving electromagnetic waves more effectively in some directions than others

An isotropic antenna radiates its energy equally in all directions with a spherical radiation pattern, namely, it exhibits omnidirectional radiation pattern in any plane. However, an isotropic antenna never physically exits; it is used as the reference antenna to compare the characteristics of actual antennas. An omnidirectional antenna radiates radio wave power uniformly in all directions in one plane, with the

DOI 10.1007/978-981-4560-75-7_52-1

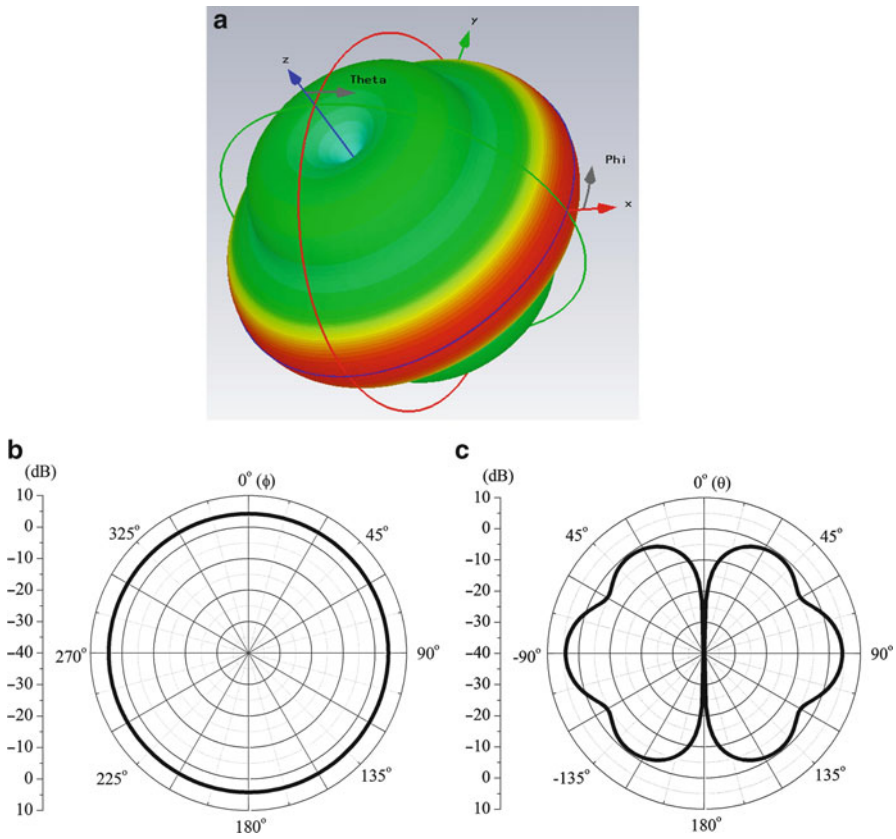


Fig. 2 Typical radiation patterns of an omnidirectional antenna, (**a**) three-dimensional pattern, (**b**) omnidirectional pattern in the azimuth plane (*x-y plane*), and (**c**) pattern in the elevation plane (*x-z* plane)

radiated power decreasing against elevation angles above or below the plane, dropping to zero on the antenna's axis, and producing a "doughnut-shaped" radiation pattern, as shown in Fig. 2.

Directivity of Omnidirectional Antenna

The directivity of an isotropic antenna is unity since its power is radiated equally in all directions. For all other antennas, the maximum directivity will always be greater than unity, and it is a relative "figure of merit" which gives an indication of the directional properties of the antenna as compared with those of an isotropic antenna. It is known that the directivity of a directional antenna can be estimated accurately using the half-power beamwidth (HFBW) of the antenna in the two main planes (Balanis 1997). However, such a formula is not valid for the omnidirectional antenna because the HFBW definition is not applicable in the plane with omnidirectional radiation.

The radiation pattern of an omnidirectional antenna in the elevation plane can be approximated by

$$U = |\sin^{n}(\theta)|, 0 \leq \theta \leq \pi, 0\,\phi \leq 2\pi \tag{1}$$

where n can be integer or non-integer values. The directivity of an antenna with pattern represented by Eq. 1 can be determined in a closed form using the definition with known n. However, an approximate formula for directivity calculation using HFBW will be more convenient for practical design.

McDonald (1978) derived a formula of antenna directivity based on the array factor of a broadside collinear array; it is given by

DOI 10.1007/978-981-4560-75-7_52-1

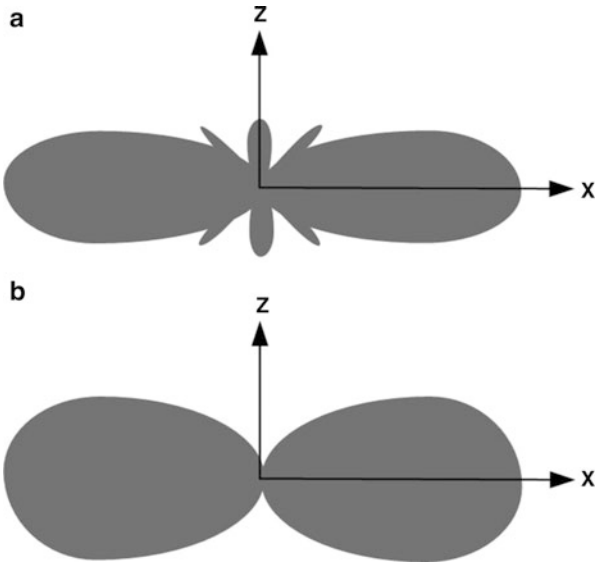


Fig. 3 Omnidirectional patterns with/without minor lobes; (**a**) with minor lobes and (**b**) without minor lobes

$$D_0 \cong \frac{101}{HPBW(\text{degrees}) - 0.0027[HPBW(\text{degrees})]^2} \tag{2}$$

Pozar (1993) derived the formula of antenna directivity based on the exact values obtained using Eq. 1 and then represented the data in a closed form using curve fitting, and it is expressed as

$$D_0 \cong -172.4 + 191\sqrt{0.818 + 1/HPBW(\text{degrees})} \tag{3}$$

In general, the formula Eq. 3 should be more accurate for omnidirectional patterns with minor side lobes of very low intensity (ideally no minor lobes), while Eq. 2 should be more accurate for omnidirectional patterns without minor side lobes, as shown in Fig. 3a, b respectively.

Omnidirectionality

An omnidirectional radiation pattern of a circle can never be achieved for practical antenna designs. As shown in Fig. 4, the omnidirectional radiation pattern of a practical antenna generally exhibits asymmetry or ripples which caused by the asymmetrical radiating structure, the fabrication tolerance, the undesired antenna assembly, the measurement error, and so on.

Omnidirectionality, or roundness, an engineering term, is not defined by the IEEE Standard Definitions of Terms for Antennas but widely used in practical applications. It is a parameter to characterize the uniformity of the omnidirectional radiation in each direction and quantified by the ratio of the maximum to minimum gain of the antenna. An ideal omnidirectional antenna radiation pattern should be a circle with consistent amplitude at any direction so that the omnidirectionality or the roundness is zero dB since the gain of the antenna keeps unchanged along the angles. The engineering requirement of omnidirectionality for an omnidirectional antenna is typically from 6 dB (or ± 3 dB) to 2 dB (or ± 1.0 dB) or even 1 dB (or ± 0.5 dB) for specific applications. Figure 4 demonstrates an omnidirectional pattern where an omnidirectionality of 5 dB is exhibited.

DOI 10.1007/978-981-4560-75-7_52-1

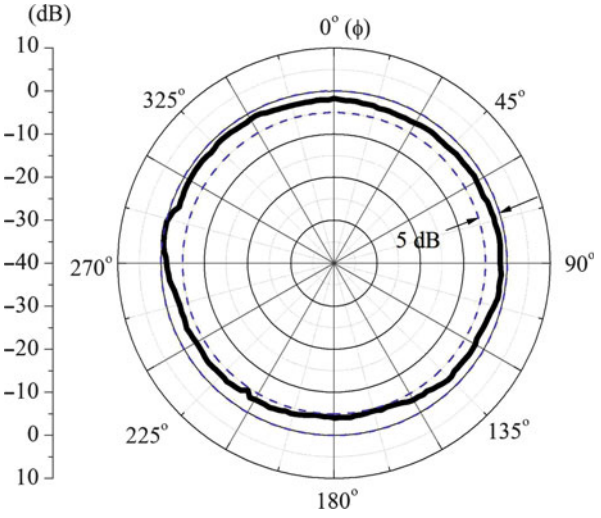


Fig. 4 The definition of the omnidirectionality

Configurations of Omnidirectional Antennas

Figure 5 shows the typical omnidirectional antenna configurations. The simplest way to realize omnidirectional radiation is to utilize a radiator with circularly symmetrically distributed current. A straight dipole is with such a current distribution and able to generate the omnidirectional radiation in the plane perpendicular to the radiator. An electrically small loop antenna exhibits similar performance wherein the current flows along the circular loop and produces omnidirectional radiation in the plane accordance with the loop.

Alternatively, the omnidirectional radiation can be realized by using multiple directional radiators. These directional radiating elements are normally positioned on the surface of a cylindrical structure or along a circle in a planar manner to generate a combined omnidirectional pattern (Knudsen 1956; Chu 1959; Jayakumar et al. 1986; Herscovici et al. 2001; Li et al. 2003; Wang et al. 2012). For multiple antenna element configurations, the positioning of the elements is the key factor for achieving a desired omnidirectional radiation pattern. In general, the omnidirectional antennas with one circularly symmetrical radiating element have better omnidirectionality than those with multiple radiating elements.

Design Considerations for Omnidirectional Antennas

The same as any other antennas, a desired omnidirectional antenna should be with specific radiating characteristics with size and cost constraints. Among the factors, the issues as following should be more important for an omnidirectional antenna design.

Polarization: The polarization is the first factor to be considered for omnidirectional antenna design, which is the starting point of the antenna configuration. A vertically positioned dipole is able to generate vertically polarized but horizontally polarized omnidirectional radiation, namely, a dipole can only produce the omnidirectional radiation pattern in *H*-planes instead of *E*-planes. To achieve horizontally polarized omnidirectional radiation, the electrically small loop antenna or multiple radiators have to be applied with proper arrangements. For dual linearly polarized or circularly polarized omnidirectional antenna, the configuration is more complicated so that orthogonal field components as well as phase delay should be considered.

Omnidirectionality: The omnidirectionality is somehow the most critical parameter for engineering applications. From a design point of view, a single radiator with circular symmetry is able to achieve desired omnidirectionality more easily than the multiple elements with combined omnidirectional radiation.

DOI 10.1007/978-981-4560-75-7_52-1

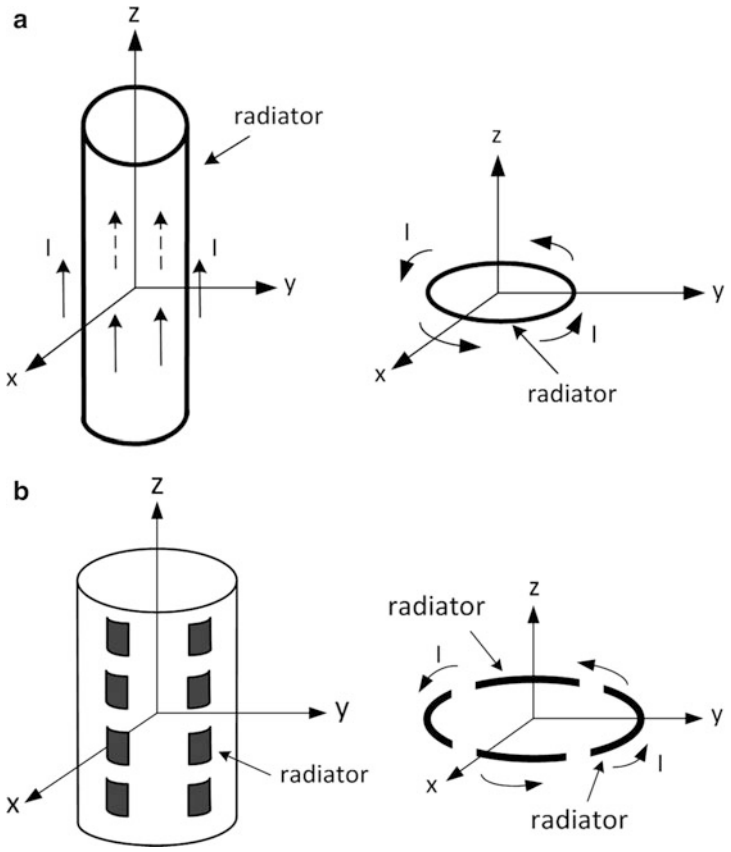


Fig. 5 Omnidirectional antenna configurations; (**a**) single radiator with circular symmetrical current distribution and (**b**) multiple radiators for omnidirectional radiation

Bandwidth: The bandwidth is another key factor for omnidirectional antennas, which can be extremely large for modern wireless systems for ultra-wideband or multiband operation. For example, an omnidirectional antenna for ultra-wideband radios is required to cover the frequency range of 3.1–10.6 GHz (109.5 %). The antenna for Long-Term Evolution (LTE) indoor access point is required to operate at a frequency band of 670 MHz to 2,690 MHz with a fractal bandwidth of 122.7 %. It is a big challenge to design such broadband antennas with constraints of omnidirectional radiation and polarization. Usually, a broadband vertically polarized omnidirectional antenna can be realized by enlarging the volume of a radiator. For a horizontally polarized omnidirectional antenna, the broadband operation can be achieved using specially configured radiator or multiple broadband elements such as tapered slots and log-period antennas.

Gain: Compared to a directional antenna, an omnidirectional antenna is normally with lower gain. To have omnidirectional antenna with higher gain, an antenna array is the most direct solution. For example, a dipole antenna exhibits gain of 2.15 dBi, while a 1×4 dipole antenna array is able to achieve gain of 6−7 dBi. In addition, using some specific configurations is able to achieve high gain as well; the omnidirectional antenna with a dual-shaped reflector exhibits the gain of more than 10 dBi.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_52-1

Vertically Polarized Omnidirectional Antenna

This section discusses a variety of vertically polarized omnidirectional antennas. When the antenna is positioned to generate vertically polarized omnidirectional radiation in a horizontal plane, or *H*-plane, the vertically polarized omnidirectional antenna is also named as an *H*-plane omnidirectional antenna.

Dipole-Based Omnidirectional Antennas

Straight Dipole

Dipoles are one of the fundamental classes of antennas that are evolved through many years. In fact, Hertz employed a dipole antenna on his spark-gap generator during the historical experiment in 1888. Being one of the fundamental types of radiators, it generates omnidirectional radiation because of the inherent circular symmetrical structure. The analysis and design principle of the dipole antenna have been fully carried out (Kraus and Marhefka 2008) and are not repeated here. It is worth to note that the shorter dipole antenna (overall length of the dipole is less than one operating wavelength) is with a main lobe only and maximum radiation in the horizontal plane ($\theta = 90^\circ$), while a longer dipole (overall length of the dipole is larger than one operating wavelength) is with multiple beams, as shown in Fig. 6.

Coaxial Collinear Dipole Antenna (CoCo Antenna)

The coaxial collinear (CoCo) antenna, introduced by H. A. Wheeler (1956), is a kind of high-gain omnidirectional antennas (Balsley and Ecklund 1972; Judasz and Balsley 1989). It radiates as a collinear array of wire dipoles driven in phase and provides a narrow broadside beam and an omnidirectional pattern in the plane perpendicular to the antenna axis. It is used both as an isolated antenna element and in large arrays. For its simple structure and easy fabrication, it has been used over the past few decades mostly in radar and communication systems.

As illustrated in Fig. 7, the CoCo antenna consists of a sequence of collinear sections of a coaxial cable that are half-wave long (in terms of the guided wavelength, λ_g), where the inner conductor of one coaxial cable section is electrically connected to the outer conductor of the adjacent one. The intent is to have the excitation voltages across all the slots in phase and equal in magnitude, so that a strong in-phase current distribution can be driven on the outer surface of the coaxial cable to obtain a high gain in the broadside direction. Therefore, every coaxial cable section acts as not only a transmission line but also a radiator.

Besides the coaxial line configuration, a microstrip version of the collinear dipole antenna was introduced by R. Bancroft and B. Bateman (2004). The geometry of the antenna is shown in Fig. 8, it consists of two metallic strips on the opposite sides of a printed circuit board (PCB). The strip on the top layer starts with a wide trace of width (W_1) and follows a narrow trace (W_2); the strips are with the same length, L. The strips alternate between a narrow section and a wide section until a wide section terminates the antenna on the bottom of the PCB. The strip on the bottom layer begins with a narrow strip and alternates from narrow to wide for complementing the wide to narrow variation of the strip on the top layer. The first and last strips on the bottom are shorted to the center of the corresponding wide strips on the top layer. The width (W_2) of the narrow strip is chosen such that it forms a 50-Ω microstrip line with the wide strip on the opposite side viewed as a ground plane. The width of the wide strip sections (W_1) is approximately five times of the narrow strip. The length L of each section is selected to be approximately $\lambda_g/2$, which is determined by the operating frequency, the dielectric constant, and the thickness of the PCB. The four-section antenna design example at 2.4 GHz band exhibits a simulated maximum gain of 6.4 dBi (measured gain of 4.6 dBi), side lobe levels of 11 dB, and VSWR ≤ 2 over a frequency band of 15.5 %.

DOI 10.1007/978-981-4560-75-7_52-1

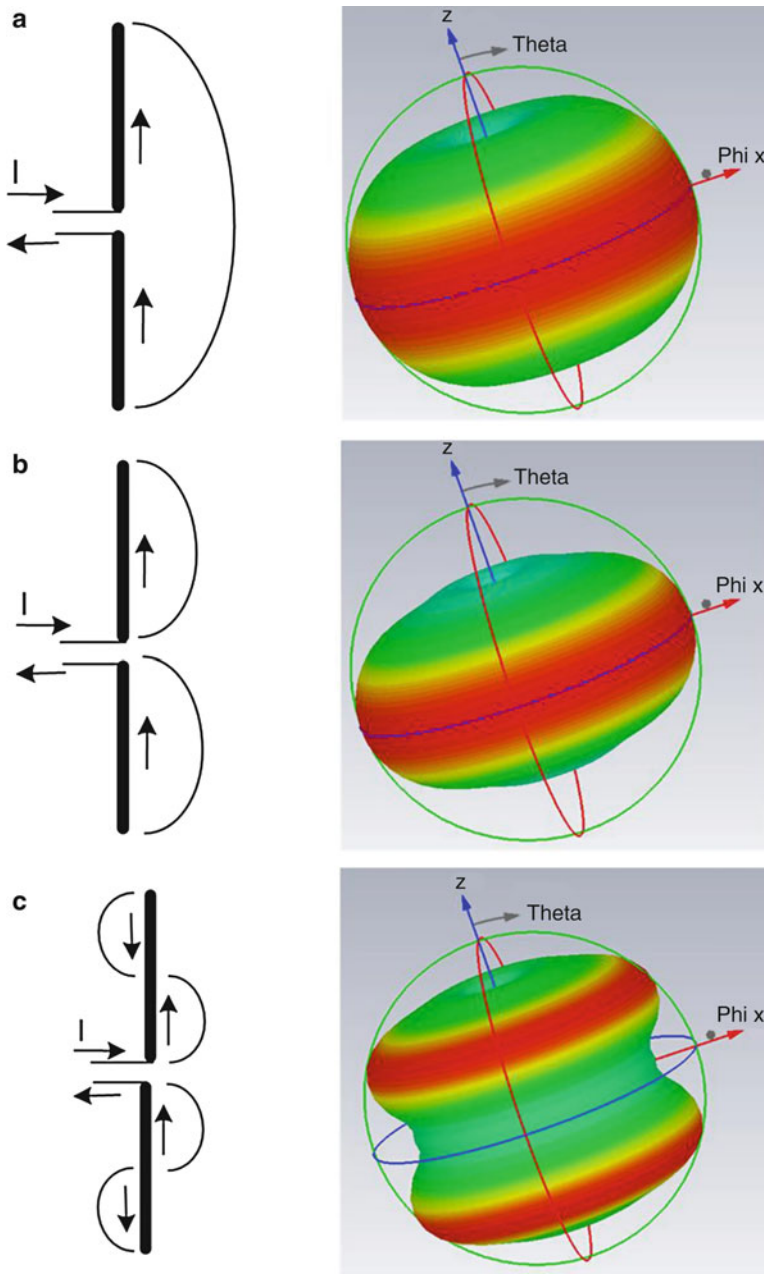


Fig. 6 Current distributions of the dipole antennas and three-dimensional radiation patterns; (**a**) half-wave dipole, (**b**) full-wave dipole, and (**c**) two-wave dipole

Biconical Antenna

Biconical antennas feature dipole-like characteristics, e.g., omnidirectional radiation in the *H*-plane, "eight figure"-shaped patterns in the *E*-plane, fixed phase center, and comparable gain, with an enormous wide bandwidth achieved by double cone elements. During the last decades, many biconical antennas have been developed with the frequency range from MHz to GHz and above with bandwidths of three or more octaves.

As shown in Fig. 9, a biconical antenna is formed by placing two metallic cones that extend opposite one another. It is typically hourglass shaped, as both the cones have a common axis. It can be thought to be a uniformly tapered transmission line (Barrow et al. 1939; Nagasawa and Matsuzuka 1988).

DOI 10.1007/978-981-4560-75-7_52-1

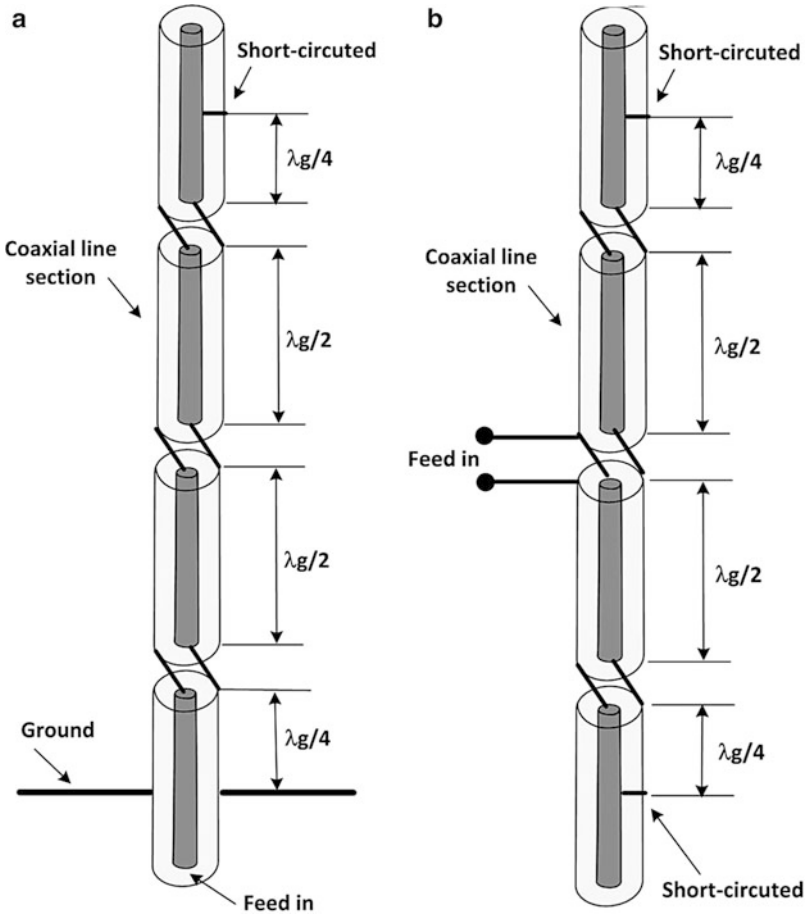


Fig. 7 Coaxial collinear dipole antennas: (**a**) bottom feed and (**b**) central feed

The input impedance of the infinite biconical antenna, Z_{in}, can be written as

$$Z_{in} = 120 \ln\left[\cot\left(\frac{\alpha}{4}\right)\right] \tag{4}$$

which is a pure resistance where α is a cone angle. For small cone angles α, Z_{in} can be expressed as

$$Z_{in} = \frac{\eta}{\pi} \ln\left[\cot\left(\frac{\alpha}{4}\right)\right] = \frac{\eta}{\pi} \ln\left[\frac{1}{\tan(\alpha/4)}\right] \cong \frac{\eta}{\pi} \ln\left[\frac{4}{\alpha}\right] \tag{5}$$

The input impedance of the biconical antenna is reasonably constant over a wide frequency range. The biconical antennas with small angles are not very practical (small α offers large Z_{in} so that the antenna may not matched with feed line easily), but wide-angle configurations ($30^\circ < \alpha/2 < 60^\circ$) are often used. Besides the cone angle, another key factor in practical antenna design is the spacing between the cones, in particular at higher frequencies. At microwave frequencies, it is a common practice to support the cones using a spacer; the property of the spacer may affect the impedance of the antenna as well. Fortunately, for practical antenna design, we do not need to calculate the impedance by all these formula; with aid of powerful electromagnetic simulation software, designers will optimize a reasonable cone angle for antennas.

DOI 10.1007/978-981-4560-75-7_52-1

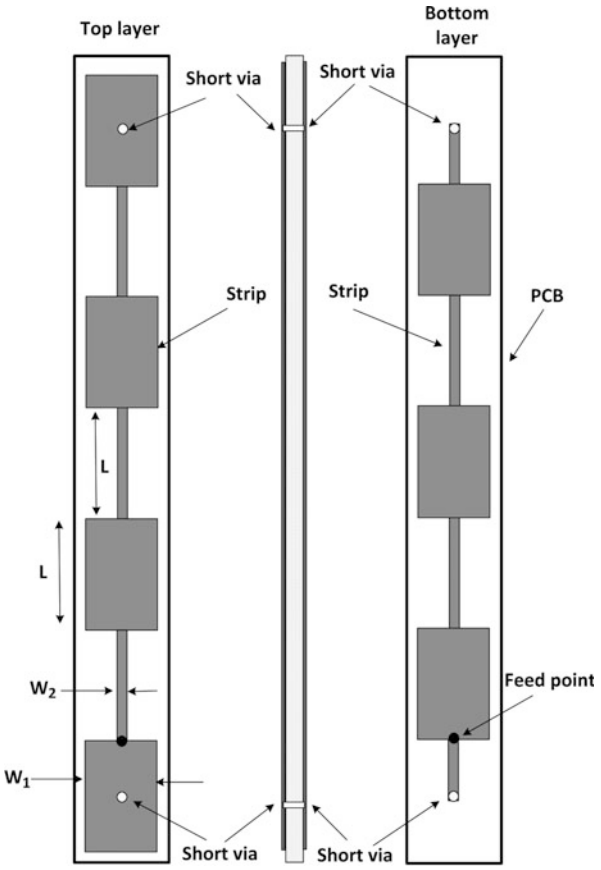


Fig. 8 Microstrip collinear dipole antenna

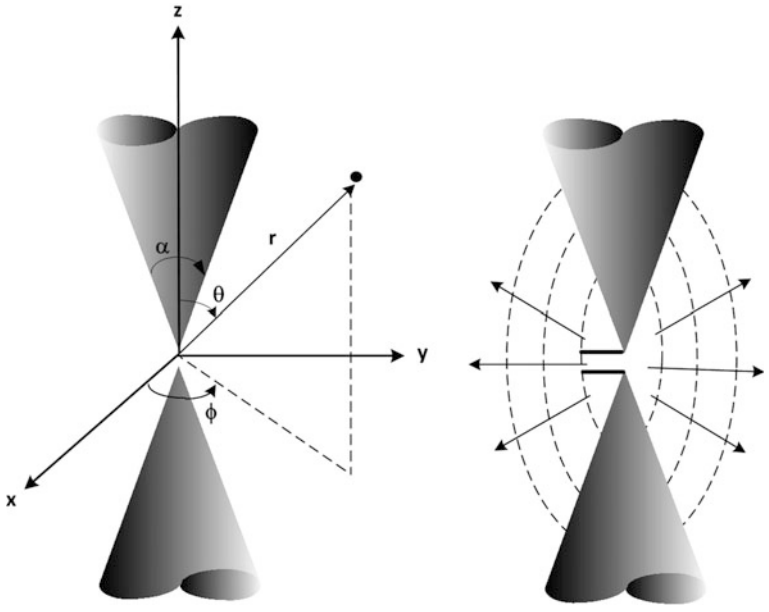


Fig. 9 Biconical antenna geometry and radiated spherical waves

DOI 10.1007/978-981-4560-75-7_52-1

Fig. 10 Different biconical antenna configurations

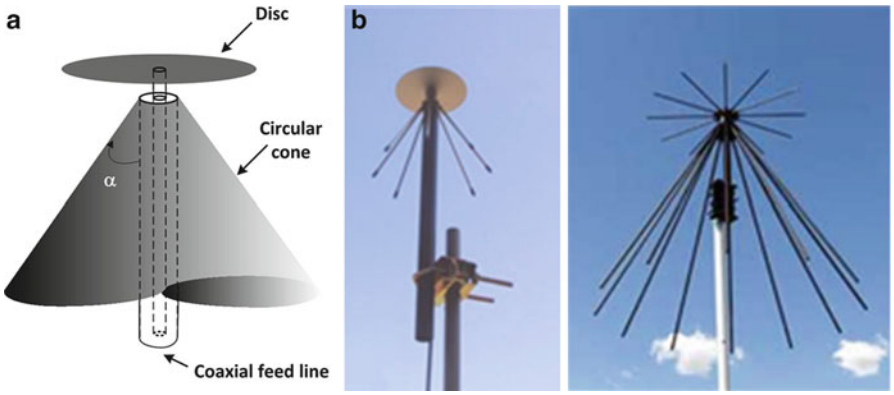


Fig. 11 Discone antenna; (**a**) antenna geometry and (**b**) practical discone antennas

The biconical antenna can be fed by a coaxial line directly or with a balun for bandwidth enhancement. Some practical biconical antennas are shown in Fig. 10; the cones can be formed by wires, the shape of the cone can be pyramid shaped, and the length of the cones can be different as well.

Discone Antenna

The discone antenna is a vertically polarized omnidirectional antenna with multi-octave bandwidth (Nagasawa and Matsuzuka 1988; Bergmann 2003; Kim et al. 2005; Qing et al. 2005; Chen et al. 2011). As illustrated in Fig. 11, a discone antenna is formed by a disk and a cone – that is where the name "discone" comes from – where the disk is on the top of the cone and perpendicular to the cone's axis. The discone antenna is easy to be fed by a coaxial line, the disk is attached to the inner conductor of the coaxial cable, and the cone is connected at its apex to the outer shield of the coaxial cable. The input impedance of the discone antenna is strongly dependent on the apex angle, as long as the cone is longer than about a quarter wavelength and the apex angle is relatively large. In general, the performance of discone antenna as a function of frequency is similar to a high-pass filter. Below a cutoff frequency, it becomes inefficient, and it produces severe standing waves in the feed line. At cutoff frequency, the slant height of the cone is approximately one quarter wavelength.

A discone antenna for ultra-wideband (UWB) applications is shown in Fig. 12 (Qing et al. 2005), and the performance of the antenna is exhibited in Fig. 13. The $|S_{11}|$ of less than -10 dB and gain of about 2 dBi (at $\theta = 90°$, $\phi = 0°$) are achieved over the frequency range of 3.0–11.0 GHz. The measured radiation patterns are shown in Fig. 13c; good omnidirectional patterns are observed in the *x*-*y* plane with the omnidirectionality of 3 dB.

DOI 10.1007/978-981-4560-75-7_52-1

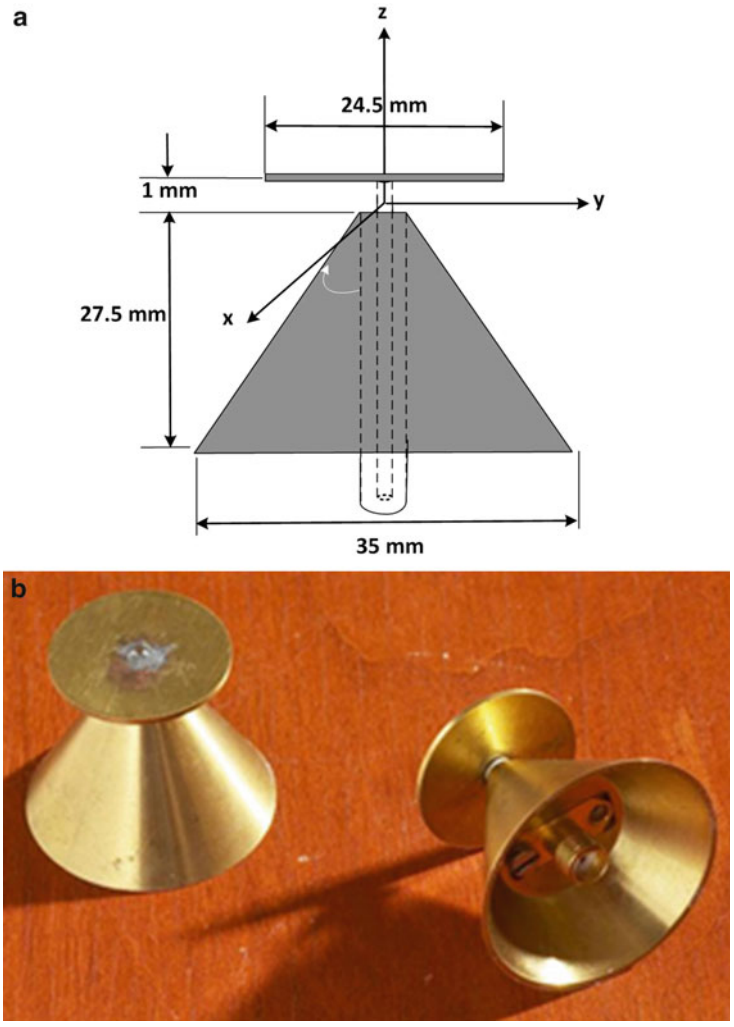


Fig. 12 Discone antenna for UWB applications; (**a**) detailed dimensions of the antenna and (**b**) antenna prototypes

Monopole-Based Omnidirectional Antenna

A conventional monopole antenna consists of a straight rod-shaped conductor, almost always mounted perpendicularly above some sort of ground planes. The monopole antenna was invented in 1895 by Guglielmo Marconi, who discovered if he attached one terminal of his transmitter to a long wire suspended in the air and the other to the Earth, he could transmit radio waves for longer distances. For this reason, it is sometimes called a *Marconi antenna*, although Alexander Popov independently invented it at nearly the same time.

As shown in Fig. 14, a monopole antenna can be visualized as being formed by replacing the bottom half of a vertical dipole antenna with a ground plane at right angle to the remaining half. From image theory, it is known that the radiation pattern of a monopole antenna with a perfectly conducting, infinite ground plane is identical to the top half of the pattern of a dipole with double length. Monopole antennas up to quarter wavelength long have a single "lobe," with field strength declining monotonically from a maximum in the horizontal direction, but longer monopoles have more complicated patterns with several conical "lobes" (radiation maxima) directed at angles into the sky. Because it radiates only into the space above the ground plane, or half the space of a dipole antenna, a monopole antenna will have a directivity of twice (3 dB over) the directivity of a full dipole antenna. In addition, the input impedance of the monopole antenna is one half of that of a full dipole antenna. For example, for a quarter-wave monopole

DOI 10.1007/978-981-4560-75-7_52-1

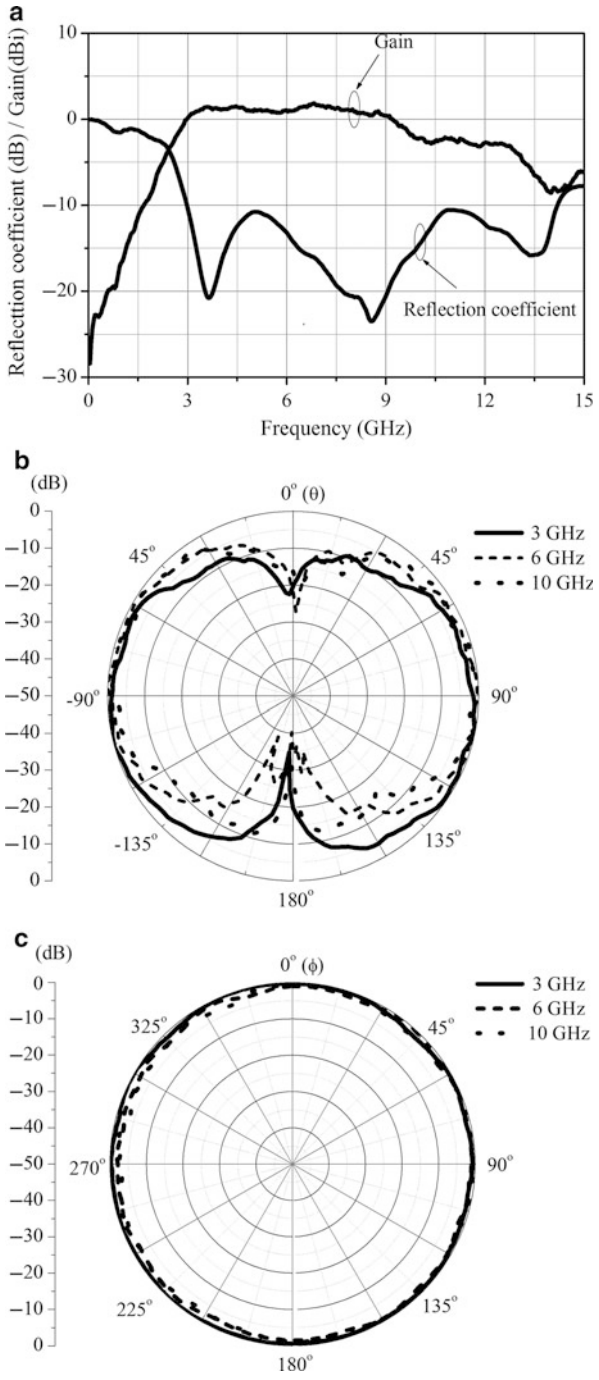


Fig. 13 Measured results of the UWB discone antenna; (**a**) reflection coefficient and gain, (**b**) normalized radiation pattern in the *x-z* plane, and (**c**) normalized radiation pattern in the *x-y* plane

antenna ($L = 0.25\lambda$) with an ideal infinite ground plane, its directivity will be 5.15 dBi and input impedance $Z_{in} = 36.5 + j21.25$ Ohms.

The ground plane used with a monopole may be the actual earth; in this case, the antenna is mounted on the ground, and one side of the feed line is connected to an earth ground at the base of the antenna. Such a design is used for the mast radiator antennas employed in radio broadcasting at low frequencies. At VHF and UHF frequencies, the physical size of the ground plane needed is smaller, so metallic plane is used to

DOI 10.1007/978-981-4560-75-7_52-1

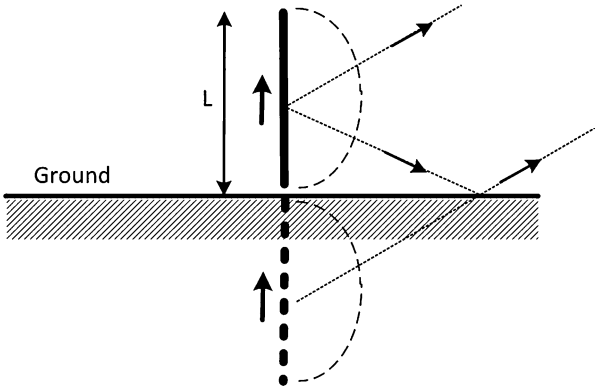


Fig. 14 Monopole analysis using image theory

allow the antenna to be mounted above it; the examples of such surfaces are the roof of a car or airplane body.

The general effect of the electrically small ground plane as well as imperfectly conducting earth ground on the performance of the monopole antenna is to tilt the direction of maximum radiation up to higher elevation angles. Figure 15 compares the radiation patterns of a quarter-wave monopole with different ground sizes.

Three-Dimensional Monopole Antennas

Unlike a dipole antenna, the monopole antenna is an unbalanced antenna so that it is very convenient to be fed using a coaxial cable. As shown in Fig. 16, the inner conductor of the coaxial connector such as subminiature version A (SMA) is connected to the lower end of the monopole and the other conductor of the SMA attached to the ground plane. Besides the basic rod shape, the radiator of the monopole antenna can be with different configurations, for example, conical cone, tapered cone, cross plate, parallel plate, and rolled plate. The radiator with larger volume exhibits the broader bandwidth over the rod-shaped one, wherein the tapered structure will provide better impedance matching (Taniguchi and Kobayashi 2002; Wong et al. 2005; Wong and Su 2005; Chen 2005). Note that all the configurations are able to offer broadband impedance matching, while only those antennas with circular symmetrical radiator can generate the desired omnidirectional radiation.

Printed Planar Monopole Antennas

A number of printed omnidirectional antennas have been reported, wherein the planar monopole and the ground are etched on a PCB. As exhibited in Fig. 17, such an antenna is composed of a radiator and a small ground plane and fed by a microstrip line or a coplanar waveguide (CPW). For microstrip-fed antenna, the radiator and the microstrip feed line are positioned on one side of the PCB while the ground plane is on the other side. For CPW-fed configuration, both the radiator and the CPW are on the same side of the PCB. The antennas are able to provide broad impedance and gain bandwidth while desired omnidirectional radiation, in particular, at high frequencies, wherein a bidirectional instead of an omnidirectional pattern is achieved. Changing the shape of the radiator is helpful to adjust the bandwidth of the antenna and realize some specific characteristic such as band-notch performance.

Slot Array Antenna

The vertically polarized omnidirectional radiation can be achieved by cutting transverse slot on the circular waveguide or coaxial line. Normally, a slot array instead of a single slot is applied for higher gain, where the slot elements are positioned with a spacing of guided wavelength (λ_g) to achieve the in-phase

DOI 10.1007/978-981-4560-75-7_52-1

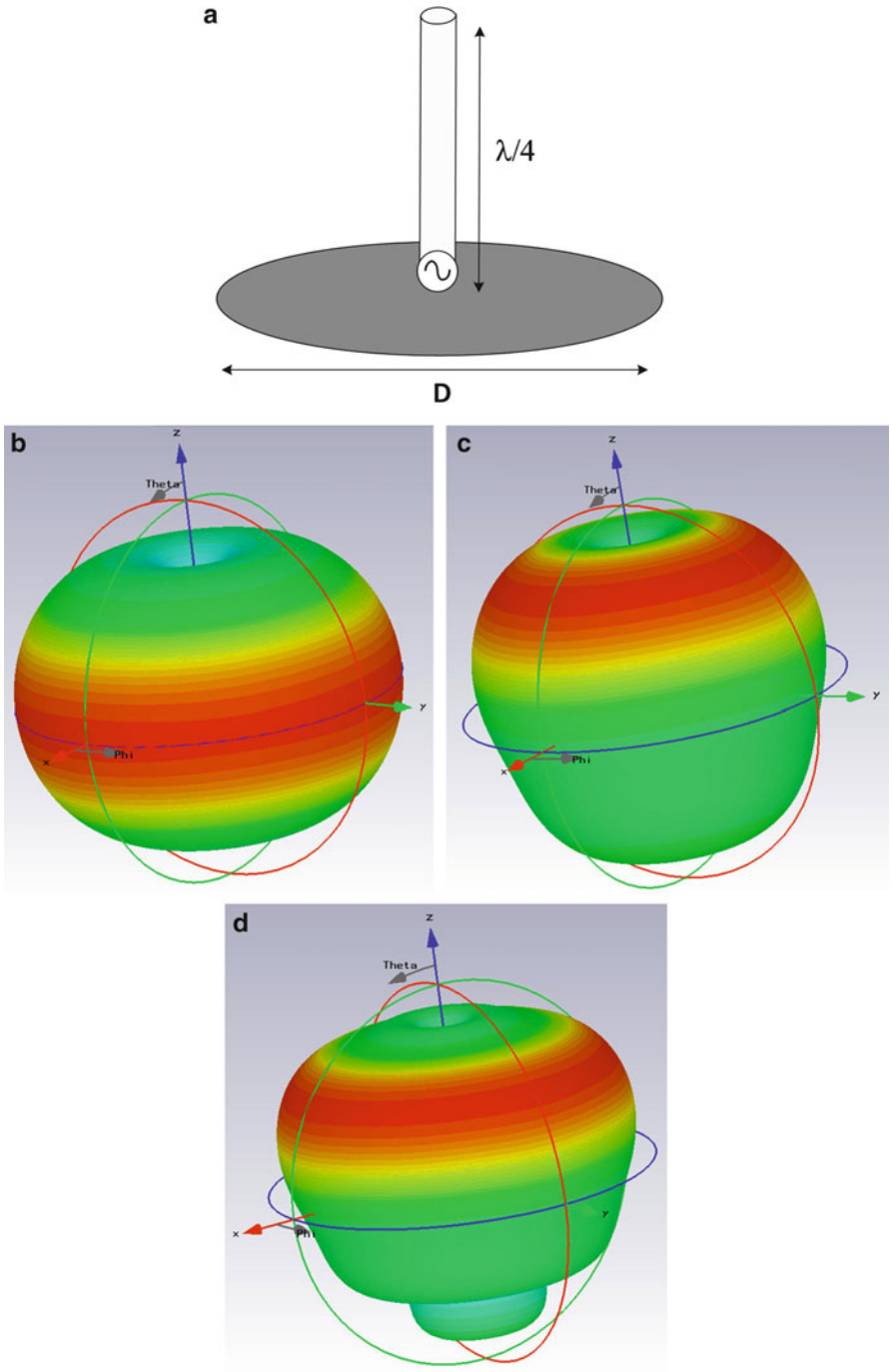


Fig. 15 Radiation patterns of a quarter-wave monopole with different sizes of ground planes; (**a**) antenna configuration, (**b**) $D = 0.25\lambda$, (**c**) $D = 1\lambda$, and (**d**) $D = 2\lambda$

excitation. A circular waveguide carrying the TM_{01} mode is suitable for implementing the slot array for omnidirectional radiation because of the circular symmetrical current distribution along the wall of the waveguide. One point to be noted is that the λ_g of a hollow circular waveguide carrying TM_{01} mode is larger than the free space wavelength (λ_0); the slot antenna array will have grating lobes. To avoid the grating lobes, λ_g should be less than λ_0. Filling dielectric into the hollow circular waveguide or utilization of the slow wave structure is suggested.

DOI 10.1007/978-981-4560-75-7_52-1

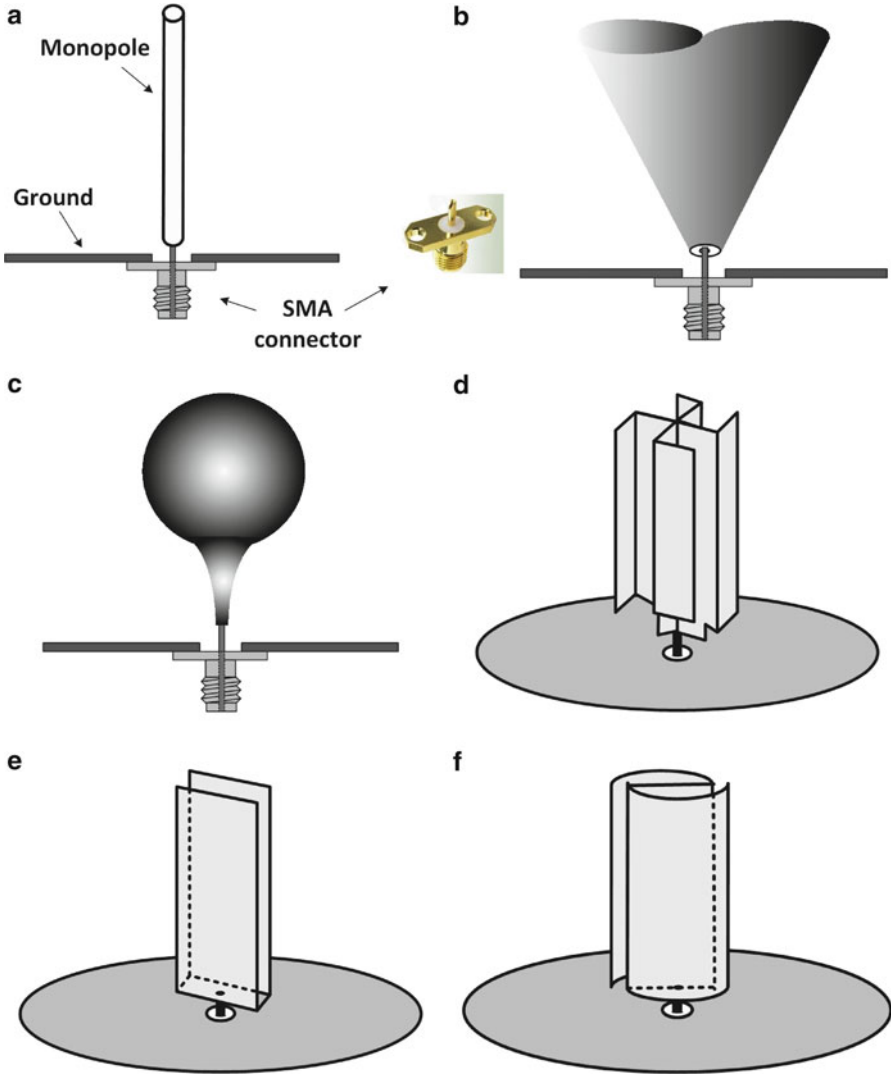


Fig. 16 Coaxial line fed monopole antenna and different three-dimensional radiator configurations; (**a**) rod-shape, (**b**) cone-shape, (**c**) taper-shape, (**d**) crossed plates, (**e**) parallel plates, and (**f**) rolled plates

Coaxial line is also suitable for omnidirectional slot array implementation. Baumer and Landstorfer (1990) demonstrated a slot array antenna as shown in Fig. 18. A slow wave structure is used as the inner conductor of the coaxial line; the diameter of the outer conductor is about a quarter of the free space wavelength at the center frequency. For implementation convenience, the nonresonant transverse slots are a little bit less than 360°. The eight-element slot array antenna is able to achieve desired omnidirectional radiation patterns in azimuth planes with the measured directivity of 10 dBi.

Chen et al. proposed a planar omnidirectional slot array antenna as shown in Fig. 19 (Chen et al. 2011). The antenna consists of a central conductor and two metal ground planes. The central conductor is embedded in a PCB and sandwiched in between two parallel ground planes to form a strip line, and one end of the central conductor is connected to the feeding coaxial cable and the other end open. A series of rectangular loop slots are etched on the top and bottom metal grounds. It is found that a nonuniform array with uneven loop slots having unequal distances between them is able to provide greater flexibility in antenna design and might achieve better performance compared to one with equal spaced slots with uniform size. The antenna is easily implemented using low-cost and precise planar printed technique. An antenna with eight back-to-back slots exhibits impedance bandwidth of 4.6 % ($|S_{11}| < -10$ dB) with

DOI 10.1007/978-981-4560-75-7_52-1

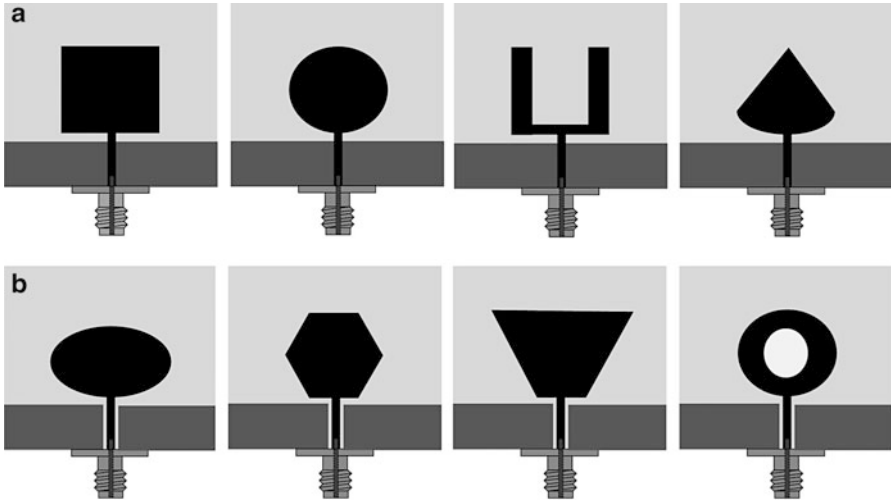


Fig. 17 Printed omnidirectional antennas; (**a**) microstrip-fed printed monopole and (**b**) coplanar waveguide-fed printed monopole antennas

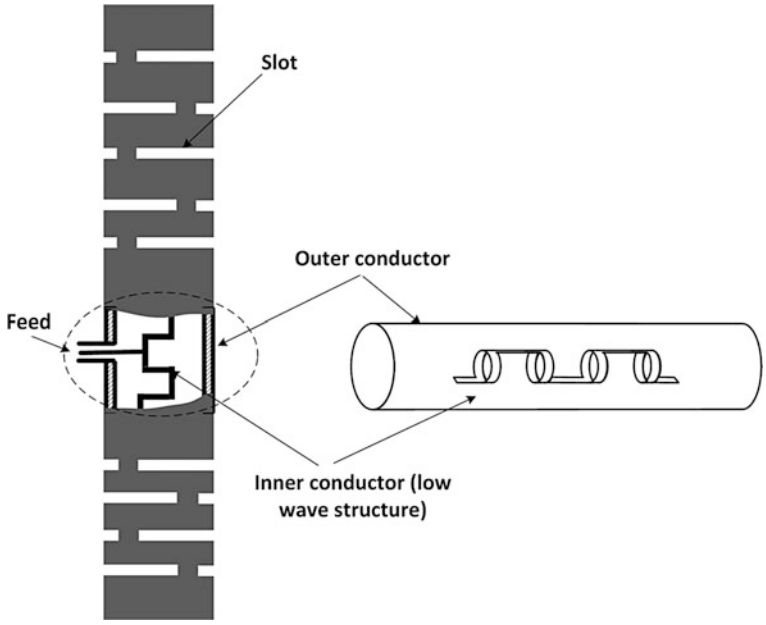


Fig. 18 Slot antenna array on coaxial line for vertically polarized omnidirectional radiation

center frequency of 5.8 GHz; it has a well-formed omnidirectional pattern with omnidirectionality of 0.5 dB in the horizontal plane, a gain of 10 dBi, side lobe levels of 10 dB below the main lobe, and no beam squint.

Dual-Reflector Omnidirectional Antenna

The coming generation of communication systems is expected to provide services through high-data-rate wireless channels. They will operate at millimeter-wave band up to 60 GHz, which requests the antenna offering the absolute operating frequency bandwidth of several gigahertz. The conventional rod-shaped dipole antenna and monopole antenna usually suffer from narrow operating bandwidth. The biconical

DOI 10.1007/978-981-4560-75-7_52-1

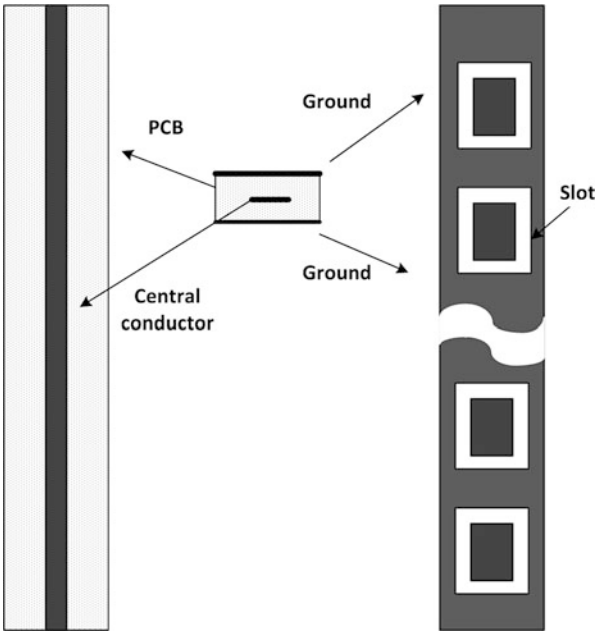


Fig. 19 Slot antenna array on strip line for vertically polarized omnidirectional radiation

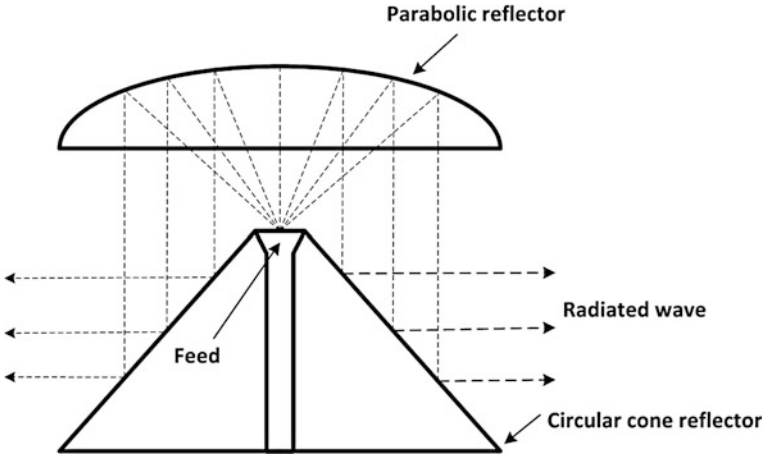


Fig. 20 Schematic diagram of dual-reflector omnidirectional antenna

antenna and discone antenna can offer broader operating bandwidths but lower gain and difficult implementation at millimeter frequencies.

The dual-reflector antenna (Willoughby and Heider 1959; Orefice and Pirinoli 1993) is a desired candidate for vertically polarized omnidirectional coverage at higher frequencies up to millimeter-wave bands. As shown in Fig. 20, an omnidirectional dual-reflector antenna is composed of two circularly symmetric reflectors with a common symmetry axis and fed by coaxial horn. The feed is located in the focus of the sub-reflector, a paraboloid. The field is firstly reflected by the paraboloid and then directed towards the main conical reflector; the plane wave is therefore eventually reflected in all the directions of the horizontal plane. The conical reflector can be shaped in order to point the maximum radiation to a specific elevation angle.

DOI 10.1007/978-981-4560-75-7_52-1

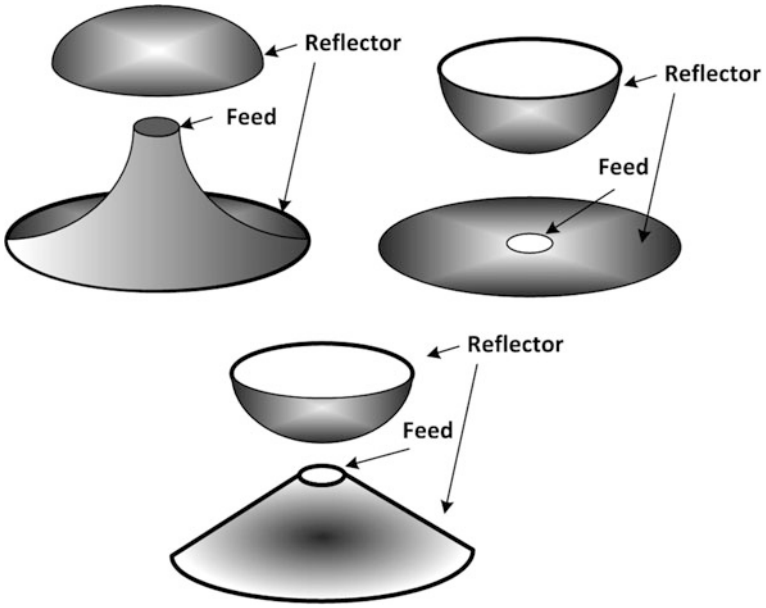


Fig. 21 Variations of the shape of the reflectors

A key feature of the antenna is the use of an unconventional feed for obtaining the vertical polarization. The feed is requested to radiate a field with only a θ-directed component and a null on the axis. A pure radial-mode aperture such as a coaxial horn or a TM_{01} excited conical horn is suitable for the antenna feed.

The configuration of the reflectors can be with different shapes as illustrated in Fig. 21, where the azimuth omnidirectional radiation is kept unchanged while the shape of the *E*-elevation radiation pattern as well as the title angle of the maximum radiation can be controlled by the profile of the reflectors (Pino et al. 2000; Bergmann and Moreira 2004; Silva and Bergmann 2005; Zang and Bergmann 2013).

Low-Profile Omnidirectional Antenna

A conventional monopole antenna with a quarter wavelength height is undesirable for today's wireless systems such as LTE and TV white space radios, where the operating frequency starts from a few hundred megahertz. Meanwhile, a low-profile antenna is always preferable, in particular, for indoor systems. The vertical height of a monopole antenna can also be reduced through antenna loading techniques using dielectric, inductive loading, and capacitive loading (Delaveaud et al. 1998; Foltz et al. 1998; Mclean et al. 1999; Liu et al. 2004; Heydari et al. 2009). Apart from the miniaturized monopole-type antennas, the other low-profile vertically polarized antennas with monopole-like omnidirectional radiation patterns include Goubau antenna (Goubau et al. 1982), a multielement antenna consisting of four electrically small vertical conductors, each one terminated in a conductive plate; quadripod kettle antenna (QKA) (Tokumaru 1976), comprising of multiple metallic plates; the spiral-mode microstrip (SMM) antenna (Wang et al. 2011), a traveling wave-based broadband antenna; and so on. These antennas are with a small antenna height of a twentieth operating wavelength.

Multielement Low-Profile Omnidirectional Antenna

A multi-plate omnidirectional antenna, namely, quadripod kettle antenna (QKA) as shown in Fig. 22, was presented by Shinobu Tokumaru in 1976 (Tokumaru 1976). The QKA is a self-standing purely multi-plate metallic structure that consists of a ground plane, a fed lower plate, and an upper plate. The lower metallic plate, in the form of cross with a central small square or circular patch, is fed by a probe at the center and shorted to the ground at the ends. The upper plate in the shape of a square or octagon/circle is connected to

DOI 10.1007/978-981-4560-75-7_52-1

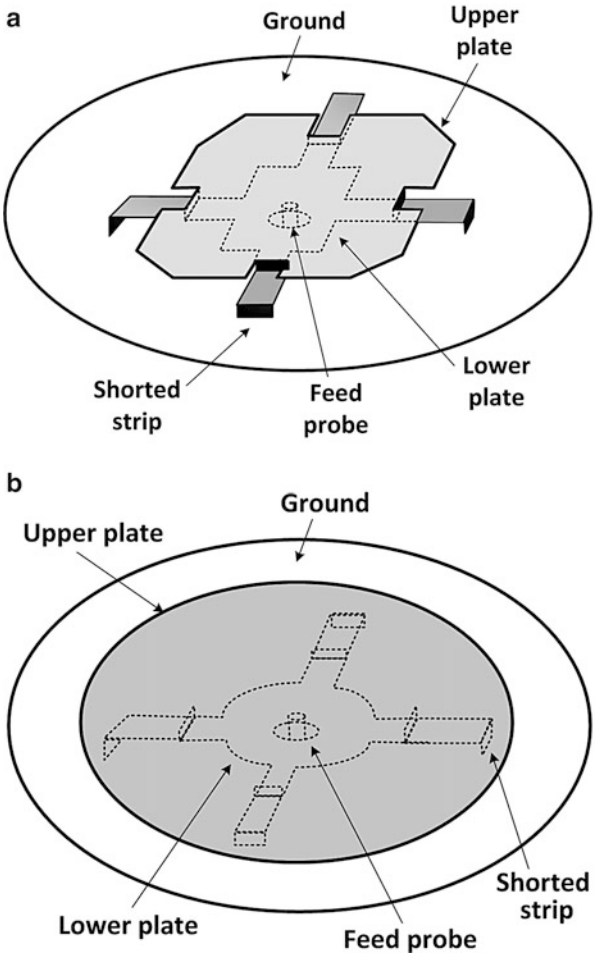


Fig. 22 Schematic diagram of quadripod kettle antenna (*QKA*), (**a**) octagon structure and (**b**) circular structure

the lower plate by means of another four pods at the edge. The antenna can be fed by a coaxial cable conveniently by connecting the external conductor of the coaxial cable to the ground plane and extending the inner conductor of the coaxial cable to the lower plate. The length of cross arms of the lower plate determines the lowest operating frequency, while the shape of the upper plate determines the impedance matching characteristic over the bandwidth. The QKA is an axially symmetrical antenna which is able to generate vertically polarized radiation similar to a monopole antenna with antenna height of less than 0.05 wavelength and bandwidth of more than 50 %.

Besides the quadripod configuration, the multi-plate antenna can be configured with tripod, hexapod, or octopod. Zurcher proposed a tripod kettle antenna (TKA) (Zurcher 2013), as shown in Fig. 23. The antenna designed at center frequency of 7.5 GHz exhibits an impedance bandwidth of 94 % ($|S_{11}| < -10$ dB) with an antenna height of 5.9 mm (0.079λ at lower edge frequency).

Another multielement low-profile vertically polarized omnidirectional antenna is shown in Fig. 24. The antenna features less than λ/20 in height and λ/5 or smaller in lateral dimension. Similar to the QTA antenna, the antenna is composed of a ground plane, a lower metallic structure with four ground ends and an upper metallic structure. Different from the QTA, the metallic plates are replaced by strips, and the upper plate is center fed by the probe together with the lower strips. Furthermore, the upper strips are open ended. The lateral dimension of the antenna is able to be miniaturized by meandering and turning the strips into form of a multi-arm spiral. The antenna is able to be matched to a 50-Ω coaxial line without the

DOI 10.1007/978-981-4560-75-7_52-1

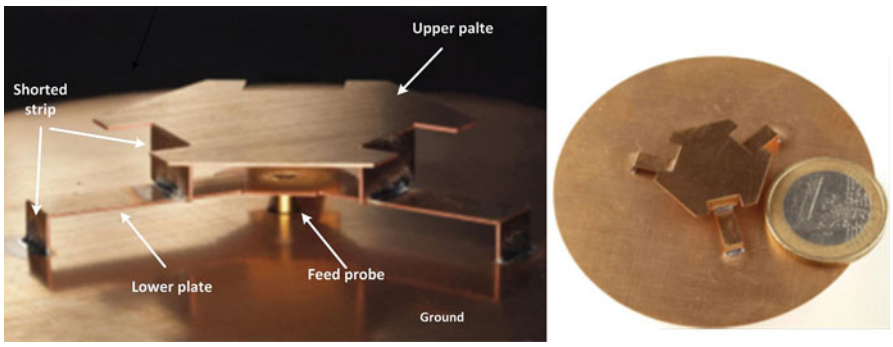


Fig. 23 Tripod kettle antenna

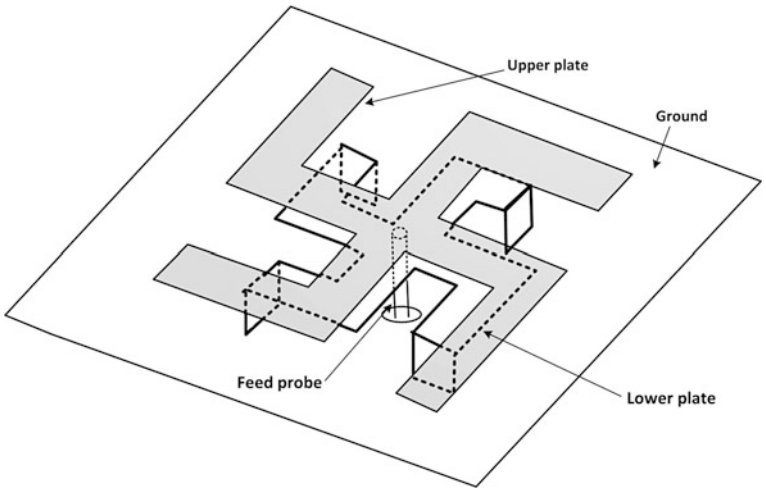


Fig. 24 Schematic diagram of miniaturized multielement monopole antennas (Hong and Sarabandi 2009)

need for external matching. The meandering between the short-circuited end and the feed point also facilitates the impedance matching.

Metamaterial-Based Low-Profile Omnidirectional Antenna

The artificial metamaterial structures can also be applied for omnidirectional antenna design. A low-profile omnidirectional zeroth-order resonator (ZOR) antenna using mushroom structures is exhibited in Fig. 25. The mushroom structure as a transmission line has negative, zero, and positive propagation constant depending on its operation frequency. Zeroth-order resonance, of which length is independent of the physical length, is naturally presented. In particular, mushroom ZOR antenna generates a uniform vertical electric field against a ground plane. The vertical electric field of mushroom ZOR is similar to that of short monopole antenna on a flat metal ground plane, so that it is able to generate the vertically polarized omnidirectional radiation in a horizontal plane. The resonant frequency of the mushroom ZOR and the number of unit cell can be determined from its dispersion curve.

Lee (Lee and Lee 2007) demonstrated an omnidirectional ZOR antenna at 7.8 GHz. The antenna is designed on a 1.57-mm thick RT/duroid 5880 PCB with dielectric constant of 2.2. The 3×2 mushroom array is with a unit cell size of 4.8×4.8 mm (patch), gap of 0.2 mm, and diameter of via of 0.3 mm. The low-profile antenna with a height of 0.04λ achieves the gain of 3.1 dBi and reflection coefficient of -15 dB at 7.8 GHz. Desired omnidirectional radiation pattern is achieved with the omnidirectionality of 2 dB.

DOI 10.1007/978-981-4560-75-7_52-1

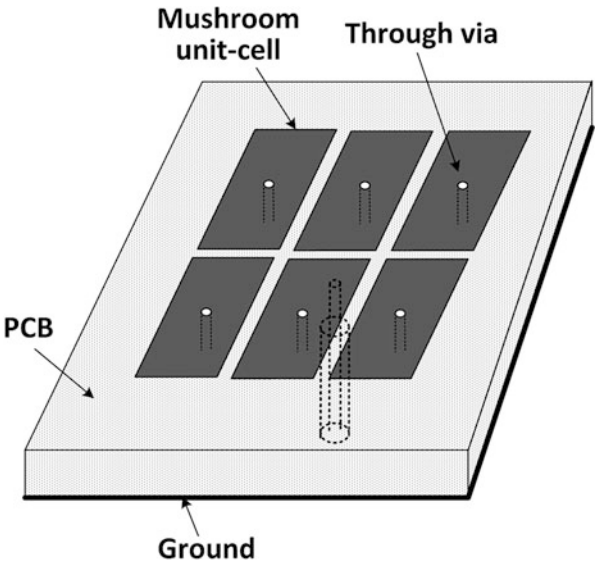


Fig. 25 ZOR antenna with coaxial probe feed

Horizontally Polarized Omnidirectional Antenna

In urban or indoor wireless environments, the polarization of propagating radio waves may change significantly because of the complicated multiple reflections or scatterings. Although vertically polarized antennas are used in many current wireless systems, it has been reported that using horizontally polarized antennas with both the transmitter and receiver can achieve a 10-dB improvement in terms of system gain as compared to vertically polarized antennas at both the ends of the link (Chizhik et al. 1998). In a third-generation (3G) network, the horizontally polarized omnidirectional antennas are often used together with vertically polarized omnidirectional antennas to form a multiple-input–multiple-output (MIMO) antenna system that provides polarization diversity in substitution for space diversity. Polarization diversity can be used to improve the reliability of a communication link, where the orthogonal polarization allows the frequency to be reused and isolation to be increased between the independent local area networks.

In contrast to the vertically polarized omnidirectional antenna which radiates equally in the plane aligning the *H*-field, the design of horizontally polarized omnidirectional antenna is more challenging since it is required to radiate equally in the plane aligning the *E*-field. The key consideration of such an antenna design is to form a structure with uniform and in-phase current distribution along a circle in the azimuth plane, either by using single loop or multielement configurations.

Loop-Based Omnidirectional Antenna

Electrically Small Loop Antenna

An electrically small solid-line loop is able to generate horizontally polarized omnidirectional radiation because of the current flowing along it features single direction and uniform distribution. However, such electrically small solid-line loop antenna has a very small radiation resistance and a large reactance, which makes the antenna very difficult to match with the excitation source. A solid-line loop antenna with perimeter comparable to one wavelength has a reasonable radiation resistance and reactance for impedance matching, while there is no longer omnidirectional radiation since the current flowing along the loop features phase inversion. In general, the small loop antenna having a circumference of less than a tenth wavelength is with uniform and in-phase current distribution. Figure 26 exhibits the radiation patterns of

DOI 10.1007/978-981-4560-75-7_52-1

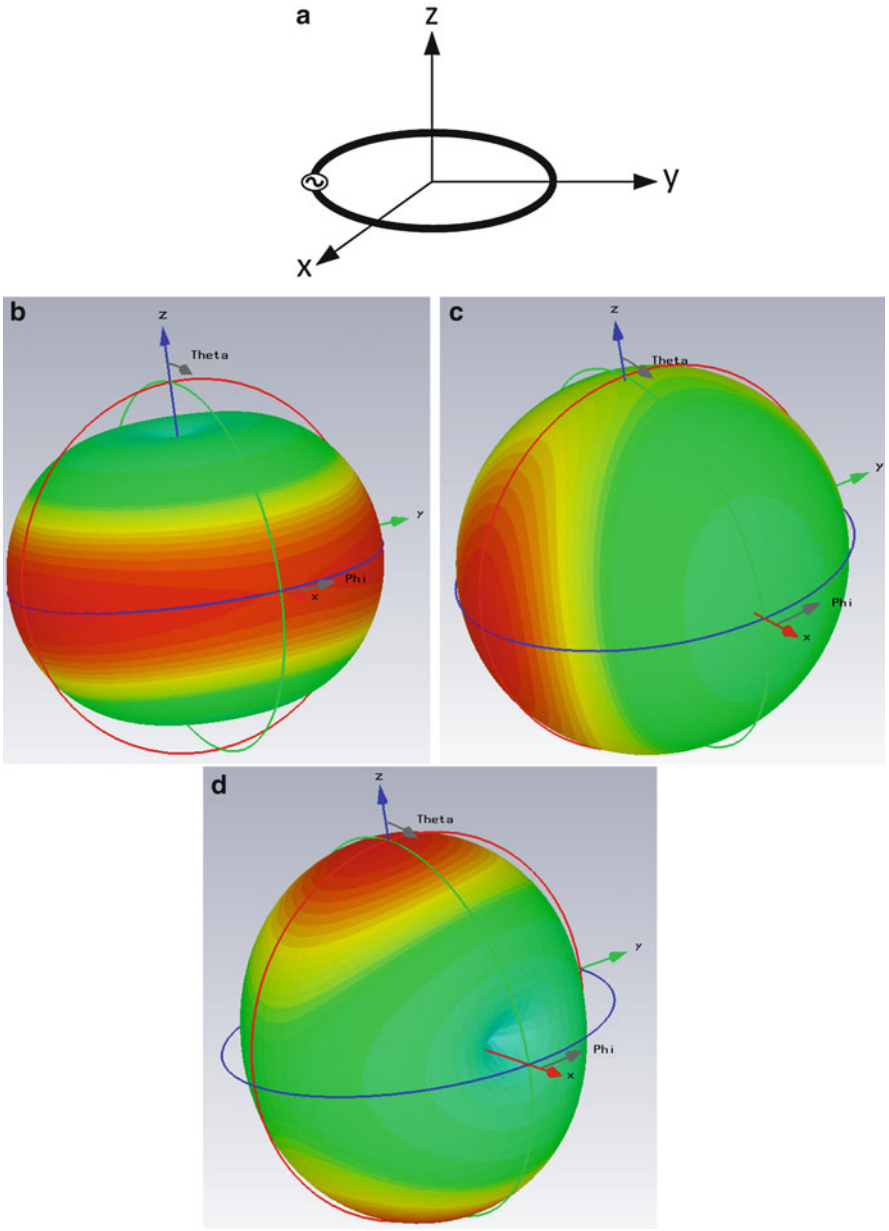


Fig. 26 Three-dimensional radiation patterns of the loop antenna with different circumference, (**a**) loop antenna with coordinate, (**b**–**d**) pattern with circumference of 0.1λ, 0.5λ, and 1λ, respectively

the loop antenna with different circumferences. It is observed that the loop antenna is able to achieve desired omnidirectional radiation when the circumference of the loop is less than 0.1λ. The loop antenna exhibits a bidirectional radiation when its circumference becomes electrically larger.

Alford Loop Antenna

The Alford loop antenna was presented by Alford and Kandoian in 1940 (Alford and Kandoian 1940). As exhibited in Fig. 27, it is composed of two Z-shaped metallic wires/strips, which are crossly positioned to form a square outline. The antenna is fed at the central portion of (K, K'); due to the symmetric structure, the current flowing on the wire sections AB and CD will have the same magnitude but opposite direction, which is the same as wire sections BC and AD. Furthermore, since the wire sections BB', DD', and AC

DOI 10.1007/978-981-4560-75-7_52-1

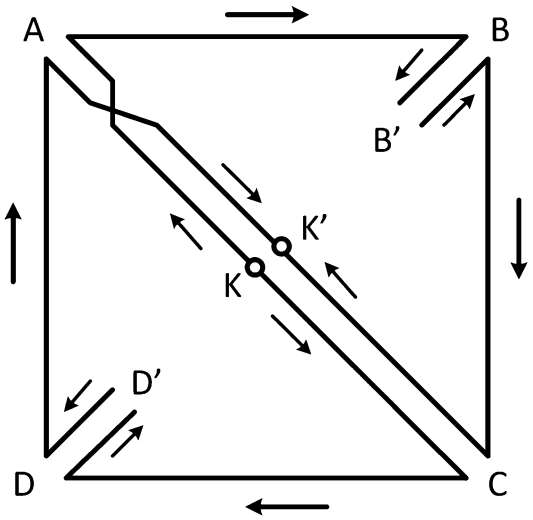


Fig. 27 Alford loop

are positioned very close to each other, the radiation from the opposite-directed current flowing along these sections cancel out each other. Therefore, the currents along the outline of the Alford loop form a square "loop"-type current distribution and achieve horizontally polarized omnidirectional radiation. In practical design, a circular instead of a square loop configuration is applied for better omnidirectional radiation performance. The circumference of the outline can be about one wavelength so that it is suitable for applications up to gigahertz with convenient implementation. A number of variations have been reported for different applications; some design examples are discussed in the following sections.

Printed Alford Loop Antenna

A printed Alford loop antenna as shown in Fig. 28a (Lin et al. 2006) was proposed for the integration with an external interface card, such as the personal computer memory card international association (PCMCIA) card and WLAN card for a laptop PC. The antenna consists of two Z-shaped strips printed on the top and bottom sides of a PCB. The bottom strip is arranged in such a manner that the "arm" is mapped to that of the top strip through the PCB substrate. A coaxial connector is used to connect the central feed point of the top and bottom strips. The "wing" length of the Alford loop is of the order of a quarter wavelength. Due to structure symmetry, the antenna current distribution on the two strips will have the same magnitude but opposite flowing directions. Since the thickness of the PCB substrate is very small, the radiation of the antenna current along the "arm" will cancel out each other. The antenna currents on the two "wings" of each Z-shaped strip establish a square "loop"-type current distribution and generate horizontally polarized omnidirectional radiation.

A windmill-shaped loop antenna (Kim et al. 2007) is exhibited in Fig. 28b. The antenna consists of two crossed Z-shaped strip lines etched on the opposite sides of a PCB. Each strip line has four arms with arc-shaped outer section which is of about one eighth wavelength. The arc-shaped outer sections on the upper side of the PCB are the mirror images of those on the bottom sides of the PCB or vice versa. Such an arrangement configures a combined loop antenna which consists of eight arc-shaped sections with a circumference of one wavelength. Feeding from the center of the antenna, the currents flowing on cross-shaped parallel strips are with opposite directions, while the currents flowing along the combined loop are with the same magnitude and single direction. Therefore, omnidirectional radiation is generated in the *E*-plane (*x*-*y* plane). An antenna using RT/duroid 5880 PCB with the dielectric constant of 2.2 mm and

DOI 10.1007/978-981-4560-75-7_52-1

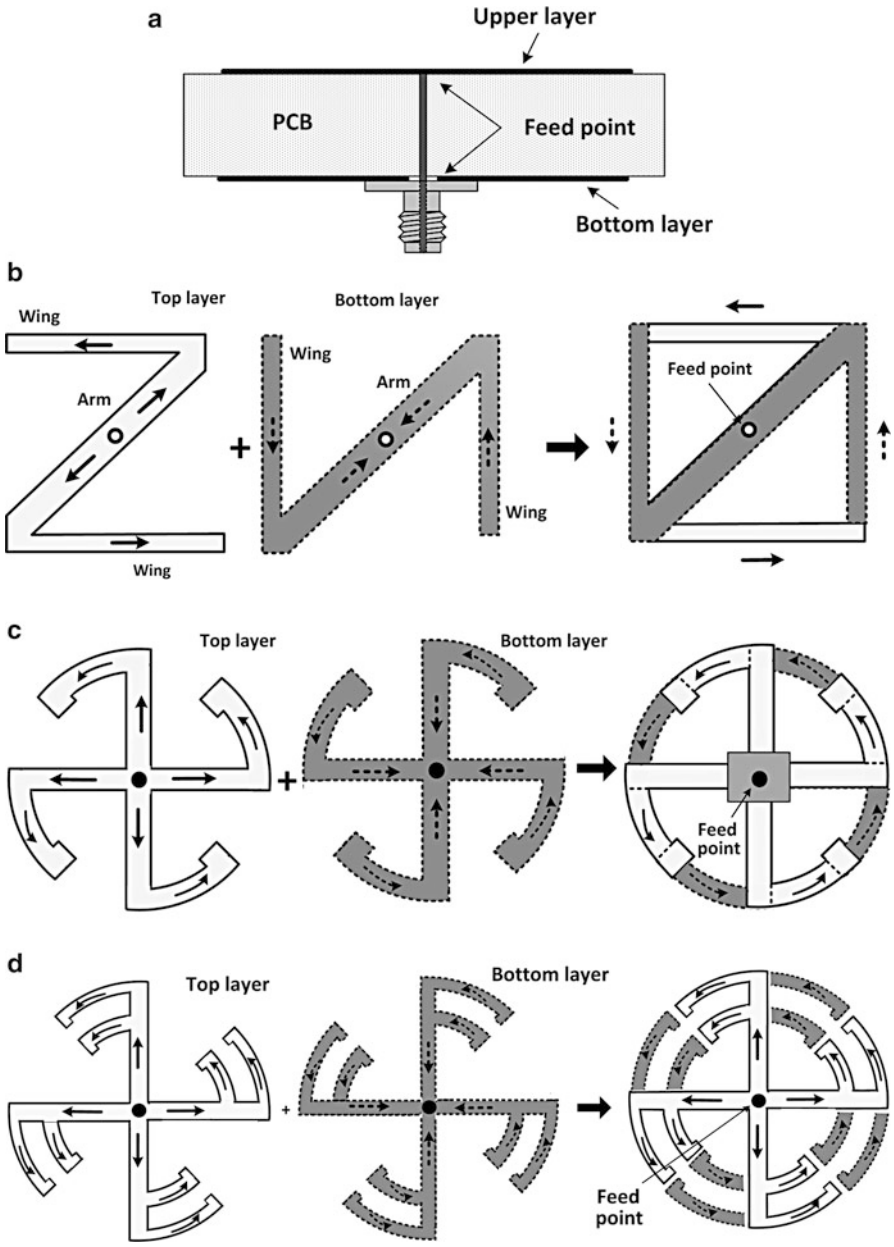


Fig. 28 Printed Alford loops, (**a**) PCB configuration, (**b**) square-shaped loop, (**c**) windmill-shaped loop, and (**d**) dual-band windmill-shaped loop

height of 1.6 mm operates at 2.6 GHz and achieves an impedance bandwidth of 6 % ($|S_{11}| < -10$ dB) and antenna gain of 1.5 dBi.

A dual-band windmill-shaped Alford loop antenna (Ahn et al. 2009) can be configured by adding one more set arc-shaped strips, as shown in Fig. 28d. The antenna is composed of two combined loops with different diameters, which resonant at different frequencies. The dual-frequency operation is achieved without any extra matching circuits or parasitic components. An antenna design has been exemplified at dual bands with the center frequencies of 2.45 and 3.9 GHz, respectively.

DOI 10.1007/978-981-4560-75-7_52-1

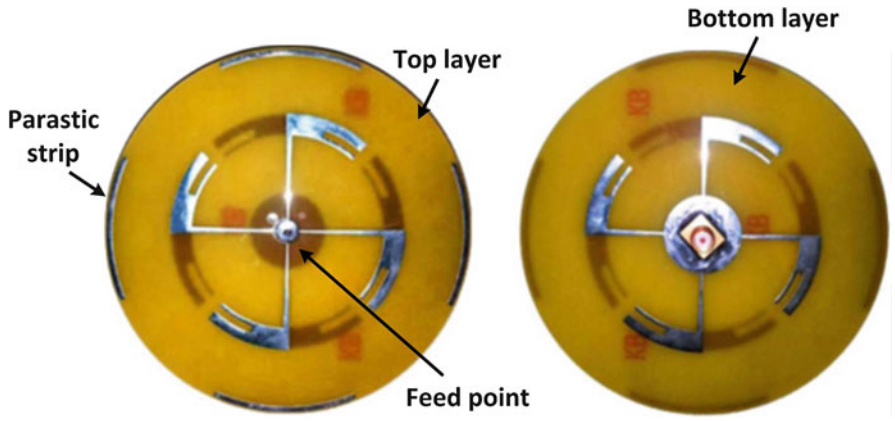


Fig. 29 Wideband Alford loop antenna for 4G LTE applications

Broadband Alford Loop Antenna

The traditional wire-type Alford loop antenna or printed strip-type Alford loop antenna is with limited bandwidth of less than 10 %, which is unable to meet the 4G LTE system requirements of high data throughput and long link range, where a peak speed of 100 Mb/s for high-mobility communication and 1 Gb/s for low-mobility communication is required.

Figure 29 shows a broadband Alford loop antenna (Yu et al. 2013). Similar to a conventional printed Alford loop antenna, the multiple arc-shaped strips on the opposite sides of a PCB are used to form the combined loop for horizontally polarized omnidirectional radiation. For bandwidth enhancement, tapered strip lines, parasitic strips, and a balun are applied.

The antenna consists of two pairs of arc-shaped radiators and a circular patch on the top and bottom sides of a PCB, respectively, as well as four parasitical strips on the top side. The four arc-shaped strips on the top layer are connected to a small circular conducting patch located at the center through four tapered strip lines, respectively. The end of the arc-shaped strip is notched to create two resonances that are close to each other. Four parasitical stripes are also printed on the top layer of the substrate to suppress the reactance of the antenna for bandwidth enhancement. The arrangement of the radiators on the bottom layer is similar to those on the top side, while the arc-shaped strips are positioned in a counter clockwise manner and connected to a larger circular conducting patch located at the center. The circular patches along with the tapered feeding lines constitute a balun for the coax feeding to the antenna. When excited, the quasi-TEM waves between two circular patches are guided to four pairs of arc-shaped radiators through the tapered strip lines. Due to the opposite-directed currents flowing on the straight strips on the opposite sides, the currents on the arc-shaped strips flow synchronously in clockwise or counterclockwise and form a circular combined loop and generate omnidirectional radiation. A practical antenna is able to operate from 1.76 to 2.68 GHz with a return loss greater than or equal to 10 dB, peak gain of 3.6–4.2 dBi, an average radiation efficiency of 83 %, and very good omnidirectional radiation pattern in *E*-plane over the entire impedance bandwidth.

Metamaterial-Based Zero-Phase-Shift Line Loop Antenna

The metamaterial-based zero-phase-shift line loop antennas have been attracted more attention recently for horizontally polarized omnidirectional antenna design. A transmission line with zero-phase-shift is ideal to configure an electrically large loop antenna for horizontally polarized omnidirectional since a uniform current distribution without phase inversion can be easily achieved. Two types of metamaterial-based transmission line, namely, the composite right-/left-handed transmission line (CRLH-TL)

DOI 10.1007/978-981-4560-75-7_52-1

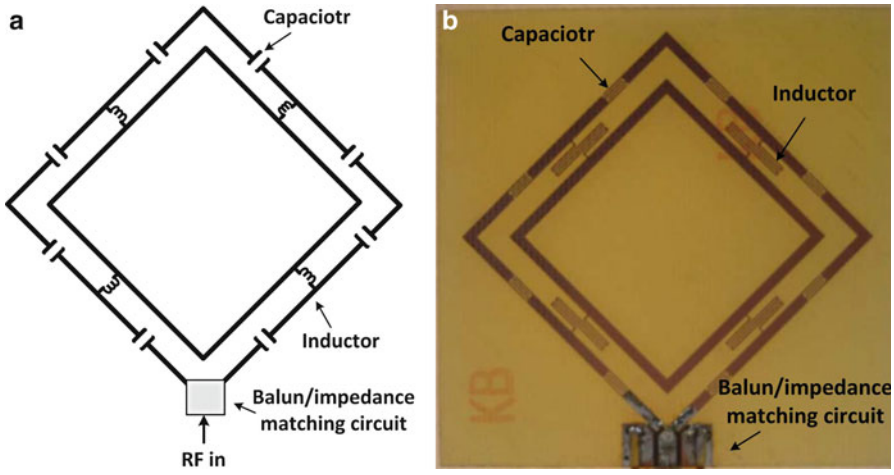


Fig. 30 CRLH line loop antennas, (**a**) antenna design with lumped components and (**b**) printed version

[16]–[18] and mu-negative transmission line (MNG-TL), have been utilized for designing the loop-type antennas for omnidirectional radiation in *E*-plane.

CRLH Line Loop Antenna

CRLH lines have been widely used for improving the performance of microwave devices (Sanada et al. 2003; Caloz et al. 2004) and leak wave antenna (Liu et al. 2002). These metamaterial-based transmission lines are characterized by the existence of three different spectral regions: the left-handed region at lower frequencies supporting a backward wave wherein a negative propagation constant β is observed ($\beta < 0$), right-handed region at higher frequencies supporting a forward wave wherein the propagation constant β is positive ($\beta > 0$), and a transition point wherein $\beta = 0$. The unique property of the zero propagation constant with nonzero group velocity at the zeroth-order resonance makes the CRLH lines suitable to form a loop antenna with omnidirectional radiation. Figure 30a exhibits a CRLH line loop antenna where surface-mounted lumped inductors and capacitors are used (Borja et al. 2007). The antenna exhibits *E*-plane omnidirectional radiation patterns but a low gain of 0.3 dBi at 500 MHz. The low gain attributes to the utilization of the lossy lumped components for CLRH line implementation. Figure 30b shows a revised design (Locatelli et al. 2012), where all the lumped components have been replaced by printed ones. The antenna demonstrates an omnidirectionality of 0.4 dB for *E*-plane pattern and maximum gain of 1.35 dBi at 2.4 GHz.

Segmented Line Loop Antenna

Segmented loop antennas have been investigated firstly for UHF radio-frequency identification (RFID) applications, where the reader antenna is required to generate strong and even magnetic field over a large interrogation zone while the conventional solid-line loop antenna cannot make it. Dobkin et al. firstly presented the segmented magnetic antenna consisting of a number of segments, and each segment is composed of a metal line and a series lumped capacitor (Dobkin et al. 2007). In this structure, segmenting and combining the parasitic inductance of each section with a lumped capacitor make the electrically large loop keeping the uniform current flowing in the same direction and generating strong magnetic field. Based on the method, several segmented loop antennas (Qing et al. 2009; Ong et al. 2010) have been presented; the lumped capacitor is replaced by either couple lines or distributed capacitors as shown in Fig. 31.

Since the electrically large segmented line loop supports the uniform and uni-directed current distribution, it has been applied to design the horizontally polarized omnidirectional antenna for 2.4 GHz and

DOI 10.1007/978-981-4560-75-7_52-1

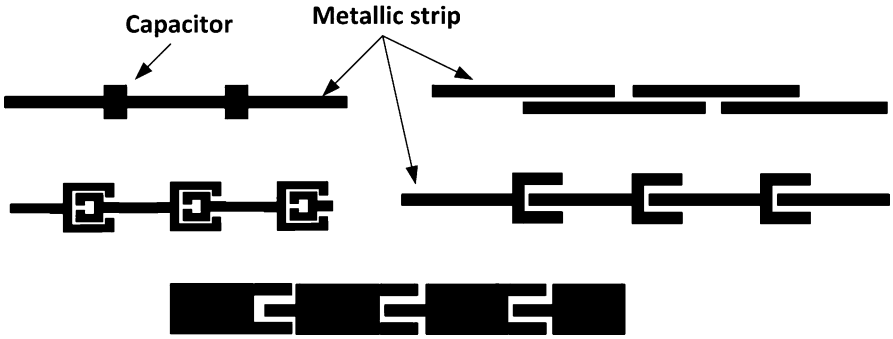


Fig. 31 Configurations of the segmented lines

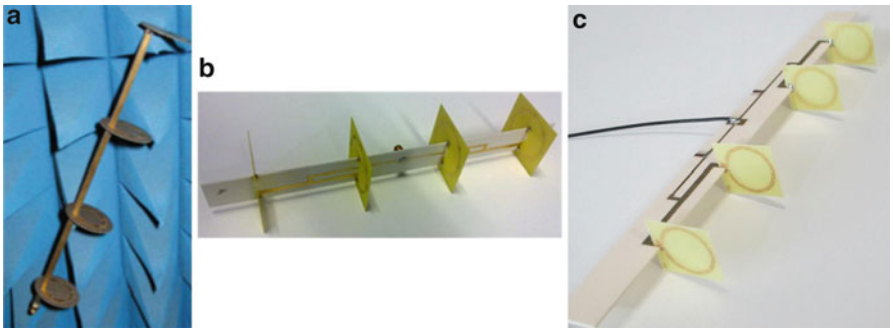


Fig. 32 Segmented line loop antenna arrays; (**a**) 2.4-GHz four-element series-fed MNT-TL loop antenna array (Wei et al. 2012), (**b**) 2.4-GHz four-element corporate-fed segmented line loop antenna array (Qing and Chen 2013), and (**c**) 5-GHz four-element corporate-fed segmented line loop antenna array (Qing and Chen 2014)

5 GHz WLANs (Qing and Chen 2012; Hasse et al. 2012; Wei et al. 2012). Wei et al. propose a circuit model to analyze the segmented line with periodically loaded parallel-plate lines (Wei et al. 2012). It is found that the effective permeability of the segmented line-based unit cell is negative, zero, and positive. When the transmission line operates at the mu-zero frequency, it has a unique property that supports a zero propagation constant with nonzero group velocity. At the zeroth-order resonance, there is no phase shift across the resonator so that the current distribution along the loop remains in phase and could yield a desired horizontally polarized omnidirectional pattern. Note that the general theoretical analysis of the segmented line is still a challenge till today since no valid transmission line mode can be applied to a single line, so that such an antenna design is mainly carried out with the help of simulation tools.

Compared to Alford loop antennas which must be fed from the center of the antenna, the side fed segmented line loop antenna is more convent to configure an antenna array for higher gain. Figure 32 shows some horizontally polarized omnidirectional segmented line loop arrays for the WLAN applications at 2.4 GHz and 5 GHz bands with series feeding or corporate feeding network.

Figures 33 and 34 demonstrate the detailed configuration and performance of the antenna array shown in Fig. 32c. The four-element antenna array is designed for 5-GHz WALN applications. The segmented line loop is printed onto a piece of FR4 PCB slab ($\varepsilon_r = 4.4$, $\tan\delta = 0.02$) with thickness of 0.5 mm. The four elements are connected to the outputs of the parallel-line feeding network, respectively. The parallel-line feeding network is etched on the opposite sides of a 0.8-mm thick RO4003 PCB ($\varepsilon_r = 3.38$, $\tan\delta = 0.0023$). Two open-circuited stubs are used to enhance the impedance matching of the antenna array. The antenna array exhibits wideband characteristics: reflection coefficient of less than −10 dB, gain

DOI 10.1007/978-981-4560-75-7_52-1

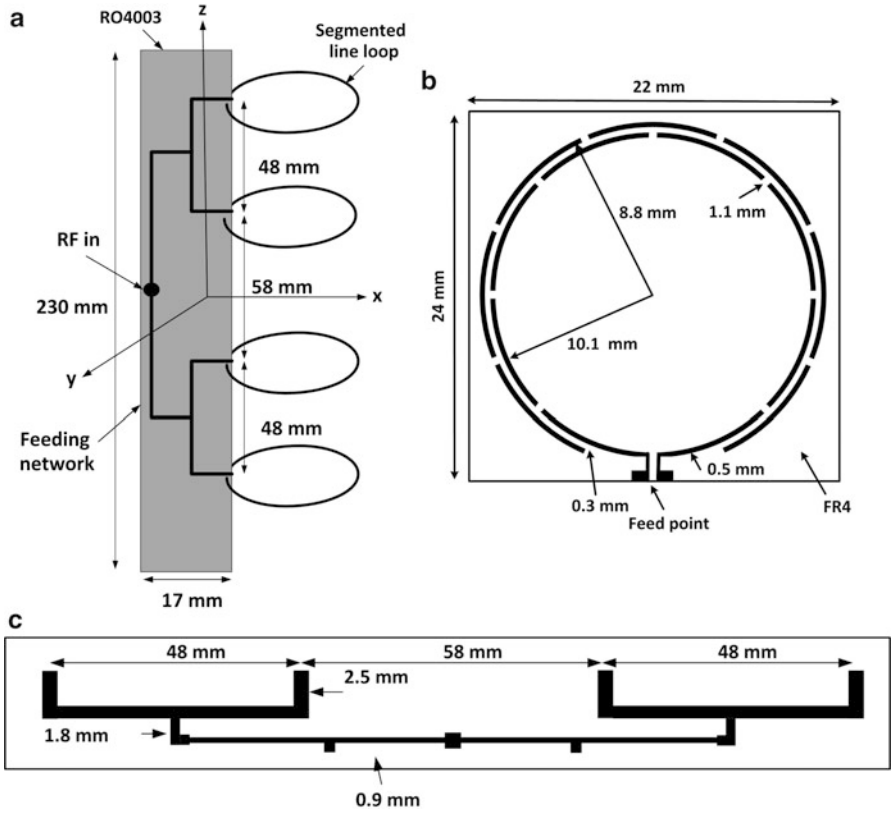


Fig. 33 Segmented line loop antenna array for WLAN applications, (**a**) array configuration, (**b**) segmented line loop element, and (**c**) feeding network

of greater than 5 dBi, omnidirectional *E*-plane radiation patterns with omnidirectionality of less than 5 dB over the frequency range of 5.05–5.9 GHz, which is desired for WALN mesh networks.

Waveguide Slot Array Antenna

Similar to vertically polarized omnidirectional slot antennas, the waveguide slot array is able to achieve the horizontally polarized omnidirectional radiation as well. A circular waveguide carrying rotationally symmetric mode TM_{01} or coaxial line with TEM mode is suitable to configure such antennas. Figure 35 shows the basic configuration of the horizontally polarized omnidirectional antenna using circular waveguide and rectangular waveguide (Sangster and Wang 1995; Grabherr and Huder 1999; Kyouichi and Tanaka 2000).

Generally, the antenna contains a number of rows of slots in its axial direction. Each row consists of several equally spaced identical half-wave slots arranged on the circumference of the waveguide. It has been found that eight slots are adequate to achieve the desired omnidirectional radiation in the azimuth planes. Because the currents on the wall of the circular waveguide carrying TM_{01} mode are axially directed, the slots must be inclined with a certain angle for efficient coupling from the waveguide to free space. The radiation pattern in the elevation plane is determined by the number and distance of the rows of slots as well as by the shape and the inclination angle of the slots themselves. Succeeding rows are generally positioned with a distance of half wavelength of the TM_{01} mode that propagates in the circular waveguide. To realize a stable broadside omnidirectional radiation, the waveguide is short circuited at quarter wavelength behind the last row of slots, which makes the antenna to be a resonant array. The inner diameter of the circular waveguide has to be chosen sufficiently small in order to allow the propagation of the TM_{01} mode, but below the cutoff for the next higher TM_{11} mode. On the other hand, the wavelength of

DOI 10.1007/978-981-4560-75-7_52-1

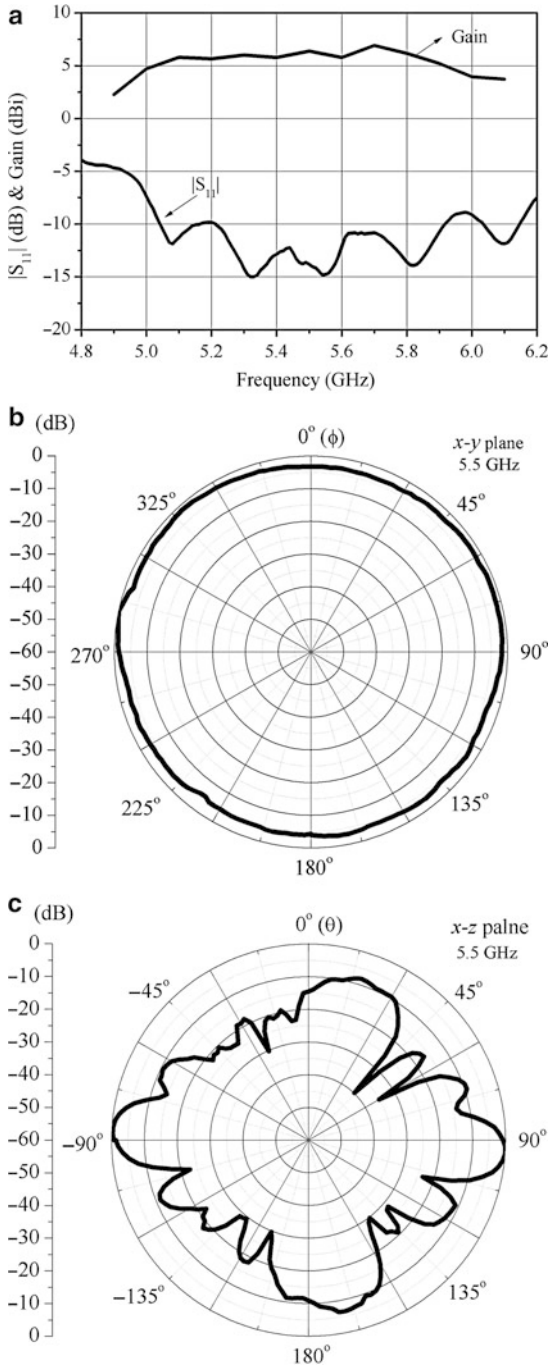


Fig. 34 Measured results of the 5-GHz segmented line loop array; (**a**) $|S_{11}|$ and gain, (**b**) normalized radiation pattern in *E*-plane, and (**c**) normalized radiation pattern in *H*-plane

the TM_{01} mode in the waveguide and the distance of successive rows of slots increase with decreasing its inner diameter. To avoid grating lobes in the radiation pattern of the elevation plane, the distance of the rows must be smaller than one free space wavelength. In general, the minimum inner diameter of the circular waveguide can be selected to be one wavelength in free space at the central operating frequency. For the slot array on the circular waveguide carrying TE_{11} mode, the slots are not needed to be inclined because the currents on the wall of the circular waveguide are ϕ directed.

DOI 10.1007/978-981-4560-75-7_52-1

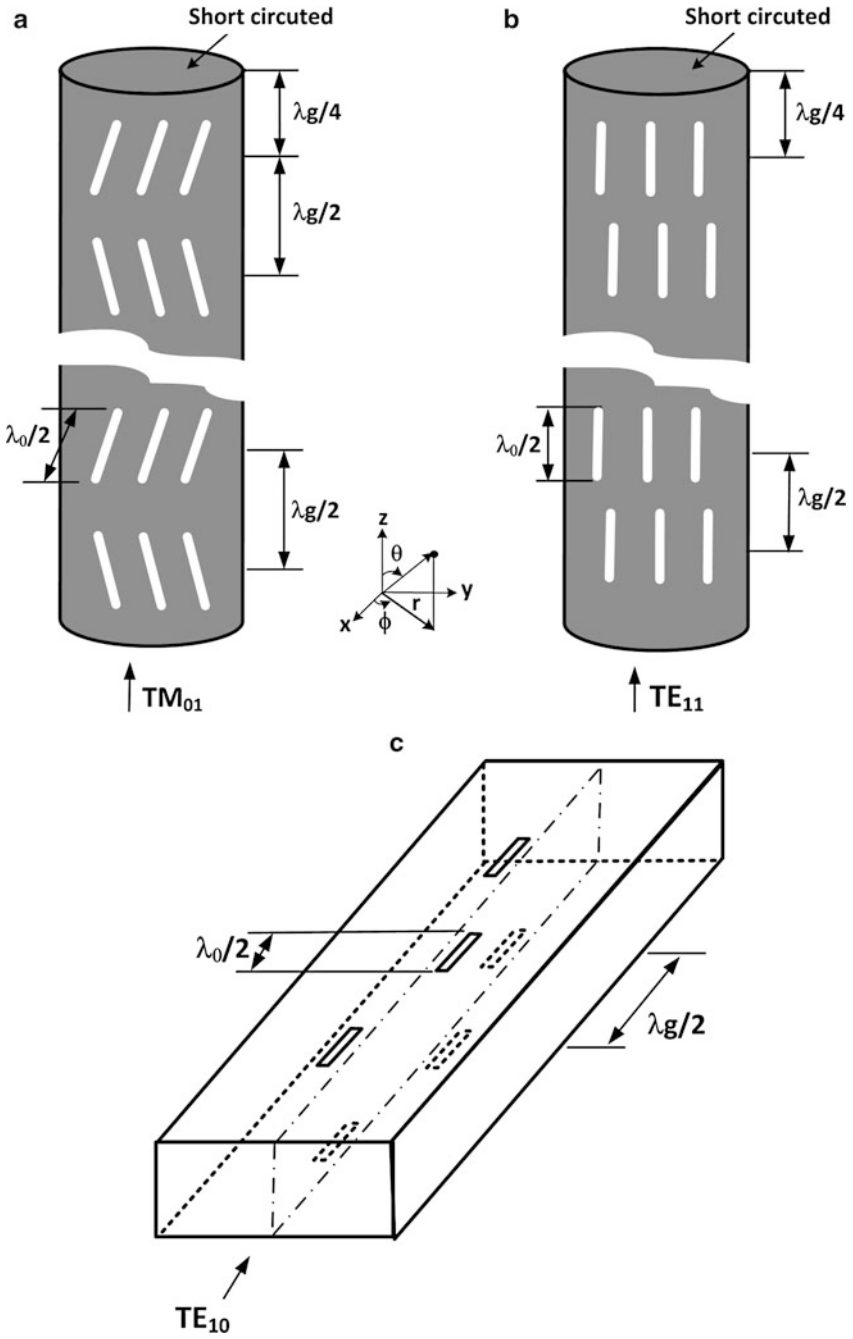


Fig. 35 Waveguide slot arrays, (**a**) slot array on circular waveguide carrying TM_{01} mode, (**b**) slot array on circular waveguide carrying TE_{11} mode, and (**c**) slot array on rectangular waveguide carrying TE_{10} mode

The rectangular waveguide carrying TE_{10} mode can be used to configure the slot array for horizontally polarized radiation as well. Furthermore, the longitudinal slot array features very low cross-polarization levels. As shown in Fig. 35c, the longitudinal slots are cut on both the broad walls of the waveguide to achieve desired omnidirectional radiation. The resonant slots are positioned at the opposite side of the central line of the waveguide alternatively with half a guided wavelength of the TE_{10} mode in the waveguide. Such an arrangement counteracts the 180° phase shift of the waveguide mode at each half-guided wavelength and ensures the slots with in-phase excitation.

DOI 10.1007/978-981-4560-75-7_52-1

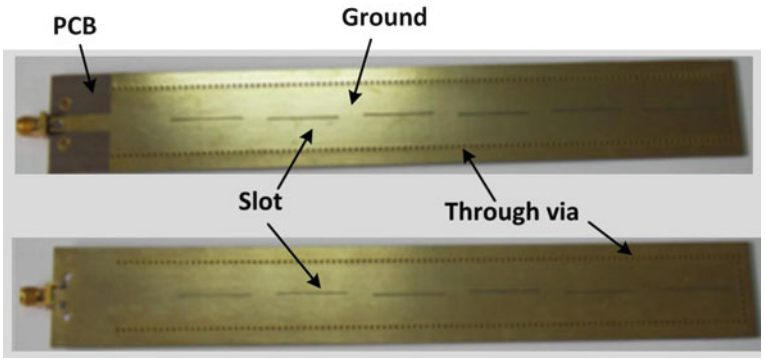


Fig. 36 SIW slot array antenna

The conventional metallic rectangular waveguide slot array antennas have been widely used in radar, navigation, and the communication systems for years. However, the volume, weight, and cost of these antennas are often puzzled for some applications. With the development of the substrate integrated waveguide (SIW) technology, it is possible to design a fully printed slot array antenna with horizontally polarized omnidirectional radiation, which is desired for WALN applications owing to the merits of small size, light weight, low manufacture cost. The SIW is a planar-guided wave structure which supports the similar TE-modes as in a conventional metallic rectangular waveguide. It can easily be integrated into microwave and millimeter-wave integrated circuits and has been extensively investigated for designing filters, couplers, power dividers, antennas, and even a passive front end (Takenoshita and Fujii 1998; Deslandes and Wu 2001; Yan et al. 2004; Hao et al. 2005).

A horizontally polarized omnidirectional SIW slot array is exemplified in Fig. 36 (Hua et al. 2008). The antenna is designed using a normal PCB ($\varepsilon_r = 2.2$, $\tan\delta = 0.003$, thickness = 1.5 mm) and fabricated with a standard PCB process, wherein longitudinal slots are etched on the upper and bottom grounds of the SIW. The antenna designed at 6GHz shows a bandwidth (VSWR $\leq$ 1.5) of 3 % and very good *E*-plane omnidirectional radiation patterns.

Apart from the waveguide slot antenna array, the horizontally polarized omnidirectional antennas based on microstrip slot configuration have been reported as well (Qing et al. 2012).

Planar Antenna Array

Consider a group of antenna elements arranged in a ring formation as shown in Fig. 37a, where N antenna elements are uniformly positioned along a circle with a radius of *r*. Assuming the antenna elements are equally excited with dominant electric current components in the direction of increasing ϕ, horizontally polarized omnidirectional radiation is expected to be generated by selecting suitable N and *r* (McEwan et al. 2003). The antenna configuration features the null radiation at zenith while varied maximum radiation at a specific elevation angle. For example, the maximum radiation will occur at θ of 90° if dipoles are used as the element; the maximum radiation will be directed to a certain elevation angle when the antenna elements are with the ground plane (such as a microstrip antenna), wherein the omnidirectional radiation with a conical beam is generated.

One antenna design is demonstrated in Fig. 38, where four broadband T-dipole elements (Sabatier 2003) are used to achieve *E*-plane omnidirectional radiation. The antenna is implemented on a circular PCB such as RF4; the T-dipoles are etched on the upper side of the PCB and are interconnected by a metallic square, which is also the ground plane for the microstrip line on the opposite side. The feed is located at the center of the antenna. The ground of the coaxial cable is connected to the ground plane of the antenna, and the inner conductor of the coaxial cable is connected to the microstrip line through the substrate. A proper impedance matching can be achieved using two stubs: a parallel stub located between

DOI 10.1007/978-981-4560-75-7_52-1

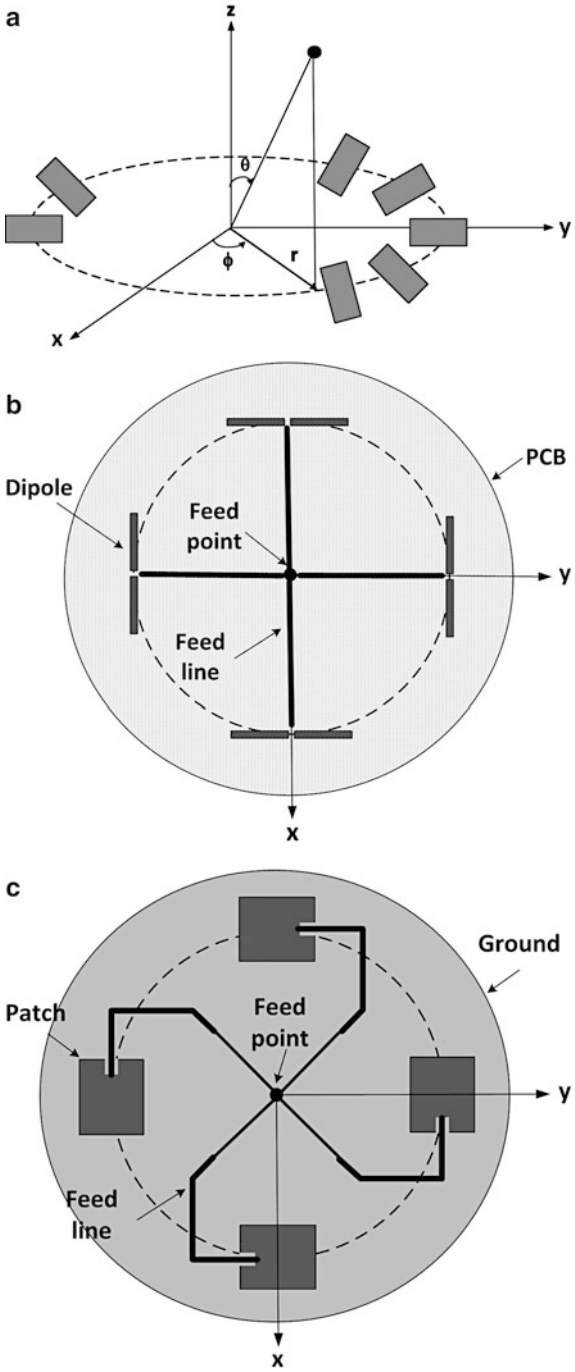


Fig. 37 Planar antenna array, (**a**) basic configuration, (**b**) antenna array with four dipole antennas, and (**c**) antenna array with four microstrip patch antennas

the two horizontal aims of the T-dipole and a serial stub at the end of the microstrip line. The four-element T-dipole antenna array (Ma and Wu 2006) achieves an impedance bandwidth of 80 % (3–7 GHz) for the reflection coefficient of less than −9 dB and gain of about 0 dBi. Good *E*-plane omnidirectional radiation patterns are achieved from 3 to 5.4 GHz with omnidirectionality of 3 dB.

Using more antenna elements may benefit the enhancement of the omnidirectionality while making the antenna configuration more complicated (Wu et al. 2007). In addition, applying broadband antenna

DOI 10.1007/978-981-4560-75-7_52-1

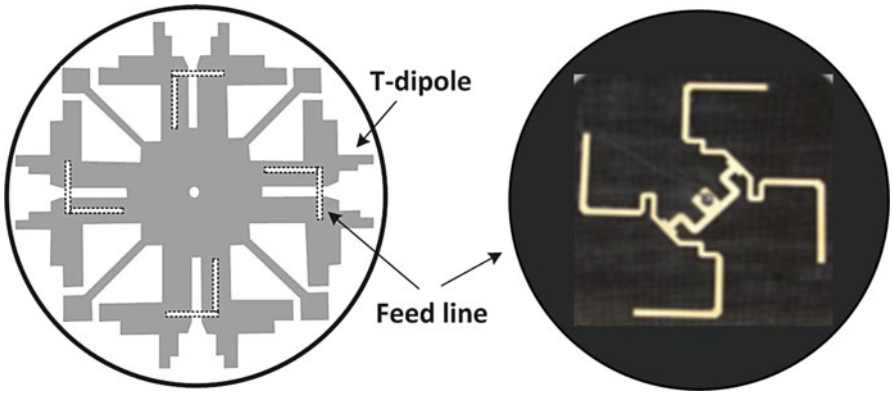


Fig. 38 Horizontally polarized omnidirectional antennas with four T-dipoles

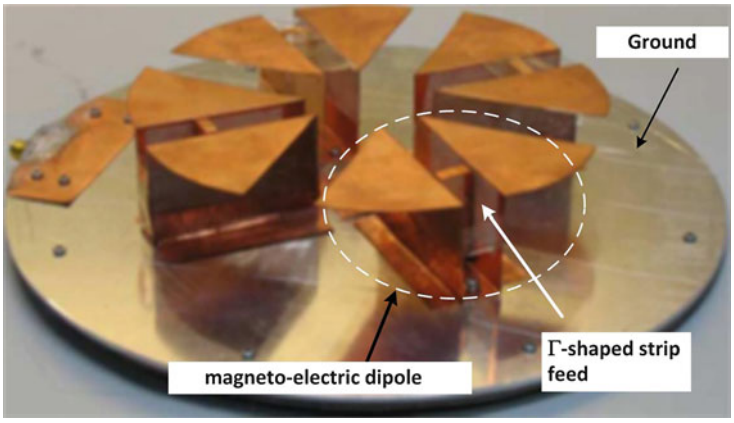


Fig. 39 Broadband omnidirectional antenna with conical beam using magnetoelectric dipoles

elements such as log-periodic antenna or tapered slot is helpful to improve the bandwidth of the antenna array (Fu and Hu 2008).

For the ceiling mounted indoor WLAN base station antennas, omnidirectional radiation with conical beam is preferable, where the antenna offers omnidirectional pattern in azimuth plane with a null at zenith but peak radiation at specific elevation angle (such as 60°). Monopole-type antennas are widely used for such purpose while they radiate vertically polarized wave and normally with high profile. Figure 39 shows a broadband horizontally polarized omnidirectional antenna with conical beam (Wu and Luk 2009). The four magnetoelectric dipoles are excited by an in-phase tapered power divider located below the ground plane. The configuration of each magnetoelectric dipole is a combination of an electric dipole and a magnetic dipole. The upper part of the antenna, which is an electric dipole, is constructed by a pair of sectorial-shape horizontal plates. The electric dipole is connected to the lower part of antenna, which is equivalent to a folded magnetic dipole. The magnetic dipole is due to a pair of vertically oriented shorted patch antennas, coupled by a Γ-shaped strip feed to produce a magnetic current along the edge of the two vertical walls. The low-profile antenna (height of $0.1867\lambda_0$) achieves an impedance bandwidth of 38 % (1.61–2.38 GHz) and average antenna gain of 5 dBi. Stable conical beam patterns are achieved across the operating band.

DOI 10.1007/978-981-4560-75-7_52-1

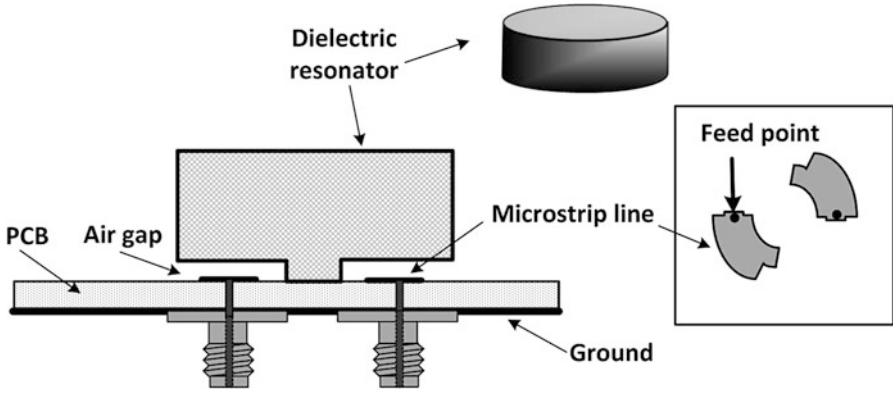


Fig. 40 Dielectric resonator antenna with $TE_{01\delta}$ mode for horizontally polarized omnidirectional radiation

Dielectric Resonator Antenna

Besides the patch array and magnetoelectric dipole array, a dielectric resonator antenna is able to generate horizontally polarized omnidirectional radiation with a conical beam as well. Regarding the first three fundamental modes in a cylindrical dielectric resonator (DR), $TM_{01\delta}$, $HEM_{01\delta}$, and $TE_{01\delta}$, the first two modes have been extensively studied for the vertically polarized omnidirectional radiation with broadside radiation patterns, respectively. In contrast, the $TE_{01\delta}$ mode is usually used as a filter because of its high-quality (Q) factor. Recently, it was reported that the $TE_{01\delta}$ mode in a cylindrical DR is able to be utilized to produce horizontally polarized omnidirectional radiation with conical beam patterns.

A design example is illustrated in Fig. 40 (Zuo and Fumeaux 2011). A cylindrical DR with ε_r of 10, diameter of 40 mm, and height of 6.62 mm is mounted on a Rogers Ultralam PCB (ε_r of 2.5 and thickness of 1.524 mm) with metalized bottom side. The $TE_{01\delta}$ mode is excited by two curved microstrip lines with identical sizes printed onto the top side of the substrate. Two SMA connectors, located at the back of the substrate, are utilized to connect the two microstrip lines to a splitter. The measured impedance bandwidth ($|S_{11}| < -10$ dB) covers a frequency range from 3.84 to 4.06 GHz. The antenna gain is above 4 dBi with a maximum of 5.2 dBi. Compared to other types of *E*-plane omnidirectional antennas, the DR antenna is attractive for its simplicity, low profile, and relatively high gain.

Dual Linearly Polarized Omnidirectional Antenna

Antennas with dual polarization are widely studied and applied in wireless communication systems because of the improved channel capacity and mitigated multipath fading. The polarization diversity that combines pairs of antennas with orthogonal polarizations has been widely used in mobile communications. For the polarization diversity scheme with a 360° full coverage, dual-polarized (vertically/horizontally) omnidirectional antennas are needed for base stations. Therefore, the design of dual-polarized omnidirectional antennas has become an interesting topic in wireless communications nowadays.

It is well known that the antenna configurations with vertically polarized omnidirectional radiation are distinct from those with horizontally polarized omnidirectional radiation. It is almost impossible to generate the orthogonal polarized omnidirectional radiation using one radiator; the most applicable solution is to combine the vertically and the horizontally polarized omnidirectional antenna elements together for achieving the dual-polarized omnidirectional radiation. An exception is the dual-polarized omnidirectional DR antenna, where different modes in one radiator are used to generate the orthogonal polarized omnidirectional radiation.

DOI 10.1007/978-981-4560-75-7_52-1

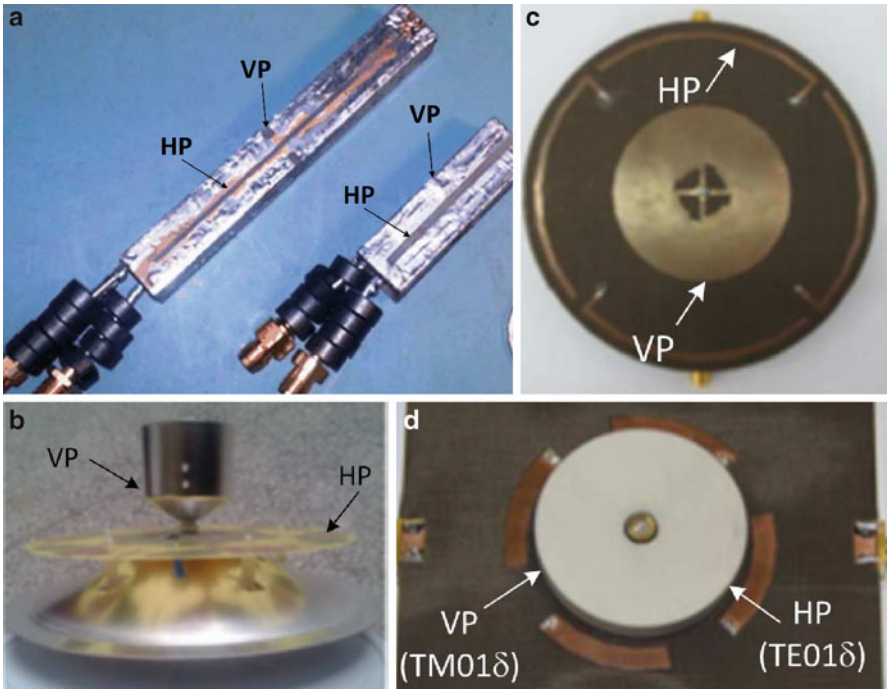


Fig. 41 Dual linearly polarized omnidirectional antennas; (**a**) compact slot antennas in slender column (Li et al. 2012, 2014), (**b**) multiband/wideband antenna (Dai et al. 2013), (**c**) low-profile antenna (Deng et al. 2012), and (**d**) dielectric resonator antenna (Zou et al. 2012)

Figure 41 illustrates some dual linearly polarized omnidirectional antennas. In Fig. 41a (Li et al. 2012, 2014), the dual polarizations are achieved by positioning two orthogonal slots cut onto the walls of a slender columnar cuboid. The antenna design at the WLAN band of 2.4–2.48 GHz achieves gains of greater than 3.17 dBi and 1.19 dBi for vertical and horizontal polarizations, respectively, and port isolation of higher than 33.5 dB with an antenna size of $0.664\lambda_0 \times 0.088\lambda_0 \times 0.088\lambda_0$. A compact design with reduced size was reported (Li et al. 2014), where a cavity-backed notch is applied for the horizontal omnidirectional polarization and a folded slot for the vertical omnidirectional polarization. With the antenna size of $0.336\lambda_0 \times 0.096\lambda_0 \times 0.096\lambda_0$, the antenna achieves isolation of higher than 32.5 dB and a maximum gain of 2.78 dBi and 1.35 dBi for vertical and horizontal polarization, respectively, from 2.4 to 2.48 GHz. The radiation patterns at 2.44 GHz show an omnidirectionality of 4.5 dB for vertical polarization and 2.4 dB for horizontal polarization.

Figure 41b exhibits a multiband/wideband dual linearly polarized antenna (Dai et al. 2013). The antenna is able to cover the frequency ranges of 806–960 MHz (17.4 %) and 1,880–2,700 MHz (35 %) for vertical polarization and 1,880–2,700 MHz (35 %) for horizontal polarization, which makes it suitable for GSM900, CDMA800, GSM1800, GSM 1990, PCS, UMTS, and LTE communication systems. The antenna is a combination of a modified asymmetric biconical antenna for the vertical polarization and six printed dipoles with concentric placement for the horizontal polarization. The element for vertical polarization consists of two modified cones with different diameters, designed to work in the multiband. The printed dipoles for horizontal polarization are fed by a six-way power splitter printed on a PCB. The horizontal polarized element is positioned between two cones of the vertical polarized element, forming a compact structure. The antenna exhibits desired performance with gain of 1.4–2.2 dBi for vertical polarization at lower band, and about 4.5 dBi for both polarizations at higher band; an isolation of greater than 25 dB; cross-polarization levels of lower than −15 dB in the azimuth plane for both polarizations; and desired omnidirectional patterns with omnidirectionality of 1.5 dB.

DOI 10.1007/978-981-4560-75-7_52-1

A low-profile broadband dual linearly polarized antenna as shown in Fig. 41c was designed to operate at 2-GHz band for the based-station applications (Deng et al. 2012). The antenna is a combination of a modified low-profile monopole for the vertical polarization and a circular planar loop for the horizontal polarization. The modified low-profile monopole is a circular folded patch shorted by four tubes, while the circular loop consists of four half-wavelength arc dipoles. The antenna achieves a bandwidth of 25 % (1.7–2.2 GHz) with an isolation of around 40 dB and omnidirectionality of 2.5 dB and 1.5 dB for the vertical and horizontal polarizations, respectively.

In contrast to the conventional dual linearly polarized antenna designs employing different radiating elements to generate the orthogonal polarized radiations, a DR antenna is able to generate dual linearly polarized radiation with single radiator. Figure 41d shows an omnidirectional cylindrical DR antenna simultaneously realizing the horizontally and vertically polarized omnidirectional radiation with low cross coupling (Zou et al. 2012). The horizontally and vertically polarized radiations are achieved by exciting the orthogonal $TE_{01\delta}$ and $TM_{01\delta}$ modes in a single cylindrical dielectric resonator. The $TE_{01\delta}$ mode corresponding to horizontally polarized radiation is with higher Q-factor so that the bandwidth of horizontal polarization is smaller than that of the vertically polarized radiation generated by the $TM_{01\delta}$ mode. To increase the bandwidth of the $TE_{01\delta}$ mode and suppress the influence of higher-order modes, two groups of four radially arranged microstrip feeding lines with two different lengths are utilized, and a dual-polarized omnidirectional radiation is realized at the overlapping operating band of the modes with a bandwidth of 7.4 % (3.78–4.07 GHz).

The other dual linearly polarized omnidirectional antenna designs are reported with different configurations; the combinations include notched disk antenna and wire antenna, planar conical cone antenna and circular loop antenna, collinear array antenna and slotted cylinder antenna, and so on. For convenience, a comparison of some antennas is tabulated in Table 1.

Circularly Polarized Omnidirectional Antenna

A linearly polarized antenna radiates wholly in one plane containing the direction of propagation. An antenna is vertically or horizontally polarized when its electric field is perpendicular or parallel to the Earth's surface. For a circularly polarized antenna, the plane of the polarization rotates in a corkscrew pattern, making one complete revolution in each wavelength. The electric field of the circularly polarized antenna lies in both the vertical and horizontal planes as well as all the planes in between in a rotating manner. Compared to linearly polarized antenna, a circularly polarized antenna features the unique advantage as follows:

- Able to establish a reliable signal link regardless of the orientation of the antennas.
- More effective to suppress the multipath interferences.
- More resistant to signal degradation due to inclement weather conditions.
- The circularly polarized wave is more effective be propagated through wall and can achieve an overall better reach throughout the building.

Circularly polarized omnidirectional antennas have been widely used in television broadcasts; mobile satellites; space vehicles such as airplanes, missiles, rockets, and spacecraft; mobile communication; and WLANs.

A circularly polarized radiation can be obtained by generating two orthogonal field components with 90° phase difference. To design a circularly polarized omnidirectional antenna is more complicated than that of the linearly polarized one since the designer has to consider the spatial orthogonal field components

Table 1 Comparison of the dual linearly polarized omnidirectional antennas

Reference	Antenna configuration	Frequency range (GHz)	Bandwidth (%)	Return loss (dB)	Gain (dBi)	Isolation (dB)	Cross-polarization (dB)	Omnidirectionality (dB)	Remark
Kuga et al. (1998)	VP: three vertical wire loop	1.688–1.838	8.5	≥10	0*	>14	N.A.	5	*Average gain
	HP: notched disk	1.648–1.878	13	≥10	0		N.A.	3	
Ando et al. (2008)	VP: planar discone	2.4–2.5	4.1	≥10	−2.5	>20	≤14.4	1.6	Dual-band design
		4.9–6.0	20.2		0.9		≤13.8	3.8	
	HP: two circular dipole	2.4–2.5	4.1	≥10	0		≤8.6	5.4	
		4.9–6.0	20.2		0.6		≤11.3	5.5	
Deng et al. (2012)	VP: annular-ring patch	4.80–5.27	9.3	≥10	5.3	>40	≤18	0.6	
	HP: combined loop	4.83–5.02	3.8	≥10	4.1		≤18	1.1	
Zou et al. (2012)	VP: annular cylindrical dielectric resonator, $TM_{01\delta}$ mode	3.55–4.30	19.1	≥10	>1.7	N.A.	≤20	N.A.	
	HP: annular cylindrical dielectric resonator, $TE_{01\delta}$ mode	3.78–4.07	7.4	≥10	>5.1		≤20	N.A.	
Quan and Li (2013)	VP: modified monopole	1.7–2.2	25	≥10	~8	>40	≤20	2.5	Eight-element array
	HP: four arc dipoles	1.7–2.2	25	≥10	~8		≤20	1.5	
Dai et al. (2013)	VP: modified conical cones	0.806–0.96	17.4	≥10	1.4–2.2	>25	≤15	1.5	Multiband design
		1.88–2.70	35	≥10	4.5				
	HP: six dipoles	1.88–2.70	35	≥10	4.5		≤15	1.5	
Li et al. (2014)	VP: folded slot	2.34–2.72	19.1	≥10	2.78	>30	N.A.	4.5	Compact design
	HP: cavity-backed notch	2.39–2.49	7.4	≥10	1.35		N.A.	2.4	

DOI 10.1007/978-981-4560-75-7_52-1

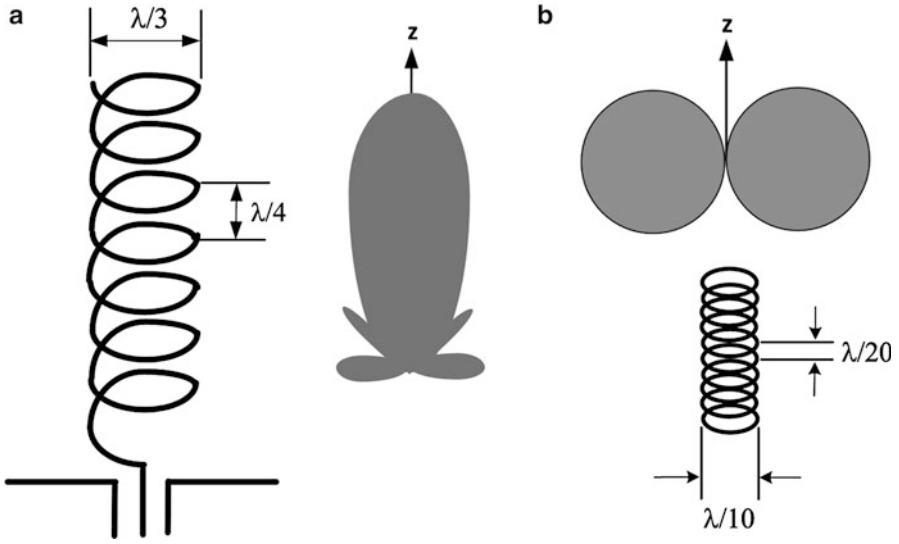


Fig. 42 Field patterns of the axial and normal radiation modes of the helical antenna; (**a**) normal mode and (**b**) axial mode

and the proper phase delay of the field components over a specific bandwidth at desired directions. The studies of circularly polarized omnidirectional antenna can be dated back to 1940s (Wheeler 1947; Lindenblad 1941; Brown and Woodward 1947). Over the last few decades, a number of circularly polarized omnidirectional antennas have been developed; the configurations can be categorized as follows:

- Single radiator
- Conformal array with circularly polarized radiator
- Combination of linear polarized radiator

Circularly Polarized Omnidirectional Antennas with Single Radiator

Helical Antenna

Helical antenna is well known for its circularly polarized radiation characteristics. When the circumference of a helical antenna is approximately one wavelength, the axial mode of radiation is dominant; when the circumference is much smaller the one wavelength, the normal mode is dominant, as shown in Fig. 42.

As illustrated in Fig. 43, the far field of a helix may be described by two components of the electric field, E_ϕ and E_θ, which are contributed by the small loop and the short dipole, respectively. For a small helix, the current is assumed to be uniform in magnitude and in phase over the entire length of the helix. The far field of the small loop has only an E_ϕ component as below (Kraus and Marhefka 2008):

$$E_\phi = \frac{120\pi^2[I]\sin\theta}{r}\frac{A}{\lambda^2} \tag{6}$$

where A is the area of the loop, $A = \pi r D^2/4$

The far field of the short dipole has only an E_θ; it can be expressed as

$$E_\theta = j\frac{60\pi[I]\sin\theta}{r}\frac{S}{\lambda} \tag{7}$$

DOI 10.1007/978-981-4560-75-7_52-1

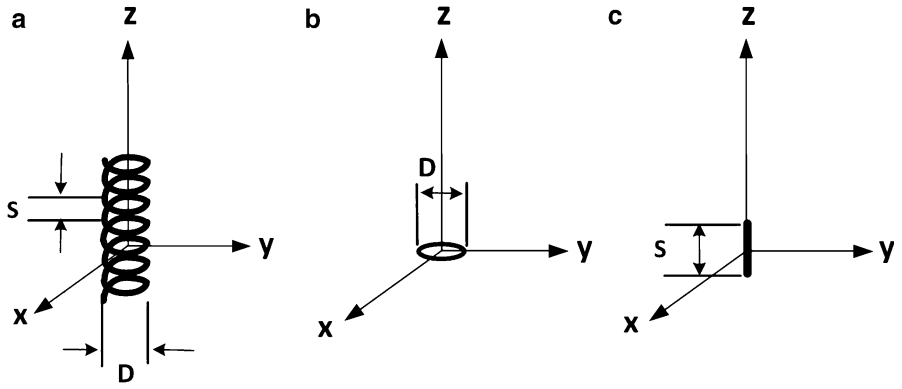


Fig. 43 Small helix for normal mode calculation; (**a**) helix, (**b**) small loop, and (**c**) short dipole

Comparing Eqs. 6 and 7, the j operator in Eq. 7 and its absence in Eq. 6 indicate that E_ϕ and E_θ are in-phase quadrature. The ratio of the magnitudes of Eqs. 6 and 7 then gives the axial ratio of the polarization ellipse of the far field

$$AR = \frac{|E_\theta|}{|E_\phi|} = \frac{S\lambda}{2\pi A} = \frac{2S\lambda}{\pi^2 D^2} = \frac{2S_\lambda}{C_\lambda^2} \tag{8}$$

where S_λ and C_λ are the pitch and circumference of the helix in wavelength, respectively.

A circularly polarized radiation can be achieved when AR is a unit; from Eq. 8, we have

$$\pi D = \sqrt{2S\lambda} \text{ or } C_\lambda = \sqrt{2S_\lambda} \text{ or } A = \frac{\lambda}{2\pi} S \tag{9}$$

The relation was first presented by Wheeler (1947) so that such antenna is also called "Wheeler coil." A monofilar normal-mode helix is a resonant and narrowband circularly polarized omnidirectional antenna. To enhance the bandwidth, the bifilar configuration can be applied (Amin et al. 2013).

Dipole-/Monopole-Type Antenna with Polarizer

It is known that a pair of 45° tilted dipole antennas with a λ/4 spacing, in orthogonal crossed disposition, radiates circularly polarized wave in both directions of the array axe if current distribution on the dipoles is identical. Figure 44 exhibits the schematic view of a dipole antenna with a polarizer for circularly polarized omnidirectional radiation; where the polarizer is formed by a concentric array of 45° tilted parasitic elements, the spacing between the dipole and the parasitic elements is set to $R = \lambda/4$. The radiation field from the vertical dipole can be divided into two orthogonally crossed electric fields (parallel and perpendicular components to the parasitic strip elements); each tilted element is excited by the parallel component of the field accordingly. Because of the λ/4 separation between the dipole and the parasitic elements, the reflection of the perpendicular component is canceled. The 90° phase difference between the two radiated field components is achieved by the interaction between the parasitic elements and the parallel component.

Generally, the parasitic elements are uniformly spaced approximately λ/2 apart. The length of the elements is around 0.5λ and the width of each element about λ/10. In practical design, the parasitic elements are normally implemented with a supporting material such as a piece of foam or PCB; the dimensions as well as the spacing between the dipole antenna and the parasitic elements are required to be adjusted for optimal performance.

DOI 10.1007/978-981-4560-75-7_52-1

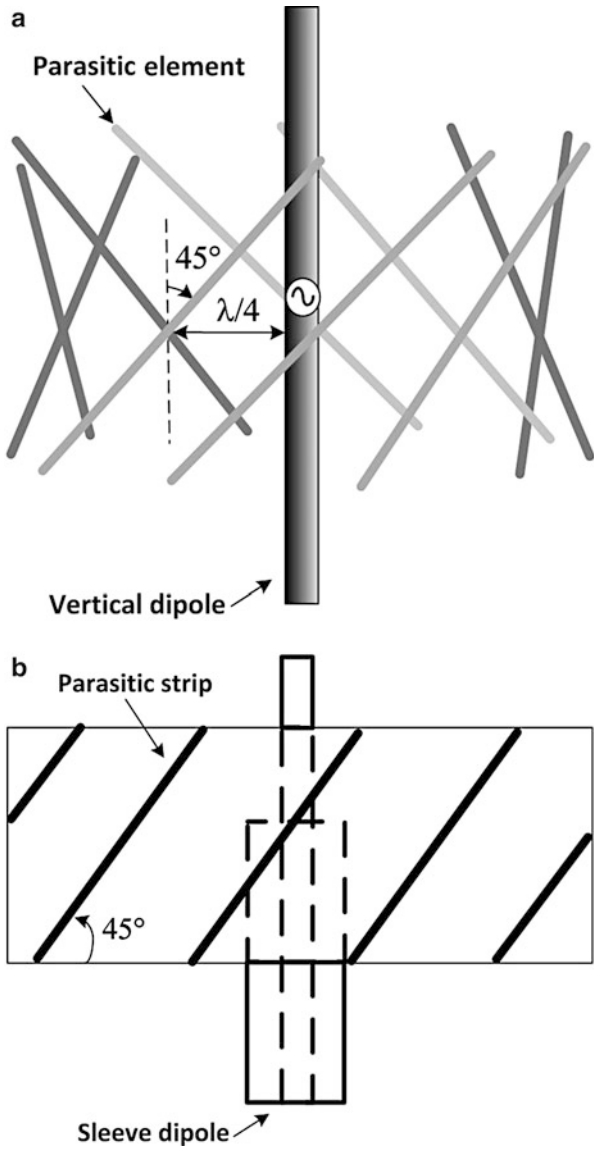


Fig. 44 Dipole/sleeve antenna with a conformal array of 45° tilted parasitic elements; (**a**) dipole configuration and (**b**) sleeve dipole configuration

Based on the basic configuration, a number of variation of the polarizer have been reported for generating circularly polarized omnidirectional radiation, where biconical cones, sleeve dipole, and a conical skirt monopole are used as the radiators while multilayer parasitic elements are applied (Kelleher and Morrow 1955; Goatley and Green 1956; Sakaguchi and Hasebe 1993; Fernandez et al. 2007).

Circularly Polarized Omnidirectional Antennas with Multiple Circularly Polarized Radiators

Conformal Cylindrical Array

The arbitrarily oriented satellites and space vehicles such as missiles and rockets require a circularly polarized omnidirectional antenna to reliably maintain a communication link with a central receiving link. A low-profile conformal cylindrical omnidirectional antenna array is desirable for such applications (Galindo and Green 1965; Wu 1995). To configure such antenna arrays, the most convenient way is to

DOI 10.1007/978-981-4560-75-7_52-1

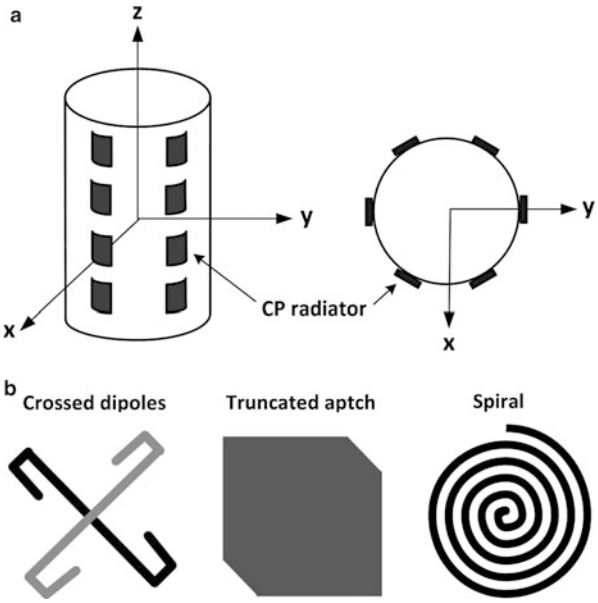


Fig. 45 Conformal cylindrical array; (**a**) basic configuration and (**b**) CP elements

position a certain number of circularly polarized radiators with equidistant intervals on the surface of a cylindrical surface. Desired omnidirectional radiation is able to be achieved by arranging adequate number of radiators on the cylindrical surface properly. A variety of low-profile radiators such as cross dipoles, truncated microstrip patch, spiral, slots, and so on can be utilized as the array elements (Dubost et al. 1979; Nakayama et al. 1999) as exhibited in Fig. 45.

Cutting slots directly on the surface of a circular waveguide is able to achieve circularly polarized omnidirectional radiation as well. Such an antenna configuration may be more preferable from a system point of view because of the merits of simplicity, robustness, and lightweight. Figure 46 shows a slot array design (Masa-Campos et al. 2007), where the conformal slot array is designed at the millimeter-wave band of 36.7–37 GHz; the eight slot pairs are symmetrically positioned on the wall of a circular waveguide carrying TM_{01} mode. Each slot pair consists of two resonant slots with a length of $\lambda_0/2$ and a separation of $\lambda g/4$ (λg is the guided wavelength of the propagated TM_{01} mode).

The conformal cylindrical antenna arrays with traditional circularly polarized radiator are with limited bandwidth for impedance matching, in particular, the circular polarization characteristic of axial ratio, which makes the antennas not suitable for modern mobile communications. For example, the second- and third-generation mobile communication systems require a base-station antenna covering the frequency range from 1.71 to 2.17 GHz; a bandwidth of at least 25 % is needed. Recently, Quan et al. (2013) presented a broadband conformal circularly polarized omnidirectional antenna as shown in Fig. 47. The four rectangular loop elements are first printed on a flexible thin dielectric substrate and then rolled into a hollow cylinder. A conducting cylinder is added inside the hollow cylinder for achieving desired omnidirectional circularly polarized performance. The antenna exhibits a bandwidth of 41 % (1.65–2.5 GHz) for axial ratio of less than 3 dB and omnidirectionality of 1 dB and 45 % (1.58–2.5 GHz) for return loss greater than 10 dB.

Back-to-Back Patches

To simplify the complicated feeding network of the conformal cylindrical array, the back-to-back coupled patch antennas are investigated for circularly polarized omnidirectional radiation; H. lwasaki and Chiba reported a coplanar waveguide (CPW) fed back-to-back patch antenna (lwasaki and Chiba 1999) as

DOI 10.1007/978-981-4560-75-7_52-1

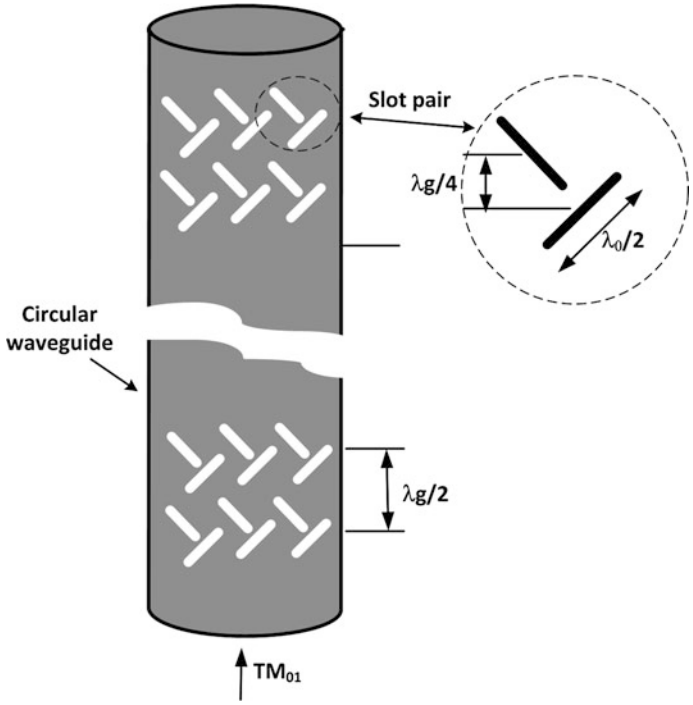


Fig. 46 Conformal circularly polarized omnidirectional antenna with slot radiators on a circular waveguide

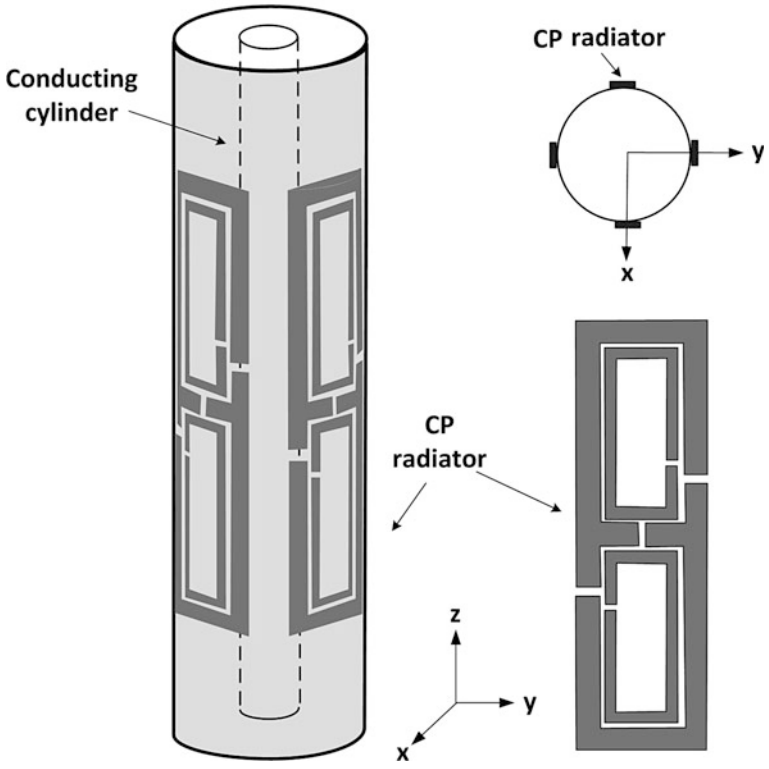


Fig. 47 A broadband omnidirectional circularly polarized antenna

DOI 10.1007/978-981-4560-75-7_52-1

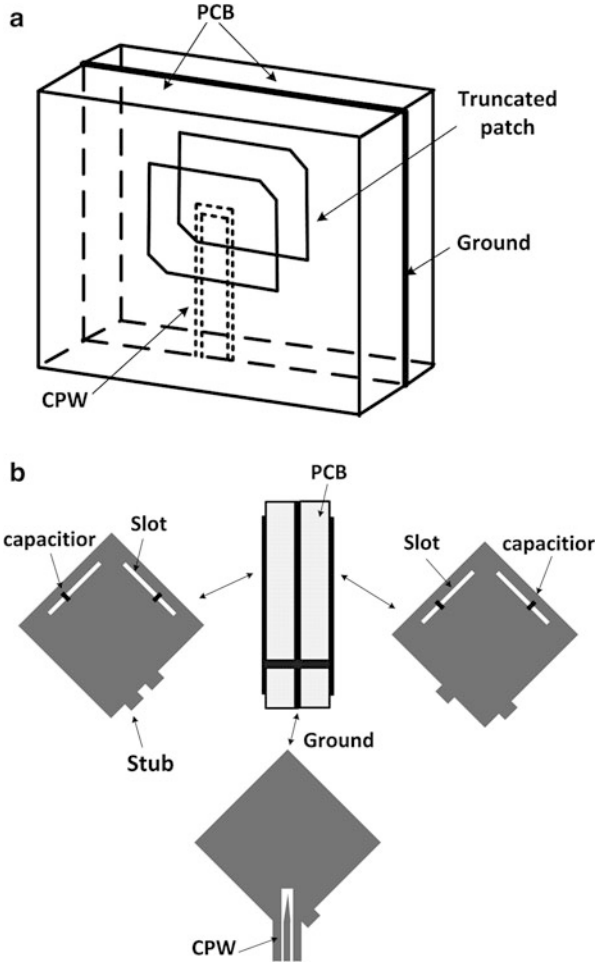


Fig. 48 Circularly polarized omnidirectional antennas with back-to-back patches, (**a**) truncated patches and (**b**) slotted patched

shown in Fig. 48a, where two truncated microstrip patch elements are positioned on the opposite sides of a PCB slab and a CPW feed line is sandwiched in between. The antenna exhibits narrow impedance bandwidth of 13 % (VSWR < 2) and axial ratio of less than 4 dB in the azimuth plane.

Narbudowicz et al. (2013) proposed a dual-band circularly polarized omnidirectional antenna using back-to-back slotted patch configuration. As illustrated in Fig. 48b, the antenna consists of two layers of substrate with three layers of metallization. The inner metallization forms the ground plane and comprises the coplanar waveguide (CPW) feeding structure, whereas the outer ones form two patches. The patches are additionally connected together by a thin copper strip. The radiators are electromagnetically coupled to the 50-Ω CPW located along the diagonal of the patches. The two slots of slightly different lengths with a lumped capacitor connected across the center are utilized to provide dual-band circularly polarized radiation. The conducting strip connected the patches, and a reduced ground plane size allows better impedance matching and reduced overall size. The dual-band antenna radiates right-handed circular polarization at both the lower and upper frequencies of 1.329 and 1.565 GHz with bandwidth of about 20 MHz, respectively. The axial ratio in the omnidirectional plane varies from 0.4 to 3.8 dB for the lower band and from 0.6 to 4.8 dB for the upper band. The omnidirectionality in the omnidirectional plane is 3.3 dB for the lower band and 3.8 dB for the upper band.

DOI 10.1007/978-981-4560-75-7_52-1

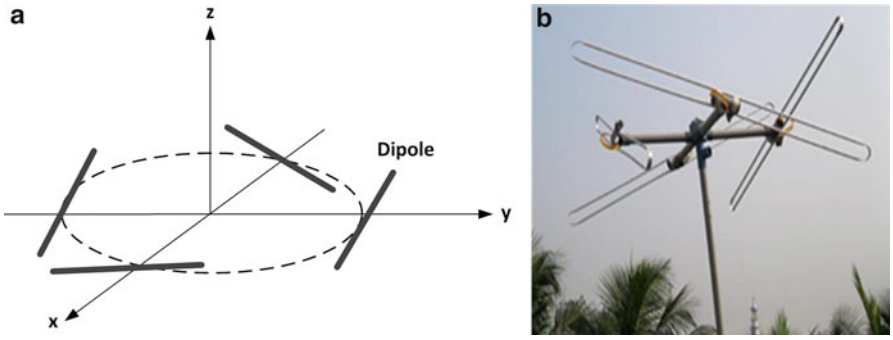


Fig. 49 Lindenblad antenna, (**a**) antenna configuration and (**b**) practical antenna design

In short, the back-to-back patch design features simple configuration with limited bandwidth and radiation performance. Furthermore, it is difficult to achieve desired omnidirectionality and axial ratio, in particular, at the directions apart from the boresight.

Circularly Polarized Omnidirectional Antennas with Combined Linearly Polarized Radiators

Lindenblad Antenna

The circularly polarized omnidirectional radiation can be realized by positioning multiple linearly polarized radiators properly. The Lindenblad antenna might be the first circularly polarized omnidirectional antenna with such a configuration (Lindenblad 1941). As shown in Fig. 49, a Lindenblad antenna uses four dipole elements that are fed in phase to create an omnidirectional circularly polarized radiation pattern. The dipoles are canted at 30° from horizontal and positioned equally around a circle of about λ/3 diameter. For easy impedance matching with a 50-Ω feed line, folded dipole instead of straight dipole is most used in practical designs. Seventy years after the invention, this antenna is still widely used in broadcasting stations, satellite communications, and amateur radios.

Monopole and Loop Combination

A dipole or monopole antenna is desired for vertically polarized omnidirectional radiation, and a loop antenna with a uniform current flowing in a single direction is able to generate horizontally polarized omnidirectional radiation. The combination of the dipole and loop antennas with an adequate arrangement to provide a 90° phase difference between the currents on the orthogonal radiators features the possibility to produce a circularly polarized omnidirectional radiation.

As shown in Fig. 50, considering the combination of the short dipole of length L and a loop with a radius of a, the dipole is positioned coincidently with the z-axis and with its center at the origin. The far field generated by the short dipole is expressed as (Kraus and Marhefka 2008)

$$E = j\frac{[I]L\omega\mu_0}{4\pi r}\sin\theta\hat{\theta} \tag{10}$$

where $[I]$ is the current on the short dipole, ω is the angular frequency, μ_0 is the permeability of the free space, and r is the distance between the origin and observation point.

The loop is placed in the x-y plane with its center at the origin; its axis is along the z-axis. The current on the loop is assumed to be the same with the dipole, which is uniform and in phase along the loop. The far field of the loop can be written as [16]

DOI 10.1007/978-981-4560-75-7_52-1

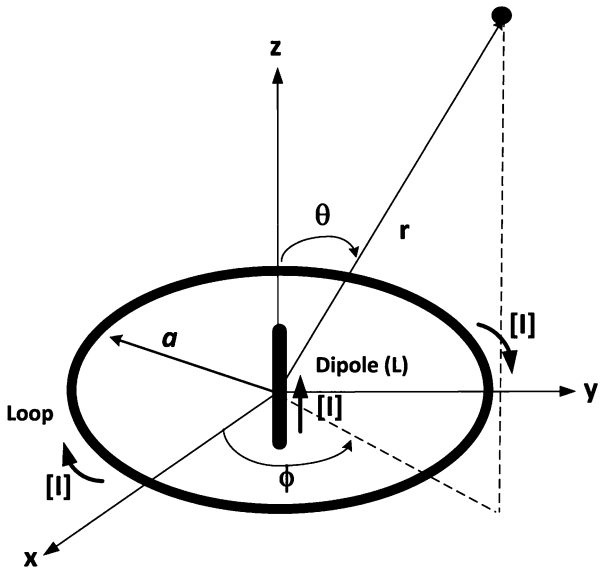


Fig. 50 Configuration of an ideal omnidirectional circularly polarized antenna combining a short dipole with the height of L and a small loop with the radius of a

$$E = \frac{[I]a\omega\mu_0}{2r} J_1(\beta a \sin\theta)\hat{\phi} \tag{11}$$

where J_1 is the Bessel function of the first order and of argument ($\beta\alpha\sin\theta$).

By superposing the far fields of the dipole and the loop, the far field of the combining dipole/loop antenna is given by

$$E = j\frac{[I]L\omega\mu_0}{4\pi r}\sin\theta\hat{\theta} + \frac{[I]\alpha\omega\mu_0}{2r} J_1(\beta a \sin\theta)\hat{\phi} \tag{12}$$

Then, the axial ratio of the combining antenna can be given by

$$AR = \left|20\log_{10}\left(\left|\frac{L}{2\pi a}\frac{\sin\theta}{J_1(\beta a\sin\theta)}\right|\right)\right| \tag{13}$$

The circularly polarized omnidirectional antenna is required to achieve good axial ratio in the azimuth plane (x-y plane). Considering $\theta = 90°$, Eq. 13 can be written as

$$AR = \left|20\log_{10}\left(\left|\frac{L}{2\pi a}\frac{1}{J_1(\beta a)}\right|\right)\right| \tag{14}$$

The AR equals 0 dB when

$$\left|\frac{L}{2\pi a}\frac{1}{J_1(\beta a)}\right| = 1 \tag{15}$$

From the above equation, it can be found that, by choosing a suitable value of L and a in Eq. 15, the axial ratio can equal 0 dB, and thus the antenna can achieve a perfect circularly polarized radiation.

DOI 10.1007/978-981-4560-75-7_52-1

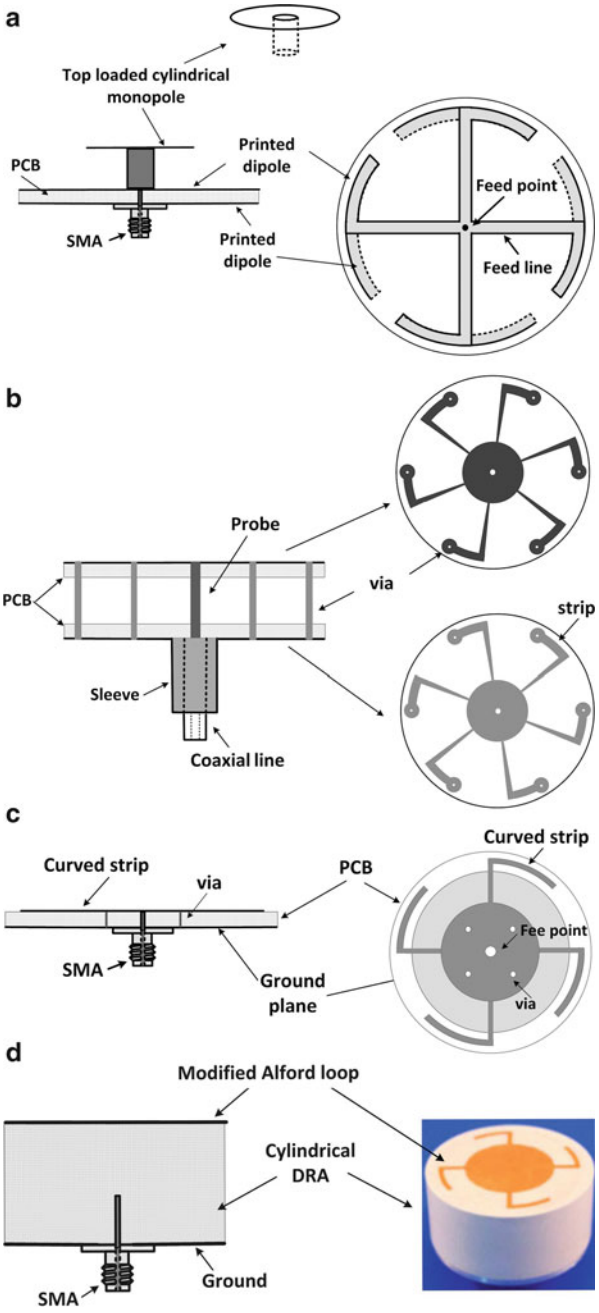


Fig. 51 CP antenna designs with combining radiators, (**a**) top-loaded monopole plus arc-shaped dipoles and (**b**) sleeve antenna plus printed spoke-like strips, (**c**) zeroth- and first-order resonance modes epsilon-negative transmission line plus curved strips, and (**d**) dielectric resonator antenna plus modified Alford loop

One of the antenna examples as shown in Fig. 51a is designed for 2.4-GHz WLAN applications, where a top-loaded cylindrical monopole generates vertically polarized omnidirectional radiation; four printed identical arc-shaped dipoles form a virtual loop to produce horizontally polarized omnidirectional radiation (Hisao and Wong 2005). The 90° phase difference between the two orthogonal components is achieved through tuning the length of the feed lines for the arc-shaped dipoles, thereby resulting in circularly polarized omnidirectional radiation. Figure 51b exhibits another example, where a metal sleeve acts as a monopole and the printed spoke-like metal strips fabricated onto two pieces of substrate function

DOI 10.1007/978-981-4560-75-7_52-1

as a loop (Li et al. 2013). Figure 51c demonstrates a metamaterial-based dual-band circularly polarized omnidirectional antenna comprising a circular mushroom structure with curved branches. The dual-band vertically polarized radiation is achieved by using the zeroth- and the first-order resonance modes of the epsilon-negative (ENG) transmission line, and the horizontally polarized omnidirectional radiation is realized by the curved branches (Park and Lee 2012).

A compact omnidirectional circularly polarized antenna combining cylindrical DR antenna and a top-loaded modified Alford loop is exhibited in Fig. 51d (Li and Leung 2013). Fed by an axial probe, the DR antenna operates in its $TM_{01\delta}$ mode, which radiates like a vertically polarized electric monopole. The modified Alford loop comprising a central circular patch and four curved branches is placed on the top of the dielectric resonator antenna and provides an equivalent horizontally polarized magnetic dipole mode. Omnidirectional circularly polarized radiation can be obtained when the two orthogonally polarized fields are equal in amplitude but 90° different in phase.

Conclusion

Omnidirectional antennas are essential for modern wireless communication systems, in particular for base-station applications for providing a wide coverage. Over the last several decades, many omnidirectional antennas have been developed for various applications. In general, the omnidirectional radiation can be achieved by using single radiator with circularly symmetrically distributed current or multiple radiators with a combined circular current distribution. This chapter has outlined the classic omnidirectional antenna configurations based on the antenna polarizations and discussed a number of antenna design examples for each topic. The relevant literature review is believed to benefit the antenna researchers, engineers, and students to further understand the omnidirectional antennas.

Cross-References

- ▶ Antenna Systems for Cellular Base Stations
- ▶ Antennas in Access Points of WLAN/WiFi
- ▶ Broadband and Multiband Planar Antennas
- ▶ Circularly Polarized Antennas
- ▶ Conformal Array Antennas
- ▶ Dielectric Resonator Antennas
- ▶ Linear Wire Antennas
- ▶ Loop Antennas
- ▶ Low-profile Antennas
- ▶ Metamaterials and Antennas
- ▶ Spiral, Helical and Rod Antennas
- ▶ Substrate-integrated Waveguide Antennas
- ▶ Ultra-wideband Antennas
- ▶ Waveguide Slot Array Antennas

DOI 10.1007/978-981-4560-75-7_52-1

References

Ahn CH, Oh SW, Chang K (2009) A dual-frequency omnidirectional antenna for polarization diversity of MIMO and wireless communication applications. IEEE Antennas Wirel Propag Lett 8:966–969

Alford A, Kandoian AG (1940) Ultrahigh-frequency loop antennas. TIEE Trans Electr Eng 59(12):843–848

Amin M, Yousaf J, Iqbal S (2013) Single feed circularly polarised omnidirectional bifilar helix antennas with wide axial ratio beamwidth. IET Microw Antennas Propag 7(10):825–830

Ando A, Kita N, Yamada W (2008) 2.4/5-GHz dual-band vertically/horizontally dual-polarized omnidirectional antennas for WLAN base station. IEEE Int Symp Antennas Propag 1–4

Balanis CA (1997) Antenna theory_analysis and design. Wiley, New York

Balsley BB, Ecklund WL (1972) A portable coaxial collinear antenna. IEEE Trans Antennas Propag 20(4):513–516

Bancroft R, Bateman B (2004) An omnidirectional planar microstrip antenna. IEEE Trans Antennas Propag 52(11):3151–3153

Barrow WL, Chu LJ, Jansen JJ (1939) Biconical electromagnetic horns. Proc IRE 27(12):769–779

Baumer C, Landstorfer F (1990) Design of omnidirectional slot antennas with vertical polarization. IEEE Int Symp Antennas Propag 2:938–941

Bergmann JR (2003) On the design of broadband omnidirectional compact antennas. Microw Opt Technol Lett 39(5):418–422

Bergmann JR, Moreira FJS (2004) An omnidirectional ADE reflector antenna. Microw Opt Technol Lett 40(3):250–254

Borja AL, Hall PS, Liu Q, Lizuka H (2007) Omnidirectional loop antenna with left-handed loading. IEEE Antennas Wirel Propag Lett 6:495–498

Brown H, Woodward OM Jr (1947) Circularly polarized omnidirectional antenna. RCA Rev 8(2):259–269

Caloz C, Sanada A, Itoh T (2004) A novel composite right-/left-handed coupled-line directional coupler with arbitrary coupling level and broad bandwidth. IEEE Trans Microw Theory Techn 52(3):980–992

Chen ZN (2005) Novel bi-arm rolled monopole for UWB applications. IEEE Trans Antennas Propag 36(9):672–677

Chen A, Jiang T, Chen Z, Su D, Wei W, Zhang Y (2011a) A wideband VHF/UHF discone based antenna. IEEE Antennas Wirel Propag Lett 10:450–453

Chen X, Huang K, Xu XB (2011b) Novel planar slot array antenna with omnidirectional pattern. IEEE Trans Antennas Propag 59(12):4853–4857

Chizhik D, Ling J, Valenzuela RA (1998) The effect of electric field polarization on indoor propagation. IEEE Int Conf Univ Pers Commun 1:459–462

Chu TS (1959) On the use of uniform circular arrays to obtain omnidirectional patterns. IRE Trans Antennas Propag 7(4):436–438

Croswell WF, Cockrell C (1969) An omnidirectional microwave antenna for use on spacecraft. IEEE Trans Antennas Propag 17(4):459–466

Dai XW, Wang ZY, Liang CH, Chen X, Wang LT (2013) Multiband and dual-polarized omnidirectional antenna for 2G/3G/LTE application. IEEE Antennas Wirel Propag Lett 12:1492–1495

Delaveaud C, Leveque P, Jecko B (1998) Small-sized low-profile antenna to replace monopole antennas. Electron Lett 34(8):716–717

Deng C, Li P, Cao W (2012) A high-isolation dual-polarization patch antenna with omnidirectional radiation patterns. IEEE Antennas Wirel Propag Lett 11:1273–1276

Deslandes D, Wu K (2001) Integrated microstrip and rectangular waveguide in planar form. IEEE Microw Wirel Components Lett 11:68–70

Dobkin DM, Weigand SM, Iye N (2007) Segmented magnetic antennas for near-field UHF RFID. Microw J 50(6):96–102

Dubost G, Samson J, Frin R (1979) Large-bandwidth flat cylindrical array with circular polarisation and omnidirectional radiation. Electron Lett 15(4):102–103

Fernandez JM, Masa-Campos JL, Sierra-P rez M (2007) Circularly polarized omnidirectional millimeter wave monopole with parasitic strip elements. Microw Opt Technol Lett 49(3):664–668

Foltz HD, McLean SJ, Crook G (1998) Disk-loaded monopoles with parallel strip elements. IEEE Trans Antennas Propag 46(12):1894–1896

Fu T, Hu B (2008) A new kind compact omnidirectional antenna for wireless communication system. In: 2nd international conference on anti-counterfeiting, Nanjing, P. R. China, pp 433–435

Galindo V, Green K (1965) A near-isotropic circulary polarized antenna for space vehicles. IEEE Trans Antennas Propag 13(6):872–877

Goatley C, Green F (1956) Circularly-polarized biconical horns. IRE Trans Antennas Propag 4(4):592–596

Goubau G, Puri NN, Schwering FK (1982) Diakoptic theory for multielement antennas. IEEE Trans Antennas Propag 30(1):15–26

Grabherr W, Huder B (1999) Omnidirectional slotted-waveguide array antenna. In: 29th European microwave conference, Munich, Germany, 3:134–137

Hao ZC, Hong W, Chen JX, Chen XP, Wu K (2005) Compact super-wide bandpass substrate integrated waveguide (SIW) filters. IEEE Trans Microw Theory Tech 53(9):2968–2977

Hasse RD, Hunsicker W, Naishadham K, Elsherbeni AZ, Kajfez D (2012) Design of a planar segmented circular loop antenna for omnidirectional radiation at 5.8 GHz. IEEE Antennas Wirel Propag Lett 11:1402–1405

Herscovici N, Sopus Z, Kildal S (2001) The cylindrical omnidirectional patch antenna. IEEE Trans Antennas Propag 49(12):1746–1753

Heydari B, Afzali A, Goodarzi HR (2009) A new ultra-wideband omnidirectional antenna. IEEE Antennas Propag Mag 51(4):124–130

Hisao FR, Wong KL (2005) Low-profile omnidirectional circularly polarized antenna for WLAN access points. Microw Opt Technol Lett 46(3):227–231

Hong W, Sarabandi K (2009) Low-profile, multi-element, miniaturized monopole antenna. IEEE Trans Antennas Propag 57(1):72–80

Hua G, Hong W, Sun XH, Zhou HX (2008) Design of an omnidirectional line array with SIW longitudinal slot antenna. Int Conf Microw Millimeter Wave Technol 3:1114–1117

Iwasaki H, Chiba N (1999) Circularly polarised back-to-back microstrip antenna with an omnidirectional pattern. IEE Proc Microw Antennas Propag 146(4):277–281

Jayakumar I, Garg R, Sarap B, Lal B (1986) A conformal cylindrical microstrip array for producing omnidirectional radiation pattern. IEEE Trans Antennas Propag 34(10):1258–1261

Judasz TJ, Balsley BB (1989) Improved theoretical and experimental models for the coaxial collinear antenna. IEEE Trans Antennas Propag 37(3):289–296

Kelleher K, Morrow C (1955) Omnidirectional circularly-polarized antennas. IRE Int Conv Rec 1:28–31

Kim KH, Kim JU, Park SO (2005) An ultrawide-band double discone antenna with the tapered cylindrical wires. IEEE Trans Antennas Propag 53(10):3403–3406

Kim DS, Ahn CH, Im YT, Lee SJ, Lee KC, Park WS (2007) A windmill-shaped loop antenna for polarization diversity. IEEE Int Symp Antennas Propag 1:361–364

Knudsen HL (1956) Radiation from ring quasi-arrays. IRE Trans Antennas Propag 4(3):452–472

DOI 10.1007/978-981-4560-75-7_52-1

Kraus JD, Marhefka RJ (2008) Antennas for all applications. McGraw-Hill, Boston

Kuga N, Arai H, Goto H (1998) A notch-wire composite antenna for polarization diversity reception. IEEE Trans Antennas Propag 46(6):902–906

Kyouichi I, Tanaka M (2000). A horizontally polarized slot-array antenna on a coaxial cylinder. Asia Pac Microw Conf 1444–1447

Lee JG, Lee JH (2007) Zeroth order resonance loop antenna. IEEE Trans Antennas Propag 35(3):994–997

Li WW, Leung KW (2013) Omnidirectional circularly polarized dielectric resonator antenna with top-loaded alford loop for pattern diversity design. IEEE Trans Antennas Propag 61(8):4246–4256

Li P, Luk KM, Lau KL (2003) An omnidirectional high gain microstrip antenna array mounted on a circular cylinder. IEEE Int Symp Antennas Propag 4:698–701

Li Y, Zhang Z, Feng Z (2012) Compact azimuthal omnidirectional dual-polarized antenna using highly isolated co-located slots. IEEE Trans Antennas Propag 60(9):4037–4045

Li B, Loia SW, Xue Q (2013) Omnidirectional circularly polarized antenna combining monopole and loop radiators. IEEE Antennas Wirel Propag Lett 12:607–610

Li Y, Zhang Z, Feng Z, Iskander MF (2014) Design of omnidirectional dual-polarized antenna in slender and low-profile column. IEEE Trans Antennas Propag 62(4):2323–2326

Lin CC, Kuo LC, Chuang HR (2006) A horizontally polarized omnidirectional printed antenna for WALN applications. IEEE Trans Antennas Propag 54(11):3551–3556

Lindenblad NE (1941) Antennas and transmission lines at the Empire State television station. Communications 21:10–14

Liu L, Caloz C, Itoh T (2002) Dominant mode leaky-wave antenna with backfire-to-endfire scanning capability. Electron Lett 38(23):1414–1416

Liu YT, Su CW, Tang CL, Fang AT, Wong KL (2004) Radiation pattern control for an on-ceiling omnidirectional monopole antenna. Microw Opt Technol Lett 41(2):106–108

Locatelli A, Capobianco A, Boscolo S, Modotto D, Midrio M, De Angelis C (2012) Low-profile CRLH omnidirectional loop antenna for mobile wireless communications. In: 42nd European microwave conference, Amsterdam, The Netherlands, pp 401–403

Ma H, Wu D (2006) A wideband E-plane omnidirectional antenna. In: 7th international symposium on antennas, propagation and EM theory, Guilin, P. R. China, pp 1–2

Masa-Campos JL, Fernandez JL, Sierra-Perez JM, Fernandez-Jambrina JLF (2007) Omnidirectional circularly polarized slot antenna fed by a cylindrical waveguide in millimeter band. Microw Opt Technol Lett 49(3):638–642

McDonald NA (1978) Approximate relationship between directivity and beamwidth for broadside collinear arrays. IEEE Trans Antennas Propag 26(2):340–341

McEwan NJ, Abd-Alhameed RA, Ibrahim EM, Excell PS, Gardiner JG (2003) A new design of horizontally polarized and dual-polarized uniplanar conical beam antennas for HIPERLAN. IEEE Trans Antennas Propag 51(2):229–237

McLean J, Foltz H, Crook G (1999) Broadband, robust, low profile monopole incorporating top loading, dielectric loading, and a distributed capacitive feed mechanism. IEEE Int Symp Antennas Propag 3:1562–2565

Nagasawa K, Matsuzuka I (1988) Radiation field consideration of biconical horn antenna with different flare angles. IEEE Trans Antennas Propag 36(9):1306–1310

Nakayama I, Kawano T, Nakano H (1999) A conformal spiral array antenna radiating an omnidirectional circularly-polarized wave. IEEE Int Symp Antennas Propag 2:894–897

Narbudowicz A, Xiu L, Ammann MJ (2013) Dual-band omnidirectional circularly polarized antenna. IEEE Trans Antennas Propag 61(1):77–83

Ong YS, Qing X, Goh CK, Chen ZN (2010) A segmented loop antenna for UHF near-field RFID. IEEE Int Symp Antennas Propag 1–4

Orefice M, Pirinoli P (1993) Dual reflector antenna with narrow broadside beam for omnidirectional coverage. Electron Lett 29(25):2158–2159

Park BC, Lee JH (2012) Dual-band omnidirectional circularly polarized antenna using zeroth- and first-order modes. IEEE Antennas Wirel Propag Lett 11:407–410

Pino G, Acuna AMA, Lopez JOR (2000) An omnidirectional dual-shaped reflector antenna. Microw Opt Technol Lett 27(5):371–374

Pozar DM (1993) Directivity of omnidirectional antennas. IEEE Trans Antennas Propag 35(5):50–51

Qing X, Chen ZN (2012) Horizontally polarized omnidirectional segmented loop antenna. In: 6th European conference on antennas and propagation (EUCAP), Prague, Czech, pp 2904–2907

Qing X, Chen ZN (2013) A compact metamaterial-based horizontally polarized omnidirectional antenna array. IEEE Int Symp Antennas Propagation 1792–1793

Qing X, Chen ZN (2014) Metamaterial-based wideband horizontally polarized omnidirectional 5-GHz WLAN antenna array. In: 8th European conference on antennas and propagation, Hague, The Netherlands, pp 605–608

Qing X, Chen ZN, Chia MYW (2005) UWB characteristic of disc cone antenna. IEEE Int Work Antenna Technol 97–100

Qing X, Chen ZN, Goh CK (2012) A horizontally polarized omnidirectional slot antenna array. IEEE Int Symp Antennas Propag 1–2

Quan X, Li R (2013) A broadband dual-polarized omnidirectional antenna for base stations. IEEE Trans Antennas Propag 61(2):943–947

Quan X, Li R, Tentzeris MM (2013) A broadband omnidirectional circularly polarized antenna. IEEE Trans Antennas Propag 61(5):2363–2370

Riblet HJ (1947) Microwave omnidirectional antennas. Proc IRE 35(5):474–478

Sabatier C (2003) T-dipole arrays for mobile applications. IEEE Antennas Propag Mag 45(6):9–26

Sakaguchi K, Hasebe N (1993) A circularly polarized omnidirectional antenna. In: 8th international conference on antennas and propagation, 1:477–480

Sanada A, Caloz C, Ito T (2003) Novel zeroth-order resonance in composite right/left-handed transmission line resonators. Asia Pac Microw Conf 3:1588–1591

Sangster AJ, Wang H (1995) Moment method analysis of a horizontally polarised omnidirectional slot array antenna. IEE Proc Microw Antennas Propag 142(1):1–6

Silva FJD, Bergmann JR (2005) Classical axis-displaced dual-reflector antennas for omnidirectional coverage. IEEE Trans Antennas Propag 53(9):2799–2908

Takenoshita T, Fujii M (1998) Development of a laminated waveguide. IEEE Trans Microw Theory Tech 46(12):2438–2443

Taniguchi T, Kobayashi T (2002) An omnidirectional and low-VSWR antenna for ultra-wideband wireless systems. IEEE Radio Wirel Conf (RAWCON) 145–148

Tokumaru S (1976) Multiplates: low profile antennas. IEEE Int Symp Antennas Propag 14:379–382

Wang JJH, Workman SC, Triplett DJ (2011) Ultra-wideband omnidirectional low-profile miniaturized 3-dimensional (3-D) traveling-wave (TW) antenna. IEEE Int Symp Antennas Propag 3301–3304

Wang Y, Yang L, Gong S, Feng X, Wang J (2012) Omnidirectional antenna with dipole array. In: 10th international symposium on antennas propagation and EM theory (ISAPE), Xian, P. R. China, pp 347–348

Wei K, Zhang Z, Feng Z, Iskander MF (2012) A MNG-TL loop antenna array with horizontally polarized omnidirectional patterns. IEEE Trans Antennas Propag 60(6):2702–2710

Wheeler HA (1947) A helical antenna for circular polarization. Proc IRE 35(12):1484–1488

Wheeler H (1956) A vertical antenna made of transposed sections of coaxial cable. IRE Int Conv Rec 4(1):160–164

Willoughby EO, Heider E (1959) Laboratory development notes – omnidirectional vertically polarized paraboloid antenna. IRE Trans Antennas Propag 7(2):201–203

Wong KL, Su SW (2005) Broadband omnidirectional ushaped metal-plate monopole antenna. Microw Opt Technol Lett 44(4):365–369

Wong KL, Wu CH, Chnag FS (2005) A compact wideband omnidirectional cross-plate monopole antenna. Microw Opt Technol Lett 52(2):492–494

Wu DL (1995) Omnidirectional circularly-polarized conformal microstrip array for telemetry applications. IEEE Int Symp Antennas Propag 2:998–1001

Wu BQ, Luk KM (2009) A wideband, low-profile, conical-beam antenna with horizontal polarization for indoor wireless communications. IEEE Antennas Wirel Propag Lett 8:634–636

Wu D, Zhao M, Fan Y, Zhang Y (2007) A wideband 8-element omnidirectional array for wireless system. Microw Opt Technol Lett 49(12):2944–2946

Yan L, Hong W, Hua G, Chen J, Wu K, Cui TJ (2004) Simulation and experiment on SIW slot array antennas. IEEE Microw Wirel Components Lett 14:446–448

Yu Y, Jolani F, Chen Z (2013) A wideband omnidirectional horizontally polarized antenna for 4G LTE applications. IEEE Antennas Wirel Propag Lett 12:686–689

Zang SR, Bergmann JR (2013) Omnidirectional dual-reflector antennas for flexible coverage. IEEE Antennas Wirel Propag Lett 12:821–824

Zou L, Fumeaux C (2011) Horizontally polarized omnidirectional dielectric resonator antenna. Asia-Pac Microw Conf 849–852

Zou L, Abbott D, Fumeaux C (2012) Omnidirectional cylindrical dielectric resonator antenna with dual polarization. IEEE Antennas Wirel Propag Lett 11:515–518

Zurcher JF (2013) Tripod omnidirectional low profile antenna: a vertically polarized antenna with 90% bandwidth. Microw Opt Technol Lett 55(3):516–521

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_53-1

Antenna Design for Diversity and MIMO Application

Zhinong Ying[a]*, Chi-Yuk Chiu[b], Kun Zhao[a,d], Shuai Zhang[c] and Sailing He[d]
[a]Research and Technology, Sony Mobile Communications AB, Lund, Sweden
[b]Sony Mobile Communications, Beijing, China
[c]Department of Electronic Systems, Aalborg University, Aalborg, Denmark
[d]Department of Electromagnetic Engineering, KTH Royal Institute of Technology, Stockholm, Sweden

Abstract

Recently, multiple-input multiple-output (MIMO) technology and diversity have attracted much attention both in industry and academia due to high data rate and high spectrum efficiency. By increasing the number of antennas at the transmitter and/or the receiver side of the wireless link, the diversity/MIMO techniques can increase wireless channel capacity without the need of additional power or spectrum in rich scattering environments. However, due to limited space of small mobile devices, the correlation coefficients between MIMO antenna elements are usually very high, and the total efficiencies of MIMO elements would be degraded severely due to mutual couplings. In addition, the human body causes high losses on electromagnetic waves. In real applications, the presence of users may result in significant reduction of total antenna efficiencies, and the correlations of MIMO antenna systems are also highly affected. In this chapter, the performance of some basic MIMO antennas as well as recent technologies to improve MIMO antenna performance of portable devices and mobile terminals are reviewed. The interactions between MIMO antennas and human body are also addressed particularly in mobile terminals application.

Keywords

MIMO; Diversity; Multiplexing; Antenna array; Mobile handset antenna; Long Term Evolution (LTE); WiFi; Over-The-Air (OTA) performance; Human body effect; Specific Absorption Rate (SAR)

Introduction

Antenna diversity techniques have been introduced in communication systems for many years since 1960 (Pierce and Stein 1960; Schwartz et al. 1965; Jakes 1974). In the late 1970s, the diversity technique was used to overcome the performance of degradation due to fading environments by using multiantenna systems (Taga 1990; Pedersen and Andersen 1999; Ogawa et al. 2001). Multiantenna systems usually require low mutual coupling loss and low pattern correlation between radiating elements in order to realize good diversity performance. Independent fading signals (diversity branches) are not only achieved by spaced receiving antennas, *space diversity*, but also attained by other techniques such as *frequency diversity, angle-of-arrival diversity, polarization diversity, time diversity, and multipath diversity.* An overall diversity performance of a system is usually a combined result of various diversity mechanisms.

*Email: Ying.zhinong@sonymobile.com
*Email: zhinong_ying@hotmail.com

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_53-1

Since 1985, mobile communication systems have been rapidly evolved from analog system (1G: first-generation system) to digital system (2G: second-generation system) and later to third-generation system (3G), which can support multimedia transmission. To further increase the data transmission rate, MIMO technology has become an important feature in LTE wireless communication systems (4G: fourth-generation system). As a great achievement in wireless communications during the last decade, MIMO can linearly increase channel capacity with an increase in number of antennas and by using space multiplexing, without requiring additional frequency spectrum or power. Moreover, most popular wireless communication systems typically operate in a rich scattering environment such as fourth-generation LTE cell phones and IEEE 802.11n, ac WiFi, in which MIMO exploits to achieve the aforesaid large performance gain.

A multiantenna system can operate as diversity or MIMO scheme depending on the signal-to-noise ratio (SNR) level under a rich scattering circumstance in a latest modern smartphone. The phone also supports multiband and multimode operations. Due to limited space of the small-size mobile terminal, the correlation coefficient between multiple antenna elements could be very high and so the corresponding bandwidths and total efficiencies would be degraded severely.

Furthermore, the human body causes high losses on electromagnetic wave energy. In practical applications, the presence of users usually results in a significant reduction on total antenna efficiency and a change of correlation in a MIMO antenna system. Mobile terminals need to fulfill different standards and requirements. For instance, the RF OTA performance of mobile phones with users' impact has been required by many telecommunication operators. In addition, SAR of a mobile terminal shows the amount of RF radiation emitted by a radio to a human body. SAR values of mobile terminals are strictly limited by governments' regulations. It is a fact that MIMO technology is able to enhance the channel capacity of a system but the complexity of SAR issue will be increased accordingly.

In this chapter, the antenna design for diversity and MIMO applications in mobile terminals will be discussed. In the section "Performance Characterization of Diversity and MIMO Antennas," general terminologies for diversity and MIMO antennas in mobile terminals will be described, such as different diversity schemes, diversity gain, correlation coefficient of multiple antennas, MIMO capacity, and multiplexing efficiency. Then, the fundamental MIMO antenna performances will be analyzed and presented in the section "Performances of certain Fundamental MIMO Antennas." In the section "Decoupling Techniques of Compact MIMO Antenna Designs," various techniques to decouple closely spaced MIMO antennas will be described. The section "Antenna Designs in LTE MIMO Applications" will provide several compact MIMO antenna designs for WiFi and LTE high and low bands in mobile terminals. Lastly, the section "Multi-Antenna SAR and OTA Performance, Field Trial" will present SAR and body loss of a MIMO terminal with different antenna arrangements, as well as some field trial data by using MIMO antennas in a 700 MHz-band LTE network.

Performance Characterization of Diversity and MIMO Antennas

Diversity Technology in Mobile Communications

Space Diversity

The most common and probably the simplest mechanism for achieving diversity branches is space or spatial diversity. Using two antennas separated at a distance, the phase delay lets signals arriving at the antenna elements differ in fading. The minimum spacing between antenna elements at a mobile terminal is usually about 0.5 wavelength, calculated using zero-order Bessel function, in order to acquire sufficient low correlation between fading signals. This does not include the effects of mutual couplings which could

DOI 10.1007/978-981-4560-75-7_53-1

be significant to the system. Antenna spacing smaller than 0.5 wavelength should incorporate other diversity mechanisms. Recently, as operating frequency of mobile systems is getting higher, space diversity has redrawn researchers' attention again.

Frequency Diversity
Frequency diversity is described as the transmission of a signal at two or more different frequency carriers, such that different versions of signals with various independent fading can be achieved. It is a costly mechanism because of the complexity to generate several transmitted signals and to combine signals at the receiver at several different frequencies.

Angle Diversity
Signals arriving at antennas are coming from different angles or directions. Being independent in their fading variations, these signals can be used for angle or angular diversity. At a mobile terminal, angle diversity can be achieved by using two omnidirectional antennas which are acting as parasitic elements to each other. The radiation patterns can be switched in order to manage the reception of signals at different angles.

Time and Multipath Diversities
Time and multipath diversities are related mechanisms which are most likely applied in the digital transmission. Time diversity is achieved by transmitting a single bit of information repetitively at different time intervals. Fading variations for these different repetitions of signal will become independent. Multipath diversity is using time diversity in multipath environments. Repetitive signals will be transmitted to the receiver at different traveling paths.

Polarization Diversity
Transmission of one polarization (depolarized by the propagation medium) gives the possibility of receiving two orthogonal polarizations with fading variations uncorrelated. This diversity mechanism is relatively practical because much smaller size of antennas can be used. It is also the mechanism that is mostly combined with spaced diversity.

MIMO in Wireless Communications

In a conventional wireless system, there is one transmitter and one receiver, which are called single-input single-output (SISO) system. In a condition of narrow band and static environment (frequency and time invariant), the scalar signal mode can be given as

$$\boldsymbol{y} = \boldsymbol{h}\boldsymbol{x} + \boldsymbol{n}, \tag{1}$$

where y is the received signal, x is the transmitted signal, and n is characterized by additive white Gaussian noise (AWGN) with zero mean and variance. For the bandwidth of 1 Hz, the channel capacity of SISO is upper bounded by Shannon (1948)and Winters (1987):

$$\boldsymbol{C} = \log_2\left(1 + \frac{\boldsymbol{E_s}}{\boldsymbol{N_0}}|\boldsymbol{h}|^2\right) \tag{2}$$

where E_s and $|\boldsymbol{h}|^2$ are the total transmit power and the power gain of the scalar channel, respectively. It is noted that the channel capacity of a SISO system is increased logarithmically with an increase in transmit power.

DOI 10.1007/978-981-4560-75-7_53-1

In a MIMO system, it allows to apply several antennas at both transmitting and receiving ends. Several pioneering works (Winters 1987; Foschini and Gans 1998; Telatar 1999) have been carried out on MIMO systems. The channel response is now described by a channel matrix H and can be expressed as

$$\begin{bmatrix} \boldsymbol{h}_{11} & \cdots & \boldsymbol{h}_{1M_R} \\ \vdots & \ddots & \vdots \\ \boldsymbol{h}_{M_T1} & \cdots & \boldsymbol{h}_{M_TM_R} \end{bmatrix} \tag{3}$$

where $\boldsymbol{h_{mn}}$ is the complex transmission coefficient from the element n at Transmitter to the element m at Receiver. In this case, the sampled vector signal model is given as

$$Y = HX + n \tag{4}$$

where Y is the received signal vector at the M_R receive antennas, X is the transmitted signal vector for the M_T transmit antennas, and n is the AWGN vector at the MR receive antennas. Without the channel state information (CSI) at the transmitter, the channel capacity of MIMO systems can be expressed by Winters (1987)

$$\boldsymbol{C} = \sum_{\boldsymbol{i}=1}^{r} \mathbf{log}_2 \left(1 + \frac{\boldsymbol{E_s}}{\boldsymbol{N_0 M_T}} \boldsymbol{\lambda_i} \right) \tag{5}$$

where r is the number of orthogonal subchannels (i.e., rank) and $\boldsymbol{\lambda_i}$s are the eigenvalues of the matrix HH^H (if $M_T < M_R$) or H^HH (if $M_T > M_R$). $(\bullet)^H$ denotes the conjugate transpose (or Hermitian) operator. Eq. 5 can reveal that the high data rate signal of a MIMO system is achieved by providing multiple streams between transmitters and receivers with the same operating frequency channel.

Multiple-Antenna Performance Evaluation

A multiple-antenna system can operate in diversity or MIMO schemes according to the SNR level in a rich scattering circumstance. If the SNR is low, the diversity can be applied to combat the fading. All antennas at the transmitter (or receiver) will send (or receive) the same signals over the same channel. Since the transmitting (or receiving) antennas are not correlated, the possibility for all antennas suffering to the fading deep would be low. So the reliability in the wireless link can be improved. In the high SNR region, the multiple antennas system will work as a MIMO scheme and utilize the fading to provide several uncorrelated channels. The different data can be simultaneously transmitted over different channels within the same operating frequency. As a result, the maximum data rate can be achieved. In this chapter, the figure of metrics for multiple-antenna systems will be addressed. This helps to evaluate the diversity and MIMO performances. Details of parameters for a single antenna (impedance bandwidth, gain, radiation pattern, efficiency, etc.) will not be discussed here but can be found in Balanis (2005).

Diversity Performance

Balanced Branch Power–Mean Effective Gain (MEG)

The imbalanced power of diversity branches will result in a diversity loss which is proportional to the imbalanced level (Plicanic et al. 2009). This imbalance will be affected by the total antenna efficiency. However, since the diversity will be used in any possible circumstances, so the antenna-channel mismatch also plays an important role. Therefore, the mean effective gain (MEG) is more desirable, in which it can

DOI 10.1007/978-981-4560-75-7_53-1

account for all effects of the total antenna efficiency, antenna gain, as well as wireless environment. MEG can be expressed by the following formula (Taga 1990):

$$\boldsymbol{MEG} = \oint\left(\frac{\boldsymbol{XPR}}{\boldsymbol{XPR}+1}\boldsymbol{G_\theta}(\Omega)\boldsymbol{P_\theta}(\Omega) + \frac{1}{\boldsymbol{XPR}+1}\boldsymbol{G_\phi}(\Omega)\boldsymbol{P_\phi}(\Omega)\right)\boldsymbol{d\Omega} \tag{6}$$

where $\boldsymbol{d\Omega} = \sin\boldsymbol{\theta d\theta d\phi}$, $\boldsymbol{G_\theta(\Omega)}$, and $\boldsymbol{G_\phi(\Omega)}$ are the $\boldsymbol{\theta}$ and ϕ components of the realized antenna gain pattern, respectively. It is noted that MEG should be normalized by antenna gains (Ogawa et al. 2001):

$$\oint\left(\boldsymbol{G_\theta}(\Omega) + \boldsymbol{G_\phi}(\Omega)\right)\boldsymbol{d\Omega} = 4\boldsymbol{\pi} \tag{7}$$

The statistical propagation properties of both vertically and horizontally polarized incident radio waves can be represented by $\boldsymbol{P_\theta(\Omega)}$ and $\boldsymbol{P_\phi(\Omega)}$. If $\boldsymbol{P_\theta(\Omega)}$ and $\boldsymbol{P_\phi(\Omega)}$ are separable in elevation and azimuth, they can be described as follows (Pedersen and Andersen 1999):

$$\boldsymbol{P_\theta}(\Omega) = \boldsymbol{P_\theta}(\boldsymbol{\theta},\phi) = \boldsymbol{P_\theta}(\boldsymbol{\theta})\boldsymbol{P_\theta}(\phi) \tag{8}$$

$$\boldsymbol{P_\phi}(\Omega) = \boldsymbol{P_\phi}(\boldsymbol{\theta},\phi) = \boldsymbol{P_\phi}(\boldsymbol{\theta})\boldsymbol{P_\phi}(\phi) \tag{9}$$

Also they satisfy the normalization relations

$$\oint\left(\boldsymbol{P_\theta}(\Omega)\right)\boldsymbol{d\Omega} = \oint\left(\boldsymbol{P_\phi}(\Omega)\right)\boldsymbol{d\Omega} = 1 \tag{10}$$

XPR (cross-polarization ratio) is the ratio of time average vertical $\boldsymbol{P_\theta(\Omega)}$ power to time average horizontal $\boldsymbol{P_\phi(\Omega)}$ power (Knudsen 2001):

$$\boldsymbol{XPR} = \frac{\boldsymbol{P_\theta}(\Omega)}{\boldsymbol{P_\phi}(\Omega)} \tag{11}$$

In an isotropic environment, it can be characterized by $XPR = 1$, $\boldsymbol{P_\theta}(\Omega) = \boldsymbol{P_\phi}(\Omega) = \frac{1}{4\boldsymbol{\pi}}$ (Kildal and Rosengren 2004), and $MEG = \boldsymbol{\eta}/2$, where $\boldsymbol{\eta}$ is the total antenna efficiency (Glazunov et al. 2009).

In order to obtain a good diversity performance, the multiple-antenna system should have balanced

$$MEG_{\text{antenna1}} \approx MEG_{\text{antenna2}} \tag{12}$$

Correlation

Correlation Coefficient Calculation from Radiation Pattern

The correlation of a multiple-antenna system is used to describe how independent across different antenna ports. The complex correlation coefficient has been given in Vaughan and Andersen (2003):

$$\boldsymbol{\rho_c} = \frac{\oint\left(\boldsymbol{XPR}\cdot\boldsymbol{E_{\theta X}}\boldsymbol{E^*_{\theta Y}}\boldsymbol{P_\theta} + \boldsymbol{E_{\phi X}}\boldsymbol{E^*_{\phi Y}}\boldsymbol{P_\phi}\right)\boldsymbol{d\Omega}}{\sqrt{\oint\left(\boldsymbol{XPR}\cdot\boldsymbol{E_{\theta X}}\boldsymbol{E^*_{\theta X}}\boldsymbol{P_\theta} + \boldsymbol{E_{\phi X}}\boldsymbol{E^*_{\phi X}}\boldsymbol{P_\phi}\right)\boldsymbol{d\Omega}\oint\left(\boldsymbol{XPR}\cdot\boldsymbol{E_{\theta Y}}\boldsymbol{E^*_{\theta Y}}\boldsymbol{P_\theta} + \boldsymbol{E_{\phi Y}}\boldsymbol{E^*_{\phi Y}}\boldsymbol{P_\phi}\right)\boldsymbol{d\Omega}}} \tag{13}$$

DOI 10.1007/978-981-4560-75-7_53-1

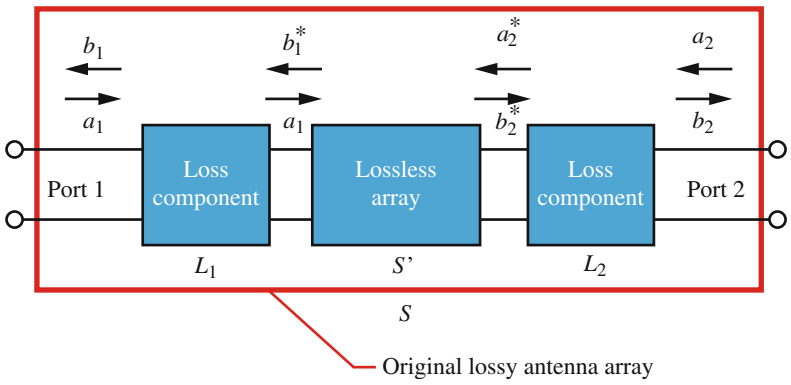


Fig. 1 Correlation coefficient calculation from S-parameters and radiation efficiency

where $\boldsymbol{E}_{\boldsymbol{\theta},\phi\mathbf{X}}$ and $\boldsymbol{E}_{\boldsymbol{\theta},\phi\boldsymbol{Y}}$ are the embedded, polarized complex electric field patterns of two antennas X and Y in a multiple-antenna system. In Rayleigh fading, the envelope correlation coefficient becomes $\boldsymbol{\rho_e} \approx |\boldsymbol{\rho_c}|^2$. The rule of thumb for good diversity/MIMO performance is $\boldsymbol{\rho_e} < 0.5$. From Eq. 13, it is shown that the correlation is also a parameter over certain wireless circumstances. In a real multiple-antenna system, the correlation can be affected by antenna element distance, directivity of antennas, ground plane (or chassis), Q factors, etc.

Correlation Coefficient Calculation from S-Parameters and Radiation Efficiency

In a lossless MIMO antenna system, the correlation coefficient can also be estimated from S-parameters (Blanch et al. 2003). Taking the radiation efficiency into account, the upper bound of the correlation can be calculated (Hallbjorner 2005) (Eq. 14). The authors in Li et al. (2013) use an equivalent circuit-based method to calculate correlation in lossy antenna arrays. The equivalent circuit is shown in Fig. 1, where a lossy component is introduced to simulate the loss. A more accurate estimation of the correlation can be obtained compared with the methods in Hallbjorner (2005) and Li et al. (2013).

$$|\rho_{ij}|_{\max} = \left| \frac{-S_{ii}^* S_{ij} - S_{ji}^* S_{jj}}{\sqrt{\left(1 - |S_{ii}|^2 - |S_{ji}|^2\right)\left(1 - |S_{jj}|^2 - |S_{ij}|^2\right)\eta_i \eta_j}} \right| + \sqrt{\left(\frac{1}{\eta_i} - 1\right)\left(\frac{1}{\eta_j} - 1\right)} \tag{14}$$

Diversity Gain

Diversity gain can be used to describe the gain enhancement of a multiple-antenna system in a combined signal over time-averaged SNR. The best antenna branch is selected as a reference antenna (Vaughan and Andersen 2003; Plicanic 2011), and then the diversity gain (DG) at the possibility $\boldsymbol{P}(\boldsymbol{r_c})$ can be expressed by Vaughan and Andersen (2003):

$$DG = \left. \frac{\left(\frac{\boldsymbol{r_c}}{\boldsymbol{\Gamma_c}}\right)}{\left(\frac{\boldsymbol{r}}{\boldsymbol{\Gamma}}\right)_{\text{Best branch}}} \right|_{\boldsymbol{P}(\boldsymbol{r_c})} \tag{15}$$

DOI 10.1007/978-981-4560-75-7_53-1

where r and r_c are instantaneous SNRs and Γ and Γ_c are mean SNRs for the combined and best antenna branch signals, respectively. Typically, the possibility $P(r_c)$ is 1 % or 50 %.

In general, the total efficiency of the best antenna branch in a compact multiple-antenna system is lower than that of a single-antenna system due to the mismatch, mutual coupling, and additional ohm loss. Effective diversity gain (EDG) is also commonly utilized to make the result more comprehensive (Kildal and Rosengren 2004):

$$EDG = \left.\frac{\left(\frac{r_c}{\Gamma_c}\right)}{\left(\frac{r}{\Gamma}\right)_{\text{Best branch}}}\right|_{P(r_c)} \cdot \eta_{\text{Best branch}} \tag{16}$$

where $\eta_{\text{Best branch}}$ is the total efficiency of the best antenna branch in a compact multiple-antenna system.

As DG and EDG give two extreme values, another parameter called actual diversity gain (ADG) is introduced to provide additional information in which it uses a real single-antenna system as reference (Plicanic 2011; Kildal and Rosengren 2004):

$$DG = \left.\frac{\left(\frac{r_c}{\Gamma_c}\right)}{\left(\frac{r}{\Gamma}\right)_{\text{Single antenna solution}}}\right|_{P(r_c)} \tag{17}$$

MIMO Performance

MIMO Capacity

In the section MIMO in Wireless Communications, it has already been discussed that without the channel state information (CSI) at the transmitter, the power will be equally allocated to the transmitter, and the channel capacity of MIMO systems can be expressed by Winters (1987):

$$\mathbf{C}_{\textbf{equal power}} = \sum_{\mathbf{i}=1}^{\mathbf{r}} \mathbf{log}_2\left(1 + \frac{E_s}{N_0 M_T}\lambda_i\right) \tag{18}$$

where r is the number of orthogonal subchannels (i.e., rank) and λ_is are the eigenvalues of the matrix HH^H (if $M_T < M_R$) or HH^H (if $M_T > M_R$). $(\bullet)^H$ denotes the conjugate transpose (or Hermitian) operator. If the CSI is available at the transmitter, the power allocation can be optimized by a water filling algorithm to the stronger subchannels rather than the weaker ones (Vaughan and Andersen 2003):

$$C_{\text{water filling}} = \sum_{i=1}^{r} \log_2(\lambda_i \cdot D) \tag{19}$$

and

$$D = \frac{1}{\lambda_i} + p_i \tag{20}$$

where D is the "water level" for each of the subchannels to be filled up to and $p_i = \mathbf{SNR}_i/\lambda_i$ is the input power to the i-th subchannel.

DOI 10.1007/978-981-4560-75-7_53-1

Multiplexing Efficiency

In order to evaluate the MIMO performance in a simple way, the multiplexing efficiency (ME) or $\eta_{\mathbf{mux}}$ is introduced according to Tian et al. (2011). *ME* is defined as the power penalty of a realistic multiple-antenna system in achieving a given capacity, compared to an ideal antenna system with 100 % total efficiency and zero correlation (Tian et al. 2011).

With the assumption of high SNR and isotropic environment (i.e., equal likelihood of impinging waves from any directions), *ME* can be expressed by Tian et al. (2011):

$$\eta_{\text{mux}} = \sqrt{\boldsymbol{\eta}_1 \boldsymbol{\eta}_2 \left(1 - |\boldsymbol{\rho}_c|^2\right)}, \tag{21}$$

where η_1 and η_2 are the total efficiencies of the MIMO antenna elements.

In a propagation channel with Gaussian distribution of the angle of arrival (AoA) given by Eq. 9 in Tian et al. (2011), the MIMO multiplexing efficiency can also be evaluated with the following assumptions: The mean incidence direction is denoted by ϕ_0 and $\boldsymbol{\theta}_0$ (as opposed to the isotropic channel, the likelihood of impinging waves is not the same in all directions but has a maximum at AoA ϕ_0 and $\boldsymbol{\theta}_0$). It is assumed that the angular spread is the same in both elevation and azimuth and approximately equal to 30°. The analysis is further restricted to channels with balanced polarizations, i.e., with cross-polarization ratio $XPR = 1$. For reference, two ideal cross-polarized antennas will give a zero correlation. Following the above conditions, the multiplexing efficiency is a function of the mean incidence direction:

$$\eta_{\text{mux}}(\phi_0, \boldsymbol{\theta}_0) = \sqrt{4\, \boldsymbol{MEG}_1(\boldsymbol{\theta}_0, \phi_0) \boldsymbol{MEG}_2(\boldsymbol{\theta}_0, \phi_0) \left(1 - |\boldsymbol{\rho}_c(\boldsymbol{\theta}_0, \phi_0)|^2\right)} \tag{22}$$

where $\boldsymbol{MEG}_1(\boldsymbol{\theta}_0, \phi_0)$ and $\boldsymbol{MEG}_2(\boldsymbol{\theta}_0, \phi_0)$ are the mean effective gains of each antenna port (Tian et al. 2012) and $\boldsymbol{\rho}_c(\phi_0, \boldsymbol{\theta}_0)$ is the complex envelope correlation of the received signal. This definition could be useful for evaluation of MIMO performance in massive MIMO and millimeter wave systems in the new-generation mobile system (Bentsson et al. 2015).

Discussion

From the introduction and analysis of metrics for multiple-antenna systems, it is found that the total efficiencies of antennas and the correlation between multiple antenna elements will affect both the diversity and MIMO performance. If the propagation channel environment is determined and the multiple antenna elements are symmetrical in a compact mobile terminal, a good diversity/MIMO performance can be achieved with high total antenna efficiencies and low correlation. In addition, multiple and wide impedance bandwidths of diversity/MIMO antenna elements are also important to make the compact terminal satisfy various standards and functions.

In a multiport antenna system, when only one port is fed and the others are terminated with 50-Ω load, the total efficiency can be evaluated by the following equations:

$$\boldsymbol{\eta}_{\text{total}} = \boldsymbol{\eta}_{\text{radiation}} \boldsymbol{\eta}_{\text{mismatching+coupling}} \tag{23}$$

$$\boldsymbol{\eta}_{\text{mismatching+coupling}} = 1 - |\boldsymbol{S_{ii}}|^2 - \sum_{\boldsymbol{j \neq i}} |\boldsymbol{S_{ji}}|^2 \tag{24}$$

where $\boldsymbol{\eta}_{\text{total}}$, $\boldsymbol{\eta}_{\text{radiation}}$, and $\boldsymbol{\eta}_{\text{mismatching+coupling}}$ are the total efficiency, radiation efficiency, and mismatch + mutual coupling efficiency, respectively. Subscripts *i, j* represent the working and terminated ports,

DOI 10.1007/978-981-4560-75-7_53-1

respectively. Mutual coupling ($\sum_{j\neq i}|S_{ji}|^2$) will also decrease total efficiency. Therefore, high efficiency and low correlation can be achieved by reducing mutual coupling between diversity/MIMO elements (Zhang et al. 2013a; Zhao et al. 2014). Typically, this method is effective when multiple antennas are operating at high frequencies (higher than 1700 MHz).

If the radiation efficiency is low in a multiple-antenna system, high losses in various forms are expected. Lower mutual coupling between antenna elements will not help to increase total efficiency. Mobile terminal MIMO/diversity antennas operating at low frequencies (less than 960 MHz) belong to this situation. In addition, the wavelength at low frequency band (700–960 MHz) is comparable with the mobile chassis size. The ground plane will become an effective radiator and make all MIMO/diversity elements have similar radiation patterns. This would degrade the multiple-antenna performance eventually. It is also noted that the low mutual coupling does not reflect the low correlation and high efficiencies in this situation.

In compact MIMO antenna design, the bandwidth requirement is one of key concerns. In fact, all antenna parameters such as antenna impedance matching, antenna efficiency, envelope correlation between antennas, diversity gain, and multiplexing efficiency have their own bandwidths. The single-antenna element bandwidth is mainly affected by the fundamental small antenna limits such as size, material loading, and Q value. However, in multiple-antenna cases, the operation bandwidth will also be affected by the size of the antenna system, antenna imbalance, and mutual coupling between the antenna elements. In the section "Performances of Certain Fundamental MIMO Antennas," the performance of several basic MIMO antennas will be discussed.

In addition, the interactions between multiple antennas and users will reduce the total efficiency of antennas by shifting resonant frequencies, and some of the radiated/received power is absorbed. The presence of human body will highly affect the correlations and efficiency in different situations (talking mode, data mode, or reading mode). More details will be discussed in the section "Multi-Antenna SAR and OTA Performance, Field Trial".

Performances of Certain Fundamental MIMO Antennas

Parallel Dipoles

Two dipole elements are placed parallel to each other and the resonant frequency set to be 850 MHz, The distance between two elements is 0.1λ, and the lengths of dipole antennas are $\lambda/6$, $\lambda/3$, and $\lambda/2$ (Fig. 2). The typically resonant length of a dipole antenna is $\lambda/2$, and thus matching networks have been added on $\lambda/3$ and $\lambda/6$ MIMO dipoles to adjust their resonant frequencies to 850 MHz.

Performances of parallel MIMO dipoles are shown in Fig. 3. Four different parameters including S-parameters, efficiencies (the radiation efficiency (RE) and the total efficiency (TE)), envelope correlation coefficient (ECC), and multiplex efficiency (MUX) are used to evaluate MIMO antenna performances. The impedance bandwidth of an antenna is generally defined with reflection coefficient S11 smaller than −6 dB. However, for MIMO applications, it is also needed to consider the bandwidth of total efficiency that the efficiency is larger than −4 dB, and the bandwidth of envelope correlation coefficient that the coefficient is smaller than 0.5. From Fig. 3a, it can be observed that larger antenna volumes are able to generate larger S11 bandwidths. Also, mutual couplings (S21) in larger antenna volumes are lower in general.

In Fig. 3b, smaller MIMO dipole antennas also have lower antenna efficiency; this is a combined effect from the strong mutual coupling and the matching network. Consequently, the total efficiency of $\lambda/6$ is always lower than −4 dB, and thus its efficiency bandwidth becomes zero.

DOI 10.1007/978-981-4560-75-7_53-1

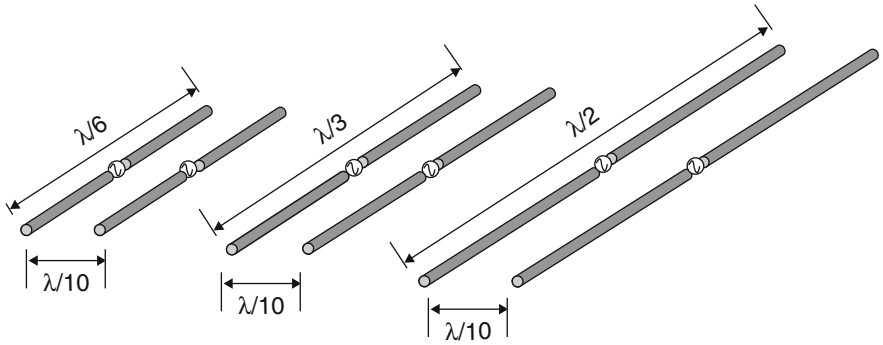


Fig. 2 Structures of fundamental MIMO dipole antennas

It is also noticed that smaller-volume MIMO dipoles have smaller ECC bandwidths. However, the smaller antenna volume leads to a lower ECC at center frequency (Fig. 3c). This is because a smaller-volume antenna always has higher Q value, which can increase the mutual scatter effect to decrease the ECC (Zhang et al. 2013a). Comparing to $\lambda/2$ MIMO dipoles, the MUX is also lower for $\lambda/3$ MIMO dipoles and $\lambda/6$ MIMO dipoles.

When the correlation of MIMO antennas is close to 1, dual-element MIMO antennas can be seen as a single antenna with a double volume. As a result, the efficiency of the antenna could be increased but the MUX remains low due to the high correlation. This phenomenon can be observed at 860 MHz of the $\lambda/6$ MIMO dipoles.

In Zhao et al. (2014) a comprehensive parameter to evaluate MIMO performance has been defined as the overlapping part of S11 bandwidth, TE bandwidth, and ECC bandwidth. The effective bandwidth is similar to MUX bandwidth but with stricter standards in general. Effective bandwidths of proposed MIMO dipole antennas are shown in Table 1. It is easy to notice that smaller-volume MIMO antennas (e.g., $\lambda/3$ MIMO dipoles and $\lambda/6$ MIMO dipoles) are able to achieve better performance at center frequency of certain parameters, but the effective bandwidth is always smaller than large-volume MIMO antennas (e.g., $\lambda/2$ MIMO dipoles). The trend of effective bandwidth coincides with MUX bandwidth but always smaller.

Cross Dipoles, Three Dipoles

Cross Dipoles

Two dipole elements are placed orthogonal to each other and their performances with different antenna volumes shown in Figs. 4 and 5, respectively.

The orthogonal MIMO dipole antennas can perform much better than the parallel dipole antennas. Results of S21 and ECC are omitted here as the two antenna elements are theoretically decoupled with each other and the value of S21 and ECC is too low. Therefore, the dual elements can be seen as two independent single antennas, and this antenna configuration can achieve a high diversity gain and large MIMO channel capacity.

As a result, the MUX bandwidth and the effective bandwidth of orthogonal arranged MIMO dipoles are much larger than parallel cases, which are shown in Table 2. These MIMO antennas perform like two single antennas because of the orthogonality. Their radiation efficiencies are high, but envelope correlation coefficients are extremely low. The effective bandwidth is mainly limited by the antenna volume.

DOI 10.1007/978-981-4560-75-7_53-1

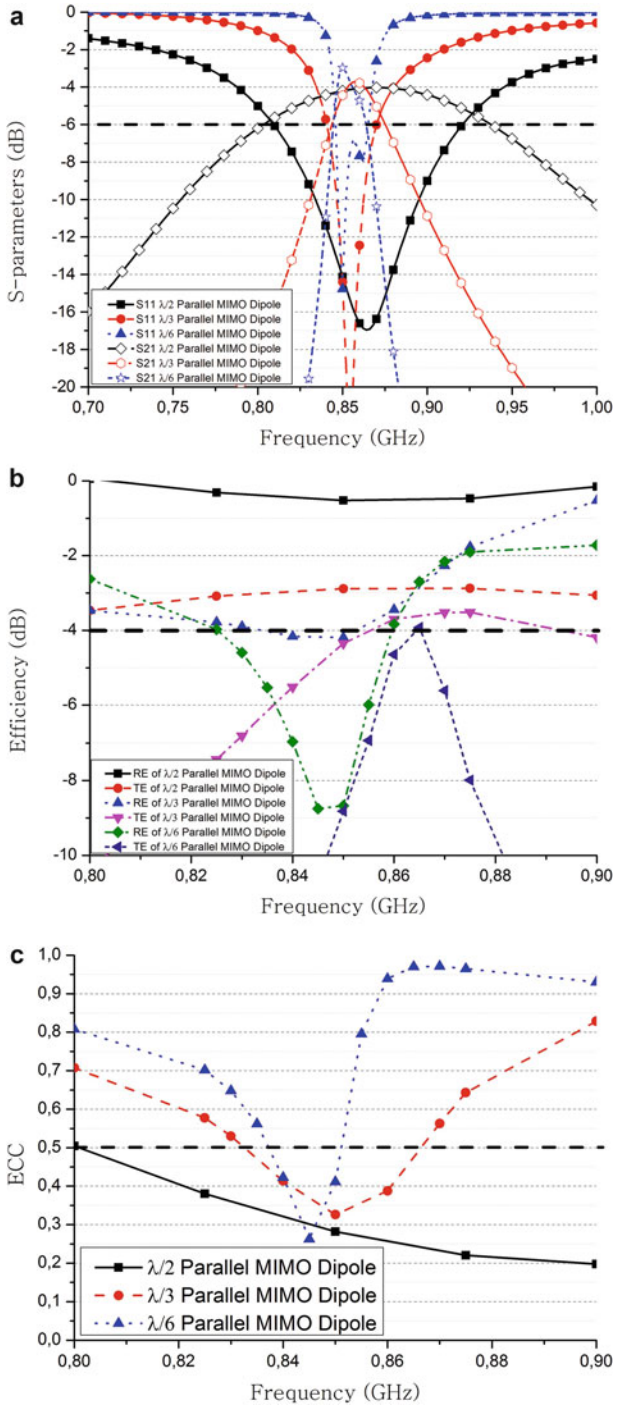


Fig. 3 (continued)

DOI 10.1007/978-981-4560-75-7_53-1

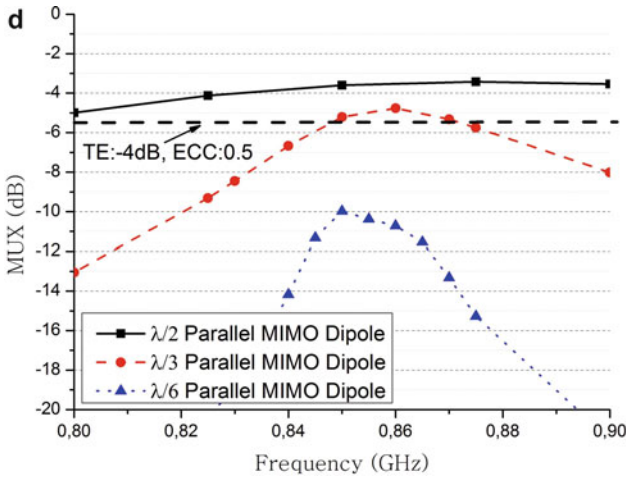


Fig. 3 (**a**) the S-paramters, (**b**) the radiation efficiencies (*RE*) and total efficiencies (*TE*), (**c**) envelope correlation coefficiencts (*ECC*) and (**d**) mutiplex effieicny (*MUX*) of MIMO dipoles with different antenna lengths

Table 1 The effective bandwidth of parallel MIMO dipoles with matching networks

	λ/2 MIMO dipoles	λ/3 MIMO dipoles	λ/6 MIMO dipoles
S11 bandwidth (<−6 dB)	800–920 MHz	840–870 MHz	840–860 MHz
TE bandwidth (>−4 dB)	770–950 MHz	850–890 MHz	Null
ECC bandwidth (<0.5)	> 800 MHz	830–860 MHz	837–850 MHz
Effective bandwidth	800–920 MHz	850–860 MHz	Null

Three Dipoles

It has been suggested by Andrews et al. (2001) that, with three orthogonal components of electric fields and magnetic fields, it is possible to obtain six independent channels at a single point with full angular spread of paths over a sphere.

In Chiu et al. (2007a), it is reported that three-port orthogonally polarized antennas can be realized with dipoles or half-slot antennas. For example, tripolarized dipoles can be constructed by integrating three dipoles in which antenna elements are fabricated on FR4 epoxy boards. The overall size of each board is 51 mm × 51 mm × 1.6 mm. It should be noted that one of the dipoles needs to be offset from the intersecting point of other two dipoles, so that a slot can be cut for assembling (Fig. 6a).

The experiment results of the tripolarized dipoles show that all antennas are able to resonate at similar frequency of 2.5 GHz. The average mutual coupling between antenna elements is below −25 dB, except mutual coupling between *P*ort 2 and *P*ort 3 which is −18 dB. This is because dipole 2 and dipole 3 are constructed on the same FR4 board on the opposite side and there is a small overlap between the two dipoles in fabrication (balun is not shown in the figure).

Three Half-Slot Antennas

A half-slot antenna is shown in Fig. 7a in which it is fabricated on an FR4 epoxy board with an overall size of 22 mm × 27 mm × 1.6 mm. The half-slot antenna is evolved from a standard slot antenna with half-wavelength resonant length. For a standard slot antenna, its midpoint appears electrically open after being transformed by quarter-wavelength from the feed. Therefore, a half-slot antenna can be formed to operate at the same frequency as a normal slot antenna but with its length reduced by half. Tripolarized antennas can then be formed by integrating three of these half-slot antennas orthogonally, as illustrated in Fig. 7b.

DOI 10.1007/978-981-4560-75-7_53-1

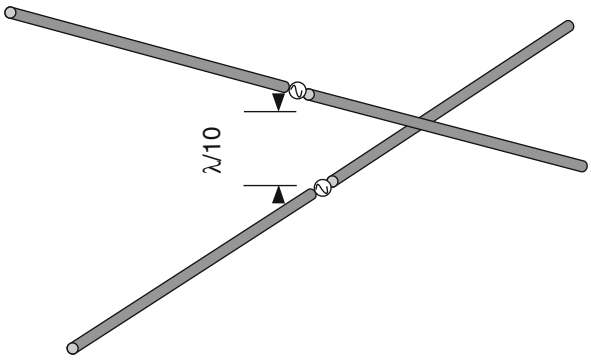


Fig. 4 The structure of orthogonal MIMO dipole antennas

Experimental results show that the worst-case mutual coupling at 2.58 GHz is −21 dB, although three antenna elements are closely packed together.

A system would have an ideal diversity performance if the signal correlation coefficients were zero and the mean received signal-to-noise ratios (SNRs) were equal (Vaughan and Anderson 1987). In case of a practical system using maximum ratio combining at the receiver, a condition for good diversity action is $\rho < 0.8$, where ρ is the signal correlation coefficient (Norklit et al. 2001). For practical antennas with nonideal radiation efficiency, the upper bound of the signal correlation coefficient can be obtained from Hallbjorner (2005), where $|\rho_{eij}|_{\max}$ is the upper bound of the signal correlation coefficient between antenna i and antenna j and η_i and η_j are the radiation efficiencies of antenna i and antenna j, respectively. From the previous expression, the maximum values of the signal correlation coefficients of the proposed antennas are computed using the measured data and the results summarized in Table 3. As seen from the table, the upper bounds of the correlation coefficients of the proposed antennas satisfy the criteria given in Norklit et al. (2001). Therefore, it can be concluded that both of the proposed antennas exhibit a good diversity performance.

Coground Monopoles

Monopole-type antennas are more popular in mobile terminal than dipole antennas because of its smaller resonant length and simple feeding structure. In this part, the MIMO performance of ideal dual-element $\lambda/4$ MIMO monopoles with center frequency around 850 MHz is studied. Radiating elements are mounted on a large round ground plane with radius larger than 2λ and the separation distance around 0.1λ, which are shown in Fig. 8.

The S-parameter, total efficiency, and envelope correlation coefficient of coground monopoles are plotted in Fig. 9. Due to the close placement, the mutual coupling is fairly high with this antenna configuration, and the total efficiency is less than −2 dB at the resonant frequency. Similar to those of parallel MIMO dipole antennas, the envelope correlation coefficient is mainly impacted by mutual scatter effects (Zhang et al. 2013a).

The effective bandwidth of illustrated MIMO monopoles is from 810 to 960 MHz, which is better than that of $\lambda/2$ parallel MIMO dipoles but worse than $\lambda/2$ orthogonal MIMO dipoles.

Decoupling Techniques of Compact MIMO Antenna Designs

In the sections "Neutralization Line," "Coupler Design," "Matching Network," and "Scatter Decoupling," coground dual-element $\lambda/4$ MIMO monopoles with different decoupling methods are

DOI 10.1007/978-981-4560-75-7_53-1

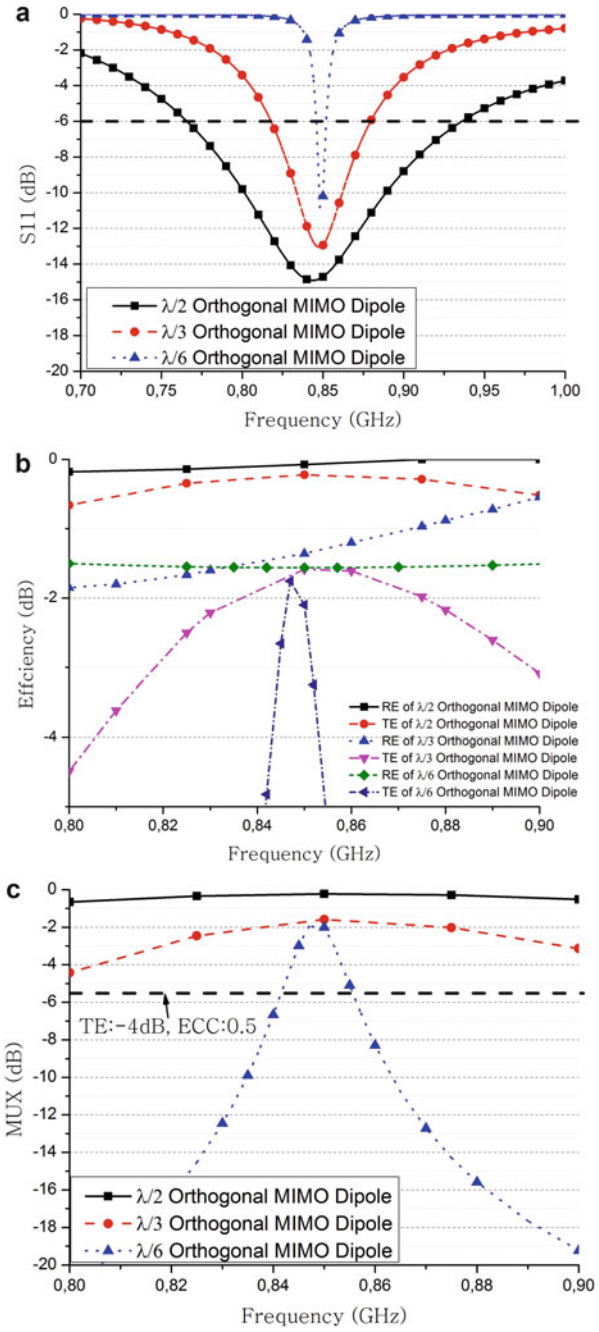


Fig. 5 (**a**) S-paramters of MIMO dipoles with different lengths, (**b**) radiation efficiencies (*RE*) and total efficiencies (*TE*) and (**c**) mutiplex effieicny (*MUX*) of orthogonal MIMO dipoles with different antenna lengths

Table 2 The effective bandwidth of orthogonal MIMO dipoles with matching networks

	$\lambda/2$ MIMO dipoles	$\lambda/3$ MIMO dipoles	$\lambda/6$ MIMO dipoles
S11 bandwidth (<-6 dB)	765–935 MHz	816–880 MHz	846–852 MHz
TE bandwidth (>-4 dB)	All	>800 MHz	845–853 MHz
ECC bandwidth (<0.5)	All	All	All
Effective bandwidth	765–935 MHz	816–880 MHz	846–852 MHz

DOI 10.1007/978-981-4560-75-7_53-1

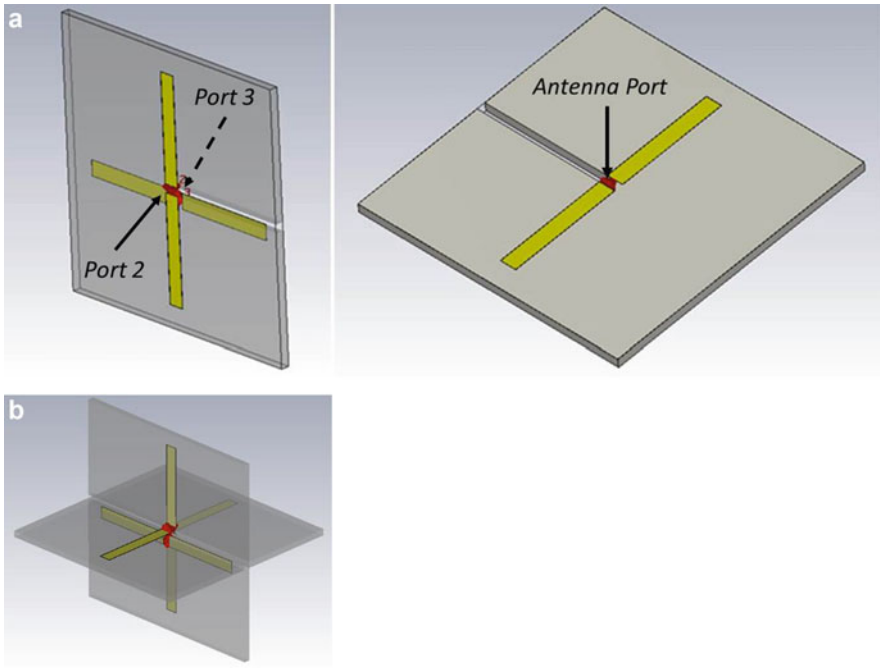


Fig. 6 (**a**) Individual dipole antenna. (**b**) 3-D view of tri-polarized dipoles

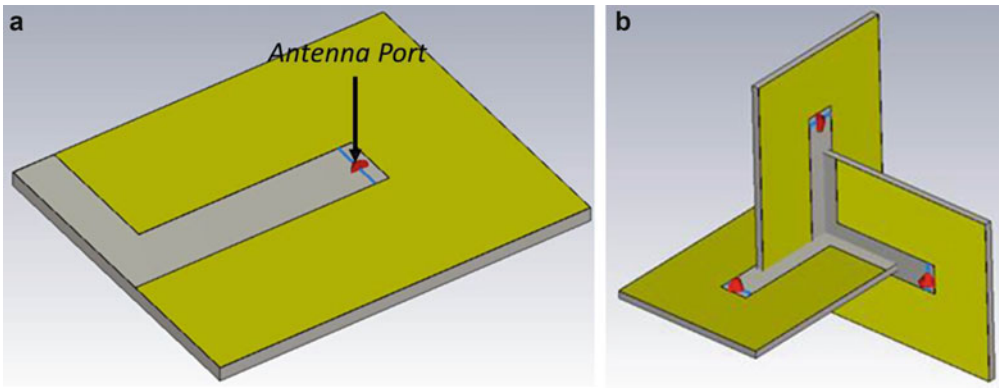


Fig. 7 (**a**) Single half-slot antenna. (**b**) 3-D view of tri-polarized half-slot antennas

studied. The designed center frequency is around 850 MHz. A raw monopole structure has been presented in Fig. 9 for reference.

Neutralization Line

Neutralization line is a metal strip that connects two antennas at some points of radiating elements or feeding lines. The idea was first proposed for PIFAs in Chebihi et al. (2008). The use of neutralization line is to compensate the coupling between two elements. If the neutralization line is positioned in an appropriate way, it is possible to achieve good decoupling bandwidth while keeping the antennas matched. In this study, a strip is used to connect two monopole antennas and antenna structures illustrated in Fig. 10.

DOI 10.1007/978-981-4560-75-7_53-1

Table 3 Calculated signal correlation coefficient of tri-polarized dipoles and half-slot antennas

Port		Dipole	Half-slot
i	j	ρ_{eij}	ρ_{eij}
1	2	0.48	0.36
1	3	0.41	0.38
2	3	0.56	0.46

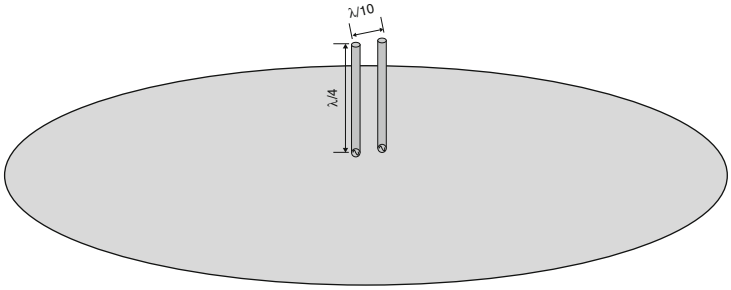


Fig. 8 The co-ground monopole antennas

As the decoupling element has to be inserted between two antennas, the separation of two radiating elements cannot be too small. However, the neutralization line can also be relocated from inner edges to outer edges of PIFAs as demonstrated in Diallo et al. (2008), such that antenna separation can be further reduced. The neutralization line method is also suitable for dual-band MIMO Wi-Fi antennas which are shown in Fig. 11. By changing the length of neutralization line, a proper decoupling frequency can be obtained. The simulation results which include a matching network are presented in Fig. 12.

Coupler Design

Hybrid coupler is a microwave component that divides power from one input to two outputs. There is a fourth port called isolated port which has no power to be delivered theoretically. The way of using hybrid coupler to decouple antennas is to connect antennas to the coupler's input and isolated ports, such that no coupling between two antennas is found ideally. In this example, a 90° hybrid coupler has been designed for operating at 859 MHz with 50 Ω characteristic impedance. The layout of the coupler and the corresponding antennas' S-parameters are shown in Fig. 13a, b, respectively.

Apart from 90° coupler, people can also use 180° coupler for antenna decoupling (Volmer et al. 2008). The 180° coupler can decompose a feeding signal into two orthogonal signals, normally referred to as "even" and "odd" modes, and thus generate a very deep and wideband decoupling. In fact, the coupler can be operated in two different ways. First, signals can be applied to either the sum port or the difference port, according to Fig. 14a. This will result in two equal parts appearing at ports 2 and 3 in phase or separated by 180°. The fourth port will then act as an isolated port. A second way of operating the coupler is to feed two separate signals into ports 2 and 3, in which signals will be adding up at the sum port and subtracting at the difference port. Decoupling antennas can be realized with the second way.

According to Fig. 14b, the coupling is relatively low, almost consistently below −20 dB for the entire simulated frequency span. This is also a direct consequence of the mode decomposition as two ports become orthogonal. The two signals will not influence each other in theory. Although perfect isolation is only strictly true for the center frequency, the decoupling effect is also achieved over a relatively large bandwidth.

DOI 10.1007/978-981-4560-75-7_53-1

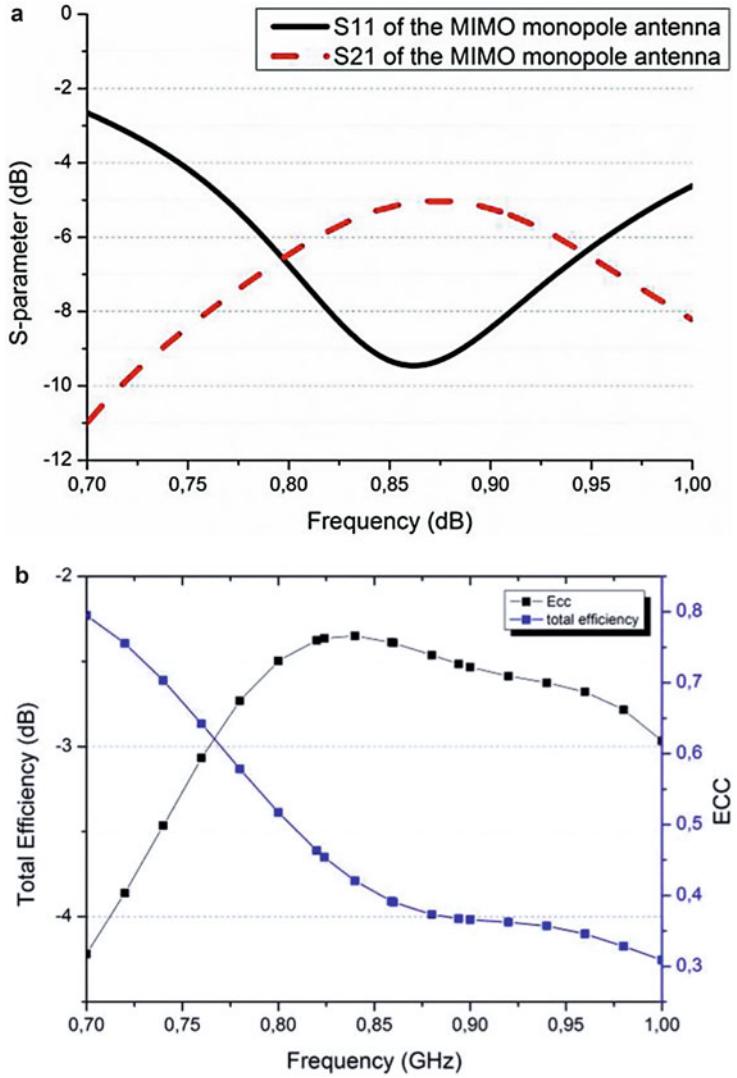


Fig. 9 (**a**) The S-parameters and (**b**) the total efficiency and envelope correlation coefficient of co-ground monopoles

Matching Network

In Pozar (2005), lumped component equivalent circuits for the 90° and 180° hybrid couplers are presented. Since these are much more compact than their distributed counterparts, it is interesting to see how well they would perform. The optimizations are carried out for both full bandwidth (824–894 MHz) and for the center frequency (859 MHz). However, only optimization for the center frequency could be found. The associated layout and S-parameters of the lumped component 90° hybrid coupler are found in Fig. 15a, b, respectively. For the 180° lumped component hybrid coupler, the layout and S-parameters are presented in Fig. 16a, b, respectively.

Scatter Decoupling

Implementing parasitic elements between two antennas is another way to reduce the mutual coupling and the correlation coefficient. The parasitic elements can act as a scatter and be formalized by chokes, parasitic stubs. In order to further reduce the mutual coupling, a new resonant mode was excited in Zhang et al. (2009a, 2010a). In Zhang et al. (2012a) a quarter-wavelength slot is excited by utilizing the edges of

DOI 10.1007/978-981-4560-75-7_53-1

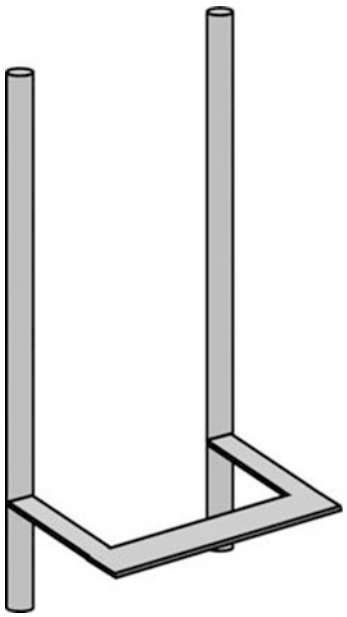


Fig. 10 The structure of neutralization line

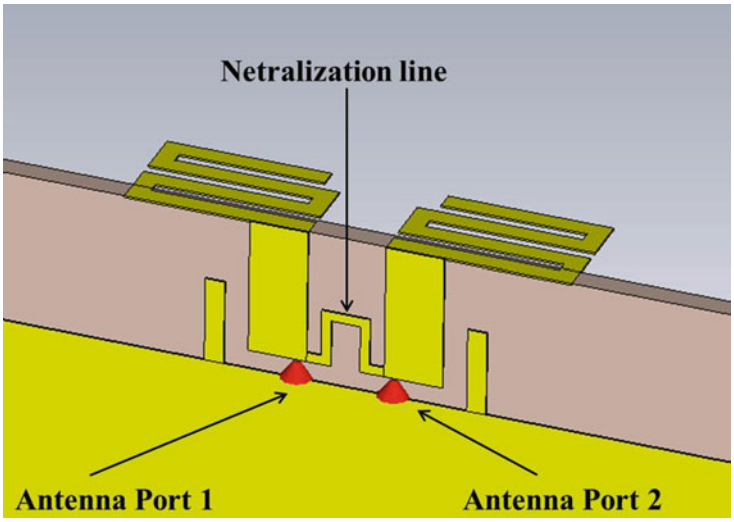


Fig. 11 The dual/multi band Wi-Fi antenna design with the decoupling neutralization line

two PIFAs. Another slot is cut on the ground plane in Ogawa et al. (2001) to form a half-wavelength U-shape slot in between. Apart from trapping the displacement coupling current between the antennas, the new mode in Zhang et al. (2010a) also traps the coupling current on the ground plane in the U-shape slot. Thus, the resonant mechanism of the slot effectively cuts off all coupling currents and allows an isolation of above 40 dB (in simulation) for an inter-PIFA spacing of 0.0016 wavelength. However, the methods in Zhang et al. (2009a, 2010b) cannot be utilized for a ground plane with arbitrary size and shape, (e.g., that of a mobile phone), because it may not be possible to match well enough the decoupling slots in these cases in order to excite them successfully. In Zhang et al. (2012a) a matching stub is added to make naturally formed decoupling slot matched on any kind of size and shape of the ground plane, as shown in

DOI 10.1007/978-981-4560-75-7_53-1

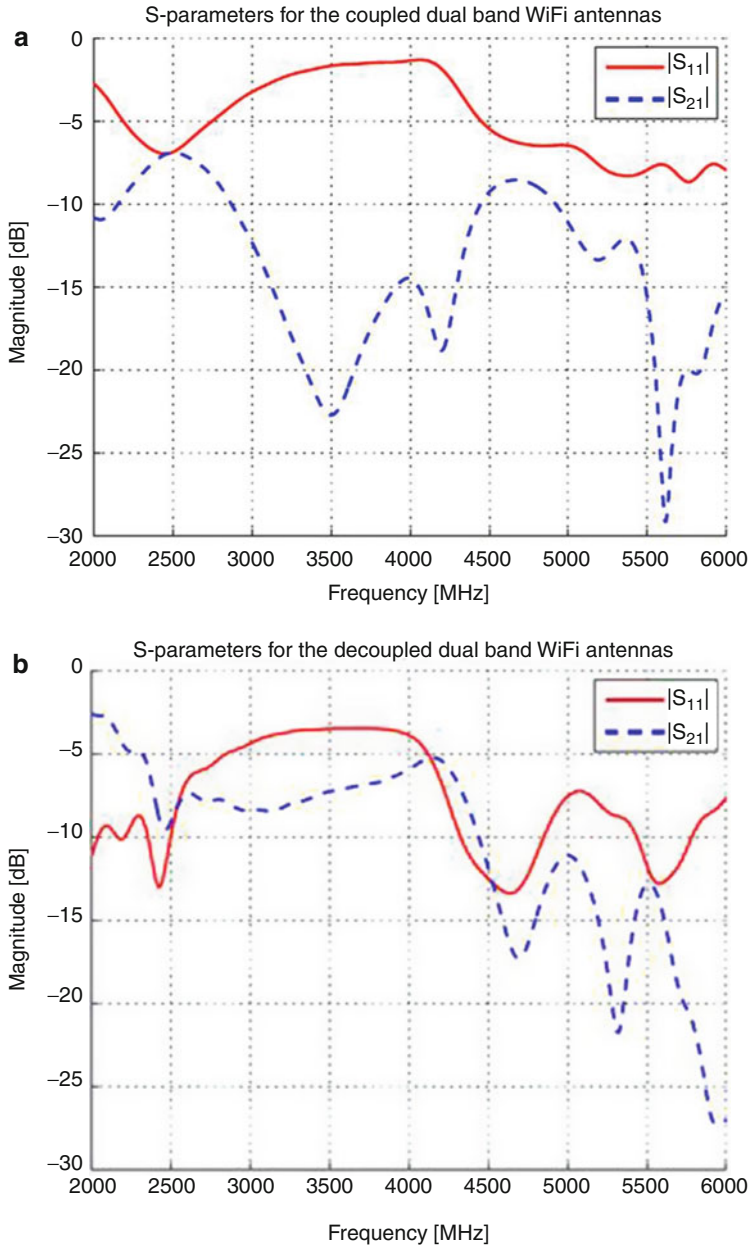


Fig. 12 (**a**) Coupled antennas S-parameters.and (**b**) decoupled antennas S-parameters

Fig. 17. In Chiu et al. (2013), a pair of parasitic stubs is added on sides of a mobile phone chassis to reduce mutual coupling of main and subcellular antenna, which is shown in Fig. 18. Furthermore, the decoupling element technique can also be used in UWB MIMO antennas in a USB dongle (Zhang et al. 2009b).

Ground Plane Modification

The previous technique was also combined with another method which consists in etching one or several thin slits in the PCB, not necessarily between radiators (Chiu et al. 2007b; Zhang et al. 2009) (Fig. 19). This technique can also be viewed as a modification to the shape of PCB ground plane. This physical modification forces the currents circulating on PCB to flow around the slits. In this way, the signal from one antenna port to another arrives with a suitable phase for overall signal cancellation. The same technique was used in Mavridis et al. (2006), Shin and Park (2007), and Chae et al. (2008) to address a

DOI 10.1007/978-981-4560-75-7_53-1

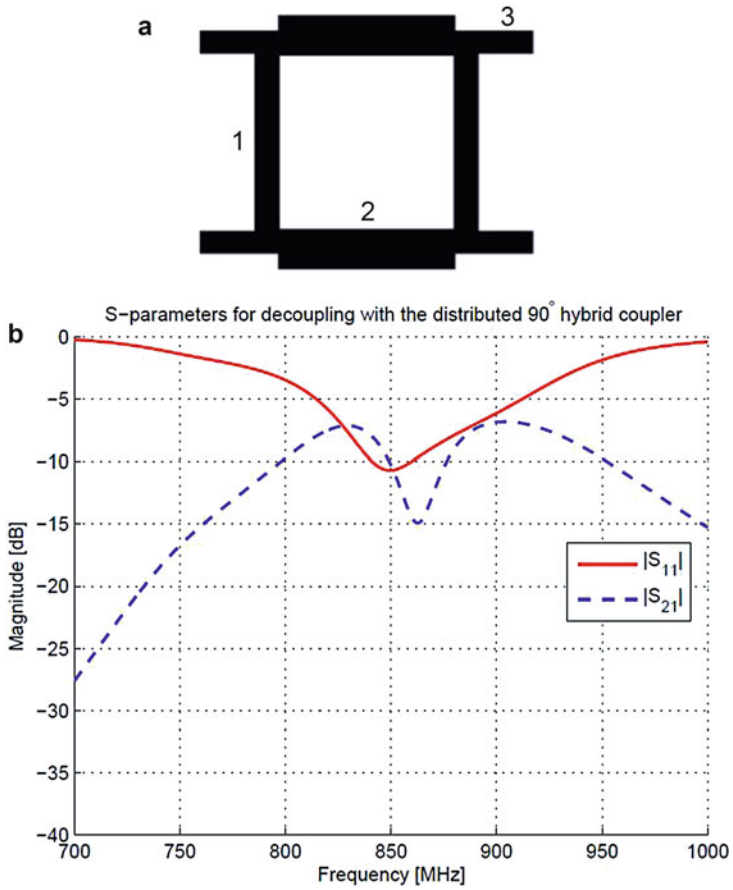


Fig. 13 The 90° hybrid couplers: (**a**) geometric layout, (**b**) S-parameters

dual-band 2.5/5 GHz case. Depending on the shape of the slot, this technique can be described in terms of N-section resonators (Kim and Ahn 2007). The authors of Chi et al. (2005) have made a protruding T-shape extension as a modification to the shape of PCB. Different and complicated extensions of PCB were presented in Liu et al. (2008)and Ding et al. (2007) to enhance the isolation between two or four monopole antennas for UMTS mobile phones. In some structures, the etched slots were even named (by the authors from RIM Blackberry and Sony-Ericsson) as the main radiators (Bolin et al. 2005; Rao and Wang 2009). Unfortunately, although the ground plane modification can achieve impressive broadband isolation enhancement between antenna elements, it may not be suitable for all real applications as every single square millimeter of PCB in smartphones nowadays is used for electronic component accommodation.

Orthogonal Mode Decoupling

Different polarizations have been applied in Tian et al. (2010) to design a very compact three-port dielectric resonant antenna (DRA) (see Fig. 20). The polarization and pattern diversity method was utilized in Zhang et al. (2009b) to realize wideband isolation for UWB MIMO antennas in a USB dongle. The designed MIMO antennas cover the UWB lower band of 3.1–5 GHz with isolation higher than 20 dB.

Another example is a pair of half-wavelength slot and narrow patch antennas stacked together to form a set of nearly collocated antennas with low mutual couplings which is shown in Fig. 21. In Chiu et al. (2009), it is reported that the antenna pair with an overall volume of $40 \times 30 \times 6.6$ mm^3 can exhibit mutual coupling better than -17 dB at around 2.5 GHz. The low mutual coupling between two

DOI 10.1007/978-981-4560-75-7_53-1

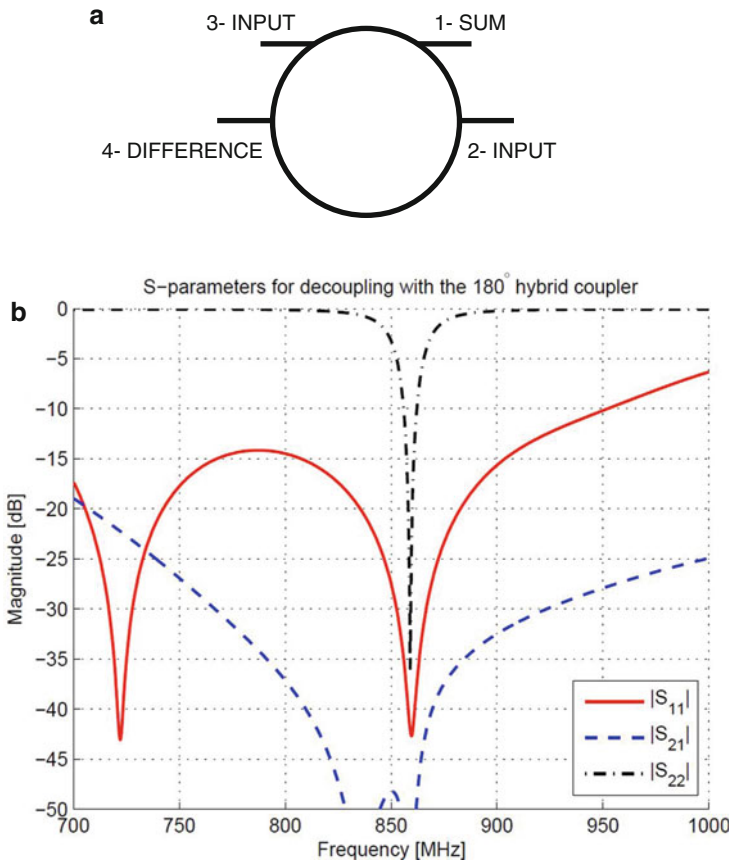


Fig. 14 The 180° hybrid coupler: (**a**) geometric layout, (**b**) S-parameters

closely packed antenna elements is due to orthogonal electric E- and magnetic H-fields received simultaneously. In fact, the idea can be further extended to exploit polarization diversity on each coordinate axis in which three pairs of half-wavelength slot and PIFA antennas are packed at a point in x, y, and z directions. This provides an example of how to realize a compact six-port antenna in which six radiating elements are at a point and are angularly spaced in the Cartesian directions.

Antenna Designs in LTE MIMO Applications

In current and future wireless telecommunication systems, such as LTE and LTE Advanced, MIMO systems are an integral part at mobile terminals. In the LTE standards, several new channels are allocated to the lower frequency bands of 700–960 MHz. As mentioned in the section "Decoupling Techniques of Compact MIMO Antenna Designs," the elements in a MIMO antenna system should have a low correlation and a high total efficiency in order to guarantee a good multiplexing MIMO performance. Unlike the higher frequency bands, the mobile handset MIMO antenna system operating at the lower frequencies will not focus on the reduction of mutual coupling but rather on the improvement of the correlation and efficiency explicitly due to the low radiation efficiency (Li et al. 2013). The wavelengths in the lower frequencies are much longer than those in the higher bands, and this poses some new challenges on the practical realization of having good MIMO performance in mobile terminals: (1) each MIMO antenna element has to be redesigned to obtain a compact structure fitting in the device; (2) the structures for decorrelation have to be small enough and still working well; (3) the MIMO elements and the

DOI 10.1007/978-981-4560-75-7_53-1

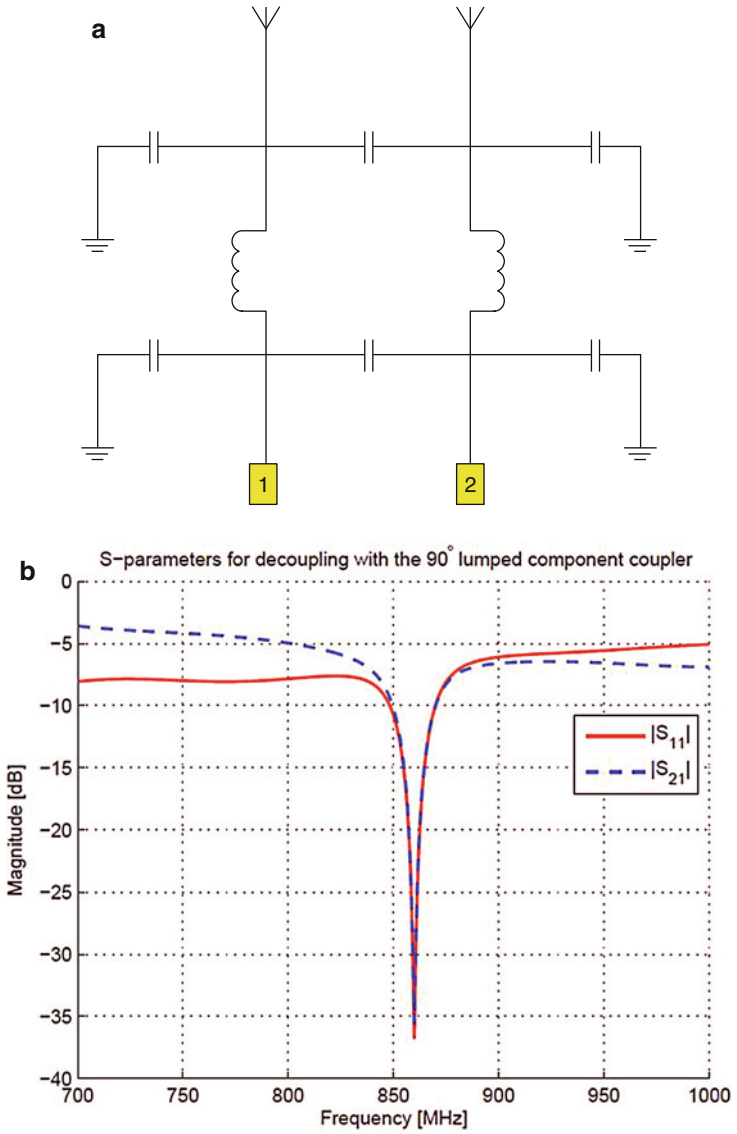


Fig. 15 (**a**) Antennas decoupled with the 90° hybrid coupler lumped component equivalent, (**b**) S-parameters

decorrelating structures are more closely positioned which cause high correlation and low efficiencies; and (4) the chassis mode will be efficiently excited, which makes the radiation pattern of each MIMO element quite similar leading to a very high correlation. An envelope correlation coefficient less than 0.5 and a total efficiency higher than 40 % are reference values for cellular LTE MIMO antennas in the lower bands according to industry researches, including field trials and mock-ups (Ying 2012). Recently, several studies have been done to solve these problems such as to use neutralization line for single-band LTE MIMO antennas in Bae et al. (2010), and Park et al. (2009) or to use decoupling networks for the lower bands in Park et al. (2011), Kim et al. (2011), andLau and Andersen (2012). However, these methods can only be used for very narrow band operations and will cause a large radiation efficiency reduction in practice. In the following, more technologies for correlation reduction and total efficiency improvement in low frequency bands will be reviewed (less than 1 GHz).

DOI 10.1007/978-981-4560-75-7_53-1

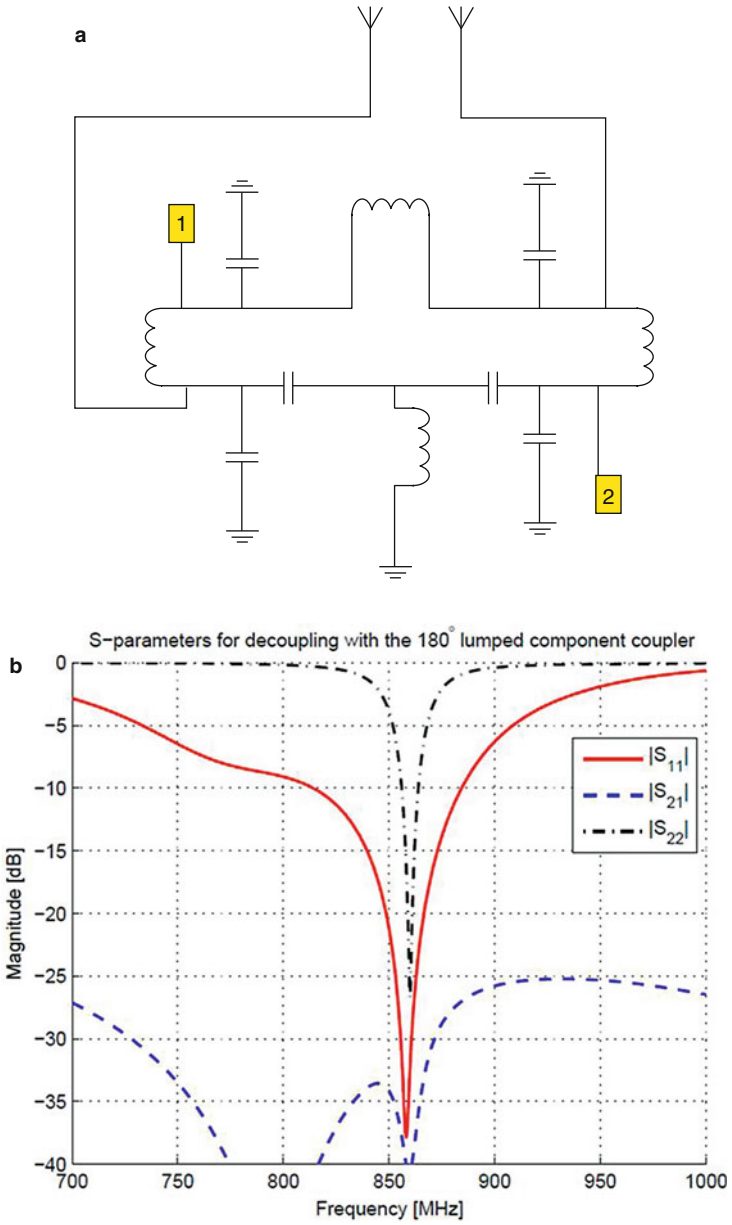


Fig. 16 (**a**) Antennas decoupled with the 180° hybrid coupler lumped component equivalent, (**b**) S-parameters

Chassis Mode and Orthogonal Chassis Mode MIMO Antennas

In Li et al. (2012b), the interactions between MIMO antennas and mobile chassis have been studied. The *E*-fields and *H*-fields of the fundamental mobile chassis characteristic mode are given in Fig. 22. The results reveal that the characteristic modes play an important role in determining the optimal placement of antennas. For example, if two electric antennas are put at the two ends of the mobile chassis, a very high mutual coupling between two antennas is expected.

In mobile terminals, the orthogonal mode MIMO antennas can be realized by combining an electrical antenna (dipole) and a magnetic antenna (slot, loop) (Li et al. 2012a), based on the analysis in Li et al. (2012a). The prototype of the designed orthogonal mode MIMO antennas is presented in Fig. 23 (Li et al. 2012a). A very low mutual coupling (<-20 dB) has been achieved in a narrow band of lower frequencies.

DOI 10.1007/978-981-4560-75-7_53-1

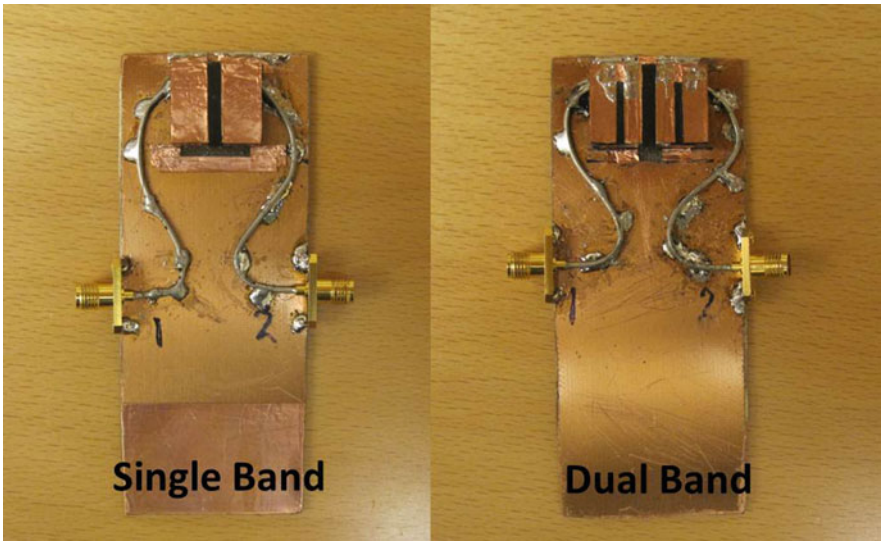


Fig. 17 Single-band and dual-band MIMO antennas with a naturally formed decoupling slot and matching stub

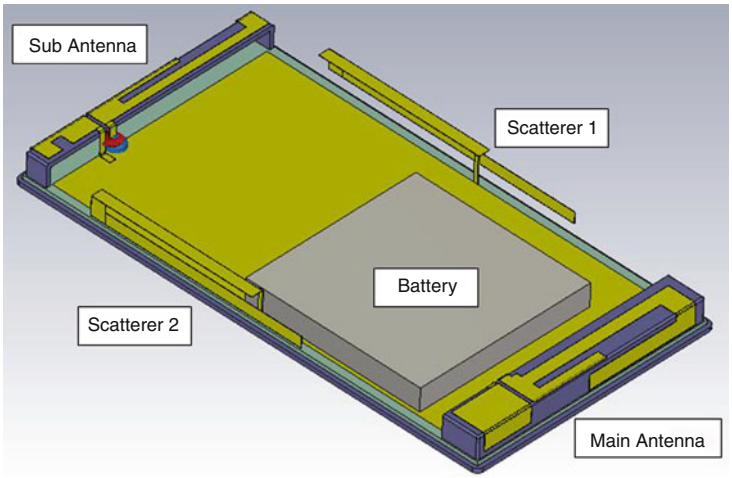


Fig. 18 Parasitic stubs for reducing the correlation coefficient of a dual MIMO antennas in mobile terminal

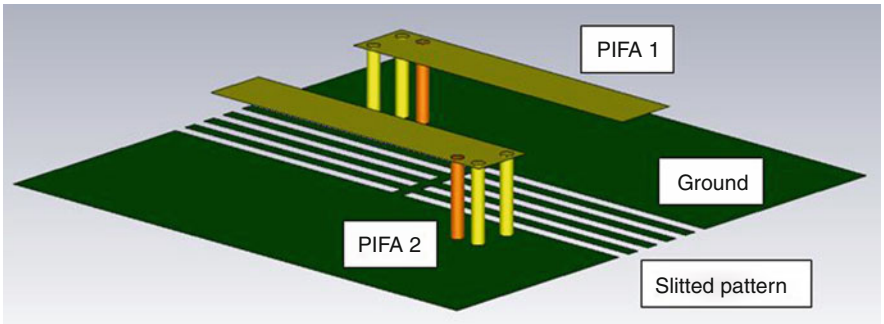


Fig. 19 MIMO antennas with fish-bone decoupling structure

DOI 10.1007/978-981-4560-75-7_53-1

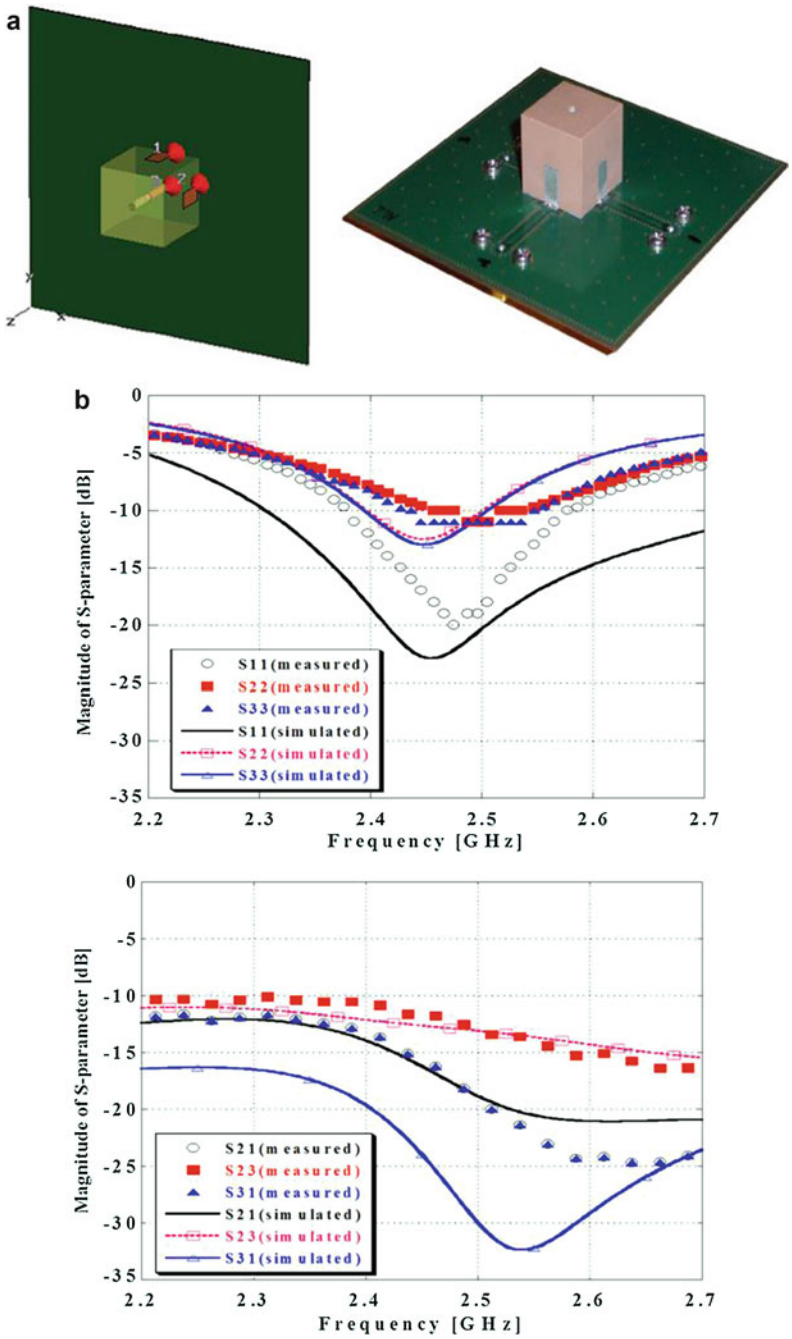


Fig. 20 (**a**) Compact 3-port DRA MIMO antennas and (**b**) measured S-parameters

Localized Mode

In some cases, when two electrical MIMO antenna elements are used in a mobile terminal, a good diversity and decoupling performance can be achieved by exciting different modes for different antennas. For example, Antenna 1 is exciting in chassis mode, whereas Antenna 2 is exciting in localized mode (Li et al. 2011). Typical directive antennas like patch, notch, and balanced dipoles are good candidates to be excited in localized mode. In Fig. 24, it shows the current distributions of a PIFA with different dielectric permittivity. With a higher dielectric permittivity, the current distribution on the PIFA is more localized. The radiation patterns of localized mode antenna will be different compared to the chassis mode

DOI 10.1007/978-981-4560-75-7_53-1

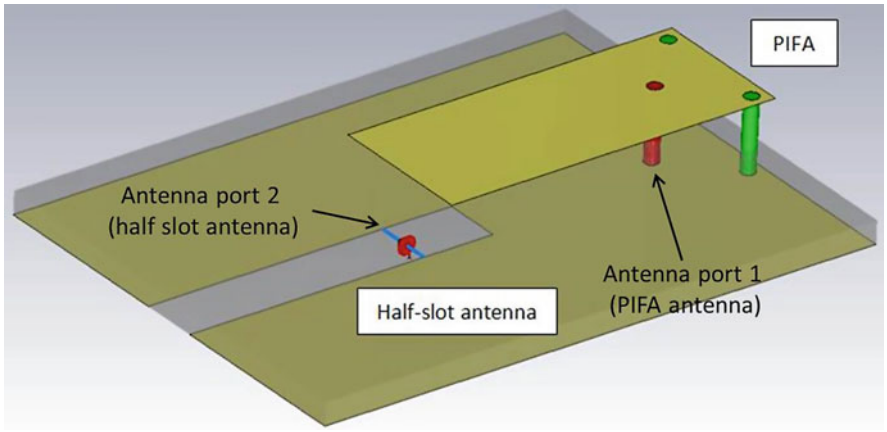


Fig. 21 Geometry of half-wavelength slot and PIFA antenna pair

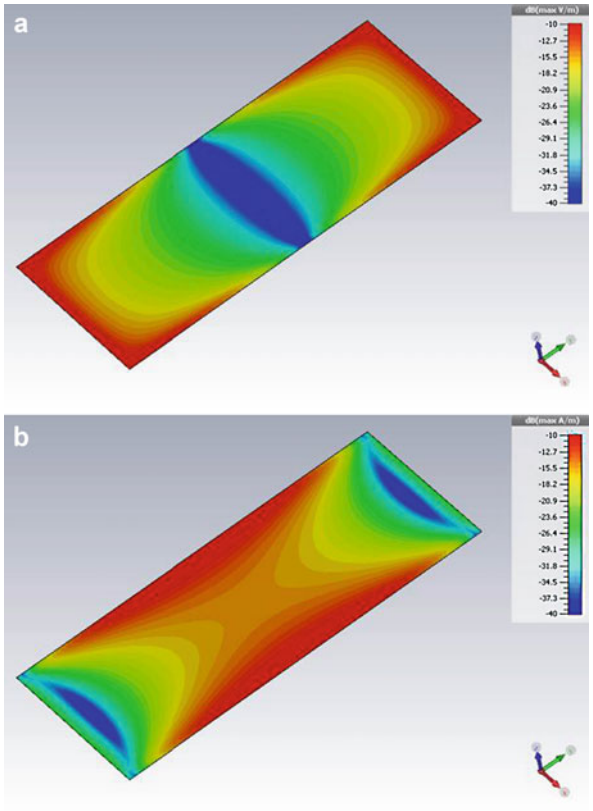


Fig. 22 Characteristic mode of mobile chassis: (**a**) *E*-fields, and (**b**) *H*-fields

antenna, thus the envelope correlation between the two antennas could be reduced. The trade-off of this design would be imbalanced MIMO antenna structures. In Li et al. (2011), current distribution with different permittivity of a MIMO antenna configuration has been given, and the impact of localized mode on MIMO performance have also been presented.

DOI 10.1007/978-981-4560-75-7_53-1

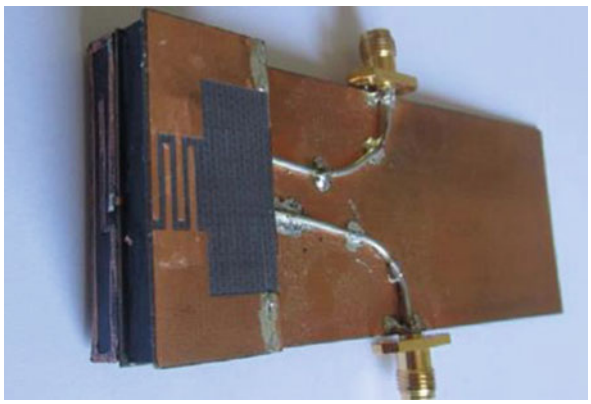


Fig. 23 Orthogonal mode MIMO antennas in a mobile terminal

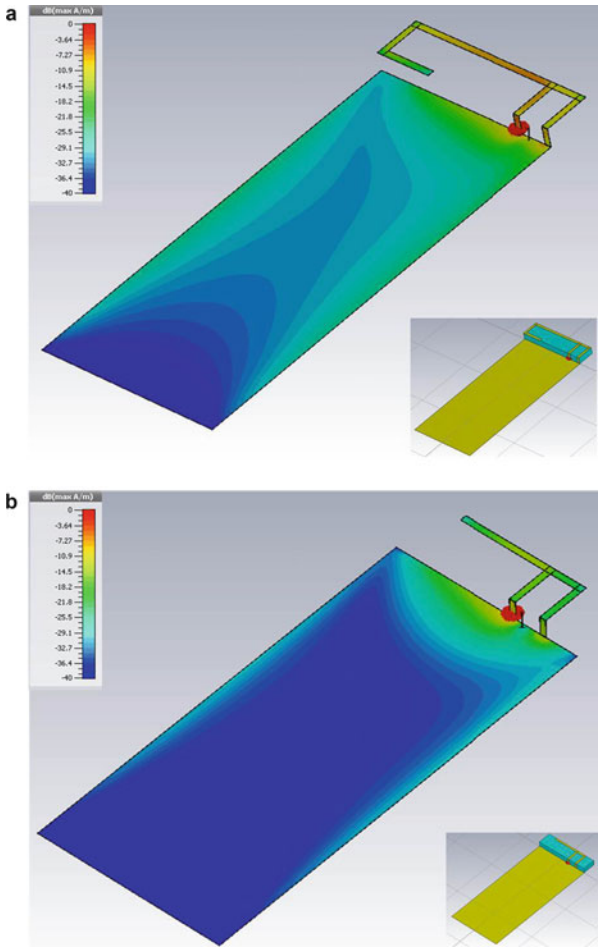


Fig. 24 Normalized magnitude of current distributions for PIFA with: (**a**) Er = 1, and (**b**) Er = 20

Mutual Scattering Mode and Diagonal Antenna-Chassis Mode for MIMO Bandwidth Enhancement

In order to realize a correlation less than 0.5 and total efficiency higher than −4 dB in a wideband operation of low frequency bands, mutual scattering mode (Zhang et al. 2012b, 2013a) and diagonal

DOI 10.1007/978-981-4560-75-7_53-1

antenna-chassis mode (Zhang et al. 2012c, 2013b, 2015a) were proposed for MIMO bandwidth enhancement.

For closely located MIMO antennas, the strong mutual scattering effect can be used to reduce correlation coefficient of the MIMO antenna effectively (Zhang et al. 2012a, b). In general, MIMO antennas with high Q values will have strong mutual scattering effect, in which this can be realized by optimizing impedance matching with lumped elements. In Fig. 25, it is revealed that when two antennas are put in the same end of the mobile terminal, higher Q factor could provide a lower correlation with an improved total efficiency. In addition, a MIMO reference antenna based on the mutual scattering mode had been proposed in Zhang et al. (2013b) for OTA applications.

To combine with a diagonal chassis mode, a better pattern diversity can be achieved in the low LTE band (less than 1 GHz) (Zhang et al. 2012c, 2013b, 2015a). When a terminal antenna operates at low frequency, the antenna element and the ground plane can be seen as two arms of a dipole antenna. Thus, the direction of the radiation pattern of this dipole antenna will be mainly determined by the current distribution on the ground plane. This can be concluded as a fact that the volume of ground plane is much larger than the antenna element. The diagonal chassis mode can be realized by putting the two antenna ports of dual-element MIMO antennas to different corners of the ground plane. As a result, the two MIMO elements can form two orthogonal dipole-like radiation patterns. In Figs. 26 and 27, it can be seen that colocated MIMO antennas have almost opposite patterns due to the effect of diagonal chassis mode.

Decorrelation Scatterer Design

Another common practice of realizing two cellular antennas in an LTE terminal is that a main antenna situates at the bottom, whereas a subantenna locates at the top of the phone. In Zhang et al. (2014), it is reported that envelope correlation can be reduced from 0.7 to 0.2 at 750 MHz after inserting scatterers. Dual-frequency operations are also possible if two scatterers are introduced on different sides of a phone chassis as depicted in Fig. 28. The associated resonant frequencies can be adjusted independently. In general, the two conformal scatterers could be printed either outside or inside of phone housing with groundings to the main PCB. It should be noted that as the scatterer can be thought as a kind of radiator, the far more away from ground plane helps enhance operating frequency bandwidth. However, hand effects are also needed to be considered in real applications.

Inserting a scatterer between two closely packed antennas can improve the isolation because an additional coupling path has been artificially created in which field cancellation occurs and consequently lowers the corresponding envelope correlation. In addition, it should be noted that according to Booker's relation (Chiu et al. 2013), dipole and slot antenna are complementary elements in terms of electromagnetism. Therefore, the scatterer design is also suitable for metallic chassis phone to lower envelope correlation in such a way that a similar slot structure is etched on the housing.

Ring Antenna and Less Body Sensitive Antenna

In Zhao et al. (2015), as the human body is insensitive to magnetic fields, the author proposed a novel mobile terminal antenna named as double-ring antenna, which is mainly working on magnetic loop mode (Fig. 29). The loop mode is enclosed by the ground plane in which ring 1 is going along the grounding element and the feed line of port 1 (follow the black line in Fig. 30a). As port 2 uses C-fed, the current is not cut off by the feeding line but is enclosed between two groundings (follow the black line in Fig. 30b). Furthermore, it can be seen that loop modes are enclosed by grounding elements, which are as isolators between the two cellular antennas. Based on this method, it is possible to integrate even more ports with a larger perimeter of a metal ring. Another advantage of using ring antenna for MIMO is the low correlation between two slot antennas.

DOI 10.1007/978-981-4560-75-7_53-1

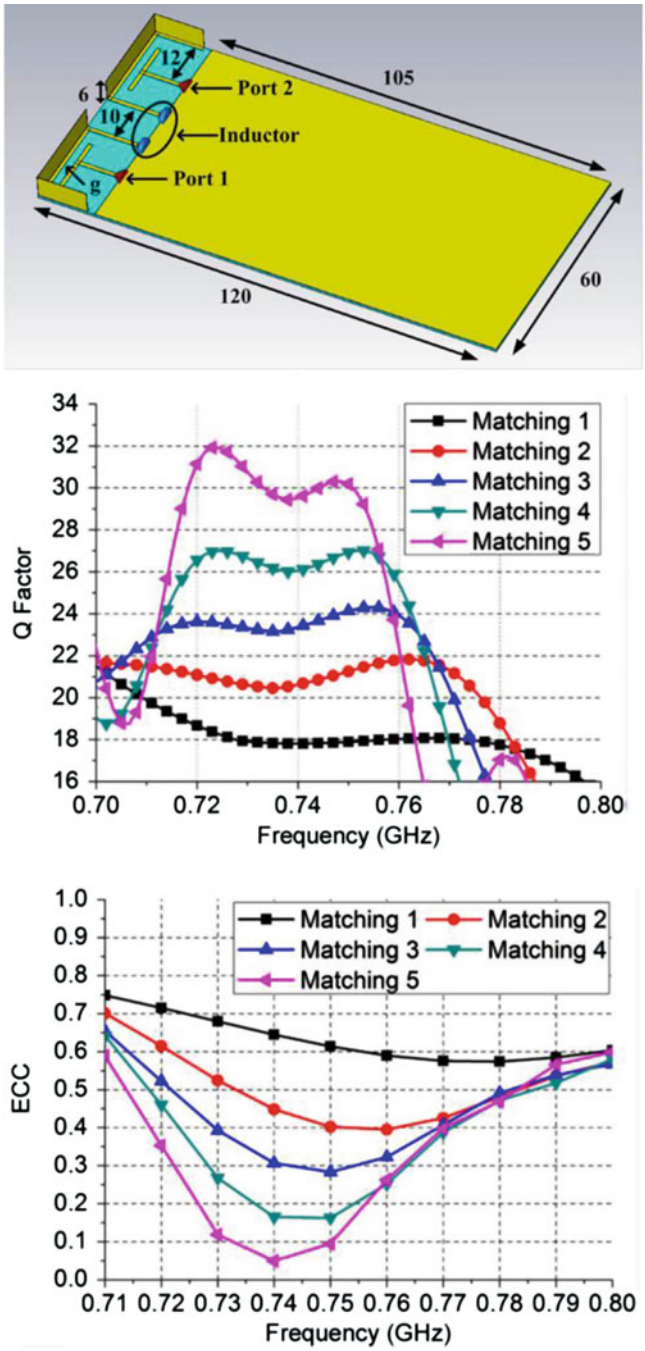


Fig. 25 Mutual scattering mode in mobile terminals

The double-ring antennas are measured with a real human (see Fig. 31a). The comparison of single hand effects on S-parameters between the double-ring antenna and a conventional design are shown in Fig. 31b. The average total efficiency loss on the single hand is presented in Table 4. All results show that the double-ring antenna is able to maintain relatively high total efficiencies in user case when compared to other conventional MIMO antennas in mobile terminals.

DOI 10.1007/978-981-4560-75-7_53-1

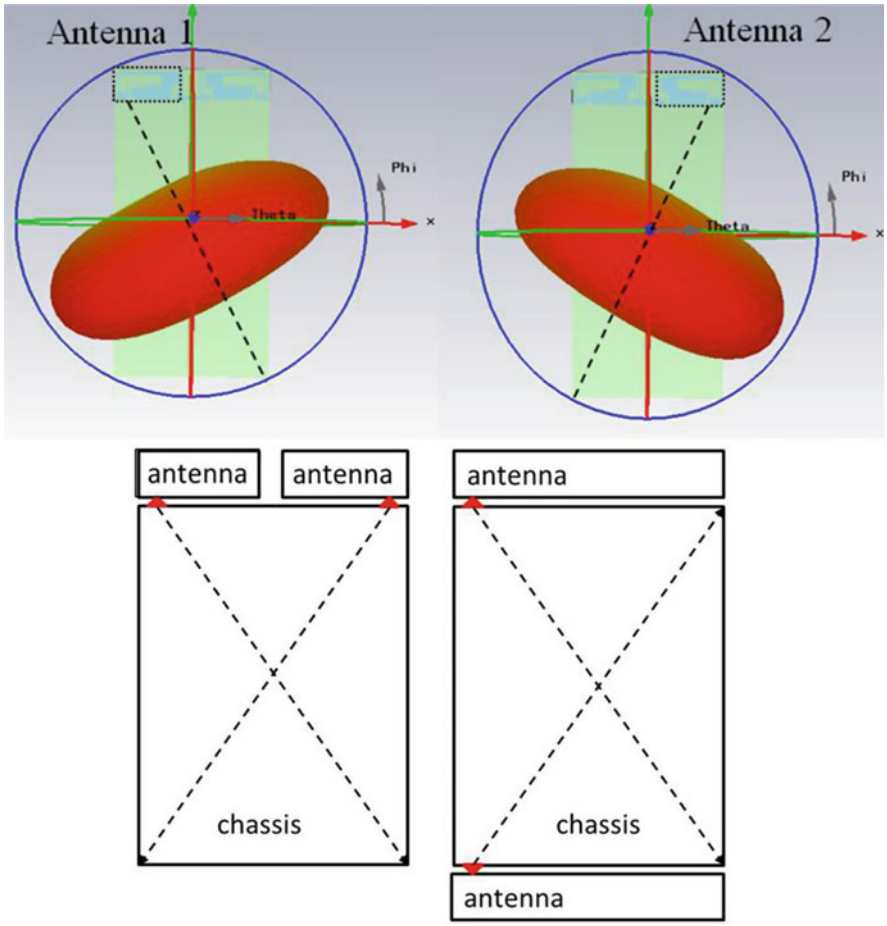


Fig. 26 Diagonal chassis mode in mobile terminals

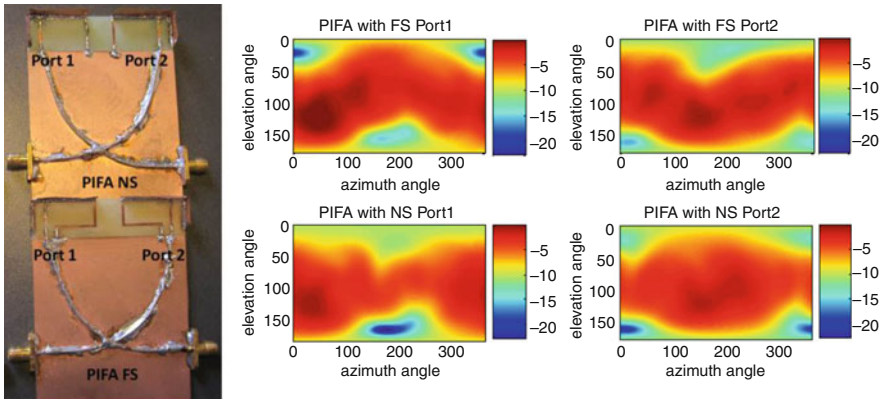


Fig. 27 Diagonal chassis mode in mobile terminals with different locations of grounding points

DOI 10.1007/978-981-4560-75-7_53-1

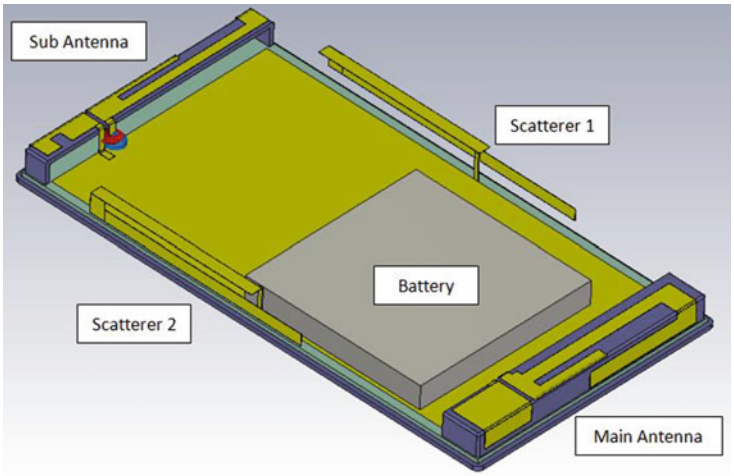


Fig. 28 Perspective view of main and sub antennas and dual scatterers arrangements in an LTE terminal

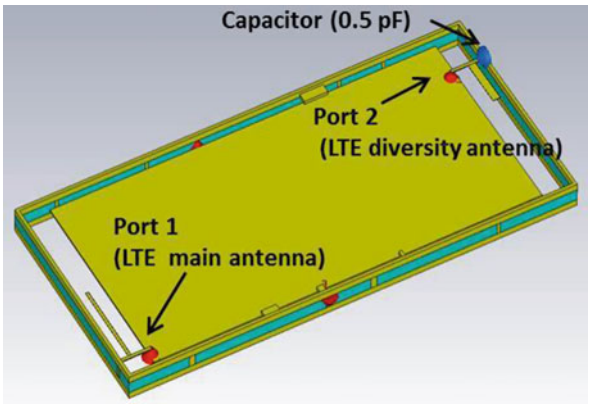


Fig. 29 Double metal ring antenna

Multiantenna SAR and OTA Performance, Field Trial

Body Effects on Compact MIMO Antennas in Mobile Terminals

The radiation performance of MIMO antennas in mobile terminals including user effects is another important issue. Test regulations for mobile terminals in the user case have been developed or are ongoing. The Cellular Telecommunications and Internet Association (CTIA)/The Wireless Association is a United States-based international organization that serves the interests of the wireless industry by lobbying government agencies and assists with regulation settings (CTIA 2011). Up to date, two user cases have been defined by CTIA in the OTA test of mobile terminals, which are the talking mode (head + hand) and the data mode (single hand) shown in Fig. 32.

For user case, both correlation coefficient and total efficiencies of MIMO terminal antennas are low in general. The substantial loss of total efficiency is mainly due to the mismatching and reduction of the radiation efficiency of the antenna (body absorption). The mismatching can be solved by tuning the operating frequency of the antenna with lumped elements, but the drop of the radiation efficiency is not easy to be compromised. Generally, the efficiency loss is related to the antenna volume, antenna type, and terminal volume. Larger antenna and terminal volume will be helpful to reduce the efficiency loss, and magnetic antennas (loop, slot) suffer less efficiency loss than electric antennas (IFA, PIFA, etc.).

DOI 10.1007/978-981-4560-75-7_53-1

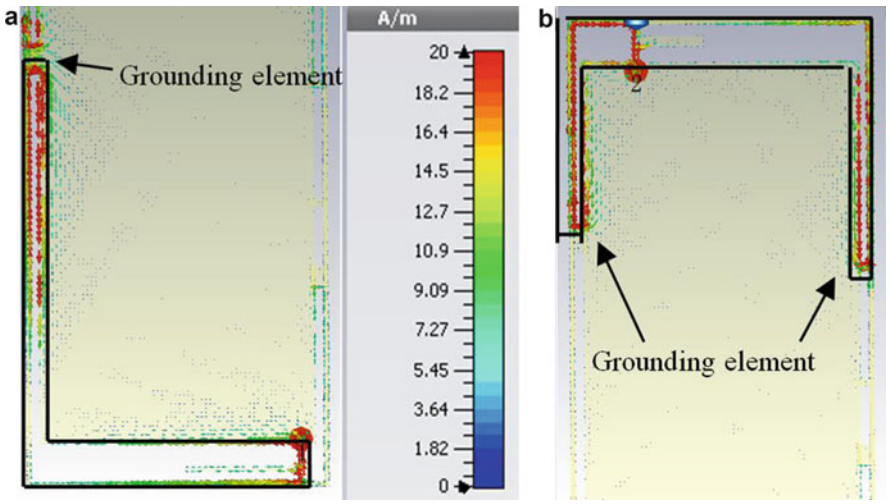


Fig. 30 Current distribution of (**a**) Port 1 at 880 MHz and (**b**) Port 2 at 880 MHz

The correlation coefficient becomes low due to body scattering and shadowing. For a good MIMO reception, the body effect to antenna performance is a main concern.

The efficiency loss of MIMO terminal antennas depends on several factors, including antenna locations, operation frequencies, phone size, etc. (Zhao et al. 2013a; Zhang et al. 2013c, 2015b), as shown in Fig. 33. Several methods have been proposed against the body loss in MIMO terminal antennas. In Zhao et al. (2013), by introducing a small space between human body and antenna, the hand-effect body loss for LTE mobile antenna can be reduced significantly in CTIA talking and data modes. In Zhang et al. (2015b), an adaptive quad element MIMO antenna array is designed (Fig. 34). The body loss can be optimized and the MIMO performance enhanced by selecting the best two elements out of four when the terminal is held by users.

Multitransmitter SAR Problem

Due to the extensive spread of mobile handsets, the radiation of mobile phones has rapidly been given increased attention. The radiation can be evaluated by SAR, which represents the time rate of microwave energy absorption inside the tissue.

$$\mathrm{SAR} = \frac{\sigma}{2\rho} E^2 \tag{25}$$

where σ and ρ are the electrical conductivity (S/m) and mass density (kg/m^3) of the tissue, respectively, and E is the internal induced electrical field (V/m). The SAR value of a mobile terminal is strictly limited by governments, which makes this parameter important to mobile antenna designs. There are two standards of SAR: In Europe, the limitation of SAR is set to be 2.0 W/kg over a 10-g cube by International Commission on Non-Ionizing Radiation Protection (ICNIRP) (ICNRP 1998); in the United States, the Federal Communication Commission (2008) set the limitation to be 1.6 W/kg over a 1-g cube (FCC 2008). The SAR value needs to be measured on Specific Anthropomorphic Mannequin (SAM) head phantom and flat phantom (Fig. 34), for the talking case and the body worn case, respectively.

For MIMO antennas, the evaluation of SAR becomes more complicated. Only the stand-alone SAR is not sufficient to evaluate SAR performances of MIMO antennas. The simultaneous SAR is required when multiple antennas are transmitting simultaneously, which needs to add up SAR fields of the two ports and calculate the combined peak SAR value.

DOI 10.1007/978-981-4560-75-7_53-1

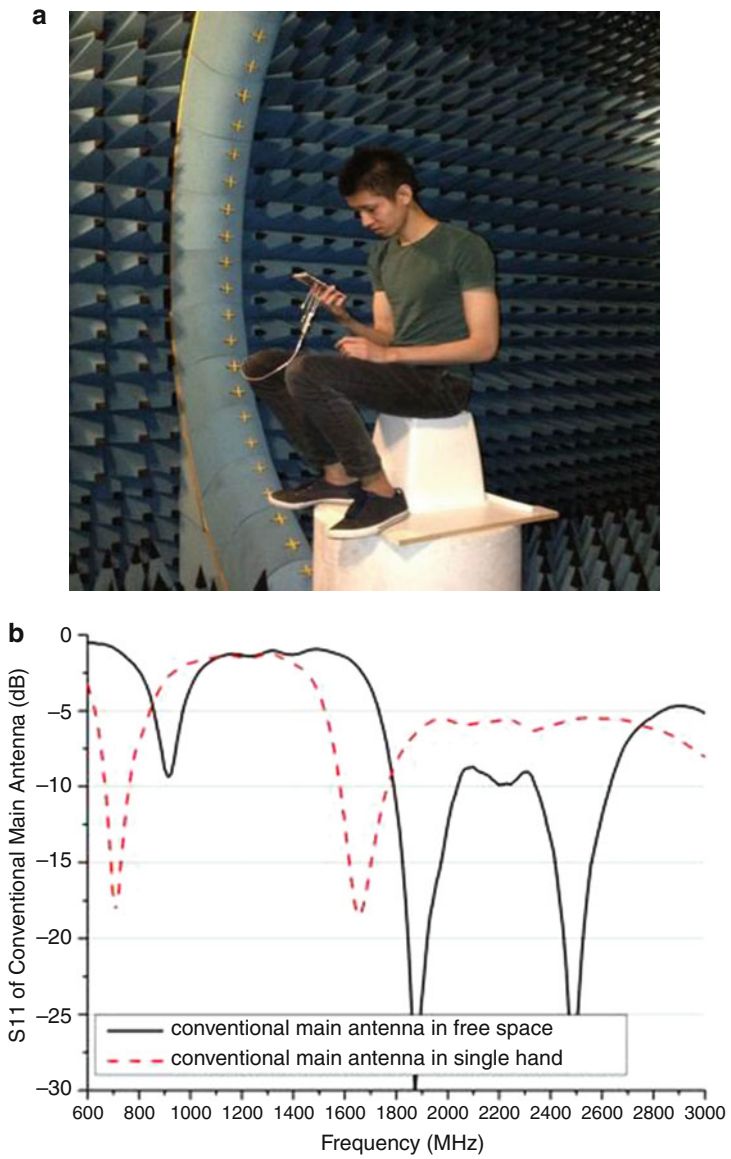


Fig. 31 Double ring antenna under test of real person: (**a**) measurement setup, and (**b**) the comparison of single hand effects on S-parameters between the double ring antenna and conventional design

Table 4 Average total efficiency loss on single hand

TE loss in talking mode (dB)	Main antenna of DR	Main antenna of CA
Left side in lower band	6.44	8.39
Right side in lower band	9.02	9.38
Left side in higher band	5.29	11.61
Right side in higher band	5.31	6.87

DOI 10.1007/978-981-4560-75-7_53-1

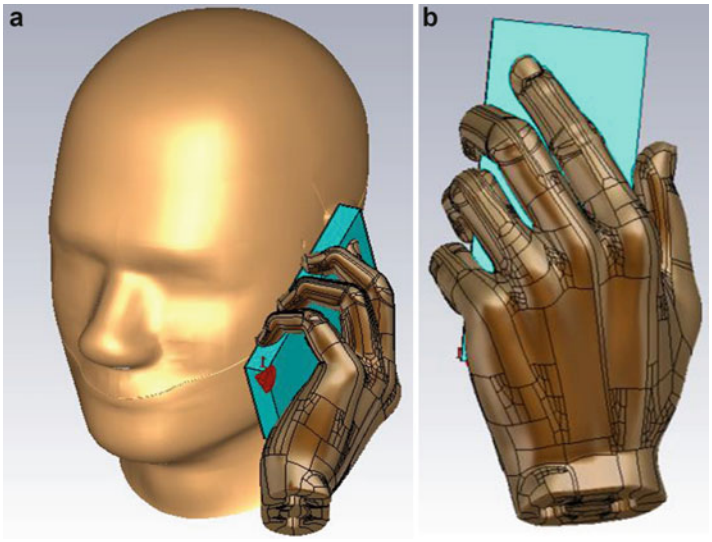


Fig. 32 CTIA user cases: (**a**) talking mode and (**b**) data mode

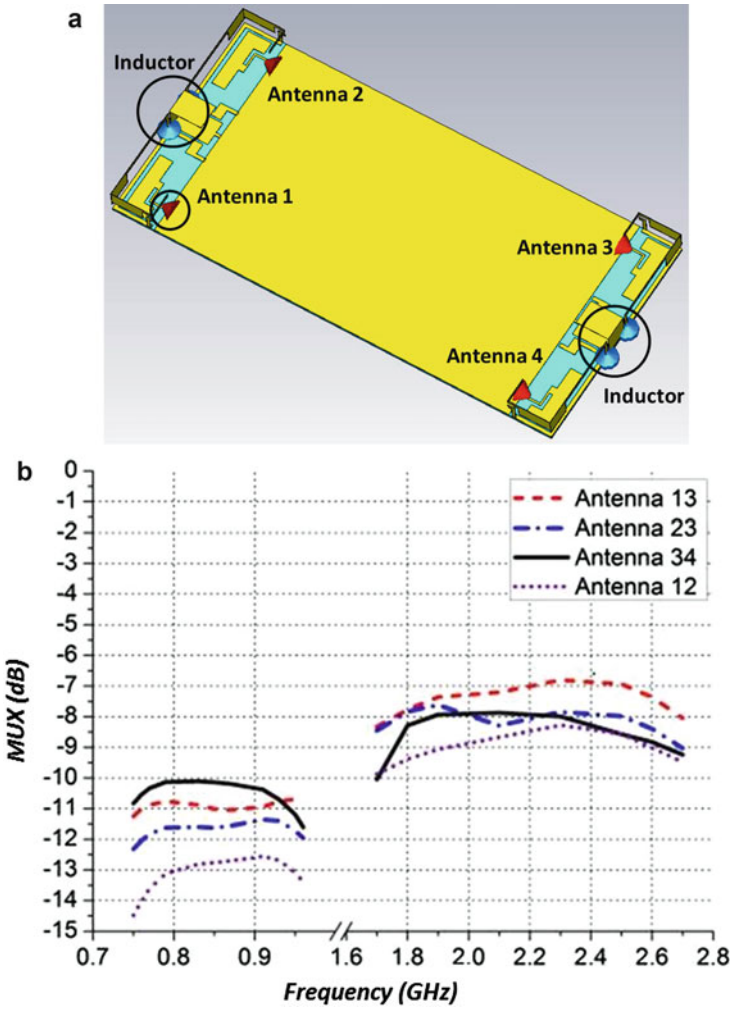


Fig. 33 User-effective adaptive quad-element MIMO antennas, (**a**) antenna configurations and (**b**) MUX of adaptive MIMO antennas in talking mode

DOI 10.1007/978-981-4560-75-7_53-1

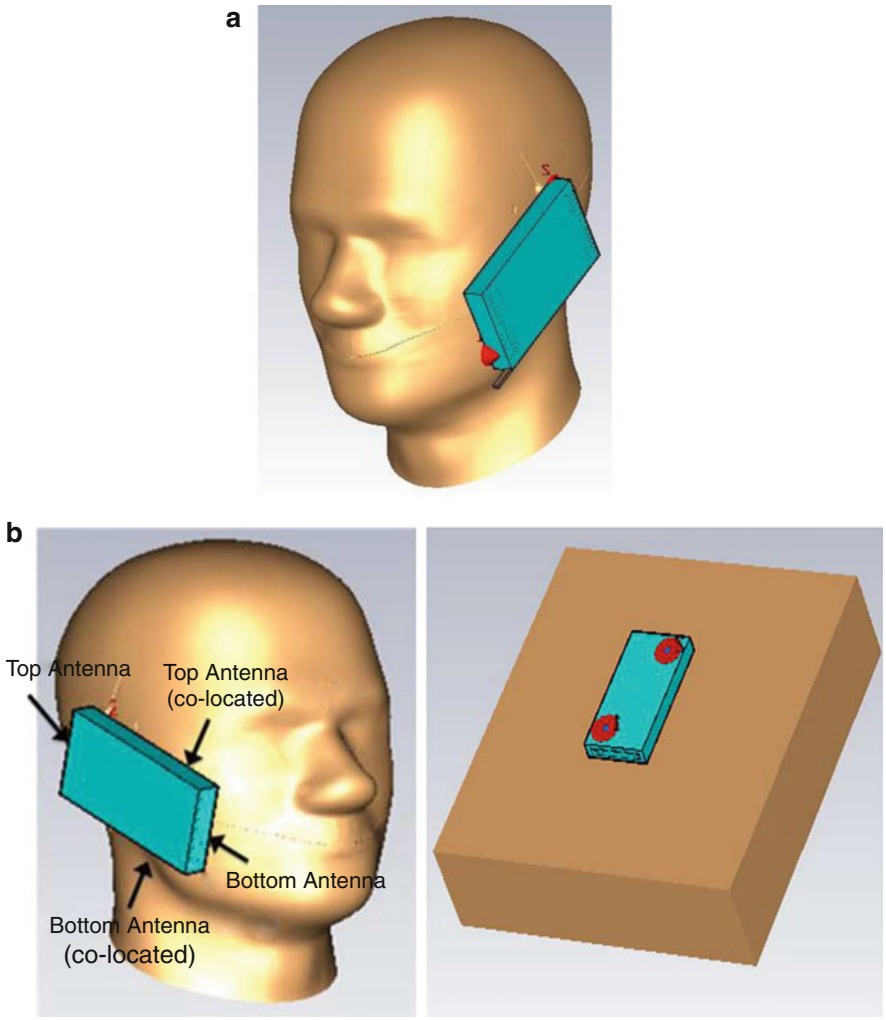


Fig. 34 SAR test positions: (**a**) SAM head phantom and (**b**) flat phantom

In order to simplify the SAR evaluation of MIMO antenna in mobile terminal, the value of SAR to PEAK location spacing ratio (SPLSR) is defined by FCC (2008):

$$\mathrm{SPLSR} = \frac{(\mathrm{SAR1} + \mathrm{SAR2})}{D} \tag{26}$$

where SAR1 and SAR2 are SAR values (W/kg) of elements 1 and 2, respectively, and D is the separation distance (cm) of the two SAR peaks, as illustrated in Fig. 35.

Subject to FCC standard, simultaneous SAR measurement of a phone is not necessary if the SPLSR is smaller than 1.6 (FCC 2008). From Eq. 26, it is clear that the length of ground plane plays an important role for simultaneous SAR. On the one hand, a proper ground plane length can excite the chassis mode, which will reduce the peak value of stand-alone SAR; on the other hand, the strong chassis mode could reduce the separation distance between the two SAR peaks due to the peak position of SAR moving the center of the ground plane, which could enhance simultaneous SAR (FCC 2008). Apart from the ground plane length, antenna types also have great impacts on SAR value, as shown in Fig. 36. From the engineering point of view, a ground free antenna is always (with high SAR) located at the bottom of the phone, whereas an on-ground antenna (with low SAR) will be put to the top of the phone to reduce

DOI 10.1007/978-981-4560-75-7_53-1

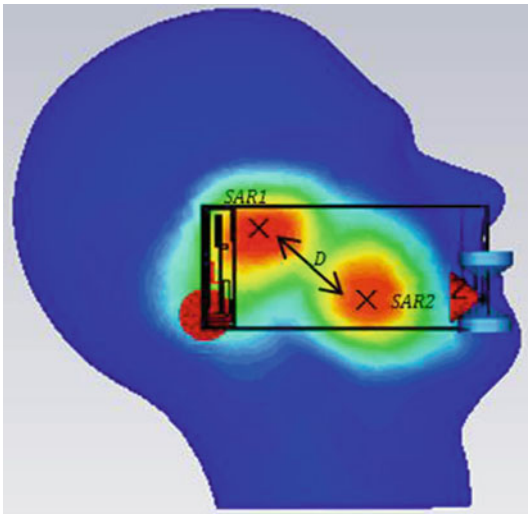


Fig. 35 The illustration of the value of SAR to PEAK location spacing ratio (SPLSR)

correlation. The head phantom can be measured through DASY system and a flat phantom SAR can be measured with iSAR system in Fig. 37. The comparison between simulated and measured SAR (dB) distribution (dB) for semi-ground-free PIFA at 10 mm above the flat phantom is presented in Fig. 38 (Zhao et al. 2013b).

LTE Field Trial at 700 MHz Band with Different MIMO Antenna Designs

For an LTE MIMO terminal, both mean effective gain and correlation will determine the MIMO throughput performance. In a typical smartphone, antennas usually have relatively low correlation for frequency bands around 2 GHz and above. So the main effort to realize a good MIMO performance is to design high efficient antennas with less coupling loss at those frequency ranges which have been discussed in the section "Decoupling Techniques of Compact MIMO Antenna Designs." However, it is quite a big challenge to design low-correlation and low-coupling MIMO antennas at low frequency bands, especially at 700 MHz band. Some detailed mock-ups in different correlation level were built and tested in Sony Mobile research laboratory and field trial with Ericsson Research. Figure 39 shows four different MIMO antenna designs for typical smartphones at 700 MHz band. Design (a) has two monopoles at top and bottom ends; the chassis mode is strong and thus the correlation is relatively high. Design (b) composes a monopole and a notch in which the two antennas are orthogonally arranged to each other. So the correlation is low in this case. Design (c) has colocated loop antenna, where one antenna element is fed by two ports. This is on-purpose design to have a very high correlation. Design (d) consists of one chassis mode monopole and one localized mode patch antenna, and it is found that the correlation is extremely low. The mock-ups were tested and characterized in Sony Mobile laboratory; the performances are summarized in Table 5 (Fig. 40).

Compared to the performance of reference antennas (separated orthogonal dipoles), it is found that a compact MIMO antenna arrangement could also have a fairly good MIMO performance. The envelope correlation can be less than 0.5 even at 700 MHz band. Good antenna efficiency of both antennas is essential to cover larger service area of LTE with good data communication quality (Hagerman et al. 2011). In user case, the human body loading and absorption are the main impacts on the mean effective gain which determine the good data rate coverage service area; the correlation are generally low in user cases (Hagerman et al. 2011; Derneryd and Ying 2010).

DOI 10.1007/978-981-4560-75-7_53-1

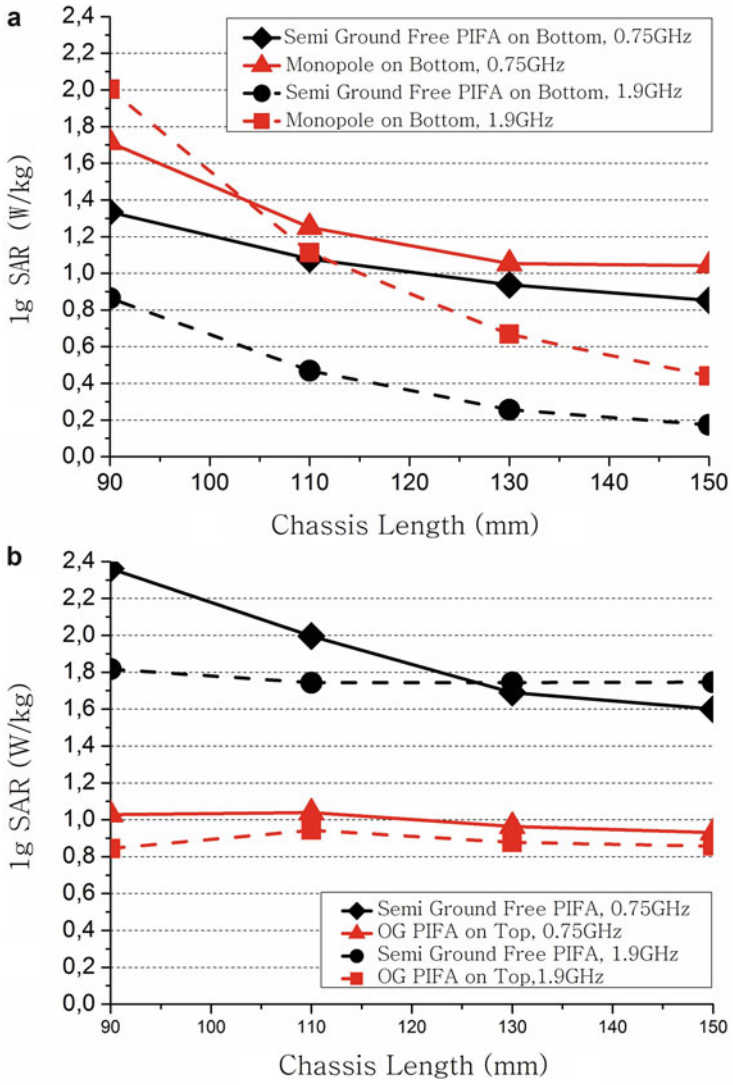


Fig. 36 Simulation SAR with different antenna type and different chassis length (**a**) *bottom* antenna (**b**) *top* antenna

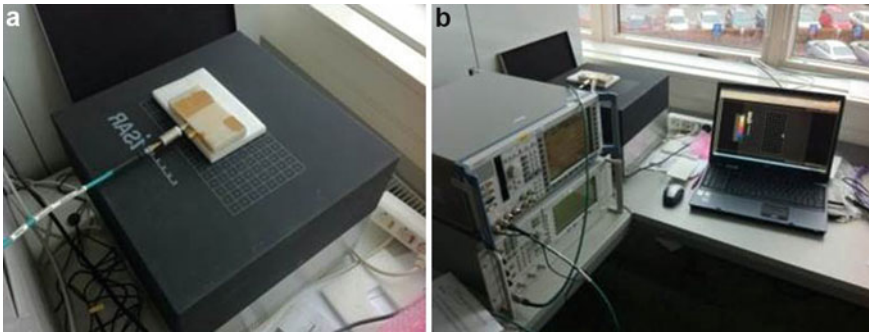


Fig. 37 The measurement setup of iSAR system

DOI 10.1007/978-981-4560-75-7_53-1

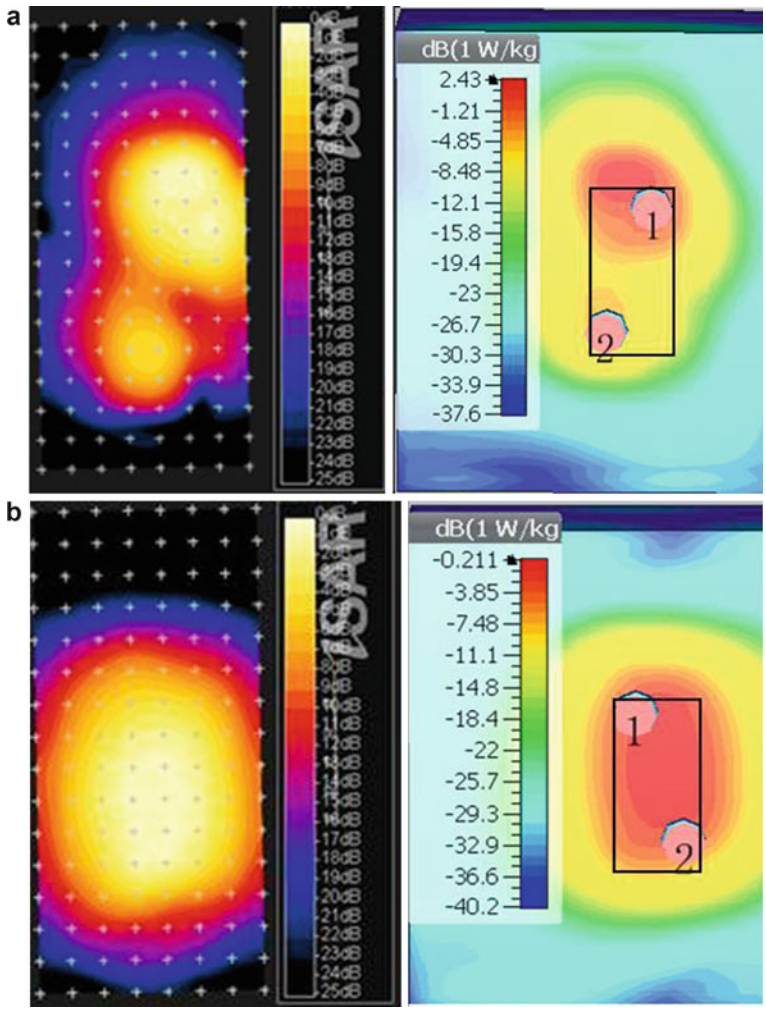


Fig. 38 The comparison between simulated and measured SAR (dB) distribution (dB) for semi ground-free PIFA at 10 mm above the flat phantom at (**a**) 0.85GHz, and (**b**) 1.9 GHz

Summary

In this chapter, certain basic parameters characterizing the diversity and MIMO antennas have been discussed. Several fundamental MIMO antennas' performance and useful decoupling techniques were presented. In addition, recent advanced technologies for compact MIMO antennas particularly for portable devices and mobile terminals were also reviewed. The interactions between MIMO antennas and human body and the multiple radio SAR problems are important for mobile industry. It is demonstrated that the antenna impact could be addressed in an LTE MIMO field trial at 700 MHz.

DOI 10.1007/978-981-4560-75-7_53-1

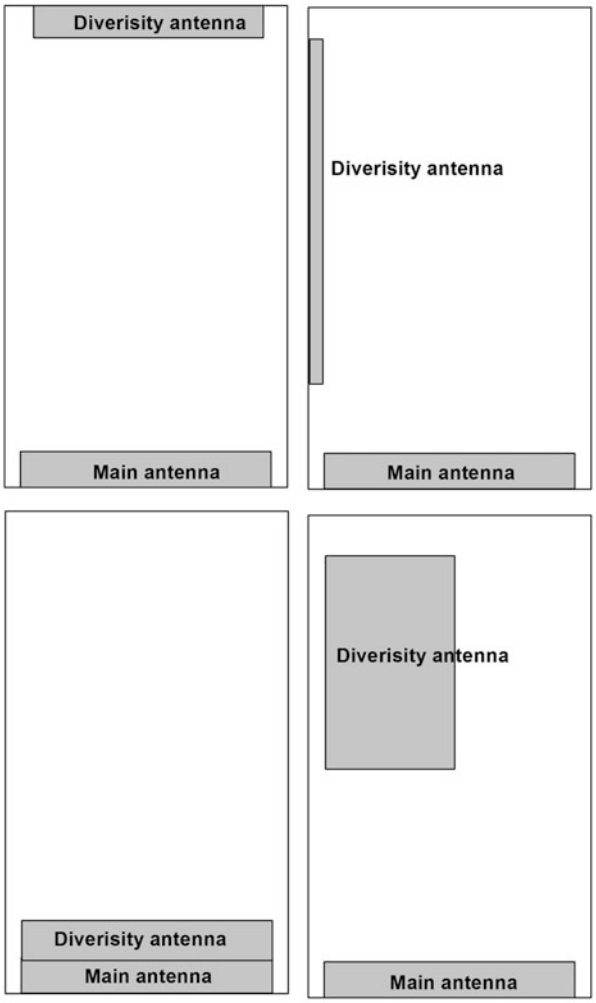


Fig. 39 Different LTE antenna designs in mobile terminal at 700 MHz bands (**a**) two monopoles; (**b**) monopole and notch; (**c**) co-located loop; (**d**) monopole and patch

Table 5 The measured MIMO antenna performance of four different mock-ups, and the field trial test results

	Mean gain of two antennas	Pattern envelope correlation	MIMO multiplexing efficiency	MIMO gain drop due to pattern correlation	Apparent diversity gain at CDF 1 % level	Field test: proportion of two streams for MIMO (separate dipole as reference) (%)
Separated dipoles	−5.5 dB	0	−5.5 dB	0	11.5 dB	100
Mock-up (a)	−5.6 dB	0.51	−7.2 dB	1.6 dB	9.5 dB	91
Mock-up (b)	−5.0 dB	0.17	−5.4 dB	0.4 dB	10 dB	95
Mock-up (c)	−5.5 dB	0.94	−11.6 dB	6.1 dB	6.5 dB	8
Mock-up (d)	−2.9 dB	0.06	−2.9 dB	0	11,5 dB	100

DOI 10.1007/978-981-4560-75-7_53-1

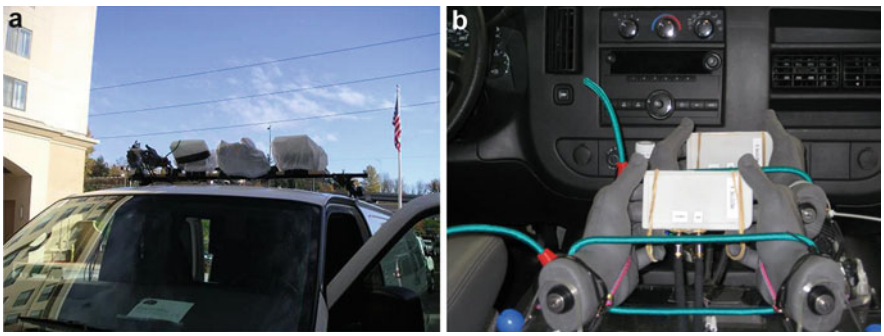


Fig. 40 (**a**) and (**b**) show the field trial in car roof and in the car, with and without hand phantoms (Courtesy of Ericsson AB)

References

Andrews MR, Mitra PP, de Carvalho R (2001) Tripling the capacity of wireless communications using electromagnetic polarization. Nature 1:316–318

Bae H, Harackiewicz FJ, Park M, Jim T, Kim N, Kim D, Lee B (2010) Compact mobile handset MIMO antenna for LTE700 applications. Microw Opt Technol Lett 11:2419–2422

Balanis CA (2005) Antenna theory: analysis and design, 3rd edn. Wiley-Interscience, Hoboken

Bentsson E, Tufvesson F, Edfors O (2015) UE antenna properties and their influence on massive MIMO system performance. In: European conference on antennas and propagation, Lisbon, accepted

Blanch S, Romeu J, Corbella I (2003) Exact representation of antenna system diversity performance from input parameter description. Electron Lett 39:705–707

Bolin T, Derneryd A, Kristensson G, Plicanic V, Ying Z (2005) Two-antenna receive diversity performance in indoor environment. IET Electron Lett 10:1205–1206

Chae SH, Yoon HS, Park SO (2008) The realization and analysis of pattern/polarization diversity in multiple input multiple output array. Microw Opt Technol Lett 3:561–566

Chebihi A, Luxey C, Diallo A, Thuc PH, Staraj R (2008) A novel isolation technique for UMTS mobile phones. IEEE Antennas Wirel Propag Lett 7:665–668

Chi G, Li B, Qi D (2005) Dual-band printed diversity antenna for 2.4/5.2-GHz WLAN application. Microw Opt Technol Lett 6:561–563

Chiu CY, Yan J, Murch RD (2007a) Compact three-port orthogonally polarized MIMO antennas. IEEE Antennas Wirel Propag Lett 6:619–622

Chiu CY, Cheng CH, Murch RD, Rowell CR (2007b) Reduction of mutual coupling between closely-packed antenna elements. IEEE Trans Antennas Propag 6:1732–1738

Chiu CY, Yan JB, Murch RD, Yun JX, Vaughan RG (2009) Design and implementation of a compact 6-port antenna. IEEE Antennas Wirel Propag Lett 7:767–770

Chiu CY, Liu, X, Gao F, Ying Z (2013) Reduction of envelope correlation on non-metallic and metallic chassis phones. In: International workshop on antenna technology, Karlsruhe, vol 3, pp 31–34

CTIA (2011) Test plan for mobile station over-the-air performance. Rev 3.1:184–189

Derneryd A, Ying Z (2010) Evaluation of different MIMO antenna designs in different correlation level for LTE 700 bands. Ericsson internal report 5

Diallo A, Luxey C, Thuc PL, Staraj R, Kossiavas G (2008) Enhanced two-antenna structures for universal mobile telecommunications system diversity terminals. IET Microwaves Antennas Propag 2:93–101

Ding Y, Du Z, Gong K, Feng Z (2007) A four-element antenna system for mobile phones. IEEE Antennas Wirel Propag Lett 6:655–658

DOI 10.1007/978-981-4560-75-7_53-1

Federal Communications Commission (2008) SAR evaluation considerations for handsets with multiple transmitters and antennas 9:4–5

Foschini GJ, Gans MJ (1998) On limits of wireless communications in a fading environment when using multiple antennas. Wirel Pers Commun 6:311–335

Glazunov AA, Molisch AF, Tufvesson F (2009) Mean effective gain of antennas in a wireless channel. IET Microwaves Antennas Propag 3:214–227

Hagerman B, Werner K, Yang J (2011) MIMO performance at 700MHz: field trials of LTE with handheld UE. In: IEEE vehicular technology conference, San Francisco, vol 9, pp 1–5

Hallbjorner P (2005) The significance of radiation efficiencies when using S-parameters to calculate the received signal correlation from two antennas. IEEE Antennas Wirel Propag Lett 4:97–99

International Commission on Non-Ionizing Radiation Protection (1998) Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz). Health Phys 4:494–522

Jakes WC (1974) Microwave mobile communications. Wiley, New York

Kildal PS, Rosengren K (2004) Correlation and capacity of MIMO systems and mutual coupling, radiation efficiency, and diversity gain of their antennas: Simulations and measurements in a reverberation chamber. IEEE Commun Mag 42:104–112

Kim KJ, Ahn KH (2007) The high isolation dual-band inverted F antenna diversity system with the small N-section resonators on the ground plane. Microw Opt Technol Lett 3:731–734

Kim J, Kim M, Jeon S (2011) Dual-antenna diversity-gain improvements for even-odd mode feed network at 755 MHz. Microw Opt Technol Lett 2:304–306

Knudsen MB (2001) Antenna systems for handsets. ATV-industrial PhD project EF-755, Aalborg University

Lau BK, Andersen JB (2012) Simple and efficient decoupling of compact arrays with parasitic scatterers. IEEE Trans Antennas Propag 2:464–472

Li H, Lau BK, Tan Y, Ying Z (2011) Impact of current localization on the performance of compact MIMO antennas. In: European conference on antennas and propagation, Rome, vol 4, pp 2423–2426

Li H, Lau BK, Ying Z, He S (2012a) Decoupling of multiple antennas in terminals with chassis excitation using polarization diversity, angle diversity and current control. IEEE Trans Antennas Propag 12:5947–5957

Li H, Tan Y, Lau BK, Ying Z, He S (2012b) Characteristic mode based tradeoff analysis of antenna-chassis interactions for multiple antenna terminals. IEEE Trans Antennas Propag 2:490–502

Li H, Lin X, Lau BK, He S (2013) Equivalent circuit based calculation of signal correlation in lossy antenna arrays. IEEE Trans Antennas Propag 10:5214–5222

Liu Q, Du Z, Gong K, Feng Z (2008) A compact wideband planar diversity antenna for mobile handsets. Microw Opt Technol Lett 1:87–91

Mavridis A, Christodoulou CG, Sahalos JN, Chryssomallis MT (2006) A planar two-branch diversity antenna for wireless applications. In: IEEE international symposium on antennas and propagation, Albuquerque, vol 7, pp 309–312

Norklit O, Teal PD, Vaughan RG (2001) Measurement and evaluation of multi-antenna handsets in indoor mobile communication. IEEE Trans Antennas Propag 3:429–437

Ogawa K, Matsuyoshi T, Monma K (2001) An analysis of the performance of a handset diversity antenna influenced by head, hand and shoulder effects at 900 MHz: part I-effective gain characteristics and part II-correlation characteristics. IEEE Trans Veh Technol 50:830–853

Park G, Kim M, Yang T, Byun J, Kim AS (2009) The compact quad-band mobile handset antenna for the LTE700 MIMO application. In: IEEE international symposium on antennas and propagation, Charleston, vol 6, pp 1–4

DOI 10.1007/978-981-4560-75-7_53-1

Park BY, Choi JH, Park SO, Yang TS, Byun JH (2011) Design of a decoupled MIMO antenna for LTE applications. Microw Opt Technol Lett 3:582–586

Pedersen GF, Andersen JB (1999) Handset antennas for mobile communications: integration, diversity and performance. Rev Radio Sci 8:119–137

Pierce JR, Stein S (1960) Multiple diversity with nonindependent fading. Proc IRE 48:89–104

Plicanic V (2011) Characterization and enhancement of antenna system performance in compact MIMO terminals. A thesis of Lund University for the degree of Doctor of Philosophy

Plicanic V, Lau BK, Derneryd A, Ying Z (2009) Actual diversity performance of a multiband diversity antenna with hand and head effects. IEEE Trans Antennas Propag 57:1547–1556

Pozar DM (2005) Microwave engineering, 3rd edn. Wiley, New York

Rao Q, Wang D (2009) Compact low coupling dual-antennas for MIMO applications in handheld devices. In: IEEE international symposium on antennas and propagation, Charleston, vol 6, pp 1–4

Schwartz M, Bennet WR, Stein S (1965) Communication systems and techniques. McGraw-Hill, New York

Shannon CE (1948) A mathematical theory of communication. Bell Syst Technol J 27:623–656

Shin YS, Park SO (2007) Spatial diversity antenna for WLAN application. Microw Opt Technol Lett 6:1290–1294

Taga T (1990) Analysis for mean effective gain of mobile antennas in land mobile radio environments. IEEE Trans Veh Technol 39:117–131

Telatar IE (1999) Capacity of multi-antenna Gaussian channels. Eur Trans Telecommun 10:585–595

Tian R, Plicanic V, Lau BK, Ying Z (2010) A compact six-port dielectric resonator antenna array: MIMO channel measurements and performance analysis. IEEE Trans Antennas Propag 4:1369–1379

Tian R, Lau BK, Ying Z (2011) Multiplexing efficiency of MIMO antennas. IEEE Antennas Wirel Propag Lett 10:183–186

Tian R, Lau BK, Ying Z (2012) Multiplexing efficiency of MIMO antennas in arbitrary propagation scenarios. In: European conference on antennas and propagation, Prague, pp 373–377

Vaughan R, Andersen JB (2003) Channels, propagation and antennas for mobile communications. Institution of Electrical Engineers, London

Vaughan RG, Anderson JB (1987) Antenna diversity in mobile communications. IEEE Trans Veh Technol 11:147–172

Volmer C, Sengül M, Weber J, Stephan R, Hein MA (2008) Broadband decoupling and matching of a superdirective two-port antenna array. IEEE Antennas Wirel Propag Lett 7:613–616

Winters J (1987) On the capacity of radio communication systems with diversity in a Rayleigh fading environment. IEEE J Sel Areas Commun 5:871–878

Ying Z (2012) Antennas in cellular phones for mobile communications. Proc IEEE 7:2286–2296

Zhang S, Xiong J, He S (2009a) MIMO antenna system of two closely-positioned PIFAs with high isolation. IET Electron Lett 45:771–773

Zhang S, Ying Z, Xiong J, He S (2009b) Ultra wideband MIMO/diversity antennas with a tree-like structure to enhance wideband isolation. IEEE Antennas Wirel Propag Lett 12:1279–1282

Zhang S, Khan SN, He S (2010a) Reducing mutual coupling for an extremely closely-packed tunable dual-element PIFA array through a resonant slot antenna formed in-between. IEEE Trans Antennas Propag 8:2771–2776

Zhang S, Zetterberg P, He S (2010b) Printed MIMO antenna system of four closely-spaced elements with large bandwidth and high isolation. IET Electron Lett 7:1052–1053

Zhang S, Lau BK, Tan Y, Ying Z, He S (2012a) Mutual coupling reduction of two PIFAs with a T-shape slot impedance transformer for MIMO mobile terminals. IEEE Trans Antennas Propag 3:1521–1531

Zhang S, Ying Z, He S (2012b) Mutual scattering mode for LTE MIMO antennas. In: Asia-pacific conference on antennas and propagation, Singapore, vol 8, pp 13–14

Zhang S, Ying Z, He S (2012c) Diagonal chassis mode for mobile handset LTE MIMO antennas and its application to correlation reduction. In: International workshop on electromagnetics; applications and student innovation, Chengdu, vol 8, pp 1–2

Zhang S, Glazunov AA, Ying Z, He S (2013a) Reduction of the envelope correlation coefficient with improved total efficiency for mobile LTE MIMO antenna arrays: mutual scattering mode. IEEE Trans Antennas Propag 6:3280–3291

Zhang S, Zhao K, Ying Z, He S (2013b) Diagonal antenna-chassis mode for wideband LTE MIMO antenna arrays in mobile handsets. In: International workshop antenna technology, Karlsruhe, vol 3, pp 407–410

Zhang S, Zhao K, Ying Z, He S (2013c) Adaptive quad-element multi-wideband antenna array for user-effective LTE MIMO mobile terminals. IEEE Trans Antennas Propag 8:4275–4283

Zhang S, Zhao K, Zhu B, Ying Z, He S (2015a) Investigation of diagonal antenna-chassis mode in mobile terminal LTE MIMO antennas for bandwidth enhancement. IEEE Antennas Propag Mag 57(2):217–228

Zhang S, Zhao K, Zhu B,Ying Z, He S (2015b) Body loss study of beam forming mode in LTE MIMO mobile terminals. In: Antennas and propagation (EuCAP), 2015 9th European conference on, pp 1–5, 13–17

Zhang S, Zhao K, Zhu B, Ying Z, He S (2014) MIMO reference antenna with controllable correlations and total efficiencies. Prog Electromagn Res 145:115–121

Zhao K, Zhang S, Ying Z, Bolin T, He S (2013a) SAR study of different MIMO antenna designs for LTE application in smart mobile handsets. IEEE Trans Antennas Propag 6:3270–3279

Zhao K, Zhang S, Ying Z, He S (2013b) Reduce the hand-effect body loss for LTE mobile antenna in CTIA talking and data modes. Prog Electromagn Res 137:73–86

Zhao K, Zhang S, Chiu C, Ying Z, He S (2014) Impact of size and decoupling element on dome fundamental compact MIMO antennas. In: International workshop on antenna technology 2014, Sydney

Zhao K, Zhang S, Ying Z, He S (2015) Body-insensitive multimode MIMO terminal antenna of double-ring structure. IEEE Trans Antennas Propag 63(5):1925–1936.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_55-1

Low-Profile Antennas

Gijo Augustin[a]*, Qinjiang Rao[b] and Tayeb A. Denidni[a]
[a]Centre for Energy, Materials and Telecommunication, National Institute of Scientific Research (INRS), Montreal, QC, Canada
[b]Blackberry Limited, Kanata, ON, Canada

Abstract

This chapter provides a broad discussion on the progress in low-profile antenna design and their applications. The antennas are divided broadly into dipole-based antenna designs, slot antennas, loop antennas, and microstrip patch antennas. In addition to the basic concepts, a literature review that outlines state-of-the-art designs along with recent trends in this field is incorporated. The objective of these discussions is to enrich the readers with conceptual knowledge in this marvelous field of engineering.

Keywords

Dipole antenna; Low-profile antenna; Compact antenna; PIFA; Loop antenna; Microstrip antenna; Printed strip dipole; Slot fed antenna; Open sleeve dipole

Introduction

In this modern era, the world is experiencing a huge technological wisdom in almost every phase of life. The developments in interdisciplinary areas of science and engineering have synergistically produced powerful tools for various applications. The magnificent field of wireless communications is not an exception; it has undergone drastic developments that resulted in a broad line of products ranging from a wireless keyboard to complex space shuttles with sophisticated communication modules. One of the outstanding revolutions in this century is the advancements in the field of wireless communication, introducing smart phones and tablet computers which are integrated with plenty of services. These handheld devices incorporate various wireless communication components that are manyfold complex in design compared to their one-decade-ago versions. The thrust of human civilization to exchange data without boundaries made further progress in this amazing field of technology. This has led most of the information technology industries to invest in wireless communication systems. For example, Apple Inc. and Google Inc. opened several research and development centers for hardware development. These advancements have resulted in huge expansion of the world of wireless communication, where most devices we meet in our day-to-day life are equipped with one or other form of wireless communication units.

In an antenna engineer's perspective, the antenna design strategies have considerably changed for the last three decades. The huge expansion of the wireless communication sector produced modern systems which require advanced low-profile antennas for newly evolved systems. One of the key differentiating facts is the integration of various communication segments to the modern mobile phones. This made them more advanced than the traditional voice communication systems. This requires careful integration of

*Email: gijo@ieee.org
*Email: augustin@emt.inrs.ca

DOI 10.1007/978-981-4560-75-7_55-1

Fig. 1 The evolution in mobile communication gadgets (**a**) First commercial handheld cellular phone from Motorola DynaTAC 800X (Reproduced from Motorola Inc.) (**b**) Smartphone from Apple Inc. iPhone 6 (Reproduced from Apple Inc.)

multiple low-profile antennas in contrast to the single RF module design in the old-generation handsets. For example, the antennas employed in current smartphones and tablet computers, such as Apple iPhone® or Google Nexus®, are manyfold advanced compared to the sleeve dipole antenna used in the first commercial handheld cellular phone, Motorola Dyna TAC 8000×, introduced in 1983 (Motorola DynaTAC). Apart from smartphones, the wireless communication sector has been enriched with numerous types of systems. Some of them are smartwatches, near-field communication (NFC) devices, chipless RFID, wearable systems, medical sensors, and so forth. These complex systems introduce a broad range of low-profile antenna design challenges that need to be solved without degrading the overall system performance. For example, modern communication units need to comply with rigorous specifications, such as specific absorption ratio (SAR), in order to pass the regulations put by various governing authorities. This chapter address four important categories of low-profile antennas, based on their fundamental working concepts, state-of-the-art designs, and recent trends. This includes dipole-based designs, slot antenna designs, loop antennas, and microstrip patch-based low-profile antennas (Fig. 1).

The Terminologies: Low-Profile Antennas and Electrically Small Antennas

- **Low-Profile Antennas:** The Oxford Dictionary defines the word profile as "a vertical cross section of a structure." Thus, it is quite straightforward to define the low-profile antenna as the one with relatively small vertical cross section. One of the earlier definitions for low-profile antennas is found in Stutzman and Davis (1998) defines as "Antennas of small physical thickness." This terminology has evolved through many years and is used to represent antennas with relatively small physical cross-sectional area with respect to the wavelength (λ) in which it operates. Some of the well-known low-profile antennas

DOI 10.1007/978-981-4560-75-7_55-1

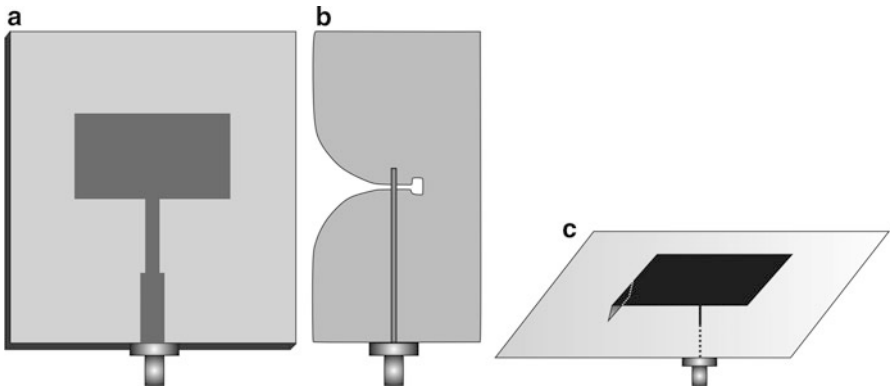


Fig. 2 Classical low-profile antennas (**a**) Microstrip patch antenna, (**b**) Tapered slot antenna (**c**) Planar Inverted F Antenna (PIFA)

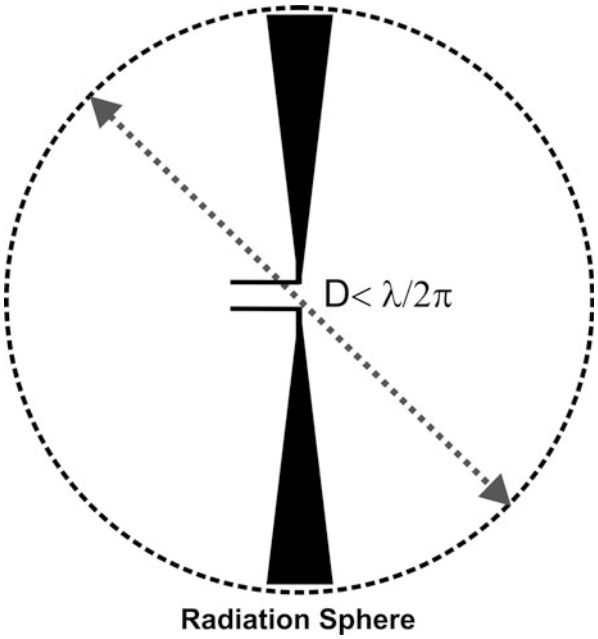


Fig. 3 Electrically small antenna definition based on Wheeler (1947)

are microstrip patch antennas, planar inverted-F antennas (PIFA), slot antennas, and loop antennas. Some classical low-profile antennas are illustrated in Fig. 2.

- **Electrically Small Antennas (ESAs):** The electrically small antennas (ESAs) were often being confused with low-profile antennas. Even though there are low-profile antennas that are electrically small, all low-profile antennas do not come under the definition of electrically small antennas. A widely adopted definition is by Harold Wheeler in his 1947 paper (Wheeler 1947). This defines ESA as an antenna that could be enclosed within a sphere of radius $r = \lambda/2\pi = 0.0796\lambda$ called the radiation sphere as shown in Fig. 3. The IEEEE standard definition of terms for antennas ("IEEE Standard Definitions of Terms for Antennas" 1993) defines electrically small antenna as "An antenna whose dimensions are such that it can be contained within a sphere whose diameter is small compared to a wavelength at the frequency of operation." These definitions indicate that electrically small antennas are not necessarily physically small. In fact, some of the historical antennas used by Marconi in the early stages of wireless

DOI 10.1007/978-981-4560-75-7_55-1

communication (1890s) were electrically small. The classical ESA is a variant of reactively loaded dipoles, loop antennas fabricated on special core material, and dielectric resonator antennas.

As a concluding remark about the terminologies, it is important to note that "low-profile" is a comparative term and the boundary between electrically small antennas and low-profile antennas is quite tough to draw. On the other hand, the electrically small antennas are clearly defined as the ones that are smaller than a radian sphere. Thus, all low-profile antennas are not necessarily electrically small antennas. This chapter discusses various low-profile antennas in which some of them can meet the standard definition of electrically small antennas as well.

Why Low-Profile Antennas Are Important?

Historical Perspective: The significance of low-profile antennas is recognized from the very beginning of the communication era, where antennas were one of the huge structures that needs to be deployed in a limited space. The invention of vacuum tubes and other relatively high power sources enabled earlier researchers to extend the technology by developing more compact antenna systems. Around the 1930s various techniques, such as magnetic material loaded antennas and slot antennas, were implemented and available for practical use. This resulted in the invention of various electrically small antennas, such as helical antennas, and microstrip patch antennas. The decades following this period, several antenna designs attracted various researchers around the globe who are working on the development of low-profile antennas and electrically small antennas (Fujimoto 1987). Some of the initial applications of this technology were the radio communication systems installed in vehicles traveling on the ground. The special situations during World Wars I and II accelerated the research in the development of antennas that are more compact and low profile. The end of the twentieth century witnessed extreme developments in the field of mobile communication which further made progress in the development of low-profile antennas. In this era, the severe increase in the demand for small and low-profile antennas was mainly because of advancements in various interdisciplinary areas of science and engineering that resulted in a system-level shrinking for almost every type of wireless communication devices. Few of the key examples include wireless communication systems such as tablet computers and smartphones. Today, consumer market also provides various wireless communication accessories for these new-generation communication devices.

Importance and design challenges: The integrated circuits inside a wireless communication equipment become highly dense in accordance with Moore's law; meanwhile the reduction in antenna size is relatively less (Schaller 1997). This is due to the fact that the size of electronic chips is based on fabrication techniques; meanwhile, the antenna size is based on the fundamental principles of physics. One of the immediate consequences of reducing the antenna size is the drop in system efficiency or bandwidth. Thus, the quest of industries to produce electrically small antennas despite this fundamental barrier makes this research field highly challenging and significantly attractive.

Design Fundamentals and Limitations

In this section, typical design strategies of various low-profile antennas are outlined along with few fundamental limitations of implementing them.

Miniaturization Techniques

Various miniaturization techniques have been employed in antenna design to develop low-profile and electrically small antennas. Most of these techniques focus on exciting fundamental resonant mode by

DOI 10.1007/978-981-4560-75-7_55-1

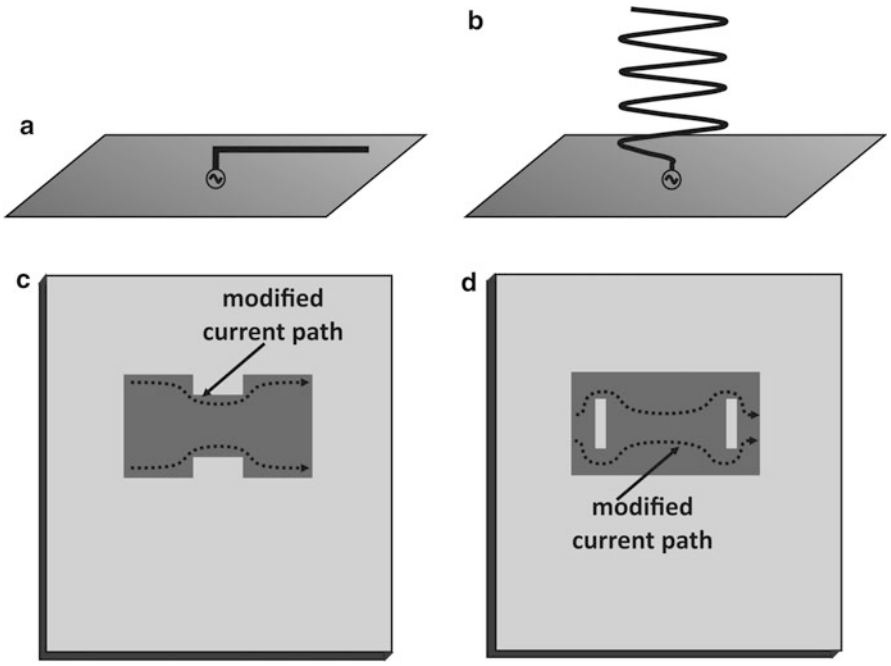


Fig. 4 Modification of resonant length in classical antennas for miniaturization (**a**) inverted L antenna (**b**) helical antenna (**c**–**d**) microstrip patch loaded with slots

incorporating some elements that help to overcome some of the structural limitations of classical antennas such as dipoles and microstrip patch antennas.

- ***Modification of Resonant Length***: This is one of the commonly employed miniaturization techniques to modify the antenna profile. In this technique, the current path is altered in such a way that it is longer than the direct distance between two points that conform the antenna structure. Design examples include inverted-L antenna, helical antenna, and slot loaded microstrip patch, as depicted in Fig. 4. Even though this provides miniaturization, extreme care must be taken during the design process to avoid cancellation of fields due to the surface currents in the opposite direction, which otherwise results in extreme reduction of radiation efficiency.
- ***Employing materials with high dielectric permittivity***: As a general design rule, an antenna structure can be excited to operate in different modes. The mode corresponding to the lowest frequency is called the dominant mode. Each resonant mode provides radiation at different frequencies with its own radiation characteristics. The antenna topologies require a resonant length of approximately $\lambda_g/2$, where $\lambda_g = \lambda_0/\sqrt{\varepsilon_{\text{eff}}}$; λ_0 is the free space wavelength, and ε_{eff} is the effective dielectric constant of the antenna material. The above relation indicates that the wavelength through the medium is inversely proportional to the effective dielectric constant of the material. This indicates that the overall antenna size can be reduced by using high permittivity substrates. For example, the dielectric resonator antennas can be realized with high-permittivity material resulting in antenna dimension reduction as small as $\lambda_0/20$. However, increasing dielectric permittivity of the antenna materials does have consequences such as reduction in the bandwidth because of the high value of the antenna quality factor (Q). It also increases the dielectric loss in the material. In practical scenarios, antenna designers find a trade-off solution between the bandwidth and the overall antenna size.
- **Reactive Loading:** This is one of the commonly adapted techniques to reduce the antenna size in mobile communication handsets. The concept is to add a capacitive or inductive effect to the resonant

DOI 10.1007/978-981-4560-75-7_55-1

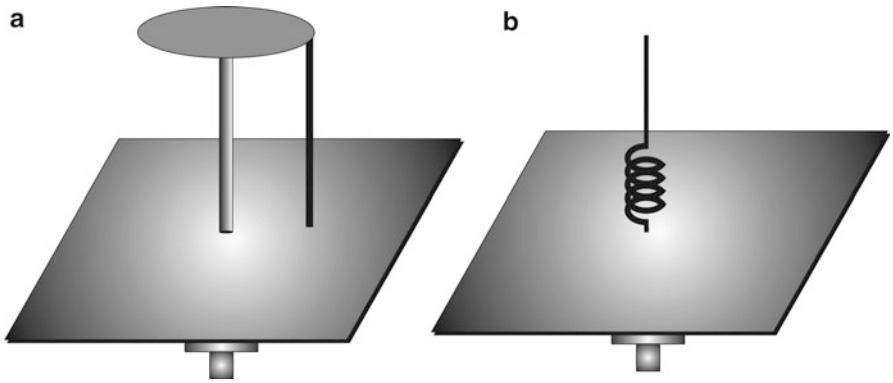


Fig. 5 Reactive loading technique to reduce the overall antenna size (**a**) capacitive loading (**b**) inductive loading

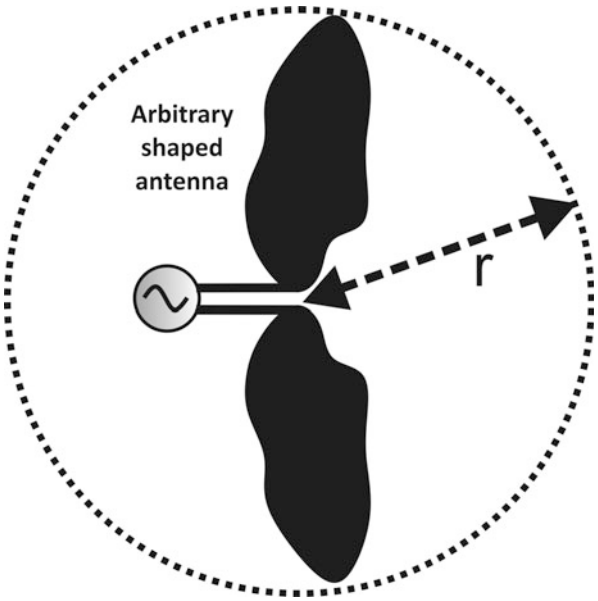


Fig. 6 The smallest sphere encloses the electrically small antenna (ESA)

structure of the antenna, which will lower the operational frequency and thus makes the antenna electrically smaller. In earlier days, the general preference was to modify the antenna structure to facilitate a reactive loading. However, with the recent technological advancements in the manufacturing of lumped reactive components, it is reliable to use lumped elements without much degradation of the overall system efficiency. For example, top-hat loading of a vertical monopole antenna and inductive loading in whip antenna shown in Fig. 5 are two of the conventional methods used for developing low-profile antennas.

Fundamental Limitations

As technology advances there is growing interest in shrinking the size of mobile communication systems. However, there are a few fundamental limitations that need to be respected while designing low-profile antennas. The concepts outlined below help antenna designers to reach a trade-off between the size and the antenna characteristics.

Chu Limit: Chu limit helps to estimate the maximum possible normalized bandwidth of an electrically small antenna from its physical size at a specific frequency. According to Chu's formula (Chu 1948), the

DOI 10.1007/978-981-4560-75-7_55-1

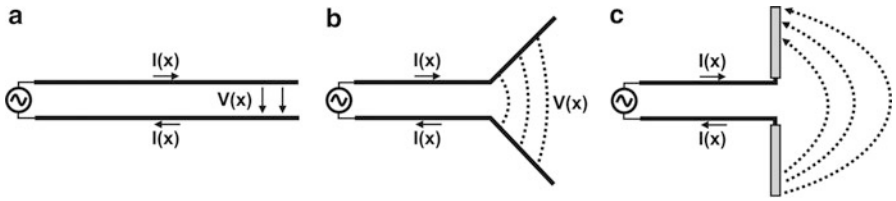


Fig. 7 Formation of a dipole antenna from a two-wire transmission line

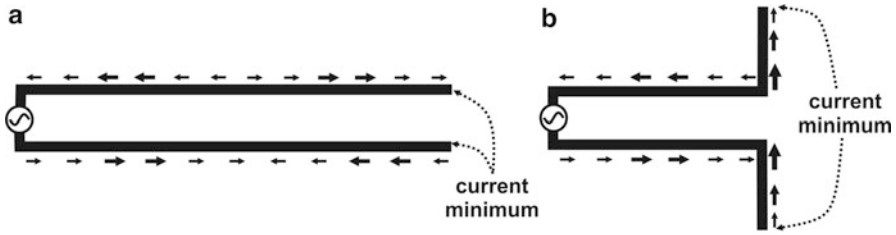


Fig. 8 Current representation in (**a**) balanced transmission line (**b**) dipole antenna

lowest quality factor (Q) of an electrically small antenna is r/λ, where r is the radius of a sphere enclosing the antenna as shown in Fig. 6. In the initial Chu's paper (Chu 1948), the radiation boundary given by Chu was based on a model without much detailed theoretical considerations so as to fit various scenarios. This is modified by other prominent scientists (Geyi 2003; McLean 1996; Wheeler 1947). The approximate value of normalized bandwidth for $Q \gg 1$ of a resonant antenna can be evaluated from (Hansen 1981)

$$\mathrm{Q-factor} = \frac{\mathrm{upper\ \ frequency\,(fu)} - \mathrm{lower\ \ frequency\,(fL)}}{\mathrm{center\ \ frequency\,(fc)}} \tag{1}$$

In practical situations, one of the challenges faced by the antenna designer would be the calculation of the minimal sphere radius. For example, in the case of an antenna mounted on a tablet computer system ground, there are chances that some part of the ground will contribute in the radiation. In that case, considering only the antenna element will result in an error in the calculation.

Types of Low-Profile Antennas

Dipole-Based Designs

Dipoles are one of the fundamental class of antennas that evolved through many years. In fact, Hertz used a dipole antenna on his spark-gap generator during the historical demonstration of electromagnetic wave propagation in 1888. Being one of the fundamental type of radiator, various designs based on a dipole have been proposed for different applications. It is true that the clear understanding of the dipole antenna concepts is necessary for further understanding various dipole-based antennas in the modern era. As shown in Fig. 7, the dipole antenna is a natural extension of a two-wire transmission line into an efficient radiator. The transmission line, shown in Fig. 7a, can be made as an antenna by bending it perpendicular to the axis of the transmission line, as indicated in Fig. 7b–c. This makes two parallel wires which when excited with an alternative signal oscillate the charges across the pair of conductors, which produces alternating electric and magnetic fields. Based on the physical dimensions of these conductors, a portion of the signal power will be radiated into space as a traveling electromagnetic wave.

DOI 10.1007/978-981-4560-75-7_55-1

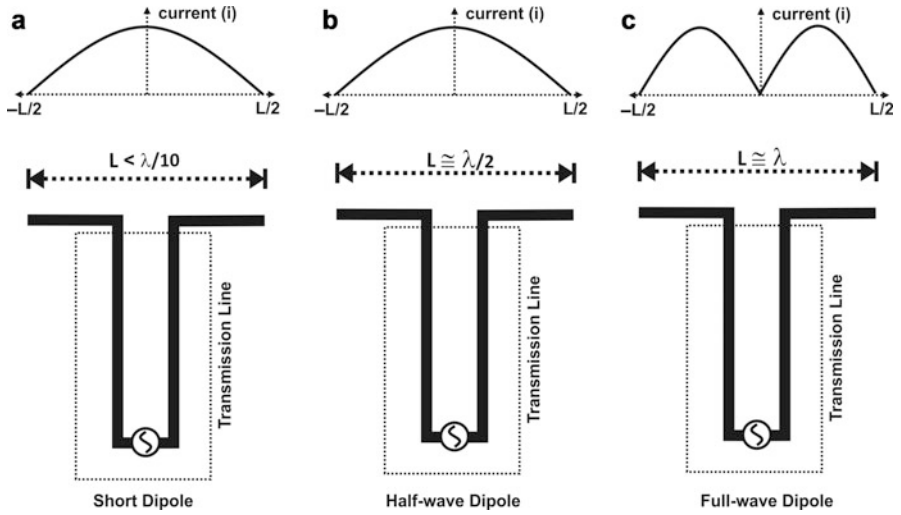
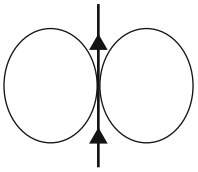
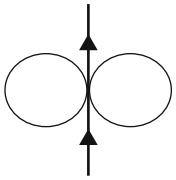
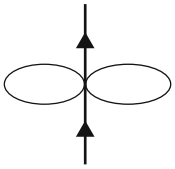


Fig. 9 General classification of dipole antennas (**a**) short dipole (**b**) half-wave dipole (**c**) full-wave dipole

Table 1 Characteristics of different dipole configurations

		Short dipole	Half-wave dipole	Full-wave dipole
Directivity, dB		1.76	2.15	3.8
Input impedance	Real	Very small	≅73 Ω	Very high
	Imaginary	Highly capacitive	≅0	≅0
Half-power beamwidth (HPBW)		∼90°	∼78°	∼47°
Polarization		Linear	Linear	Linear
Radiation pattern				

- **Why dipoles are called so:** This can be better understood by observing the current distribution in the transmission line and of a dipole, as shown in Fig. 8. In an open-circuit transmission line, as illustrated in Fig. 8a, the current at the end of the transmission line is zero (minimum). When we bend the transmission line to form a dipole, at a specific frequency, the current distribution will self-arrange as depicted in Fig. 8b. It is observed from the current directions in Fig. 8b that at a specific instant of time, the current is flowing from the bottom to the top of the dipole. This leads to an assumption that negative charges will be accumulated at the bottom end, while positive charges will be accumulated at the top end of the bended transmission line (dipole). This is in turn observed as two charges or poles (positive and negative) separated by a distance, and therefore it is a di-pole.
- **Basic classifications:** Based on the total length of the radiator(L), the dipole antennas can be classified as short dipole, half-wave dipole, or full-wave dipole as illustrated in Fig. 9. This classification is based on how much longer or shorter the dipole length (L) is with respect to the wavelength (λ) of the signal fed to the antenna. The characteristics of various dipole configurations are outlined in Table 1.
- **Input Impedance:** Impedance (z) is a concept that relates voltage (v) to the current (I) at the input of an antenna terminal. The real part of the impedance is a representation of the part of energy that is either

DOI 10.1007/978-981-4560-75-7_55-1

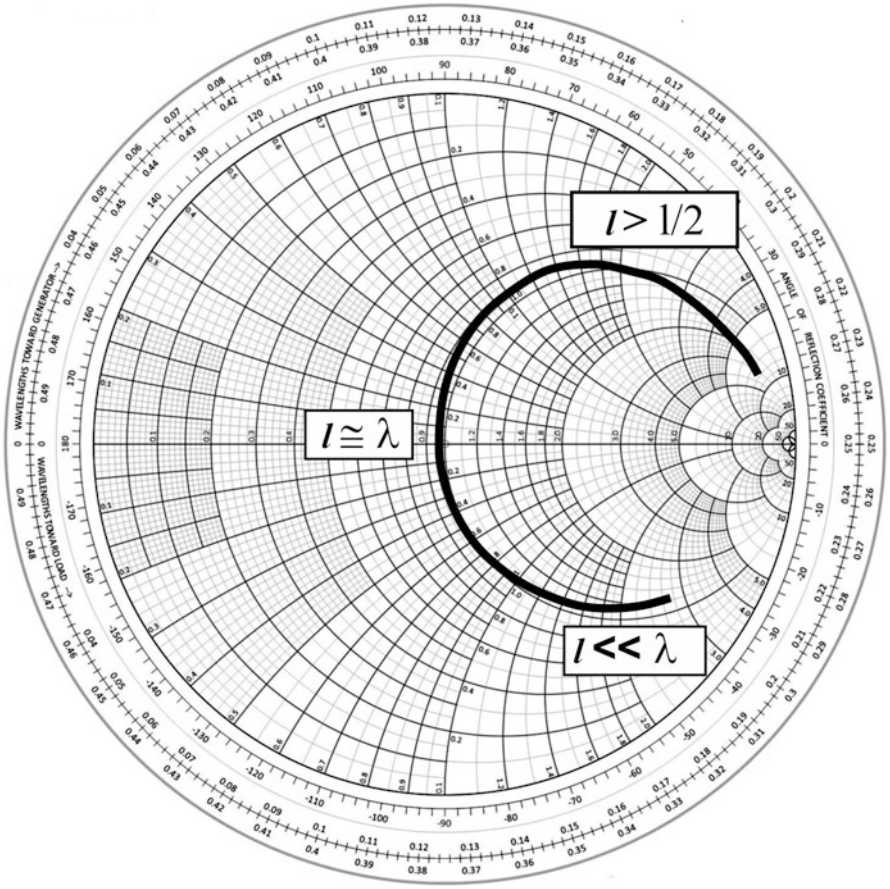


Fig. 10 Smith chart of a dipole antenna

radiated or absorbed by the antenna, while the imaginary part is a function of the energy stored in the field in the close surroundings of the antenna (near-field). An antenna can be made resonant when the antenna impedance is purely real (zero imaginary part). A Smith chart (invented by Philip H. Smith) is a powerful tool to analyze multiple antenna parameters including input impedance. A closer look at the Smith chart shown in Fig. 10 provides a clear view of the input impedance as function of dipole length (L).

It is interesting to note that when the length of the dipole is equal to half the wavelength, the antenna is at its natural resonance, meaning the imaginary part of the impedance is equal to zero. At its natural resonance, the real part of the impedance for a half-wave dipole is approximately 73 Ω. The dipole behaves like a lossy capacitor as the lengths of the poles become very small; this can be understood from the fact that the shorter the dipole, the more it acts as an open-ended transmission line, which is highly capacitive. As the dipole length (L) increases beyond half-wave length, the impedance increases and reaches an antiresonance when L reaches around λ with very high input impedance. When L = ½ λ, the input impedance is independent of the conductor thickness whereas for all other cases, the input reactance is highly sensitive to the diameter of the dipole wire. This makes the half-wave dipole one of the most popular antenna configurations.

- **Current Distribution:** The approximate current distribution on a thin dipole antenna can be mathematically expressed as

DOI 10.1007/978-981-4560-75-7_55-1

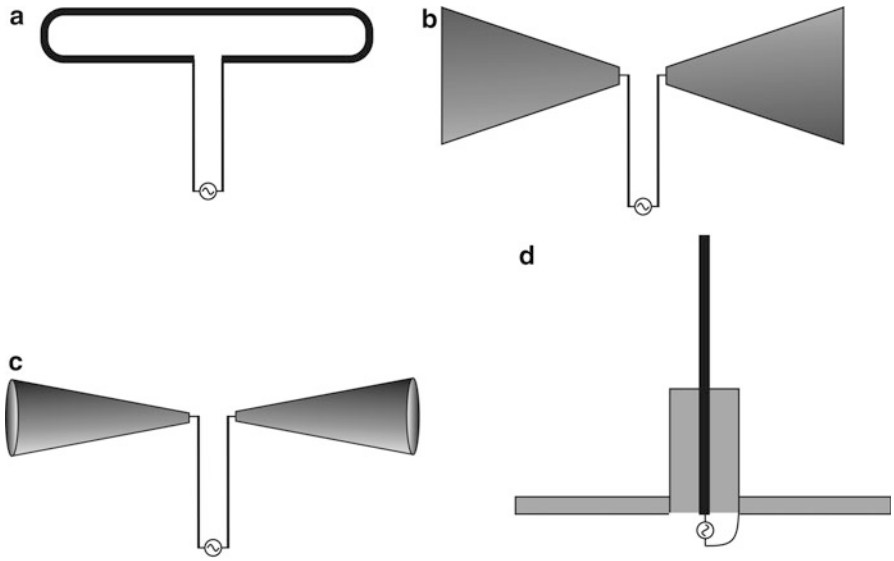


Fig. 11 Design variations of dipole antenna

$$I(z) = I_0 \sin(\beta(l - Z)) \quad \text{for} \quad 0 \le Z \le l \tag{2}$$

$$I(z) = I_0 \sin(\beta(l + Z)) \quad \text{for} \quad -l \le Z \le 0 \tag{3}$$

where I_0 is the peak current on the wire forming the dipole, which is obviously different from the current at the feed point. It reveals that the current at the ends of the dipole always approaches zero. As it can be seen from the equations, the current distribution is also a periodic distribution of one wavelength.

The key features of the half-wave dipole are

- Omnidirectional radiation pattern, which is suitable for most wireless communication scenarios.
- Relatively good directivity suitable for broad coverage in the horizontal plane.
- The input impedance is around 73 Ω with zero reactive part, which is in a reasonable range relatively easy to match with standard impedance matching techniques.
- The reactive impedance is insensitive to the wire diameter, which gives more flexibility and design freedom.

Popular design variations of dipole antennas are outlined in Fig. 11, including a biconical, bow-tie, half-wave, folded, and coaxial sleeve dipole.

State-of-the-Art Low-Profile Dipole Antennas: The advancements in the area of printed circuit technology have provided researchers a good platform to develop various printed antennas. Many antenna designs, including the printed dipole antennas, have evolved from its basic wire antenna counterpart. Table 2 lists some of the state-of-the-art designs in this field. There are various design strategies that have been used for implementing low-profile antennas, such as printing dipole elements on either side of a high-frequency laminate. The thickness can be kept in the range of $0.003\ \lambda_0 < h < 0.012$, while maintaining impedance bandwidths up to 109 % in dipole-based printed antenna designs. The artificially engineered materials have been extensively utilized for enhancing the overall performance of antennas. For example, the zeroth-order resonator (ZOR) can be embedded into multiband low-profile monopole

DOI 10.1007/978-981-4560-75-7_55-1

Table 2 State-of-the-art designs of low-profile dipole antennas

Center freq. (Fc), GHz	Er	Maximum antenna dimension	Thickness	Bandwidth (%)	Average gain in the band	Number of metallic layers	References
2.1	2.2	1.2 λ_0	0.006 λ_0	38.1	5	2	(Tefiku and Grimes 2000)
2.4	2.3	1.1 λ_0	0.004 λ_0	18.8	4.2	2	(Suh and Chang 2000)
2.5	4.4	0.4 λ_0	0.003 λ_0	19.7	3.8	1	(Su et al. 2002)
2.3	2.2	0.8 λ_0	0.012 λ_0	27.9	–	1	(Tilley et al. 1994)
6.9	2.2	2.1 λ_0	0.036 λ_0	109.5	4	2	(Ma and Jeng 2005)
EBG-based designs							
2.0 (dipole with EBG)	3.38 (dipole) 4.5 (EBG)	0.7 λ_0 (with EBG)	0.03 λ_0 (with EBG)	35.2	6.8	4 (dipole, balloon, and EBG layers)	(Azad and Ali 2008)
2.7 (dipole with EBG)	2.2	0.9 λ_0 (with EBG)	0.02 λ_0 (with EBG)	6	6.7	4 (dipole and EBG layers)	(Abedin and Ali 2005)

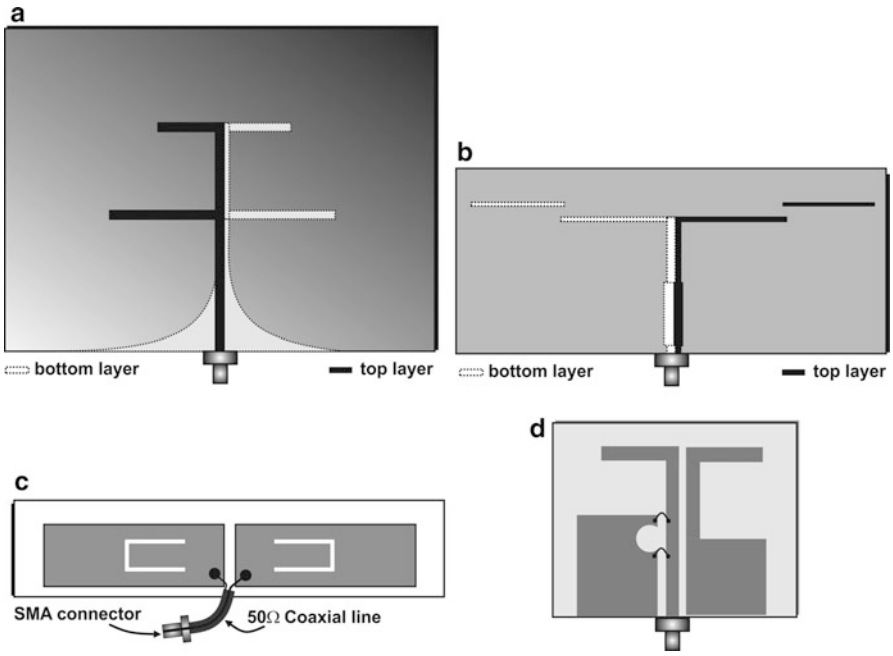


Fig. 12 Some of the popular examples of low-profile dipole-based designs

antennas to reduce the overall size up to 60 % (Mehdipour et al. 2014a). Another interesting design approach is to integrate electromagnetic bandgap structures (EBG) in the ground plane of the patch antenna. This approach provides size reduction while improving the impedance bandwidth and gain (Siya and Yee Hui 2013). Examples of dipole-based low-profile designs are shown in Fig. 12.

Recent Trends: Being one of the popular antenna design, dipole-based antennas are still an interesting design strategy for many researchers around the world. Some of the recent works are outlined in Table 3,

DOI 10.1007/978-981-4560-75-7_55-1

Table 3 Recent developments of dipole based low-profile antennas

Center frequency (Fc), GHz	Er	Maximum antenna dimension	Thickness	Bandwidth (%)	Average gain in the band	Number of metallic layers	References
3.2	2.5	0.93 λ_0	0.017 λ_0	27.3	21.4	2	(Qudrat-E-Maula et al. 2014)
2.4	3.4	0.73 λ_0	0.004 λ_0	71.1	6.7	2	(Lu et al. 2014)
1.7	2.2	0.70 λ_0	0.004 λ_0	103	7.75	2	(Ge and Luk 2014)
2.4	10.2	0.95 λ_0	0.010 λ_0	6.8	3.35	2	(Cure et al. 2014)
1.5	4.4	0.96 λ_0	0.005 λ_0	14.4	5.34	1	(Chen et al. 2014)

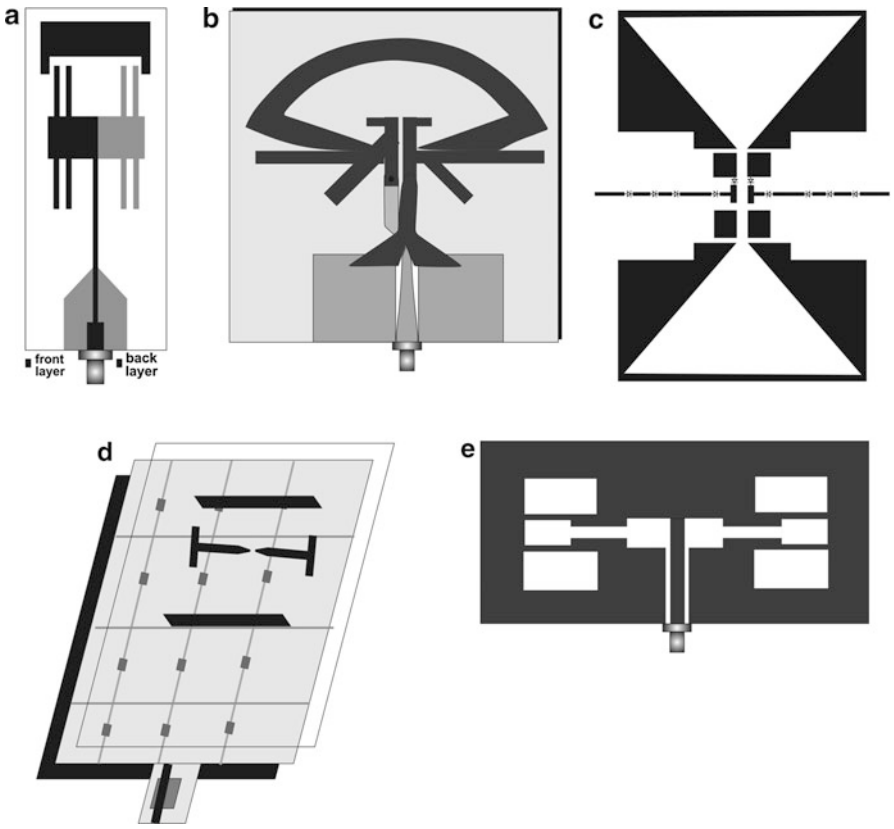


Fig. 13 Examples of recent dipole-based low-profile antennas (**a**) Printed dipole with corrugated elements (Qudrat-E-Maula et al. 2014), (**b**) Loop dipole for broadband applications (Lu et al. 2014), (**c**) Electronically reconfigurable dipole antenna (Ge and Luk 2014), (**d**) Tunable very low-profile dipole antenna loaded with barium strontium titanate (BST) (Cure et al. 2014) and (**e**) CPW-fed broadband printed dipole antenna (Chen et al. 2014)

There are numerous applications which require antennas with equal beamwidth in principal radiation planes. Recent reports have indicated that very low-profile dipole-based directive antennas with thickness of 0.017 λ_0 have been proposed. These antennas provide an average gain up to 21 dB for a corrugated microstrip dipole antenna with integrated reflector (Qudrat-E-Maula et al. 2014). Some of the recent examples of dipole-based designs are shown in Fig. 13. The recent approaches also include effective

DOI 10.1007/978-981-4560-75-7_55-1

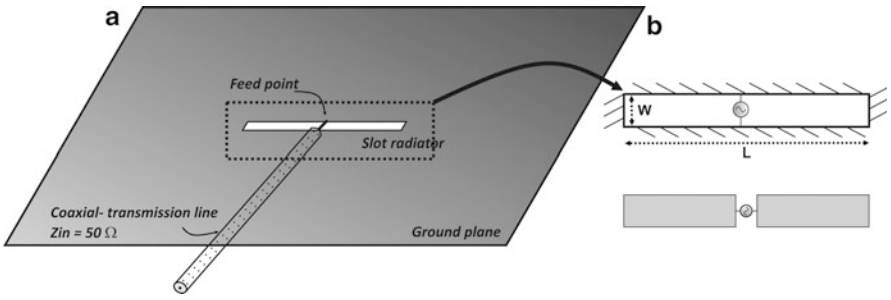


Fig. 14 Structure of a slot antenna (**a**) Slot antenna (**b**) Complementary dipole

utilization of artificially engineered materials for reconfiguring dipole antennas. In a recent work (Zarghooni et al. 2014), a dipole antenna operating at 3.3 GHz with reconfigurable radiation pattern has been reported. This novel approach uses a combination of double-split-ring resonator and spiral resonator as a metamaterial media to facilitate beam steering of 35° in the E-plane. Modified printed dipole antennas are suitable for mobile phone applications. One of the design variants is monopole based, in which one half of a dipole is mounted above the system ground plane. In addition, another attractive design using a uniplanar folded monopole antenna has been proposed (Augustin et al. 2014b). In this design two folded monopole elements and a capacitively coupled **parasiti**c element are integrated to a footprint of 22 × 34 × 0.8 mm^3. The single metal layer design provides three resonant bands, which are broad enough to cover major LTE/GSM/UMTS/WiFi bands.

Slot Antenna Designs

Slot antennas are another canonical design widely considered for the realization of low-profile antennas. It comes under a broad category of "aperture antennas," which are typically used in a wide frequency spectrum from 200 MHz to more than 25 GHz. As the name implies, slot antennas are formed by making a cutout in a relatively large aperture such as the system ground plane. The key design features are

- Ability to form low-profile, conformal antennas
- Ability to form linear/circular polarization
- Nearly omnidirectional radiation patterns

The structure of a slot antenna is outlined in Fig. 14, which shows a slot created in a plain metal sheet. This can be made on one wall of a waveguide or in a printed circuit board. The slot can be excited through various methods, including electromagnetic coupling and coaxial feed lines. The structural parameters that influence the radiation characteristics are

1. Shape of the slot
2. Material for fabrication (e.g., the dielectric constant of the high-frequency laminate)
3. Dimensions of the slot

- **Babinet's Principle:** Babinet's principle in optics states that a source of light behind a complementary thin conducting sheet produces lit regions on the source-free side that when superposed will give a completely lit region, just the same as that would exist without the sheet. In this perspective, the complementary structure of a slot antenna is shown in Fig. 14b, which is a dipole. An extension of this principle from optical to low frequency can be applied to calculate the input impedance of complementary antennas (Lang 1973). Thus, for a self-complementary antenna,

DOI 10.1007/978-981-4560-75-7_55-1

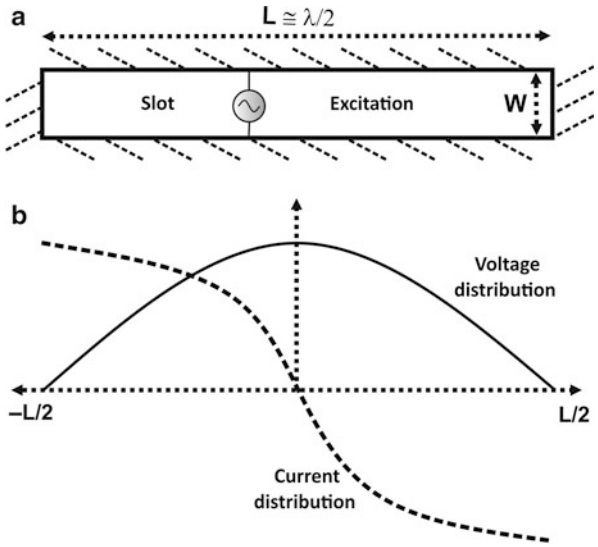


Fig. 15 Voltage and current distribution on a half-wavelength slot antenna

$$Z_{metal}Z_{air} = \frac{\eta^2}{4} \tag{4}$$

where Z_{metal} is the impedance when metal is present, Z_{air} is the impedance when air is present, and η is the intrinsic impedance of the medium, which is 377 Ω for air, and is also given by $\sqrt{\mu_0/\varepsilon_0}$. This implies that, based on the concept of self-complementary structure, a slot antenna will behave as a dipole antenna with electric and magnetic fields interchanged. This will result in interchanging the polarization of the slot antenna compared to a dipole. This means, if a dipole is horizontally polarized, its complementary slot antenna will be vertically polarized.

Half-wave slot antenna: A half-wave (λ/2) slot antenna is shown in Fig. 15a. The antenna is fed from one side of the slot to the other side, and the feed point will be slightly offset from the center in order to have relatively low input impedance. As evident from the voltage (V) and current (I) profile of the antenna, if the slot is fed at the center, a high impedance will result ($I \rightarrow 0$; $V/I \rightarrow \alpha$). It is important to note that the radiation in a slot antenna is due to the time-varying electric fields across the slot which is in contrast to the current variation in a wire dipole. When the antenna is excited, the currents flow in the metal sheet and spread out to the entire sheet. This is complementary to a wire dipole in which the current is more confined. The current vectors around the slot in a slot antenna are equal in magnitude but opposite in phase, therefore they cancel each other. This reveals the fact that, compared to a slot antenna, relatively higher amount of current flow is required in a wire dipole to produce a given wattage of radiation. The slot antenna can take different shapes ranging from simple square to an exponential curve. The shape of the slot influences the radiation characteristics of the antenna. Examples of various slot antenna structures are outlined with their key merits in Table 4.

- **State-of-the-art slot antennas:**
  There are various low-profile slot antenna designs that were quite popular during the last few decades. This includes various slot shapes with different excitation mechanisms. The progress in low-cost printed circuit board materials along with modern fabrication techniques has facilitated extreme growth in this sector. Some examples of low-profile slot antenna designs along with their

DOI 10.1007/978-981-4560-75-7_55-1

Table 4 Various slot shapes and their key benefits

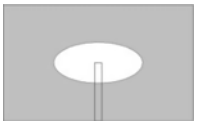
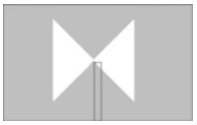
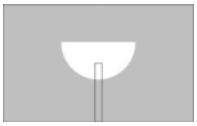
Slot type	Key merits	Shape
Tapered slot	Directive pattern, broad bandwidth	
U slot	Broad bandwidth	
Rectangular slot	Omnidirectional pattern	
Elliptical slot	Broad bandwidth	
Bow-tie	Broad bandwidth	
Annular	Multiband resonance	

Table 5 State-of-the-art designs of low-profile slot antenns

Center frequency (F_c), GHz	Er	Slot shape	Maximum Antenna dimension	Thickness	Bandwidth (%)	Average gain in the band	Number of metallic layers	References
2.37	4.4	Square	0.90 λ_0	0.006 λ_0	46.1	4.7	2	(Sze and Wong 2001)
2.22	4.4	Square	0.75 λ_0	0.012 λ_0	59.5	4.5	1	(Chen 2003)
6.85	4.7	Elliptical	0.48 λ_0	0.034 λ_0	109.5	4.3	1	(Li et al. 2006)
6.14	3.4	Square	2.89 λ_0	0.017 λ_0	109.0	3.5	1	(Chair et al. 2004)
5.45	4.4	Triangular	2.83 λ_0	0.015 λ_0	111.2	8.3	2	(Liu et al. 2004)

characteristics are outlined in Table 5. These examples are selected based on the low-profile characteristics in which the thickness of the antenna is between 0.006 and 0.034 λ_0. The overall dimensions,

DOI 10.1007/978-981-4560-75-7_55-1

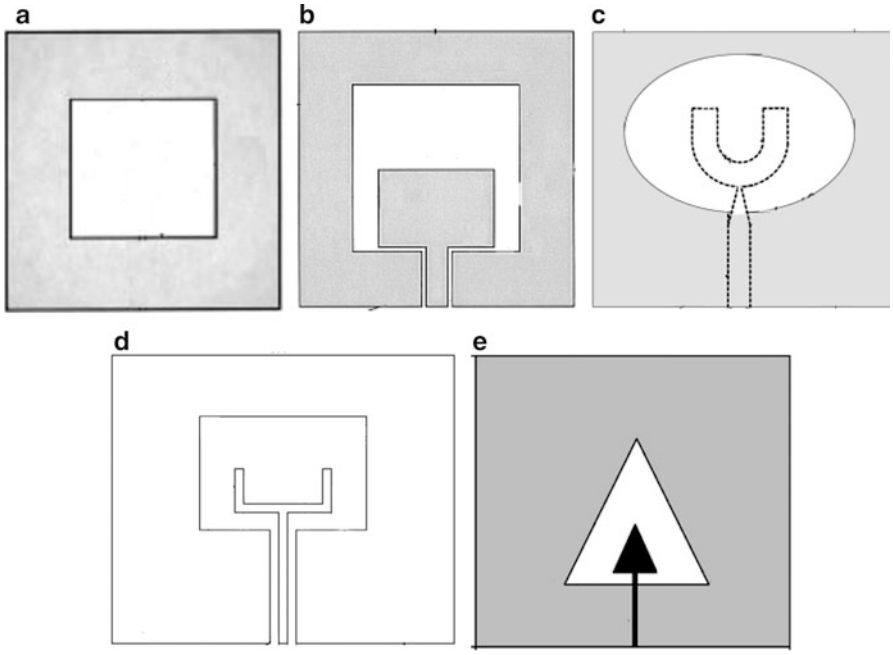


Fig. 16 Examples of most popular slot antenna designs (**a**) Broadband microstrip-fed square slot antenna (Sze and Wong 2001), CPW-fed uniplanar slot antenna (Chen 2003), Ultrawideband elliptical slot antenna (Li et al. 2006), U-stub-loaded rectangular slot antenna (Li et al. 2006), and Triangular slot antenna (Li et al. 2006)

Table 6 Examples of recent slot antenna designs [$\lambda_0 = C/Fc$]

Center frequency (F_c), GHz	Er	Slot shape	Overall Antenna dimension	Thickness	Bandwidth %	Average gain in the band dBi	Number of metallic layers	References
2.0	4.4	Stepped	0.85 λ_0	0.005 λ_0	10.0	3.46	2	(Wang and Chen 2014)
2.5	4.4	Annular ring	0.60 λ_0	0.014 λ_0	5.4	1.2	1	(Sim et al. 2014)
10.2	2.2	Bow-tie	0.81 λ_0	0.027 λ_0	9.4	3.49	1	(Mukherjee et al. 2014)
3.0	4.4	Square	0.58 λ_0	0.010 λ_0	38.8	4.255	1	(Li et al. 2014a)
5.8	2.2	Rectangular	1.58 λ_0	0.232 λ_0	9.8	7.75	2	(Chen and Shen 2014)

including the ground plane and feeding, range from 0.5 to 2.89 λ_0 in these designs. The geometries of these designs are outlined in Fig. 16.

- **Recent Trends:**

  The ability of slot antennas to use a broad surface made them popular for various applications. The recent technological advancements in the field of wireless communication demand conformal antennas for various communication modules. Some of the designs such as tapered slot/elliptical slots facilitate simultaneous excitation of various modes in slot antenna designs. This feature makes them popular for antenna designs that require a broad VSWR bandwidth. Some of the recent slot antenna designs are listed in Table 6.

DOI 10.1007/978-981-4560-75-7_55-1

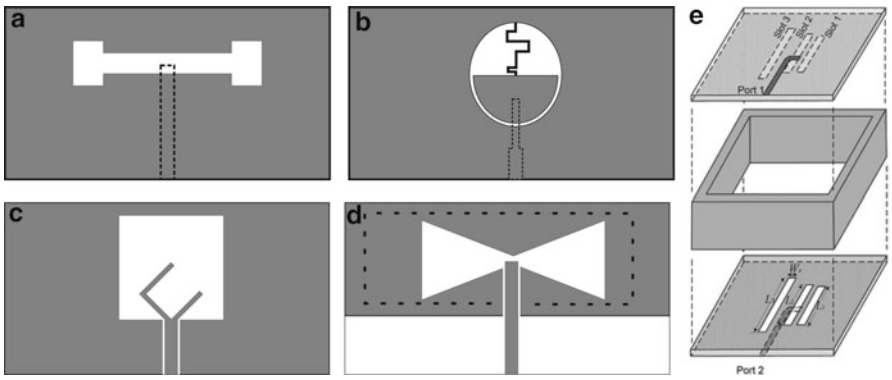


Fig. 17 Examples of most recent slot antenna designs. (**a**) Stepped slot antenna (Wang and Chen 2014), (**b**) Annular slot with frequency configurability (Sim et al. 2014), (**c**) Square slot antenna with U stub (Li et al. 2014b), (**d**) Broadband bow-tie slot antenna (Mukherjee et al. 2014), and (**e**) Cavity-backed end-fire slot antenna (Chen and Shen 2014)

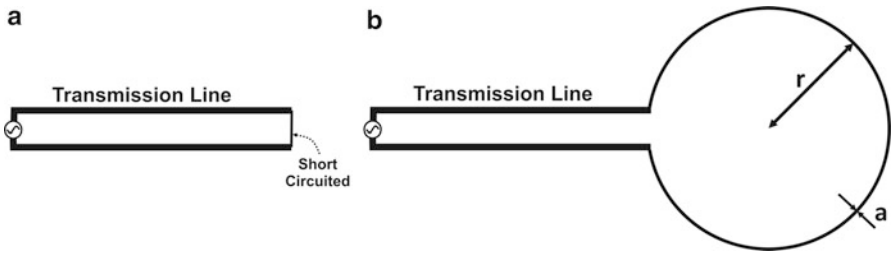


Fig. 18 Short-circuit transmission line and a loop antenna

The classical shapes such as annular ring and bow tie are still attractive candidates for many researchers due to their key design characteristics. The recent developments in very thin-profile slot antenna design indicate that there are designs with thickness as low as 0.005 λ_0, which has an overall maximum antenna dimension of 0.85 λ_0. Some interesting slot antenna shapes have recently been developed and are outlined in Fig. 17. The slot antenna topology also provides a good platform for developing dual-port integrated antennas. For example, in Augustin and Denidni (2012), a coplanar waveguide-fed slot antenna is integrated at a space between two tapered slots. This dual-port (ultrawideband/narrowband) design provides a single metallic-layer antenna structure with good interport isolation without degrading the frequency- and time-domain characteristics of the ultrawideband antenna. Another feature is the easy integration of RF circuits for antenna reconfigurability. An interesting application of this method is in cognitive radios (Augustin et al. 2014a; Chacko et al. 2015), in which the communication system can electronically change its transmitter parameters by detecting the environment. Another exciting application of dual-port slot antenna is for enhancing the diversity performance of multiple-input-multiple-output (MIMO) wireless communication systems (Ayatollahi et al. 2012). An orthogonal feeding in a modified annular slot can provide good polarization diversity with low envelope correlation coefficient and nearly omnidirectional radiation pattern (Chacko et al. 2013). The above reports indicate that the slot antennas provide attractive features that have potential applications in modern wireless communication systems.

Loop Antennas

The historical root of loop antennas goes back to Hertz's radio system, where a dipole antenna was employed in the transmitter and a loop antenna in the receiver module. In contrast to the dipole antenna, which is evolved from an open-circuit transmission line, the loop antenna can be seen as developed from a

DOI 10.1007/978-981-4560-75-7_55-1

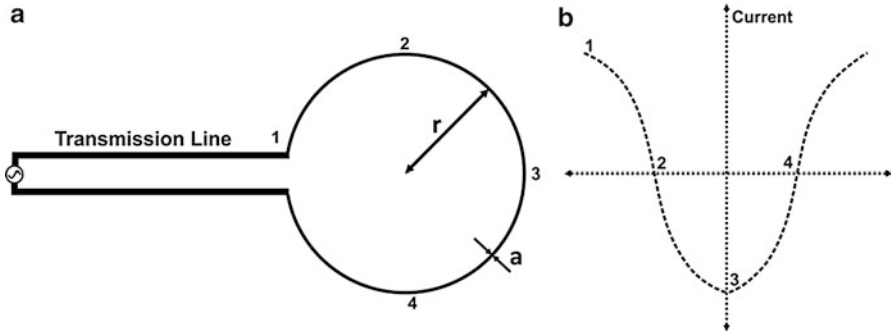


Fig. 19 Current in a resonant loop

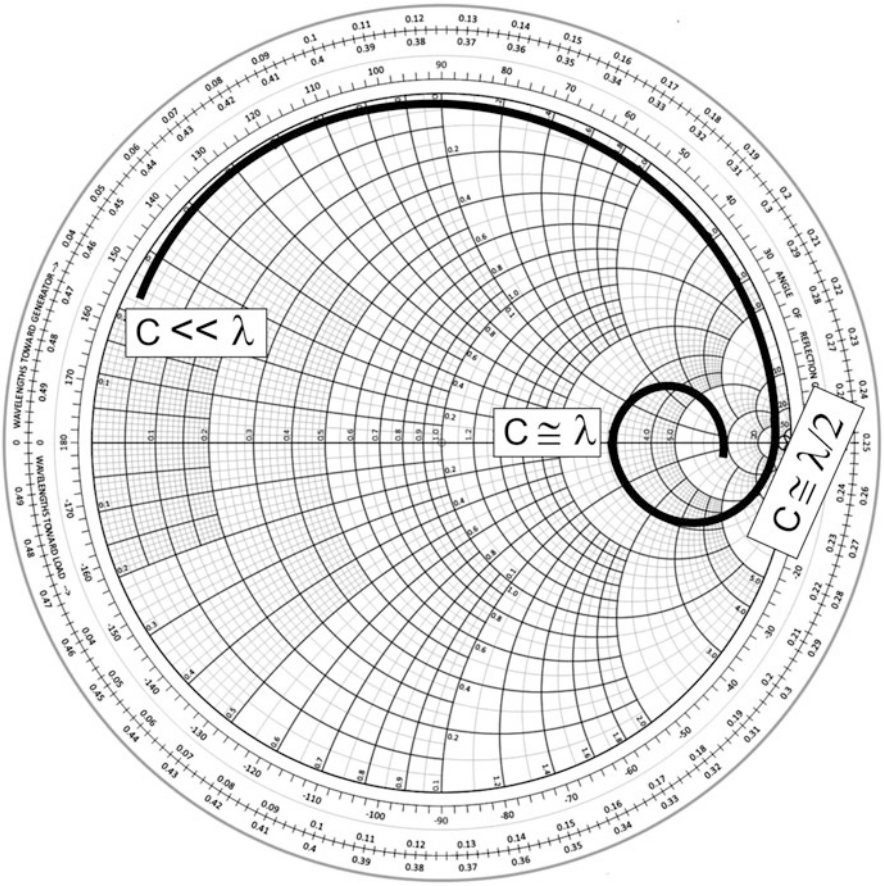


Fig. 20 Smith chart of a loop antenna demonstrating impedance characteristics at different loop circumferences

short-circuit transmission line, as illustrated in Fig. 18. In its simple form, the loop antenna consists of a balanced transmission line feeding a single or multiple loop(s) of electrical conductor. The loop antenna becomes resonant (imaginary part of the impedance approaches zero), when its circumference approaches one wavelength. For example, an IEEE 802.11 b/g WiFi antenna, operating at 2.4 GHz, will only need to be 12.5 cm circumference. It also can form different shapes such as circular, triangular, square, elliptical, and so forth. This structural flexibility makes them capable of becoming electrically small as well as electrically large with loop circumference ranges from 0.1 to 1.0 λ. Another important design parameter is the number of loops, which is directly proportional to the radiation resistance. This infers that electrically

DOI 10.1007/978-981-4560-75-7_55-1

Table 7 Examples of popular low-profile loop antenna designs

Center frequency (F_c), GHz	Er	Loop structure	Overall antenna size λ_0	Thickness λ_0	Bandwidth %	Average gain in the band dBi	Number of metallic layers	References
0.893	4.4	Rectangular	0.35	0.030	15.9	1.25	2	(Wong and Huang 2008)
0.9875	4.4	Meandered	0.41	0.003	33.92	1.25	2	(Wong and Chen 2009)
4.80	3.5	Square	1.36	0.075	25	8.65	4	(Deo et al. 2010)
1.25	4.4	Rectangular	0.18	0.042	8	–	2	(Lee et al. 2008)
1.5	1	Circular	0.89	0.064	74.67	7	2	(Nakano et al. 2000)

small loops with a single turn have relatively small radiation resistance. For example, if the circumference of the loop is smaller than $\lambda/3$, the current around the loop is much higher than the voltage across the loop. This results in smaller values of the input impedance, which is quite difficult to match with the transmitter. However, the antenna designer needs to choose the shapes in such a way that it will not result in the cancellation of current and thereby reduction in the antenna efficiency and bandwidth. For example, a widely used technique of meandering can introduce capacitance of the loop and cause the currents to cancel.

Duality – Loop and Dipole: Dipole antenna is considered as the "dual" of the loop antenna because it is possible to predict the electric/magnetic field of the other by knowing the field of either one of the antennas. Based on the duality theorem, if the basic principle behind radiation in one antenna is magnetic current then the other's would be electric current. This implies that the electric far field of a loop antenna is identical to the magnetic far field of the dipole antenna. Thus both dipole and loop antennas have identical donut-shaped radiation patterns, but the nulls in the dipoles are in the direction of the wire in the dipole, while the nulls in the loop are aligned perpendicular to the plane of the loop. It is also important to note that being a modified short-circuit transmission line, the loop antennas can induce high current in the coil. This is in contrast with the dipole antennas, which are modified open-ended transmission lines. In fact, the current in a resonant loop antenna is a cosine function as illustrated in Fig. 19. However, the overall efficiency of a loop antenna depends on the relative values of the dissipation resistance and radiation resistance.

Radiation resistance: The radiation resistance of a an electrically small loop antenna is relatively smaller than that of a short dipole. This reduces the efficiency of these antennas and limits them for those applications, where the overall system efficiency is the critical design criterion. The radiation resistance of a small loop antenna is given as

$$R_r = 20\pi^2 \left(\frac{C}{\lambda}\right)^4 \tag{5}$$

where C is the circumference of the loop and λ is the free-space wavelength. The Smith chart representation of the input impedance of a loop antenna is illustrated in Fig. 20. At lower frequencies (loop circumference $<< \lambda$), the loop behaves almost like a short circuit, meaning as frequency increases, the inductive impedance of the loop increases, which leads to increased radiation resistance. When the loop

DOI 10.1007/978-981-4560-75-7_55-1

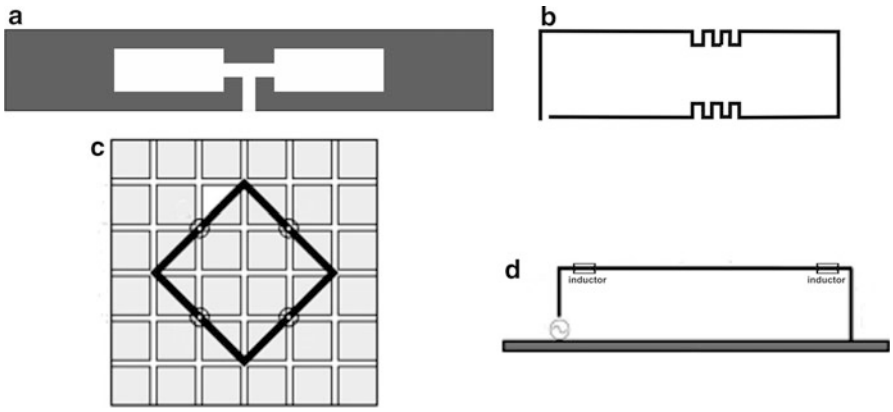


Fig. 21 Examples of various loop shapes in low-profile antenna designs. (**a**) multiband loop structure (Wong and Huang 2008), (**b**) meandered square loop structure (Wong and Chen 2009), (**c**) square loop above a high impedance surface (Deo et al. 2010), (**d**) inductor loaded

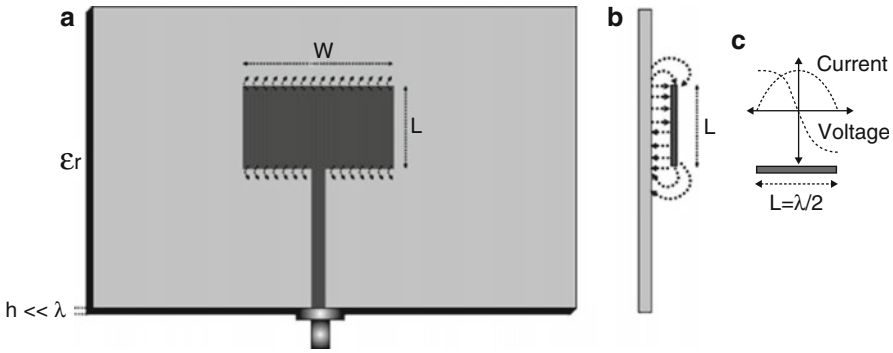


Fig. 22 Microstrip patch antenna (**a**) *top* view (**b**) side view (**c**) voltage and current in a half-wave patch

circumference is close to half wavelength, the trace crosses the real axis on Smith chart (reflection coefficient close to 1). The resonance condition occurs when C is approximately a wavelength long.

Loop antennas and human body: It is well known that the human body tends to be conductive and possesses high permittivity (Salonen et al. 2004). This makes antennas that have strong electric field in the close vicinity (near field) highly sensitive when they are placed near a human body. For example, a dipole antenna is not a good selection for body-centric wireless communication systems because permittivity acts on electric field and in turn tunes the radiation characteristics. On the other hand, a human body is relatively less magnetic. In addition, the magnetic fields are stronger than the electric fields in the near field of a loop antenna. This implies that the loop antennas are highly robust since they are relatively insusceptible to the human body. Hence loop antennas are frequently used in wearable systems such as a hearing aid implant or smartwatch. Some of the key benefits of the loop antennas are outlined below:

- Possibility of low-profile structural design
- Omnidirectional/directional radiation pattern
- Design flexibility for unbalanced or balanced feeding
- Linear/circular polarization
- Robust design for wearable antenna applications

DOI 10.1007/978-981-4560-75-7_55-1

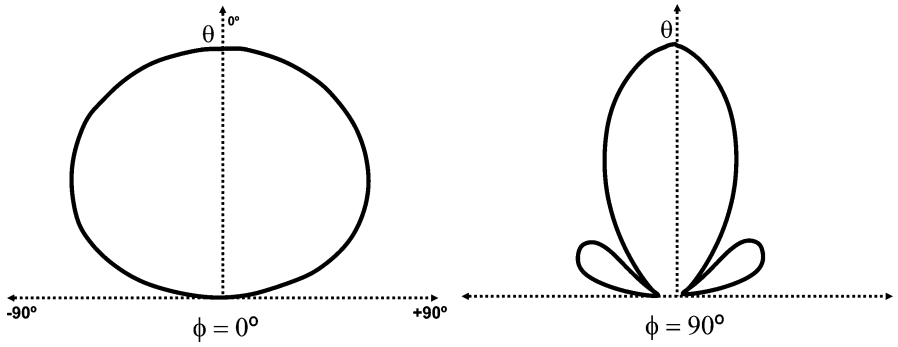


Fig. 23 Radiation pattern of a microstrip patch antenna in two principal planes

- **Popular Loop Antenna Designs:**
  There are various design modifications of low-profile loop antennas. They are very promising candidates for various applications, including mobile communication handsets (Denidni et al. 2005; Wang et al. 2008). For this purpose, various designs have been proposed to operate as full-wavelength structures. For instance, some popular low-profile antennas are outlined in Table 7. A vast majority of the literature focuses on the development of low-profile loop antennas for various commercial applications. For example, most of the designs outlined in Table 7 use loop shapes such as meander line, square, and circular above a system ground plane. These designs are prototyped on a very thin FR4 laminate (Lee et al. 2008; Wong and Chen 2009; Wong and Huang 2008). Another attractive strategy is a coplanar waveguide-fed design (Denidni et al. 2005) that demonstrates a wideband impedance bandwidth (1.68–2.68 GHz) without degrading the radiation efficiency (Fig. 21).

Microstrip Patch/Planar Inverted F Antenna

Microstrip patch antennas are one of the most popular low-profile designs commercially available today. The initial concepts were introduced in the early 1950s (Deschamps 1953). These ideas faced some challenges for experimental realization due to the unavailability of low-cost standard dielectric material. The technological advancements during the mid-1970s made significant progress to these concepts. This attracted large numbers of researchers who were focused on the development of patch antennas for various wireless communication applications. In the modern era, cutting-edge technology allows the patch antennas to be printed directly on a circuit board. The key attractions include ease of fabrication and low cost.

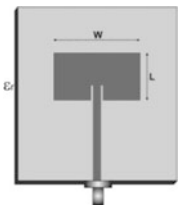
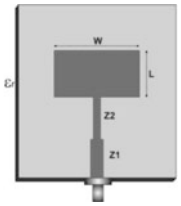
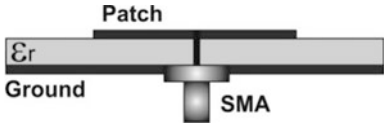
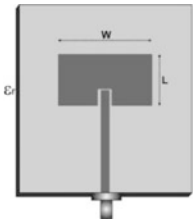
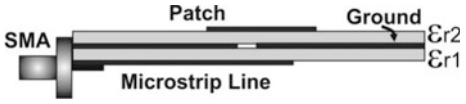
- **Basic Concepts:**
  As shown in Fig. 22a the patch antenna in its basic form consists of a piece of metal plate over a large ground plane. The antenna structure incorporates a cavity with a dielectric material or air. The center frequency of the patch antenna can be approximately calculated as

$$F_c \cong \frac{C}{2L\sqrt{\varepsilon_r}} \tag{6}$$

  When excited with an alternating signal, the voltage measured across the patch and the ground plane and its relation to the current over the patch for a half-wavelength patch are shown in Fig. 22c. On the other hand, the current on the patch adds up in phase, but equal and opposite currents over the patch and

DOI 10.1007/978-981-4560-75-7_55-1

Table 8 Various feeding mechanisms

Feed type	Illustration	Comments
Insert feed		The insert length controls the input impedance which is a function of cos(πD/L), where D is the insert length
Stepped impedance txn line		The concept is to match impedance using a stepped impedance transformer
Probe feed		In this method, the feed has more freedom to alter the feed position and in turn to control the impedance
Capacitively coupled feed		The capacitive gap introduces an extra degree of design flexibility by providing a reactive element
Aperture coupled feed		The key benefit is that the transmission line is shielded by the ground plane and in turn avoids spurious radiation from the microstrip line

ground plane cancel each other. Thus a microstrip patch antenna is a voltage radiation in which the radiation arises from the fringing fields as illustrated in Fig. 22b. This is in contrast to the wire antennas which are current radiators. The radiation pattern of a patch antenna can be expressed as

$$E\theta = \frac{\sin\left(\frac{kW\sin\theta\sin\varphi}{2}\right)}{\frac{kW\sin\theta\sin\varphi}{2}}\cos\left(\frac{kL}{2}\sin\theta\cos\varphi\right)\cos\varphi \tag{7}$$

DOI 10.1007/978-981-4560-75-7_55-1

Table 9 Design parameters and influence on various antenna characteristics

Parameter	Influence
L	Mainly controls the resonant frequency $F_c \cong \frac{C}{2L\sqrt{\varepsilon_r}}$
W	Controls the input impedance (in general, the wider the patch, the lower the input impedance) Influence on radiation pattern
h	Influences the bandwidth (higher the height, the broader the bandwidth)
ε_r	It controls the fringing of the fields and thereby the radiation pattern. Lower the dielectric constant, wider the fringing and broader the radiation pattern The input impedance increases with higher permittivity Higher the permittivity, the smaller the patch size (it is evident from the basic equation for Fc that increasing ε_r four times results in size reduction by two times)

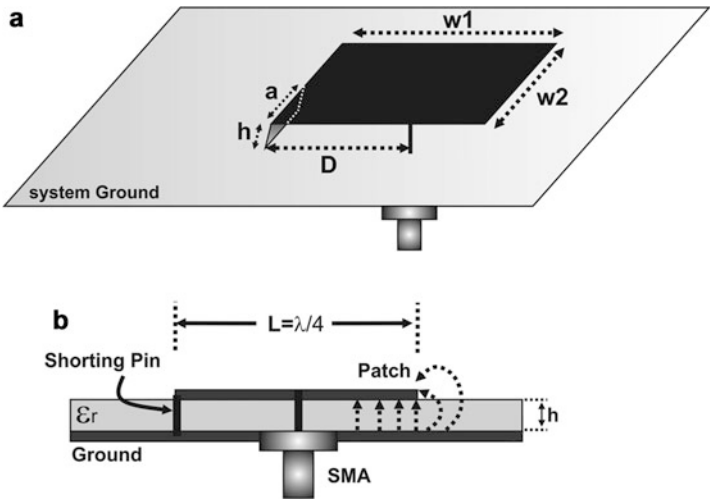


Fig. 24 Planar inverted F antenna (**a**) perspective view (**b**) side view

$$E\varphi = -\frac{\sin\left(\frac{kW\sin\theta\sin\varphi}{2}\right)}{\frac{kW\sin\theta\sin\varphi}{2}}\cos\left(\frac{kL}{2}\sin\theta\cos\varphi\right)\cos\theta\sin\varphi \tag{8}$$

where k = free space wave number = $2\pi/\lambda$. The radiation pattern in two principal planes are depicted in Fig. 23.

- **Feeding Techniques and Design Parameters:**

The excitation techniques of microstrip antennas are outlined in Table 8. This design flexibility of having broad range of feeding options is one of the reasons for their popularity in the industrial sector. On the other hand, the structural parameters of the antenna have strong influence on their radiation characteristics. For example, the dielectric constant does have an impact on how bowed the fringing of the field is, which means the higher the dielectric constant, the less bowed the fields become. The width of the patch antenna controls the input impedance and bandwidth. For example, in a square patch, the input impedance is approximately 300 Ω. However, by increasing the width, the input impedance can be reduced. Table 9 summarize structural parameters of a microstrip patch along with their influence on various antenna characteristics.

- ***Planar Inverted F Antenna:***

Table 10 State-of-the-art microstrip patch/PIFA antennas

Antenna type	Center frequency (F_c), GHz	Er	Patch structure	Overall antenna dimension λ_0	Thickness λ_0	Bandwidth %	Average gain in the band dBi	Number of metallic layers	References
Microstrip patch antenna	2.405	1	E-shaped	0.68	0.12	21.21	8	2	(Yang et al. 2001)
	4.475	1	Rectangular	0.65	0.04	32.4	7.25	2	(Lee et al. 1997)
	1.537	4.8	Circular	0.06	0.007	1.041	2.9	2	(Waterhouse 1995)
	4.6	1	Square	0.59	0.101	36.52	7.5	2	(Luk et al. 1998)
Planar inverted F antenna	1.8	1	Rectangular	0.16	0.03	12.22	-	2	(Rowell and Murch 1997)
	0.905	1	Rectangular	0.11	0.012	9.945	-	2	(Rowell and Murch 1998)
	0.892	4.4	Rectangular	0.13	0.002	15.7	1.85	2	(Chang and Wong 2009)
	0.925	4.6	Rectangular	0.13	0.019	22.7	0	2	(Cabedo et al. 2009)

DOI 10.1007/978-981-4560-75-7_55-1

The basic concept behind a planar inverted F antenna has originated from an effort to further reduce the size of a microstrip patch antenna. At the center of a half-wave patch antenna, the current peaks and an additional short will reduce the overall size of the patch by half, while maintaining the same current voltage distribution. This can be obtained by short-circuiting a half-wave patch at the center as shown in Fig. 24.

The shorting distance, D, controls the input impedance of the PIFA. The closer the feed to the shorting pin the smaller will be the input impedance. Therefore, distance D acts as a tuning parameter for the PIFA. The resonant frequency of the PIFA also depends on the sorting pin dimension "a." An approximate relation that determines the frequency of a PIFA is given by Eq. 9,

$$F_{\mathrm{GHz}} = \frac{30}{4 \cdot [W_{1\ \mathrm{cm}} + W_{2\ \mathrm{cm}} - a_{\mathrm{cm}}]} \tag{9}$$

As depicted in Fig. 24b, this results in the disappearance of fringing fields at one end of the patch due to the shorting to the ground. Meanwhile, the fringing on the other end of the patch remains the same and contributes to the radiation. This provides the benefit of overall size reduction up to 50 % while maintaining the basic properties of the patch antenna. However, one of the immediate consequences is the reduction in the gain.

State-of-the-Art Designs

There are various microstrip antenna designs that become conceptual foundations for various modern antenna designs. The present technology enables us to prototype these printed low-profile designs through various techniques including laser-based techniques. Some of the fundamental microstrip/PIFA antennas are tabulated in Table 10.

Most of these designs that evolved from the basic patch antenna design do have various patch shapes such as "E" shape, rectangular, circular, etc. The overall patch dimensions vary from 0.11 to 0.68 λ_0 with thickness ranges from 0.007 to 0.12 λ_0. It is interesting to note that most of these foundational designs do have bandwidth limitations. Some designs such as Yang et al. (2001)and Cabedo et al. (2009) employed bandwidth enhancement techniques such as slot loading or L-probe feeding. Patch antennas also provide electronic reconfigurability by tuning the surface current path using RF switching diodes. One of the simple methods is to load a square microstrip patch with a hexagonal slot and switch between resonant modes TM_{10} and TM_{01} with a PIN diode (Shynu et al. 2005, 2006). The recent technological advancements in the field of microelectromechanical systems (MEMS) are also utilized to develop reasonably efficient patch antennas with improved reconfigurability (Rajagopalan et al. 2014). The design simplicity of the patch antennas makes them easily adaptable to broad range of applications. For instance, the reinforced continuous carbon fiber composite (RCCFC) materials are widely used in aviation industry because of their outstanding mechanical and chemical properties (Mehdipour et al. 2013). The patch antenna design can be made reconfigurable by loading RCCFC composite which makes them an excellent candidate for military applications (Mehdipour et al. 2014b). Despite their attractive features of simple design and low-cost fabrication, they suffer from relatively low gain. One of the conventional techniques is to form an antenna array at the cost of large size and complex feed network. The use of artificially engineered materials is a modern technique to resolve this issue (Augustin et al. 2013; Chacko et al. 2014; Niroo-Jazi et al. 2013).

DOI 10.1007/978-981-4560-75-7_55-1

Conclusion

This chapter has outlined low-profile antenna designs that evolved through the last few decades. The fundamental low-profile antenna designs have been outlined that include dipole antennas, slot antennas, loop antennas, microstrip patches, and planar inverted F antennas. The objective of this chapter is to help the early-stage antenna designers with fundamental concepts along with recent updates of this magnificent field of antenna engineering.

Cross-References

- ▶ Antennas in Hand-Held Devices
- ▶ Dielectric Resonator Antennas
- ▶ Small Antennas

References

Abedin MF, Ali M (2005) Effects of EBG reflection phase profiles on the input impedance and bandwidth of ultrathin directional dipoles. IEEE Trans Antennas Propag 53(11):3664–3672. doi:10.1109/TAP.2005.858584

Augustin G, Chacko BP, Denidni TA (2013) A zero-index metamaterial unit-cell for antenna gain enhancement. Paper presented at the IEEE Antennas and Propagation Society, AP-S international symposium (digest)

Augustin G, Chacko BP, Denidni TA (2014a) Electronically reconfigurable uni-planar antenna for cognitive radio applications. IET Microwaves Antennas Propag 8(5):367–376. doi:10.1049/iet-map.2013.0287

Augustin G, Chacko BP, Denidni TA (2014b) Uniplanar folded monopole antenna for mobile phone applications in LTE/GSM/UMTS/WiFi band. Paper presented at the IEEE Antennas and Propagation Society, AP-S international symposium (digest)

Augustin G, Denidni TA (2012) An integrated ultra wideband/narrow band antenna in uniplanar configuration for cognitive radio systems. IEEE Trans Antennas Propag 60(11):5479–5484. doi:10.1109/TAP.2012.2207688

Ayatollahi M, Rao Q, Wang D (2012) A compact, high isolation and wide bandwidth antenna array for long term evolution wireless devices. IEEE Trans Antennas Propag 60(10):4960–4963

Azad MZ, Ali M (2008) Novel wideband directional dipole antenna on a mushroom like EBG structure. IEEE Trans Antennas Propag 56(5):1242–1250. doi:10.1109/TAP.2008.922673

Cabedo A, Anguera J, Picher C, Ribó M, Puente C (2009) Multiband handset antenna combining a PIFA, slots, and ground plane modes. IEEE Trans Antennas Propag 57(9):2526–2533

Chacko BP, Augustin G, Denidni TA (2013) Uniplanar slot antenna for ultrawideband polarization-diversity applications. IEEE Antennas Wirel Propag Lett 12:88–91. doi:10.1109/LAWP.2013.2242841

Chacko BP, Augustin G, Denidni TA (2014) High-gain metamaterial antenna for point-to-point communication systems. Paper presented at the 8th European conference on antennas and propagation, EuCAP 2014

Chacko BP, Augustin G, Denidni TA (2015) Electronically reconfigurable uniplanar antenna with polarization diversity for cognitive radio applications. IEEE Antennas Wirel Propag Lett 14:213–216. doi:10.1109/LAWP.2014.2360353

Chair R, Kishk AA, Lee KF (2004) Ultrawide-band coplanar waveguide-fed rectangular slot antenna. IEEE Antennas Wirel Propag Lett 3(1):227–229

Chang CH, Wong KL (2009) Printed λ/8-PIFA for penta-band WWAN operation in the mobile phone. IEEE Trans Antennas Propag 57(5):1373–1381

Chen HD (2003) Broadband CPW-fed square slot antennas with a widened tuning stub. IEEE Trans Antennas Propag 51(8):1982–1986

Chen SW, Wang DY, Tu WH (2014) Dual-band/tri-band/broadband CPW-Fed stepped-impedance slot dipole antennas. IEEE Trans Antennas Propag 62(1):485–490

Chen Z, Shen Z (2014) A compact cavity-backed endfire slot antenna. IEEE Antennas Wirel Propag Lett 13:281–284

Chu LJ (1948) Physical limitations of omni-directional antennas. J Appl Phys 19(12):1163–1175

Cure D, Weller TM, Price T, Miranda FA, Van Keuls FW (2014) Low-profile tunable dipole antenna using barium strontium titanate varactors. IEEE Trans Antennas Propag 62(3):1185–1193

Denidni TA, Lee H, Lim Y, Rao Q (2005) Wide-band high-efficiency printed loop antenna design for wireless communication systems. IEEE Trans Veh Technol 54(3):873–878

Deo P, Mehta A, Mirshekar-Syahkal D, Massey PJ, Nakano H (2010) Thickness reduction and performance enhancement of steerable square loop antenna using hybrid high impedance surface. IEEE Trans Antennas Propag 58(5):1477–1485

Deschamps GA (1953) Microstrip microwave antennas. Paper presented at the third USAF symposium on antennas

Fujimoto K (1987) Small antennas. Wiley Online Library, USA

Ge L, Luk KM (2014) A band-reconfigurable antenna based on directed dipole. IEEE Trans Antennas Propag 62(1):64–71

Geyi W (2003) Physical limitations of antenna. IEEE Trans Antennas Propag 51(8):2116–2123

Hansen RC (1981) Fundamental limitations in antennas. Proc IEEE 69(2):170–182

IEEE Standard Definitions of Terms for Antennas (1993) IEEE Std 145-1993 1–32. doi:10.1109/IEEESTD.1993.119664

Lang K (1973) Babinet's principle for a perfectly conducting screen with aperture covered by resistive sheet. IEEE Trans Antennas Propag 21(5):738–740. doi:10.1109/TAP.1973.1140576

Lee DH, Chauraya A, Vardaxoglou Y, Park WS (2008) A compact and low-profile tunable loop antenna integrated with inductors. IEEE Antennas Wirel Propag Lett 7:621–624

Lee KF, Luk KM, Tong KF, Shum SM, Huynh T, Lee RQ (1997) Experimental and simulation studies of the coaxially fed U-slot rectangular patch antenna. IEE Proc Microwaves Antennas Propag 144(5):354–358

Li G, Zhai H, Li L, Liang C (2014a) A nesting-L slot antenna with enhanced circularly polarized bandwidth and radiation. IEEE Antennas Wirel Propag Lett 13:225–228

Li P, Liang J, Chen X (2006) Study of printed elliptical/circular slot antennas for ultrawideband applications. IEEE Trans Antennas Propag 54(6):1670–1675

Li WM, Liu B, Zhao HW (2014b) The U-shaped structure in dual-band circularly polarized slot antenna design. IEEE Antennas Wirel Propag Lett 13:447–450

Liu YF, Lau KL, Xue Q, Chan CH (2004) Experimental studies of printed wide-slot antenna for wide-band applications. IEEE Antennas Wirel Propag Lett 3(1):273–275

Lu WJ, Liu GM, Tong KF, Zhu HB (2014) Dual-band loop-dipole composite unidirectional antenna for broadband wireless communications. IEEE Trans Antennas Propag 62(5):2860–2866

Luk KM, Mak CL, Chow YL, Lee KF (1998) Broadband microstrip patch antenna. Electron Lett 34(15):1442–1443

Ma TG, Jeng SK (2005) A printed dipole antenna with tapered slot feed for ultrawide-band applications. IEEE Trans Antennas Propag 53(11):3833–3836

McLean JS (1996) A re-examination of the fundamental limits on the radiation Q of electrically small antennas. IEEE Trans Antennas Propag 44(5):672

Mehdipour A, Denidni TA, Sebak AR (2014a) Multi-band miniaturized antenna loaded by ZOR and CSRR metamaterial structures with monopolar radiation pattern. IEEE Trans Antennas Propag 62(2):555–562. doi:10.1109/TAP.2013.2290791

Mehdipour A, Denidni TA, Sebak AR, Trueman CW (2014b) Reconfigurable TX/RX antenna systems loaded by anisotropic conductive carbon-fiber composite materials. IEEE Trans Antennas Propag 62(2):1002–1006. doi:10.1109/TAP.2013.2293784

Mehdipour A, Denidni TA, Sebak AR, Trueman CW, Rosca ID, Hoa SV (2013) Mechanically reconfigurable antennas using an anisotropic carbon-fibre composite ground. IET Microwaves Antennas Propag 7(13):1055–1063. doi:10.1049/iet-map.2013.0115

Motorola DynaTAC. https://www.motorola.com/us/consumers/about-motorola-us/About_Motorola-History-Timeline/About_Motorola-History-Timeline.html#1980

Mukherjee S, Biswas A, Srivastava KV (2014) Broadband substrate integrated waveguide cavity-backed bow-tie slot antenna. IEEE Antennas Wirel Propag Lett 13:1152–1155

Nakano H, Fujimori K, Yamauchi J (2000) A low-profile conical beam loop antenna with an electromagnetically coupled feed system. IEEE Trans Antennas Propag 48(12):1864–1866

Niroo-Jazi M, Erfani E, Denidni TA (2013) On the antenna gain enhancement using artificial materials. Paper presented at the IEEE Antennas and Propagation Society, AP-S international symposium (digest)

Qudrat-E-Maula M, Shafai L, Pour ZA (2014) A corrugated printed dipole antenna with equal beamwidths. IEEE Trans Antennas Propag 62(3):1469–1474

Rajagopalan H, Kovitz JM, Rahmat-Samii Y (2014) MEMS reconfigurable optimized e-shaped patch antenna design for cognitive radio. IEEE Trans Antennas Propag 62(3):1056–1064. doi:10.1109/TAP.2013.2292531

Rowell CR, Murch RD (1997) A capacitively loaded PIFA for compact mobile telephone handsets. IEEE Trans Antennas Propag 45(5):837–842

Rowell CR, Murch RD (1998) A compact PIFA suitable for dual-frequency 900/1800-MHz operation. IEEE Trans Antennas Propag 46(4):596–598

Salonen P, Rahmat-Samii Y, Kivikoski M (2004) Wearable antennas in the vicinity of human body. Paper presented at the Antennas and Propagation Society international symposium, 20–25 June 2004. IEEE

Schaller RR (1997) Moore's law: past, present and future. IEEE Spectr 34(6):52–59

Shynu SV, Augustin G, Aanandan CK, Mohanan P, Vasudevan K (2005) A reconfigurable dual-frequency slot-loaded microstrip antenna controlled by pin diodes. Microw Opt Technol Lett 44(4):374–376. doi:10.1002/mop.20639

Shynu SV, Augustin G, Aanandan CK, Mohanan P, Vasudevan K (2006) C-shaped slot loaded reconfigurable microstrip antenna. Electron Lett 42(6):316–318. doi:10.1049/el:20060238

Sim CYD, Han TY, Liao YJ (2014) A frequency reconfigurable half annular ring slot antenna design. IEEE Trans Antennas Propag 62(6):3428–3431

Siya M, Yee Hui L (2013) Design of a compact patch antenna using UC-EBG structure. Paper presented at the antennas and propagation society international symposium (APSURSI), 7–13 July 2013 IEEE

Stutzman WL, Davis WA (1998) Antenna theory. Wiley Online Library, USA

Su CM, Chen HT, Wong KL (2002) Printed dual-band dipole antenna with U-slotted arms for 2.4/5.2 GHz WLAN operation. Electron Lett 38(22):1308–1309

Suh YH, Chang K (2000) Low cost microstrip-fed dual frequency printed dipole antenna for wireless communications. Electron Lett 36(14):1177–1179

DOI 10.1007/978-981-4560-75-7_55-1

Sze JY, Wong KL (2001) Bandwidth enhancement of a microstrip-line-fed printed wide-slot antenna. IEEE Trans Antennas Propag 49(7):1020–1024

Tefiku F, Grimes CA (2000) Design of broad-band and dual-band antennas comprised of series-fed printed-strip dipole pairs. IEEE Trans Antennas Propag 48(6):895–900

Tilley K, Wu XD, Chang K (1994) Coplanar waveguide fed coplanar strip dipole antenna. Electron Lett 30(3):176–177

Wang CJ, Chen LT (2014) Modeling of stepped-impedance slot antenna. IEEE Trans Antennas Propag 62(2):955–959

Wang D, Wen G, Rao Q, Pecen M (2008) A quad-band loop PIFA Antenna for wireless applications. Paper presented at the 2008 world automation congress, WAC 2008

Waterhouse R (1995) Small microstrip patch antenna. Electron Lett 31(8):604–605

Wheeler HA (1947) Fundamental limitations of small antennas. Proc IRE 35(12):1479–1484

Wong KL, Chen WY (2009) Small-size printed loop antenna for penta-band thin-profile mobile phone application. Microw Opt Technol Lett 51(6):1512–1517

Wong KL, Huang CH (2008) Printed loop antenna with a perpendicular feed for penta-band mobile phone application. IEEE Trans Antennas Propag 56(7):2138–2141

Yang F, Zhang XX, Ye X, Rahmat-Samii Y (2001) Wide-band E-shaped patch antennas for wireless communications. IEEE Trans Antennas Propag 49(7):1094–1100

Zarghooni B, Dadgarpour A, Denidni TA (2014) Reconfigurable metamaterial dipole antenna. Paper presented at the IEEE Antennas and Propagation Society, AP-S international symposium (digest)

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_56-1

On-Chip Antennas

Tianwei Deng[a]* and Yue Ping Zhang[b]
[a]Temasek Laboratories, National University of Singapore, Singapore, Singapore
[b]School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore, Singapore

Abstract

On-chip antennas feature the integration of antennas with other front-end circuits on the same chip in mainstream silicon technologies such as CMOS and SiGe. This chapter describes the basics, technology, and applications of on-chip antennas with particular emphasis on their circuit modeling, radiation efficiency, and integration as well as their crosstalk mechanisms with other circuit elements.

Keywords

On-chip antenna; CMOS technology; RF circuit; Radiation efficiency; Communication

Introduction

Silicon technology sustains the prosperous development of semiconductor and integrated circuits industries, providing the possibility of integrating an entire system on a single chip. A key issue, or an ultimate wish, for a radio-frequency (RF) system-on-chip, is the integration of antennas with other circuits. In 2002, Floyd et al. demonstrated on-chip dipole antennas with integrated transmitter and receiver in a standard 0.18-μm CMOS technology for intra-chip wireless interconnect at 15 GHz (Floyd et al. 2002). In 2003, Chan et al. characterized an on-chip monopole antenna fabricated on silicon wafer with an optimized proton implantation process over 100 GHz (Chan et al. 2003). In 2005, Zhang et al. reported the impedance and radiation characteristics of on-chip inverted-F and quasi-Yagi antennas at 60 GHz (Zhang et al. 2005). Meanwhile, the propagation mechanisms and communication performances between on-chip antennas have received attention. Kim et al. observed that the transmission between on-chip antennas improves with the frequency over an intra-chip channel (Kim and Kenneth 1998; Kim et al. 2001). Zhang et al. found that a frequency window for better transmission exists and that the guided propagation of surface waves plays an important role over the intra-chip channel (Zhang et al. 2007). Zhang also analyzed the bit-error-rate performance of a wireless interconnect between on-chip antennas over intra-chip channels (Zhang 2004). Lin et al. showed wireless communication between on-chip antennas over free space (Lin et al. 2004). The abovementioned and other studies revealed that the key problem of on-chip antennas is poor radiation efficiency (Cheema and Shamim 2013; Zhang and Liu 2009). Hence, new techniques were devised to enhance the radiation efficiency, such as proton implantation (Chan et al. 2003), micromachining (Chu et al. 2012), and the use of artificial magnetic conductors (Kang et al. 2010) and superstrates (Ou and Rebeiz 2012). Recently, much effort has been paid to on-chip antennas for applications in the bands of millimeter and terahertz frequencies (Uzunkol et al. 2013; Ojefors et al. 2009).

*Email: tsldt@nus.edu.sg

DOI 10.1007/978-981-4560-75-7_56-1

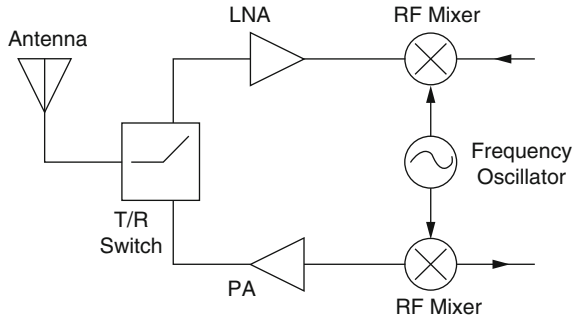


Fig. 1 Single-chip RF direct-conversion transceiver

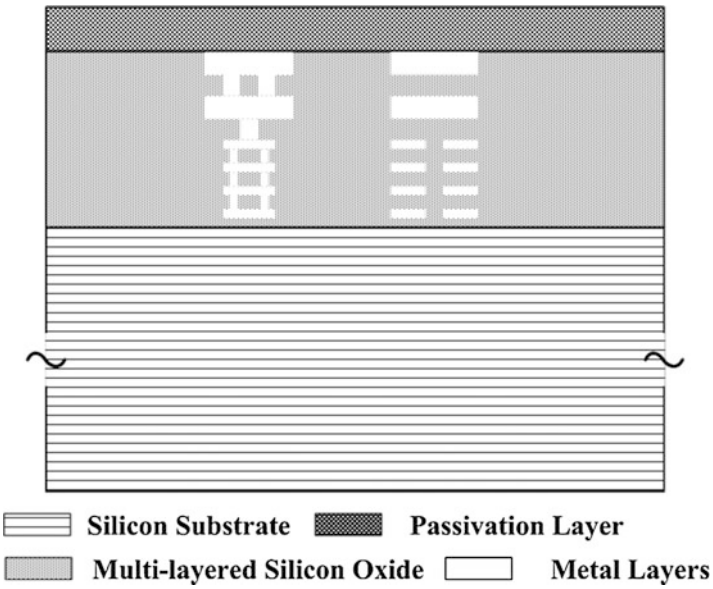


Fig. 2 The multiple-layered structure of standard CMOS chip

On-Chip Antenna Technology

The sophisticated integration of circuits/systems in CMOS technology brings new challenges to antenna designs. These challenges come from the complexity of integrated chip and co-design with circuits compared with single antenna structure optimization. Figure 1 shows a typical RF front-end of a direct-conversion transceiver suitable for a single-chip solution. The transceiver uses an antenna by incorporating a T/R switch to connect with the low-noise amplifier (LNA) or power amplifier (PA). In addition, it should be mentioned that a differential architecture that helps improve the signal-to-noise ratio (SNR) is extensively adopted in RF-integrated transceivers. In such cases, a differential antenna is required.

The standard CMOS process involves multiple metal and dielectric layers as shown in Fig. 2. The silicon substrate has thickness $t_s = 250 \sim 700$ μm, dielectric constant $\varepsilon_{rs} = 11.9$, and low resistivity $\rho_s = 10\Omega$.cm. Transistors are constructed at the top region of the silicon substrate by doping this area to be N- or P-well. Above the silicon substrate, there are layers of silicon oxide, or its variations, with layer thickness $t_i = 0.2 \sim 1.5$ μm and dielectric constant $\varepsilon_{ri} = 2.2 \sim 4.2$. The silicon oxide layers act as the insulators between metal layers. Currently, the number of metal layers is $4 \sim 8$, and the separation from

DOI 10.1007/978-981-4560-75-7_56-1

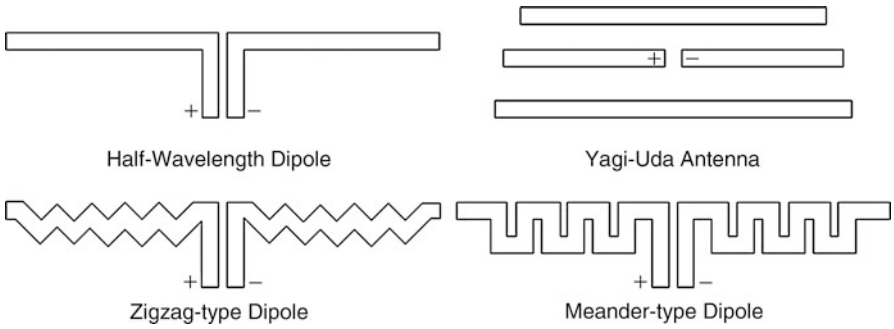


Fig. 3 Illustration of on-chip dipoles

the top to the bottom metal layers rarely exceeds 15 μm. Between metal layers, metal via can be used to connect them. These metal layers are used for interconnect or passive components, such as inductors and capacitors. The on-chip antennas are usually realized using the top metal (copper) layer with thickness $t_{\mathrm{d}} = 1 \sim 4$ μm and conductivity $\sigma_d = 5.8 \times 10^6$S/m. At the top of the chip, the passivation layer is silicon nitride with thickness $t_{\mathrm{p}} = 1 \sim 5$ μm and dielectric constant $\varepsilon_{rp} = 4.2 \sim 8.1$.

Design rule checking (DRC) is the most significant process before fabrication. Basically, there are four rules that impact the passive components design, especially for on-chip antennas. First, the metal layer design rules define the maximum width (w_{mi}), minimum separation space (w_{msi}), and minimum diagonal length of enclosed area (l_{mei}) of metal line/patch. Here, the subscript "*i*" denotes the layer number. The requirements vary for different layers. Generally, the maximum width (w_{mi}) of metal lines is less than 20 μm, which limits the antenna design greatly. Second, the metal via design rules describe the maximum width (w_{vi}), minimum separation space (w_{vsi}), and minimum enclosure width (w_{vei}) of the metal via array. Third, the metal density design rules provide restrict requirements on the metal density at each metal layer and in every checking window region. Rectangle metal fills are recommended by foundries to enlarge the physical metal density. The metal fills accordingly increase the dielectric loss of the silicon oxide layers and induce other design concerns for those cases in which the fills are placed near or under/above the passive components. Fourth, the off-chip connection rules require a connection window in the passivation layer and a connection pad on the aluminum cap above the top metal layer. Bonding wires and bump bolder, which are conductive and bulky, are used for off-chip connection. The impact of off-chip connection on on-chip antennas should be studied.

The dipole is a popular on-chip antenna structure. Generally, a half-wavelength dipole with a width of only several micrometers is used. Hence, it occupies a quite small area and is easy to be arranged with other circuit elements. To further reduce its size, a dipole can have a zigzag or meander shape, as shown in Fig. 3. A Yagi-Uda antenna is often used to increase the directivity of a normal dipole at the cost of larger antenna size.

Figure 4 shows the configuration of on-chip loop antennas. A loop antenna takes up a larger area compared with a dipole. It is usually placed around circuits for easy integration (Shamim et al. 2007; Huang and Wentzloff 2011; Ojefors et al. 2007). However, the passive components, such as the inductors of a low-noise amplifier, have a relatively large size. Thus, their layout inside the loop antenna needs to be carefully planned.

The microstrip antenna is another popular candidate for on-chip antennas. When this type of antenna is adopted for on-chip systems, two concerns should be considered. First, the typical microstrip patch is nearly a half-wavelength long, which is much larger than the requirement in the metal layer design rules. Designers have to slot the patch to meet the rules, as shown in Fig. 5 (Seok et al. 2008). Second, the ground plane should be kept as far as possible from the patch.

DOI 10.1007/978-981-4560-75-7_56-1

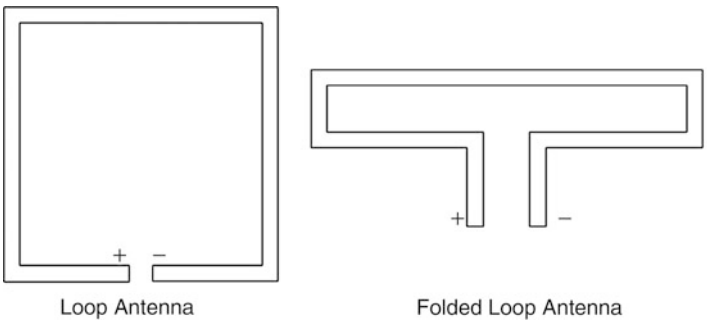


Fig. 4 Illustration of on-chip loop antennas

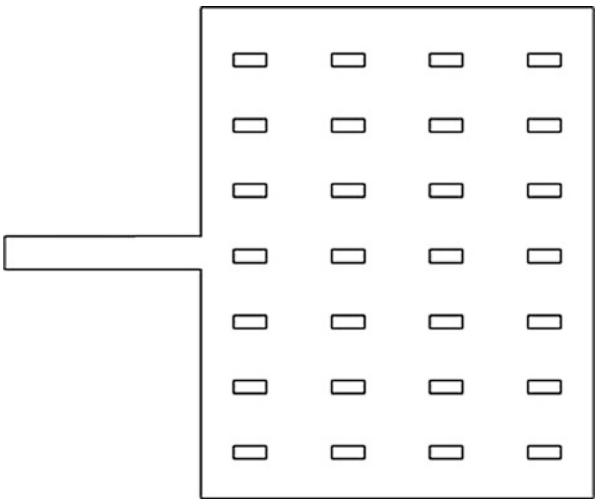


Fig. 5 Illustration of on-chip microstrip antenna

The slot antenna is also used in on-chip antenna designs (Jiang et al. 2006). Normally, a dielectric resonator or superstrate layer is placed above the slot antenna to increase the radiation efficiency (Edwards and Rebeiz 2012). Monopole and inverted-F antennas also find applications in on-chip antenna designs (Chan et al. 2003; Zhang et al. 2005). Table 1 lists key data of some reported on-chip antennas.

On-Chip Antenna Compact Circuit Model and Radiation Efficiency

The design of on-chip antennas relies heavily on numerical field simulators (e.g., HFSS or CST), which unfortunately require large computation power. Upon integration of antennas on a chip with circuits, co-design of antennas and circuits is necessary to evaluate the impact of impedance matching and mutual coupling on each other. However, the field simulators cannot be easily incorporated with circuit simulators (e.g., Cadence or Mentor Graphics). Designers need compact circuit models of on-chip antennas to obtain not only reduced computation power and time but also rapid circuit optimization. As previously mentioned, the dipole is widely used in the design of on-chip antennas. Figure 6 shows a compact circuit model of an on-chip dipole. The compact circuit model is divided into two parts: the shunt branch estimates the effect of the silicon substrate and the silicon oxide layer, and the series branch evaluates the radiation performance and the losses of the antenna. In the model, the series branch comprises R_r, R_c, R_{sur}, L_d, and C_d. The resistances R_r, R_c, and R_{sur} account for the radiation, conductor, and surface-wave losses

DOI 10.1007/978-981-4560-75-7_56-1

Table 1 Reported on-chip antennas

References	Architecture	Process	Frequency (GHz)	Gain (dBi)	Efficiency (%)	Aperture size (mm)
Guo and Chuang (2008)	Inverted-F	0.18-μm CMOS	57 ~ 64	−15.5	6	0.4 * 0.3
Hsu et al. (2008)	Quasi-Yagi	0.18-μm CMOS	55 ~ 65	−10.6	10	1.1 * 0.95
Kuo et al. (2008)	Triangular monopole	0.18-μm CMOS	50 ~ 65	−9.4	12	1 * 0.81
Pan and Capolino (2011)	Slot antenna with cavity	CMOS	140	−2	18	1.2 * 0.6
Edwards and Rebeiz (2012)	Elliptical slot + superstrate	0.13-μm CMOS	89	0.7	30	2.5 * 1.5; t = 0.4
Ou and Rebeiz (2012)	Superstrate	0.13-μm CMOS	86 ~ 100	3	63	1.2 * 1.24; t = 0.1
May et al. (2010)	Superstrates	0.13-μm SiGe	92 ~ 98	3	57	1.5 * 18; t = 0.125
Hou et al. (2012)	Meander slot + DRA	0.18-μm CMOS	123 ~ 137	4.7	43	0.8 * 0.9; t = 1.27
Bijumon et al. (2008)	DRA	0.5-μm SiGe	28	1	45	2 * 2.3; t = 0.93
Nezhad-Ahmadi et al. (2010)	H-slot + DRA	0.13-μm SiGe	33 ~ 37	1	48	1.1 * 1.6; t = 0.5
Kang et al. (2010)	Monopole + AMC	90-nm RF CMOS	57.5 ~ 66	−2.1	19.60	1.3 * 1.1

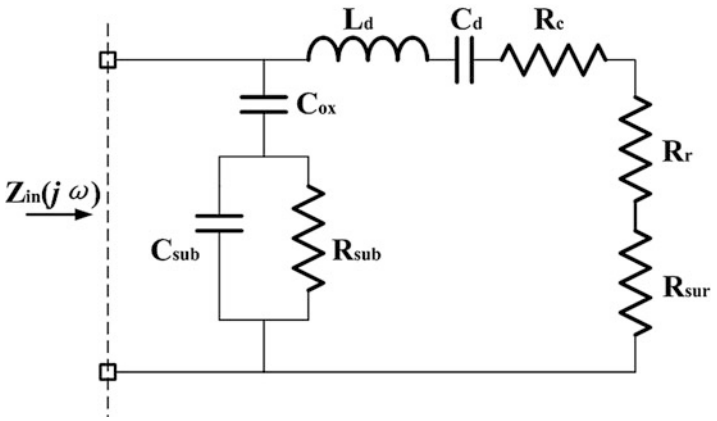


Fig. 6 Compact circuit model of on-chip dipole antenna with an off-chip ground shield

of the dipole, respectively. L_d and C_d are the inductance and capacitance of the dipole, respectively. The shunt branch consists of C_{ox}, C_{sub}, and R_{sub}. C_{ox} represents the oxide capacitance between the dipole and the silicon substrate. C_{sub} and R_{sub} are the silicon substrate capacitance and resistance, respectively. Regarding other on-chip antenna structures, the compact circuit models are similar. The differences that should be revised are that the shunted resonance tank (L_d and C_d) is used for the shunted resonance antenna and the oxide capacitance, C_{ox}, is shorted to ground without the silicon substrate capacitance and resistance C_{sub} and R_{sub}, respectively, when on-chip ground is applied.

DOI 10.1007/978-981-4560-75-7_56-1

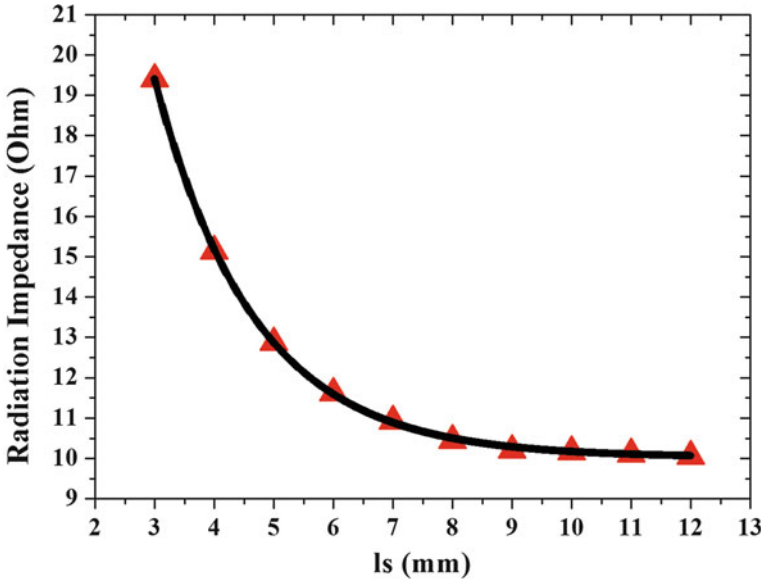


Fig. 7 The radiation resistance of on-chip dipole versus the side length of the chip (*ls*)

For an on-chip dipole with a multilayered substrate, the radiation resistance R_r is sensitive to the properties of the substrate layers. This can be expressed approximately similar to that of a microstrip dipole (Perlmutter et al. 1985):

$$R_r = \frac{120}{\pi}\varepsilon_{eff}\int_0^{\pi/2}\int_0^{2\pi}\frac{\cos^2\left(\pi\sin\theta\cos\phi/\left(2\sqrt{\varepsilon_{eff}}\right)\right)}{\left(\sin^2\theta\cos^2\phi-\varepsilon_{eff}\right)^2}\times\left[\frac{\cos^2\theta\sin^2\phi}{\left(\varepsilon_{eq}-\sin^2\theta\right)\cot^2\left(hk_0\sqrt{\varepsilon_{eq}-\sin^2\theta}\right)+\cos^2\theta}\right.$$
$$\left.+\frac{\cos^2\theta\cos^2\phi\left(\varepsilon_{eq}-\sin^2\theta\right)}{\left(\varepsilon_{eq}-\sin^2\theta\right)+\varepsilon_{eq}^2\cos^2\theta\cot^2\left(hk_0\sqrt{\varepsilon_{eq}-\sin^2\theta}\right)}\right]\times\sin c^2\left(w_dk_0\sin\theta\sin\phi/2\right)\sin\theta\cdot d\theta d\phi \tag{1}$$

where h, w_d, and k_0 denote the total thickness of the substrate, the width of the microstrip dipole, and the free space wave vector, respectively. ε_{eq} is the equivalent permittivity of the multilayered substrate (Yoon and Kim 2000). ε_{eff} is the effective permittivity of the multilayered substrate, which can be derived as the well-known equation:

$$\varepsilon_{eff} = \frac{\varepsilon_{eq}+1}{2}+\frac{\varepsilon_{eq}-1}{2\sqrt{1+10h/w_d}} \tag{2}$$

A particular concern is the chip size, which is usually less than 5 mm × 5 mm. Considering the possibility of integrating more circuits such as the baseband processor on the same chip, one can expect that the chip size may increase to 10 mm × 10 mm. However, the dimensions are not electrically large to ignore its influence on the radiation resistance (Gutierrez et al. 2009). For example, Fig. 7 shows that the radiation resistance R_r of an on-chip dipole will decrease from 19.5 to 10.5 Ω when the chip side length increases from 3 to 10 mm at 60 GHz; this is due to the surface-wave radiation at the edge of the chip (Deng et al. 2011).

DOI 10.1007/978-981-4560-75-7_56-1

The conductor loss of the microstrip antenna is usually negligible at low frequency but significant in the millimeter wave region. The thickness of the available metal layer on the chip is very thin, possible less than 5 μm. The eddy current incorporates an inevitable conductor loss R_c. The skin depth δ of the metal layer can be given as

$$\delta = \sqrt{\frac{R_{sh} \cdot t_d}{\pi \cdot \mu_0 \cdot f}} \tag{3}$$

where R_{sh} and t_d denote the sheet resistance and the thickness of the metal layer, respectively, and f is the frequency. The conductor resistance R_c of a rectangular metal patch with length l_d and equivalent radius w_{eff} can be expressed as

$$R_c = \frac{l_d}{w_{eff}} \sqrt{\frac{\mu_0 \cdot f \cdot R_{sh} \cdot t_d}{\pi \cdot \exp\left(-\frac{t_d}{\delta}\right)}} \tag{4}$$

For example, at 60 GHz, when the sheet resistance $R_{sh} = 25$ mohm/sq, the thickness $t_d = 2$ μm, and the ratio of $l_d/w_{eff} = 60$, the conductivity is $\sigma_d \approx 2.35 \times 10^6$, and the conductor resistance is $R_c \approx 15.3\Omega$.

The surface-wave resistance R_{sur} denotes the losses of surface waves, which varies with different silicon substrate thicknesses ($t_s = 300 \sim 700$ μm) and operation frequencies. The surface-wave power p_{sw} of the basic TM_0 mode is given by (Pozar 1990)

$$P_{sur} = 30\pi k_0^2 \frac{\varepsilon_{eq}\left(\chi_0^2 - 1\right)}{\varepsilon_{eq}\left[\frac{1}{\sqrt{\chi_0^2 - 1}} + \frac{\sqrt{\chi_0^2 - 1}}{\varepsilon_{eq} - \chi_0^2}\right] + \left[1 + \frac{\varepsilon_{eq}^2\left(\chi_0^2 - 1\right)}{\varepsilon_{eq} - \chi_0^2}\right] k_0(t_s + t_i)} \tag{5}$$

where $\beta_0 = x_0 \cdot k_0$ is the TM_0 surface-wave propagation constant and k_0 is the wave number of air. The result of p_{sw} is exact if x_0 is exact. For a high-dielectric-constant substrate, x_0 can be given as

$$x_0 = 1 + \frac{-\varepsilon_{eq}^2 + \alpha_0\alpha_1 + \varepsilon_{eq}\sqrt{\varepsilon_{eq}^2 - 2\alpha_0\alpha_1 + \alpha_0^2}}{\varepsilon_{eq}^2 - \alpha_1^2} \tag{6}$$

where

$$\alpha_0 = \sqrt{\varepsilon_{eq} - 1}\tan\left(k_0 t_s\right)$$
$$\alpha_1 = \frac{-1}{\sqrt{\varepsilon_{eq} - 1}}\left[\tan\left(k_0 t_s\right) + \frac{k_0 t_s}{\cos^2(k_0 t_s)}\right]$$

With the surface-wave power p_{sw}, the surface-wave resistance R_{sur} can be expressed as

$$R_{sur} = \frac{Z_0^2}{2P_{sur}} \tag{7}$$

DOI 10.1007/978-981-4560-75-7_56-1

The silicon substrate resistance R_{sub} denotes the losses in the low-resistive silicon substrate, which are further enlarged because $C_{ox} << C_{sub}$. Floating metal-shielding structures were designed and placed between the metal layer and silicon substrate (Kim et al. 2001), which increases the value of C_{ox} and blocks EM waves propagated into the silicon substrate. However, this results in the disadvantage of induced losses on the metal-shielding structures.

The main concern regarding on-chip antennas is their poor radiation efficiency. This is due to the combined effect of the low resistivity and high permittivity of the silicon substrate (Zhang and Liu 2009). The low resistivity causes loss due to heating in the presence of an electric field in the substrate, and the high permittivity causes loss due to power trapped in surface-wave modes in the substrate. The radiation efficiency of on-chip antennas is the key factor in evaluating their radiation performance, which could be expressed as

$$\eta = \frac{R_r}{\mathrm{Re}[Z_{in}(j\omega_d)]} \tag{8}$$

where $Re[Z_{in}(j\omega_d)]$ is the real part of the input impedance of the on-chip antenna at the resonant angular frequency ω_d. The radiation resistance is actually quite low, and the other resistances are large, leading to low radiation efficiency. There are several effective methods of improving the radiation efficiency of on-chip antennas.

Modifying the Silicon Substrate

Silicon substrate has low resistivity, which causes large substrate loss and is a key reason for the low radiation efficiency of on-chip antennas. A proton implantation process has been developed to increase the resistivity from 10 to 10^6 Ω cm. To avoid the latch-up issue, the implantation is done only in certain selected areas underneath the antenna (Chan et al. 2003). A micromachining technique has also been developed to remove silicon substrate and create an air cavity under the on-chip antennas. It has been shown that the radiation efficiency of a single on-chip monopole can be increased to 93.9 % (Chu et al. 2012). A thick silicon substrate supports more surface-wave modes and results in high surface-wave loss. Substrate thinning is beneficial to radiation efficiency.

Implementing Artificial Magnetic Conductor

Because the losses are mostly from the silicon substrate, designers can place an on-chip ground at the bottom metal layer to block the propagation of electric field into the silicon substrate and to reduce the surface-wave power. However, the closely placed on-chip ground induces an image current to diminish the radiation of the antenna. Usually, the distance between on-chip ground and antenna is less than 10 μm. At THz frequencies, the on-chip ground is effective and widely used in on-chip antenna design (Seok et al. 2008). At millimeter frequencies, artificial magnetic conductor (AMC) can be implemented to avoid the impact of the image current (Kang et al. 2010; Kuo et al. 2013; Jiang et al. 2012; Bao et al. 2012; Takahagi and Sano 2011). This provides an in-phase reflection of electromagnetic waves as a perfect magnetic conductor (PMC). The radiation is enhanced rather than canceled. An efficiency improvement from 5 % to 10 % was achieved by implementing an AMC structure under a Yagi-Uda antenna.

Adding Superstrate and Dielectric Resonator

Adding a superstrate above an antenna can improve the radiation efficiency (Alexopoulos and Jackson 1984). For instance, a slot-coupled antenna achieves 30 % radiation efficiency with a quartz superstrate,

DOI 10.1007/978-981-4560-75-7_56-1

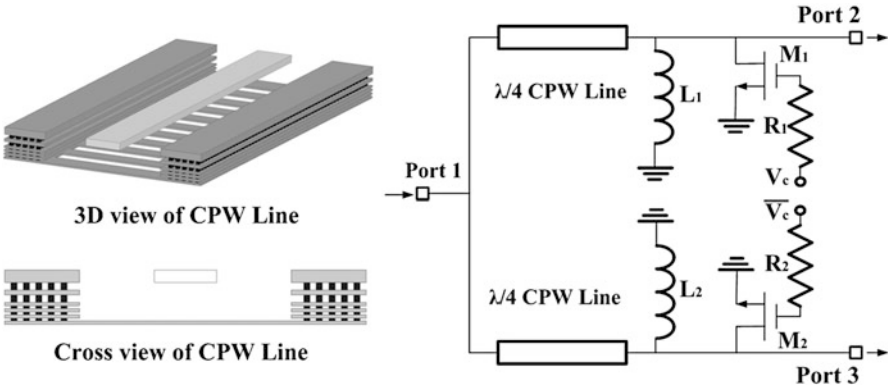


Fig. 8 Schematic of the SPDT switch with the CPW line (Uzunkol and Rebeiz 2010)

compared with 6 % radiation efficiency without the superstrate (Edwards and Rebeiz 2012). The use of on-chip dielectric resonator antenna (DRA) is another approach. The radiation efficiency of DRAs can be very high because the silicon substrate is isolated by an on-chip ground, and there is no image current issue to cancel out the antenna radiation. However, one concern is that adding the DRA on a chip with high precision is difficult.

Using Substrate Lens

A hemispherical silicon lens with a matching layer can convert the surface-wave power to a useful radiated power. Therefore, such lens is used to increase the radiation efficiency and directivity of on-chip antennas (Babakhani et al. 2006). Usually, the bottom metal layers are used for the antennas, and the silicon substrate is thinned to minimize the path length through which the radiated wave travels inside the low-resistivity silicon substrate.

On-Chip Antennas and Circuits

This section introduces the integration of an antenna with circuits on the same chip and discusses the crosstalk mechanisms and effects between the antenna and passive circuit components.

Integration of Antennas with Circuits

For a single transmitter or a receiver, an on-chip antenna is directly integrated with a PA or LNA (Uzunkol et al. 2013; Babakhani et al. 2006). For a transceiver, the antenna is usually integrated with a single-pole double-throw (SPDT) switch. At higher frequencies, the SPDT switch topology shown in Fig. 8 is often adopted. This has three ports, with port 2 to the PA, port 3 to the LNA, and port 1 to the antenna. When Vc is high, transistor M_1 is on, and transistor M_2 is off, the signal picked up by the antenna is fed to the LNA. When Vc is low, transistor M_1 is off, and transistor M_2 is on, the signal from the PA is fed to the antenna. Two quarter-wavelength transmission line sections are used for impedance matching and two shorted transmission line stubs acting as two inductors are used to tune out the capacitances of the transistors (Uzunkol and Rebeiz 2010).

Figure 9 shows the 3D view of a slot-folded dipole antenna with a CPW feed line used in the integration (Chen et al. 2012). The folded-slot structure is favored because it provides a wide bandwidth, a CPW-friendly interface, low metallic loss, and high isolation.

DOI 10.1007/978-981-4560-75-7_56-1

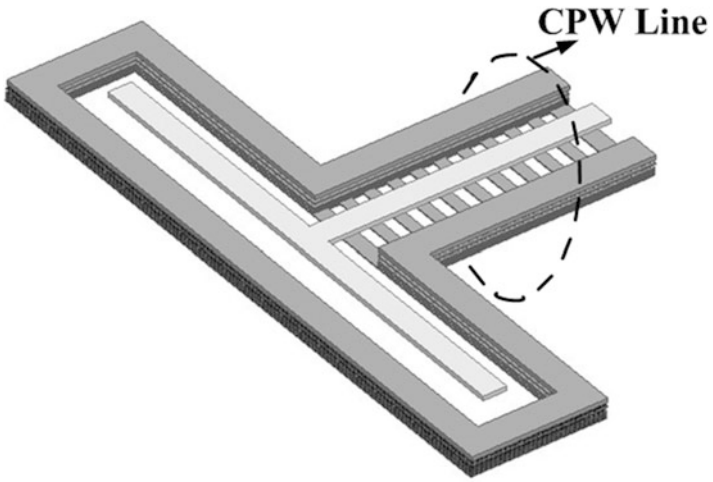


Fig. 9 3D view of slot-folded dipole antenna (Chen et al. 2012)

Crosstalk Between On-Chip Antennas and Circuits

A major concern with on-chip antennas is the risk of crosstalk with the on-chip circuits such as an LNA and a voltage-controlled oscillator (VCO) (Deng et al. 2013). The noise currents are injected from the on-chip antennas into the substrate and coupled to the active region of the on-chip circuits through the substrate medium (Briaire and Krisch 2000). Figure 10 shows the crosstalk from an on-chip antenna to active region of the on-chip circuits. The crosstalk between on-chip antenna and MOS transistor may become significant at high frequencies due to the drain and source junction capacitances and at low frequencies due to the body effect. Higher-resistivity doping on silicon substrate can obtain lower substrate noise coupling (Soens et al. 2006). Besides, isolation via and guard rings show good noise-isolation properties (Chu et al. 2012; Kang et al. 2010; Uemura et al. 2012). Correspondingly, thermal and switching noise currents generated by transistors flow from the substrate to the on-chip antennas. These are common mode in nature and could be adequately rejected by applying differential antennas instead of single-ended ones.

The passive elements are more susceptible to crosstalk from the on-chip antennas than are the active devices. This is because the sizes of passive elements, such as inductors, capacitors, and transmission lines, are much bigger than those of active devices, such as transistors in the LNA and varactors in the VCO circuits. Normally, the crosstalk between on-chip antenna and passive elements involves different mechanisms, depending on the frequency of operation and the distance of separation. It also considers that on-chip antenna and passive elements are placed in the near-field region. Figure 11 shows the equivalent circuit of an on-chip monopole with an inductor. C_{ant}, L_{ant}, and R_{ant} are the capacitance, inductance, and resistance of the monopole, respectively. L_{ind} and R_{ind} are the inductance and resistance of the inductor, respectively. C_{oxi_ant} (or C_{oxi_ind}) is the oxide capacitance between the monopole (or the inductor) and silicon substrate. R_{sil_ant} and C_{sil_ant} (or R_{sil_ind} and C_{sil_ind}) represent the silicon substrate resistance and capacitance, respectively, of the monopole (or the inductor). C_m and R_m denote, respectively, the capacitance and resistance between the monopole and the inductor. The magnetic coupling coefficient K_m estimates the magnetic coupling between the monopole and inductor.

The discussion of the crosstalk mechanism between the monopole and inductor is conveniently classified as electric coupling, magnetic coupling, and resistive coupling. The electric coupling through the mutual capacitance (C_m) is dominant at low frequency, which shows a high-pass frequency response; the magnetic coupling from the magnetic coupling coefficient K_m varies with the different shapes and

DOI 10.1007/978-981-4560-75-7_56-1

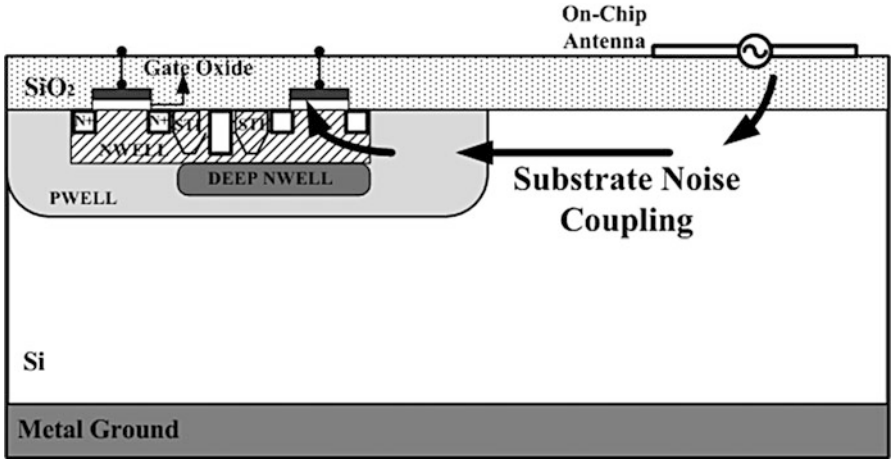


Fig. 10 The substrate crosstalk is due to electrical fluctuations propagated in substrate medium

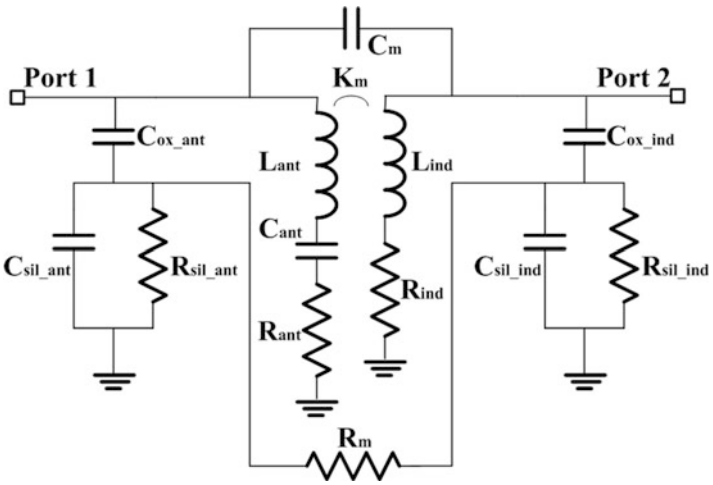


Fig. 11 The equivalent circuit of crosstalk between on-chip antenna and inductor

arrangements of inductors; and the resistive coupling through the substrate due to R_m is inevitable, the level of which is invariable versus the frequency. However, the resistive coupling is weak compared with the EM coupling. There is a resonance due to the combined effect of the magnetic (inductive due to the monopole and inductor, i.e., L_{ant}, L_{ind}, and K_m) and electric (capacitive C_m) coupling. The response frequency is given from the equivalent circuit:

$$f_{dip} = \frac{1}{2\pi\sqrt{L_{eq}C_{eq}}} \approx \frac{1}{2\pi\sqrt{\left(\frac{L_1L_2}{L_3} + L_1 + L_2\right)C_m}} \tag{9}$$

where

$L_1 = L_{ant} - L_m, L_2 = L_{ind} - L_m, L_m = K_m\sqrt{L_{ind}L_{ant}}$

DOI 10.1007/978-981-4560-75-7_56-1

$$L_3 = L_m - \frac{R_{ind}^2}{\omega^2 C_{ant}\left[(R_{ant} + R_{ind})^2 + 1/\omega^2 C_{ant}^2\right]} \approx L_m$$

Equation 9 predicts a low mutual coupling null/window. The response frequency of this null/window shifts with the different shielding and guide structures placed around or under the inductor. In addition, the oxide and substrate capacitances (C_{oxi_ant}, C_{sil_ant}, C_{oxi_ind}, and C_{sil_ind}) might decrease the coupling level with increasing frequencies, because the signal is shorted to substrate or ground from the oxide and substrate capacitances, rather than coupled to another port. Regarding the crosstalk between antennas and transmission lines, the coupling mechanism is similar at these frequencies. Note that coplanar waveguide (CPW) lines have ground lines adjacent to the signal line. The coupling drops because the oxide and substrate capacitances do not occur in this case.

Crosstalk effects are important to optimize the layout of circuits with antennas. The passive elements, including inductors, capacitors, and transmission lines, act as parasitic patches in on-chip antennas for the case they are placed near each other. Normally, the electric field is terminated at the off-chip ground through the silicon substrate. When passive elements are placed, they tend to interact with the monopole to induce observation of asymmetric electric field distribution and terminate the electric field instead of the off-chip ground.

The far-field patterns of the monopole shown in Fig. 12 verify this. The nearby inductor or CPW line shifts the maximum radiation direction of the monopole to $10^\circ \sim 20^\circ$ in the H plane. The passive elements can be used as parasitic reflectors similarly to that in Yagi-Uda antennas. However, they increase the cross-polarization at least 20 dB in the E plane, which decreases the radiation performance of the monopole. The on-chip monopole has limited influence on the passive elements, especially when floating shielding structures are added to protect these elements.

Examples of On-Chip Antenna Application

Wireless interconnect has been proposed to communicate among cores within a chip for ultra large-scale integration (ULSI) or system-on-chip (SoC). On-chip antennas are used to construct the intra-chip communication systems, and the signal coupling between on-chip antennas is based on the propagation of radio waves rather than R-C coupling. Figure 13 shows a wireless interconnect for delivering a global clock signal on a chip at 15 GHz (Floyd et al. 2002). As seen in the figure, one clock transmitter and one clock receiver are placed on different locations on the chip. On-chip dipole antennas are used in both the transmitter and receiver, and a 1.875-GHz local clock signal from the clock transmitter can be detected by a clock receiver over a 5.6-mm distance.

Conventional wireless communication systems that use on-chip antennas have also been shown. For example, the wireless communication range between a single-chip radio with the on-chip antenna and a 20-dBi gain horn antenna can be as far as 95 m at 24 GHz (Cao et al. 2008). The wireless communication range between another single-chip radio on a high-resistivity substrate and an 11-dBi gain patch antenna can achieve 100-kb/s data transmission at 5.8 GHz over a separation distance of 310 m (Lin et al. 2007).

On-chip antennas are also used in highly integrated radars and imagers. In a range-sensing radar (Khanh et al. 2013); a 180-nm CMOS transceiver with on-chip antennas is implemented with a 9–11-GHz damping-pulse transmitter and a receiver including a mixer and a three-stage LNA. The transmitter and receiver antennas are the same and placed in an orthogonal direction to minimize crosstalk. Besides, a

DOI 10.1007/978-981-4560-75-7_56-1

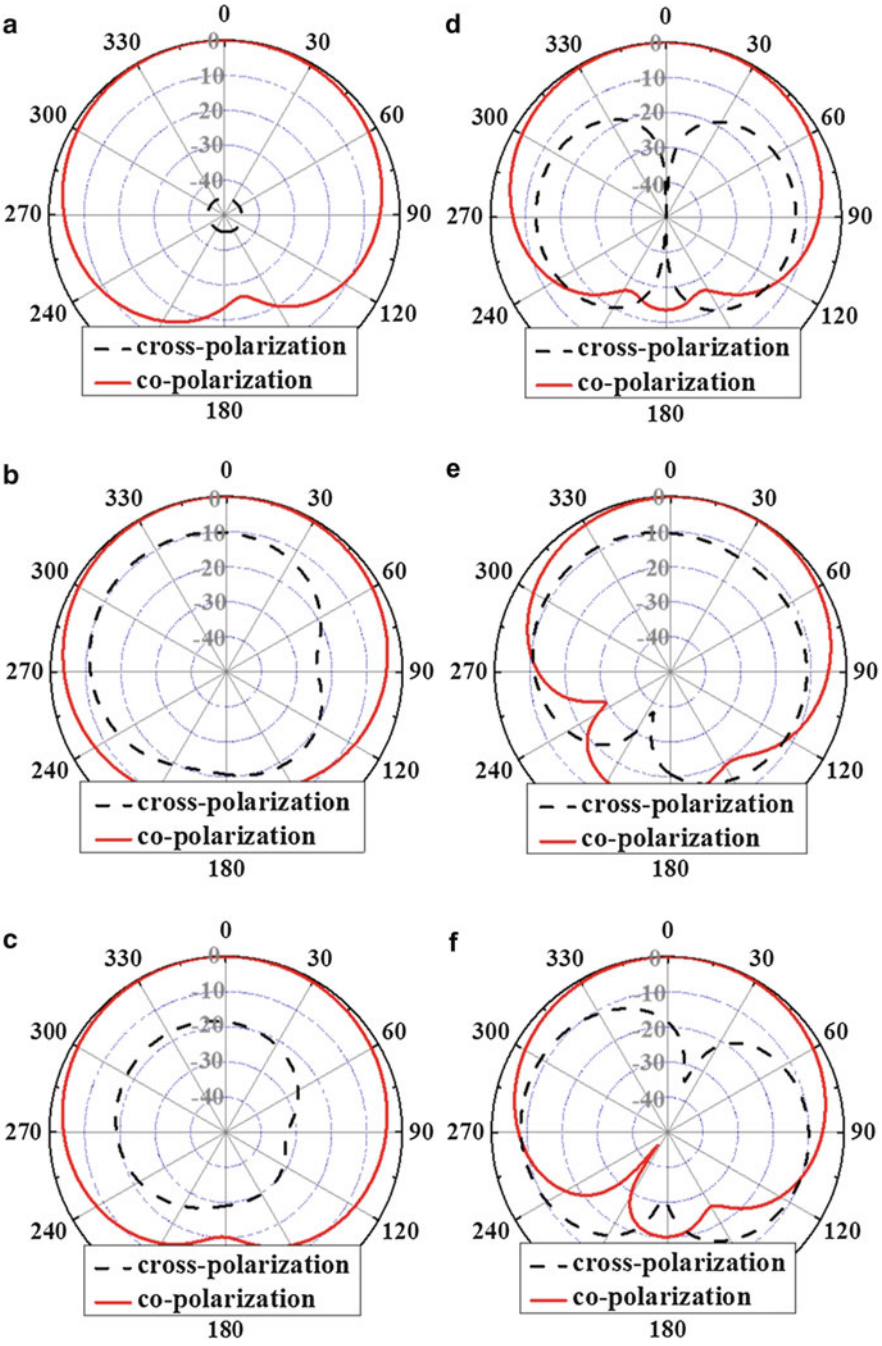


Fig. 12 The far-field radiation patterns of the monopoles at resonant frequency: (**a**) as a stand-alone monopole in the E plane and (**b**) in the H plane, (**c**) as a monopole with inductor separated 100 μm at 40 GHz in the E plane and (**d**) in the H plane, (**e**) as a monopole with CPW line separated 100 μm at 40 GHz in the E plane, and (**f**) in the H plane

focal-plane array for room-temperature detection of 0.65-THz radiation fully integrated into a low-cost 0.25-μm CMOS process technology. The circuit architecture is based on the principle of distributed resistive self-mixing and facilitates broadband direct detection well beyond the cutoff frequency of the technology. The 3 × 5-pixel array consists differential on-chip patch antennas, NMOS direct detectors, and integrated 43-dB voltage amplifiers. At 0.65 THz the FPA achieves a 80-kV/W responsibility and a noise equivalent power of 300 pW/sqrt(Hz) (Ojefors et al. 2009).

DOI 10.1007/978-981-4560-75-7_56-1

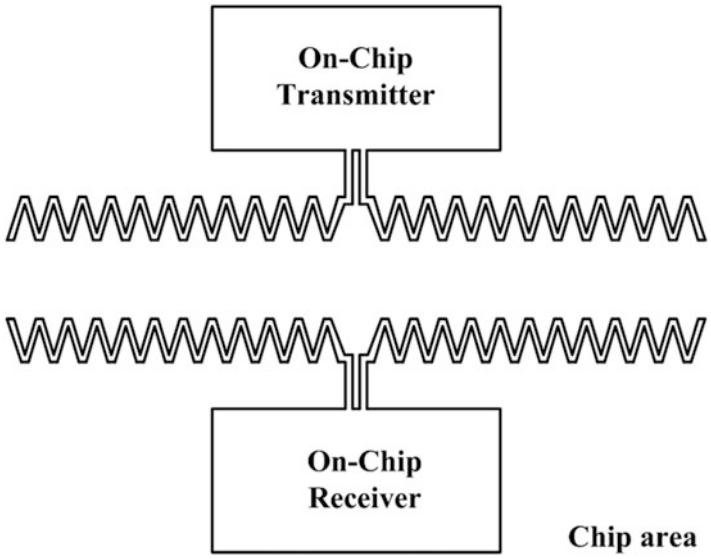


Fig. 13 Intra-chip wireless interconnect using on-chip dipole antennas for clock signal distribution

Summary

Driven by the development of the RF system-on-chip, there have been some attempts to implement on-chip antennas in mainstream silicon technologies. In this chapter, the basics, technology, and applications of on-chip antennas are described, with emphasis on their circuit modeling, radiation efficiency, and integration, as well as their crosstalk mechanisms with other circuit elements. Future studies should focus on the challenges of fast simulation and optimization of on-chip antennas and co-design of on-chip antennas with other circuits. Also, how to test on-chip antennas at THz frequencies, particularly, their radiation efficiencies should be made? Furthermore, an effective solution to improve radiation efficiency is always the main concern in on-chip antenna design.

References

Alexopoulos N, Jackson DR (1984) Fundamental superstrate (cover) effects on printed circuit antennas. IEEE Trans Antennas Propag 32:807–816

Babakhani A, Guan X, Komijani A, Natarajan A, Hajimiri A (2006) A 77-GHz phased-array transceiver with on-chip antennas in silicon: receiver and antennas. IEEE J Solid-State Circuits 41:2795–2806

Bao XY, Guo YX, Xiong YZ (2012) 60-GHz AMC-based circularly polarized cn-chip antenna using standard 0.18-μm CMOS technology. IEEE Trans Antennas Propag 60:2234–2241

Bijumon PV, Antar YMM, Freundorfer AP, Sayer M (2008) Dielectric resonator antenna on silicon substrate for system on-chip applications. IEEE Trans Antennas Propag 56:3404–3410

Briaire J, Krisch KS (2000) Principles of substrate crosstalk generation in CMOS circuits. IEEE Trans Comput Aided Des Integr Circuits Syst 19:645–653

Cao C, Ding Y, Yang X, Lin JJ, Wu HT, Verma AK, Lin J, Martin F, Kenneth KO (2008) A 24-GHz transmitter with on-chip dipole antenna in 0.13-μm CMOS. IEEE J Solid-State Circuits 43:1394–1402

DOI 10.1007/978-981-4560-75-7_56-1

Chan KT, Chin A, Lin YD, Chang CY, Zhu CX, Li MF, Kwong DL, McAlister S, Duh DS, Lin WJ (2003) Integrated antennas on Si with over 100 GHz performance, fabricated using an optimized proton implantation process. IEEE Microwave Wirel Compon Lett 13:487–489

Cheema HM, Shamim A (2013) The last barrier: on-chip antennas. IEEE Microw Mag 14:79–91

Chen Z, Wang CC, Yao HC, Heydari P (2012) A BiCMOS W-band 2 × 2 focal-plane array with on-chip antenna. IEEE J Solid-State Circuits 47:2355–2371

Chu H, Guo YX, Lim TG, Khoo YM, Shi X (2012) 135-GHz micromachined on-chip antenna and antenna array. IEEE Trans Antennas Propag 60:4582–4588

Deng T, Zhang Y, Guo Y (2011) Modeling of radiation resistance of 60 GHz on-chip dipole. In: Proceeding of the Asia-Pacific microwave conference, Melbourne, Australia

Deng T, Chen Z, Zhang YP (2013) Coupling mechanisms and effects between on-chip antenna and inductor or coplanar waveguide. IEEE Trans Electron Devices 60:20–27

Edwards JM, Rebeiz GM (2012) High-efficiency elliptical slot antennas with quartz superstrates for silicon RFICs. IEEE Trans Antennas Propag 60:5010–5020

Floyd BA, Hung CM, Kenneth KO (2002) Intra-chip wireless interconnect for clock distribution implemented with integrated antennas, receivers, and transmitters. IEEE J Solid-State Circuits 37:543–552

Guo PJ, Chuang HR (2008) A 60-GHz millimeter-wave CMOS RFIC-on-chip meander-line planar inverted-F antenna for WPAN applications. In: IEEE antennas and propagation society international symposium, San Diego, USA

Gutierrez F, Agarwal S, Parrish K, Rappaport TS (2009) On-chip integrated antenna structures in CMOS for 60 GHz WPAN systems. IEEE J Sel Areas Commun 27:1367–1378

Hou D, Xiong YZ, Goh WL, Hu S, Hong W, Madihian M (2012) 130-GHz on-chip meander slot antennas with stacked dielectric resonators in standard CMOS technology. IEEE Trans Antennas Propag 60:4102–4109

Hsu SS, Wei KC, Hsu CY, Huey RC (2008) A 60-GHz millimeter-wave CPW-fed Yagi antenna fabricated by using 0.18-μm CMOS technology. IEEE Electron Device Lett 29:625–627

Huang KK, Wentzloff DD (2011) A 60 GHz antenna-referenced frequency-locked loop in 0.13 μm CMOS for wireless sensor networks. IEEE J Solid-State Circuits 46:2956–2965

Jiang L, Mao JF, Yin WY (2006) High transmission gain slot antennas on silicon substrate for wireless interconnect. In: IEEE proceedings of Asia-Pacific microwave conference, Yokohama, Japan

Jiang I, Mao JF, Leung KW (2012) A CMOS UWB on-chip antenna with a MIM capacitor loading AMC. IEEE Trans Electron Devices 59:1757–1764

Kang K, Lin F, Pham DD, Brinkhoff J, Heng CH, Guo YX, Yuan X (2010) A 60-GHz OOK receiver with an on-chip antenna in 90 nm CMOS. IEEE J Solid-State Circuits 45:1720–1731

Khanh M, Ngoc N, Asada K (2013) A 0.18-μm CMOS fully integrated antenna pulse transceiver with leakage-cancellation technique for wide-band microwave range sensing radar. In: IEEE radio frequency integrated circuits symposium (RFIC), Seattle, USA, pp 97–100

Kim K, Kenneth KO (1998) Characteristics of integrated dipole antennas on bulk, SOI, and SOS substrates for wireless communication. In: Proceedings of the IEEE 1998 international interconnect technology conference, San Francisco, USA, pp 21–23

Kim K, Bomstad W, Kenneth KO (2001) A plane wave model approach to understanding propagation in an intra-chip communication system. In: 2001 I.E. antennas and propagation society international symposium, vol 2, Boston, USA, pp 166–169

Kuo PC, Hsu SS, Lin CC, Hsu CY, Chuang HR (2008) A 60-GHz millimeter-wave triangular monopole antenna fabricated using 0.18-μm CMOS technology. In: 3rd International Conference on Innovative Computing Information and Control (ICICIC '08), Dalian, China

Kuo HC, Yue HL, Ou YW, Lin CC, Chuang HR (2013) A 60-GHz CMOS sub-harmonic RF receiver with integrated on-chip artificial-magnetic-conductor Yagi antenna and balun bandpass filter for very-short-range gigabit communications. IEEE Trans Microwave Theory Tech 61:1681–1691

Lin JJ, Gao L, Sugavanam A, Guo X, Li R, Brewer JE, Kenneth KO (2004) Integrated antennas on silicon substrates for communication over free space. IEEE Electron Device Lett 25:196–198

Lin JJ, Wu HT, Su Y, Gao L, Sugavanam A, Brewer JE, Kenneth KO (2007) Communication using antennas fabricated in silicon integrated circuits. IEEE J Solid-State Circuits 42:1678–1687

May JW, Alhalabi RA, Rebeiz GM (2010) A 3 G-Bits W-band SiGe ASK receiver with a high-efficiency on-chip electromagnetically-coupled antenna. In: IEEE radio frequency integrated circuits symposium (RFIC), Anaheim, USA

Nezhad-Ahmadi MR, Fakharzadeh M, Biglarbegian B, Safavi-Naeini S (2010) High-efficiency on-chip dielectric resonator antenna for mm-wave transceivers. IEEE Trans Antennas Propag 58:3388–3392

Ojefors E, Sonmez E, Chartier S, Lindberg P, Schick C, Rydberg A, Schumacher H (2007) Monolithic integration of a folded dipole antenna with a 24-GHz receiver in SiGe HBT technology. IEEE Trans Microwave Theory Tech 55:1467–1475

Ojefors E, Pfeiffer UR, Lisauskas A, Roskos HG (2009) A 0.65 THz focal-plane array in a quarter-micron CMOS process technology. IEEE J Solid-State Circuits 44:1968–1976

Ou YC, Rebeiz GM (2012) Differential microstrip and slot-ring antennas for millimeter-wave silicon systems. IEEE Trans Antennas Propag 60:2611–2619

Pan S, Capolino F (2011) Design of a CMOS on-chip slot antenna with extremely flat cavity at 140 GHz. IEEE Antennas Wirel Propag Lett 10:827–830

Perlmutter P, Shtrikman S, Treves D (1985) Electric surface current model for the analysis of microstrip antennas with application to rectangular elements. IEEE Trans Antennas Propag 33:301–311

Pozar DM (1990) Rigorous closed-form expressions for the surface wave loss of printed antennas. Electron Lett 26:954–956

Seok E, Cao C, Shim D, Arenas DJ, Tanner DB, Hung CM, Kenneth KO (2008) A 410 GHz CMOS push-push oscillator with an on-chip patch antenna. In: IEEE international Solid-State circuits conference (ISSCC), San Francisco, USA

Shamim A, Arsalan M, Roy L (2007) 5 GHz monolithic CMOS transmitter and antenna for short-range communications. In: The second European conference on antennas and propagation (EuCAP), Edinburgh, UK

Soens C, Van Der Plas G, Badaroglu M, Wambacq P, Donnay S, Rolain Y, Kuijk M (2006) Modeling of substrate noise generation, isolation, and impact for an LC-VCO and a digital modem on a lightly-doped substrate. IEEE J Solid-State Circuits 41:2040–2051

Takahagi K, Sano E (2011) High-gain silicon on-chip antenna with artificial dielectric layer. IEEE Trans Antennas Propag 59:3624–3629

Uemura S, Hiraoka Y, Kai T, Dosho S (2012) Isolation techniques against substrate noise coupling utilizing through silicon via (TSV) process for RF/mixed-signal SoCs. IEEE J Solid-State Circuits 47:810–816

Uzunkol M, Rebeiz GM (2010) A low-loss 50–70 GHz SPDT switch in 90 nm CMOS. IEEE J Solid-State Circuits 45:2003–2007

Uzunkol M, Gurbuz OD, Golcuk F, Rebeiz GM (2013) A 0.32 THz SiGe 4 $\times$ 4 imaging array using high-efficiency on-chip antennas. IEEE J Solid-State Circuits 48:2056–2065

Yoon YJ, Kim B (2000) A new formula for effective dielectric constant in multi-dielectric layer microstrip structure. In: IEEE conference on electrical performance of electronic packaging, Scottsdale, USA, pp 163–167

DOI 10.1007/978-981-4560-75-7_56-1

Zhang YP (2004) Bit-error-rate performance of intra-chip wireless interconnect systems. IEEE Commun Lett 8:39–41

Zhang YP, Liu D (2009) Antenna-on-chip and antenna-in-package solutions to highly integrated millimeter-wave devices for wireless communications. IEEE Trans Antennas Propag 57:2830–2841

Zhang YP, Sun M, Guo LH (2005) On-chip antennas for 60-GHz radios in silicon technology. IEEE Trans Electron Devices 52:1664–1668

Zhang YP, Chen ZM, Sun M (2007) Propagation mechanisms of radio waves over intra-chip channels with integrated antennas: frequency-domain measurements and time-domain analysis. IEEE Trans Antennas Propag 55:2900–2906

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_57-1

Substrate Integrated Waveguide Antennas

Tarek Djerafi*, Ali Doghri and Ke Wu
Poly-Grames Research Center, École Polytechnique de Montréal, Montreal, QC, Canada

Abstract

Research and development progress in the area of millimeter-wave antennas has been associated since recent years with substrate integrate waveguide (SIW). SIW was proposed and studied as a class of efficient integrated transmission lines compatible with planar technologies, offering incomparable self-consistent shielding and high-quality factor performances. SIW can be designed as open wave-guiding structures and energy leakage will take place when the uniformity of those guides is perturbed or they are not excited in an appropriate mode. This leakage effects may be used positively for the design of antennas by deliberately introducing perturbations in these guides so that they radiate in a controlled fashion. With the advantages of broadband, efficiency, and high gain, a specific benefit of such antennas is their compatibility with SIW from which they are derived, thus facilitating an integrated design. In this chapter, H-plane horns, leaky-wave structures (long and periodic) and tapered slot antennas are presented. Special attention is given for tapered slot antennas and arrays with their application in the design of passive imaging systems.

Keywords

Antenna array; Antipodal structure; Imaging system; Horn antenna; Leaky-wave; Millimeter-wave; Polarization; Slot antenna; Rod antenna; Substrate integrated waveguides (SIW); Tapered slot antenna; Twist; Directional coupler; Vivaldi antenna

Introduction

The frequency range from 30 to 300 GHz is generally defined as millimeter waves or extremely high-frequency waves. Wavelengths over this frequency band are ranged from 10 to 1 mm, thus naming it the millimeter-wave band or simply millimeter-wave (MMW or mmW). Compared to lower frequency bands, electromagnetic waves in this band are subject to high atmospheric attenuations and they are absorbed by atmospheric gases. Beyond this millimeter-wave range, electromagnetic radiation is referred to as terahertz (THz) radiation. Millimeter-wave bands offer exciting opportunities for various bandwidth-demanding and spectrum-sensitive applications such as short-range communications in the 60 GHz band and automotive radar and imaging systems in the frequency windows of 77 and 94 GHz (Meinel 1995; Solbach and Schneider 1999; Schoebel and Herrero 2010). Since a number of years, emerging applications are being reported such as imaging sensors for security and biomedical applications (Salmon 2004). There is also a great deal of interest in millimeter imaging arrays for other applications such as remote sensing, radioastronomy, and plasma measurement (Nanzer 2012).

In commercial system developments, low-cost, mass-producible, high-performance and high-yield microwave, and millimeter-wave technologies are critical for developing successful and sustainable

*Email: tarek.djerafi@polymtl.ca

DOI 10.1007/978-981-4560-75-7_57-1

commercial broadband systems, in particular, front end and antenna modules. Specifications of antenna components and systems for those applications should satisfy not only desired electrical performances as gain, low side lobe level (SLL), and low cross-polarization, but also stringent mechanical criteria of size, weight, and shielding. The integrability with other active and passive system elements is also considered as a main key criterion, which generally defines the ultimate cost and consequently the commercial success. The millimeter-wave bands are useful for enabling the design of densely packed systems integrated with high-gain antenna arrays.

Antennas in standard technology or in SIW can be classified as either resonant or nonresonant types. Resonant antennas, such as microstrip patches and dipoles, resonate at discrete frequencies. Nonresonant antennas or traveling-wave antennas do not resonate but rather leak energy out along their wave propagating path, thereby producing more efficient broadband structures. The resonant structures do not provide an easy-to-accept performance in the millimeter-wave range. This is because of very high conductor losses arising from immense current densities (singularities) at the strip edges, especially in the feeding network. The substrate becomes electrically thicker at higher frequency and more surface wave modes (parasitic modes) may exist. This can be deleterious for standard microstrip antennas with radiation efficiency reduction and radiation pattern perturbations, which result from the diffraction of surface waves at the edges of the antenna structure in an uncontrolled manner (Pozar 1983).

In this chapter, different constraints and limits of SIW antennas are examined with focus on no-resonance structures. At millimeter-wave, the energy is much easier to radiate out from open transmission lines. As such, a number of leaky-wave or surface-wave antennas can be synthesized and developed. They include tapered slot antenna (TSA), printed Yagi-Uda antenna or log-periodic dipole array, dielectric rod antenna, long slot leaky-wave, partially reflective surface-based antenna, etc. For more background information and development history, readers are referred to (Walter 1965; Oliner et al. 2007; Suntinjo et al. 2008; Caloz et al. 2011). Tapered slot antenna (TSA), also known as flared notch or Vivaldi antenna, is among the most promising candidates satisfying all the requirements described above (radiation, mechanical, integrabilty, and cost). TSA has some unique characteristics of symmetrical patterns in two planes, high gain in addition to having wide bandwidth characteristics in terms of radiation performance and impedance characteristics. TSA has proved to be of particular interest due to its potential as effective radiating elements in an array, and the possibility of using it at very high frequencies.

The integration of structure is a crucial issue for front-end and antenna system developments. The conservative waveguide solutions, which have been widely deployed for numerous applications, present a number of challenging problems in connection with size, cost, weight, and fabrication complexity of feed networks and antenna systems. Various discontinuities and lossy structures within the feed networks should be avoided or reduced or used with caution since noise emitted or interference created by such components into the receiver could affect the overall performance. Substrate integrated waveguide (SIW) with the advantages of self-consistent field shielding, low cost, and light weight is an excellent candidate to substitute the traditional solutions for antenna array design as well for whole system development (Djerafi and Wu 2013).

This chapter presents an overview and a summary of recent developments for leaky-wave antennas (LWAs). An LWA makes use of a guiding structure that supports wave propagation along the length of the structure with the wave radiating or “leaking” continuously along the structure. Such antennas may be uniform, quasi uniform, or periodic in topology. After reviewing the basic physics and operating principles behind, a summary of some recent advances for those types of structure is given and discussed. Horn antenna based on SIW presents some constraint in connection with thickness. Recent advances on SIW horn antenna will be presented.

The performance of TSAs with three different profiles will be compared over the Ka-band and corrugation effects will be detailed. Mutual coupling effects are critical on the performance of antenna

DOI 10.1007/978-981-4560-75-7_57-1

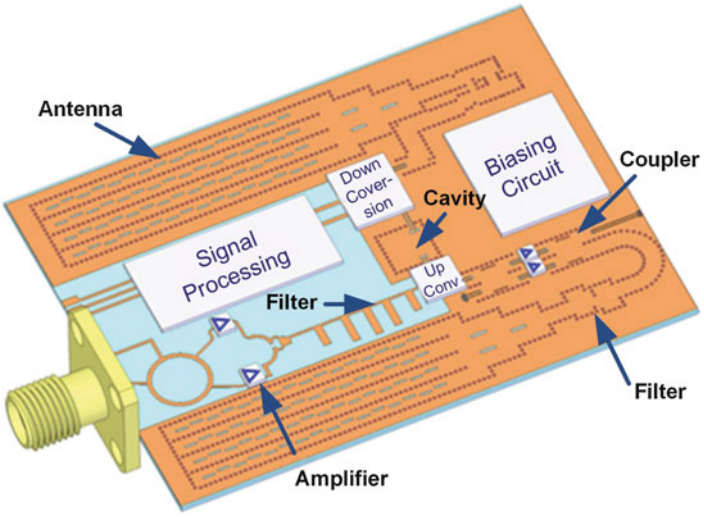


Fig. 1 A potential SIW-based platform in which SIW antennas, couplers, cavities, filters are all combined with active and DSP circuits on the same substrate or master board

arrays. The coupling in H and E-plane arrangement of two TSAs is studied as well as 90° configuration. This configuration is useful to build up antenna arrays with dual or circular polarization capabilities.

Substrate Integrated Waveguide

A new generation of high-frequency integrated circuits called "substrate integrated circuits – SICs" was proposed and demonstrated many years ago (Wu 2001). This disruptive concept has unified the hybrid and monolithic integrations of various planar and nonplanar circuits that are made in single substrate and/or multilayer platforms. Basically, any nonplanar structures can be made or synthesized in planar form, which presents the foundation of SICs. Being synthesized on a planar substrate, the substrate integrated image guide (SIIG), for example, can be combined in a hybrid way with the SIW and the Substrate Integrated Nonradiative Dielectric guide (SINRD) on the same substrate, operating with different modes and/or different frequency ranges. These guides can also be combined with standard waveguide, microstrip, Coplanar waveguide (CPW) or slotline, leading to attractive hybrid schemes of planar and nonplanar structures. In fact, an easy-to-handle low-cost hybrid design strategy is of critical importance for the development of high-volume millimeter-wave ICs and systems as illustrated in Fig. 1. These technologies have well been studied in the millimeter wave range. Since the transmission line technology is instrumental for developing high-frequency electromagnetic hardware, the choice of an appropriate waveguide or line structure is critical for millimeter-wave developments and applications. The transmission lines of choice should allow high-density integration and mass-production scheme at low cost. Rectangular waveguides have widely been used in the development of microwave- and millimeter-wave components and systems with their salient features such as low insertion loss, high quality factor (Q-factor) and high power capability, etc. However, they are also characterized by their bulky size, stringent manufacturing precision and nonplanar geometry. Therefore, it is impossible to design and develop microwave and millimeter-wave integrated circuits with this technological platform. Benefiting from the properties of low profile, easy fabrication, and low cost, microstrip-like circuits including coplanar waveguides (CPW) and strip lines are presently the principal choices of integration for the development of microwave and millimeter wave circuits. Unfortunately, such printed circuits suffer

DOI 10.1007/978-981-4560-75-7_57-1

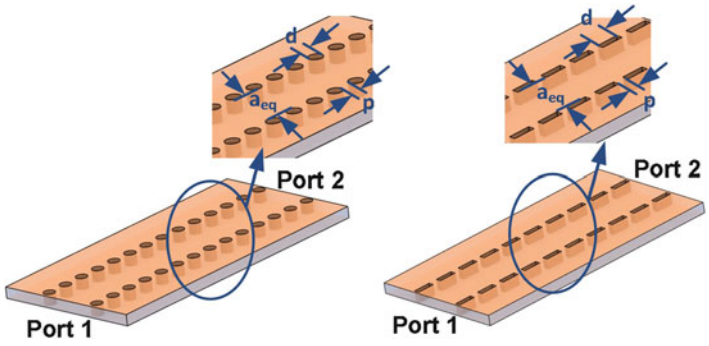


Fig. 2 The topology of a typical single-layered SIW. Metalized via and slot arrays for creating equivalent metallic fence or wall configurations with: (**a**) cylindrical via arrays, (**b**) long slot trenches

from significant losses and packaging problems. In fact, the performances of microstrip-like circuits are fundamentally limited by physical properties such as field or current singularities occurring at the stripline (Zhu and Wu 2002).

SIW structure preserves most of the advantages associated with conventional metallic waveguides, namely high Q-factor (low loss) and high power-handling capability with self-consistent electromagnetic shielding. The most significant advantage of SIW technology is its capacity of enabling a possible complete integration of all the components on the same substrate, including passive components, active elements, and even antennas as presented in Wu et al. (2003, 2012), Deslandes and Wu (2006), Bozzi et al. (2011). SIW techniques can be used to solve a series of headache problems, which renders it hugely popular in the community today (Vye 2011). A remarkable problem arising at high frequency is the appearance (trapping) of surface waves that generally decrease the antenna efficiency. The SIW can effectively control this phenomenon. Since SIW components are bounded by conducting surfaces on both sides of the substrate, they exhibit the merits of extremely low (completely negligible) radiation/leakage loss and insensitive to outer interferences. The SIW technology has already experienced a rapid development over more than one decade. This development allows the demonstration and applications of innovative passive and active circuits, antennas, and systems at microwave and millimeter-wave frequencies covering a very broad frequency range from subgegahertz to subterahertz. In addition, the SIW technique can be combined with other SICs platforms to create multiformat and multifunction devices and systems.

SIW Techniques and Design Basics

The SIW (also known as postwall waveguide or a laminated waveguide (Uchimura et al. 1998)) is a rectangular waveguide-like structure in an integrated planar form, which can be synthesized and fabricated by using two rows of conducting cylinders, vias or slots embedded in a dielectric substrate that is electrically sandwiched by two parallel metal plates as illustrated in Fig. 2.

The operating frequency range is delimited by the monomode propagation of quasi-TE_{10} wave as its cut-off frequency is only related to equivalent width a_{eq} of the synthesized waveguide as long as the substrate thickness or waveguide height is smaller than this width. This equivalent width will be discussed in the following section.

Via Configurations

Round metalized via holes are used to create the electric sidewalls or fences of SIW through two parallel via arrays. The discontinued current flow along the via- or slot-synthesized metalized sidewalls does not allow the propagation of TM modes. In addition, a large width-to-height ratio of SIW supports the

DOI 10.1007/978-981-4560-75-7_57-1

propagation of TE_{m0} modes. Since these techniques are amenable to arbitrarily shaped perforations, the limitation of circular via shapes is no longer mandatory. Rectangular slot trenches were found to be advantageous for lower leakage and better definition of the SIW sidewalls. This is important for some designs such as those iris and window coupling geometries found in the filter design. Rounded corners increase the overall mechanical stability, allowing for a better metallization, which often cannot be avoided in the fabrication process due to the finite diameter of laser beams in the case of using a laser-based micromachining process. Figure 2 shows two different slot trench (SIW) configurations: (a) cylindrical via arrays and (b) longer slots for SIW operation between the first and second stop bands. Note that the SIW structure is a periodic geometry, which is subject to the guided-wave phenomena of all periodic waveguides such as bandgap (stopband) effects. Such stopbands, caused by the distributed Bragg reflection, occur around frequencies at which the periodic spacing p is equal to a multiple of half a guided wavelength (Patrovsky et al. 2007).

SIW Propagation Modes

A three-dimensional view of the SIW is illustrated in Fig. 2. The PCB's ground planes form the broad walls of the SIW and a linear array of metalized via holes is used for the narrow walls. Since only vertically oriented currents can flow on the via-hole walls, this structure is capable of containing mode patterns with only vertical components of the electric field vector. Mode patterns with horizontal components of the electric field vector have horizontal components of the current distribution on the via hole walls, which will radiate. The SIW is thus ideally suited for containing the dominant mode pattern in a rectangular waveguide. The thickness h of waveguide does not have an effect on the propagation constant. This gives the designer certain degree of freedom in the choice of the SIW substrate.

The cut-off frequency of modes TE_{m0} can be derived using:

$$f_{c_{m0}} = \frac{m}{2a_{eq}\sqrt{\mu\varepsilon}} \tag{1}$$

and the corresponded guided wavelength is:

$$\lambda_g = \frac{\lambda_0}{\sqrt{\varepsilon_r\left(1 - \left(\frac{f_c}{f}\right)^2\right)}} \tag{2}$$

This expression gives a good approximation of the field propagation in the SIW. This equation shows that there are two important parameters: ε that represents the dielectric constant inside the waveguide and a the width of the metallic waveguide that is not defined here in the case of the SIW.

The modeling and design of SIW interconnects and circuits can be made on full-wave analysis tools, either general purpose commercial packages or specialized numerical codes. Among the most common techniques adopted for the design of SIW structures, there are the Finite-Difference Time Domain (FDTD) method or the Finite-Difference Frequency Domain (FDFD) (Xu et al. 2003), the Boundary Integral-Resonant Mode Expansion (BI-RME) method (Cassivi et al. 2002), the Method of Lines (MoL) (Yan et al. 2005), and the Transverse Resonance Method (Deslandes and Wu 2006).

Using those different numerical methods, the propagating parameters of the dominant mode, like wavenumber k or complex propagation constant γ can be derived.

The SIW propagation constant of the mode TE_{10} can be obtained using a simulation software like HFSS or CST or using measurement. To characterize the propagation and attenuation constants of the SIW line,

DOI 10.1007/978-981-4560-75-7_57-1

simulations are conducted using the HFSS software. The permittivity and height of the substrate used are respectively 2.94 and 0.762 mm in the calculations. The width of the line is 3.75 mm, $p = 0.4$ mm, and $d = 0.4$ mm according to Fig. 2.

By considering two SIW lines with different length l_2 and l_1, and after generating simulated S-parameters, phase constant β and attenuation constant α are calculated using:

$$\beta = \frac{\left|\angle\left(S_{21,l_2}\right) - \angle\left(S_{21,l_1}\right)\right|}{\Delta l} \tag{3}$$

and

$$\alpha = -\frac{ln\left|\left(S_{21,l_1}\right) / \left(S_{21,l_2}\right)\right|}{\Delta l} \tag{4}$$

and

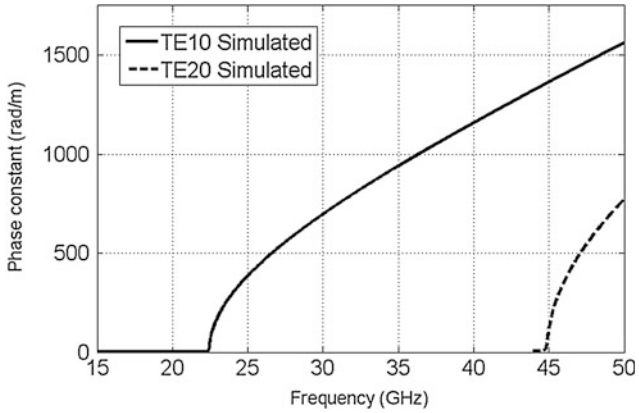


Fig. 3 Phase constants of TE_{01} and TE_{20} modes simulated using a full wave simulator

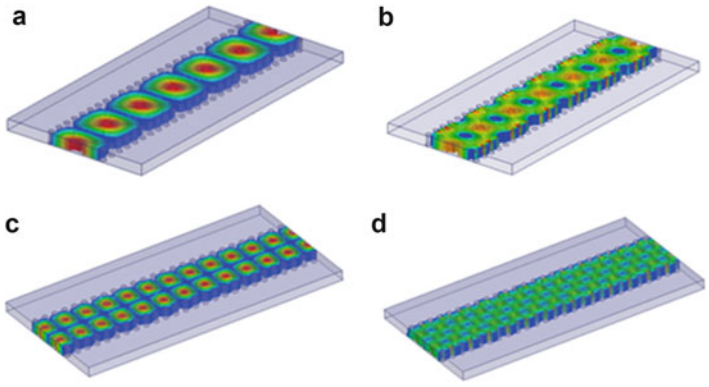


Fig. 4 (**a**) E-field distribution of the first mode, (**b**) H-field distribution of the first mode, (**c**) E-field distribution of the second mode, and (**d**) H-field distribution of the second mode

DOI 10.1007/978-981-4560-75-7_57-1

$$\gamma = \alpha + i\beta \tag{5}$$

Figure 3 shows the phase constant for the first and second modes; the second mode's cut-off frequency is twice that of the first mode just like the standard rectangular waveguide. The corresponding E and H field distributions are illustrated in Fig. 4.

Equivalent Width

The SIW can be modeled by a conventional rectangular waveguide (RW) through the so-called equivalent width a_{eq}. This parameter is calculated such that the resulting dielectric-filled rectangular waveguide has the same cut-off frequency of the fundamental TE_{10} mode as its corresponding SIW structure. This determines the propagation characteristics of the TE_{10} mode. Physical parameters of via-holes d and p are set to minimize the radiation (or leakage) loss as well as the return loss (Deslandes and Wu 2003). The equivalent rectangular waveguide width can be approximated according to the geometrical parameters illustrated in Fig. 2a as follows:

$$a_{eq} = a_{SIW} - \frac{d^2}{0.95p} \tag{6}$$

The cut-off frequency of the first mode is defined by:

$$f_c(\mathrm{TE}_{10}) = \frac{\mathrm{c}}{2 \cdot \sqrt{\varepsilon_r}} \cdot \left(a_{SIW} - \frac{d^2}{0.95p} \right)^{-1} \tag{7}$$

And for the second mode by:

$$f_c(\mathrm{TE}_{20}) = \frac{\mathrm{c}}{\sqrt{\varepsilon_r}} \cdot \left(a_{SIW} - \frac{d^2}{1.1p} - \frac{d^3}{6.6p^2} \right)^{-1} \tag{8}$$

Loss Considerations

The energy in transmission may be lost or dissipated through different physical mechanisms including dielectric losses, conductor losses, and radiation losses. Since the inner part of SIW is filled with a dielectric material, an adequate choice of dielectric material and conductor quality can reduce the contribution of the first two loss mechanisms (Bozzi et al. 2014).

- Attenuation because of the dielectric losses is related to tanδ of the dielectric substrate:

$$\alpha_d = \frac{k^2 \tan\delta}{2\beta} \tag{9}$$

- Attenuation because of the conductive loss:

$$\alpha_c = \frac{R_s}{a^3 b \beta k \eta} \left(2b\pi^2 + a^3 k^2 \right) \tag{10}$$

with:

DOI 10.1007/978-981-4560-75-7_57-1

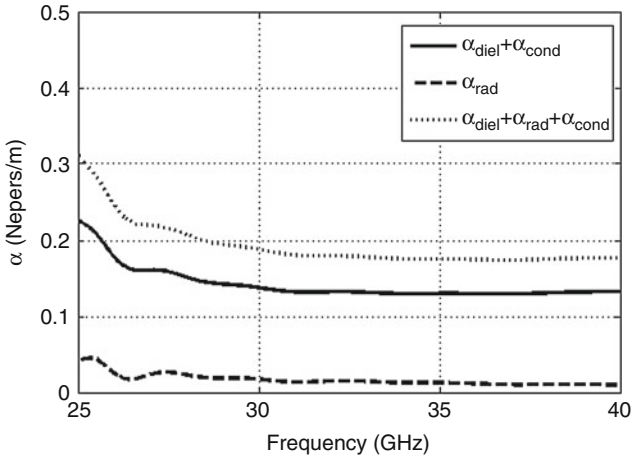


Fig. 5 Attenuation constant versus frequency for an SIW structure

$$R_s = \sqrt{\frac{\omega\mu_0}{2\sigma}} \text{ (surface resistance)} \tag{11}$$

The radiation leakage is caused by the gaps between the metalized via and slots. Radiation or leakage leads to two consequences, namely additional signal losses and undesired interferences.

$$\alpha_R = \frac{\frac{1}{a_{SIW}}\left(\frac{d}{a_{SIW}}\right)^{2.84}\left(\frac{p}{d}-1\right)^{6.28}}{4.85\sqrt{\left(\frac{2a_{SIW}}{\lambda_g}\right)^2-1}}\left[\frac{dB}{m}\right] \tag{12}$$

In order to ensure that the synthesized waveguide section becomes radiationless or free from leakage loss, parametric effects of p and d were studied on those issues (Deslandes and Wu 2003). To simplify the analysis, dielectric and conductor losses are not considered, the loss solely comes from radiation. It is found that the following requirements can be put forward to minimize the return and leakage losses, that is, the diameter of hole should satisfy some geometric constraints:

$$d < \frac{\lambda_g}{5} \tag{13}$$

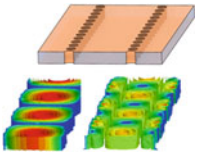
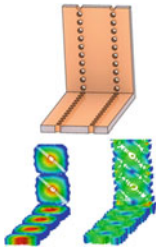
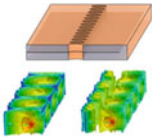
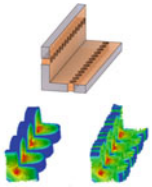
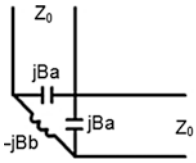
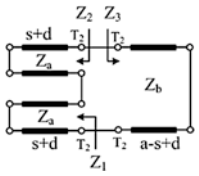
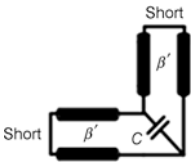
$$p \leq 2d \tag{14}$$

At millimeter and submillimeter frequencies, planar circuits usually suffer from radiation originating at bends and discontinuities. 90° SIW and microstrip bend losses were analyzed (Djerafi and Wu 2012a) in the 77 GHz band. The major part of losses in microstrip is related to the radiation compared to an insignificant radiation loss in SIW bend.

Using Eq. 5 and the same lines used to estimate the phase propagation constant below, Fig. 5 shows the attenuation versus frequency with only the contribution of the radiation loss and with the metallic and dielectric loss. The dielectric loss represents the major part of the losses. The attenuation constant decreases when frequency increases, especially the radiation loss.

DOI 10.1007/978-981-4560-75-7_57-1

Table 1 SIW line configurations allowing for a spatial arrangement

Type	H-plane SIW line	Two plane SIW line	C-type SIW line	L-type SIW line
Structure				
Circuit model				

Principal SIW Parameters

Substrate Materials

Low-loss material is the foundation for developing high-performance integrated circuits and systems. This becomes more critical for power budget as frequency increases to the millimeter-wave ranges and beyond. This is because it is relatively difficult to amplify over those ranges. Thermal effect, dielectric nonuniformity and metallic surface roughness may have to be taken into account for better and accurate design. This is especially important for antenna developments. The SIW can theoretically be constructed with any available substrate. The most used ones are Rogers RT/duroid®5880 glass microfiber reinforced PTFE composite and RT/duroid®6002 for conventional PCB processing, which are easily sheared with laser and machined to the required shape. The holes can easily be drilled mechanically into these machinable materials compared to ceramics which can only be processed by the laser perforation and other special techniques. All these materials have an excellent dimensional stability. Of course, a good thermal stability of the material of choice should also be considered in the design. These selections will not just affect the performances but also defines power handling capabilities (Cheng et al. 2008).

SIW Line Configurations

The SIW structures can be arranged in different configurations. Depending on applications in term of allowed space, integrability, task and performance, one of them can specially be preferred. Table 1 summarizes some attention-grabbing configurations. The H-to E-plane interconnection shown in (b) is very useful in the design of multipolarization antenna feeding. To realize the corner, a vertical line is inserted into a horizontal line. The vertical and horizontal layers making up the corner were fabricated using a conventional printed circuit board (PCB) process, before manually introducing the vertical layer in the horizontal one (Djerafi et al. 2012a).

Comparing to other types of planar transmission lines, a large form factor of SIW is a bottleneck for its integration with other conventional integrated circuits. Miniaturization techniques for SIW therefore become very crucial in some designs (Ding and Wu 2008; Djerafi et al. 2012b). The proposed structures present transversely folded waveguides in different shape, namely, C-type, T-type, and L-type. In 2005, a 2-layered T-type structure was proposed for the miniaturization of an SIW for the first time with SIW in

DOI 10.1007/978-981-4560-75-7_57-1

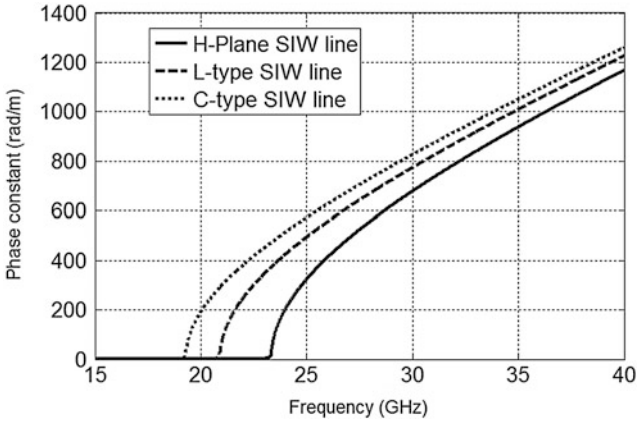


Fig. 6 Dispersion characteristics of different SIW lines

multilayered substrates (Grigoropoulos and Young 2004). Subsequently, C-type folding approaches as described in column (b) were also implemented with SIW in multilayered substrates (Grigoropoulos et al. 2005). L-FSIW was proposed (Doghri et al. 2013) to build a series of couplers. The cross section of the L-FSIW is equivalent to 2-D transmission line with a shunt capacitance that models the effect of the corner region.

For the C-type SIW line, the propagation constant is obtained by solving:

$$\tan(s+d)k_x + \tan(a-s-d)k_x = 0 \tag{15}$$

and for L-type SIW line by:

$$\beta_{10}^{f} = \sqrt{(2\pi f\sqrt{\mu\varepsilon})^2 - \beta'^2} \tag{16}$$

where

$$\beta'_{10} = \frac{2}{a'}cot^{-1}\left[8a'tf^2\mu\varepsilon\ln\left(\frac{2*\left(1-\frac{\sqrt{2}t}{a'}\right)}{\sqrt{2}C_f}\right)\right] \tag{17}$$

In Fig. 6, dispersion characteristics of the dominant mode in the two types of folded waveguide are compared with that of the conventional version of the original unfolded SIW (Ding and Wu).

All three lines have the same substrate filling. It is clearly observed that the cut-off frequencies of the L-type SIW line and C-type SIW line are lower than the classic SIW line. The shift in frequency is due to the effect added by the corners.

Characteristic Impedance

The impedance is an important characteristic especially when the SIW is connected to other planar and nonplaner structures. One definition of impedance in a rectangular waveguide is the wave impedance. This impedance, however, does not take into account the geometry of the transmission line. Another impedance definition more often employed in the design of matching networks is the characteristic

DOI 10.1007/978-981-4560-75-7_57-1

impedance. The characteristic impedance in a transmission line supporting a TEM wave can be defined in one of the three following ways (D'Orazio 2004):

$$Z_{PV} = \frac{VV^*}{2P_t} \tag{18}$$

$$Z_{PI} = \frac{2P_t}{II^*} \tag{19}$$

$$Z_{VI} = \frac{V}{I} \tag{20}$$

In a rectangular waveguide, however, the choice of voltage and current is not unique and these definitions do not produce the same results. The choice of which definition is more suitable depends on the targeted applications in question. The three definitions of characteristic impedance in an SIW are given by:

$$Z_{PV} = \frac{2h}{a_{eq}} Z_{TE} \tag{21}$$

$$Z_{PI} = \frac{\pi^2 h}{8a_{eq}} Z_{TE} \tag{22}$$

$$Z_{VI} = \frac{\pi h}{2a_{eq}} Z_{TE} \tag{23}$$

where Z_{TE} is the wave impedance of the TE mode.

SIW Transitions

Transitions from SIW to microstrip or coplanar waveguide or from SIW to air-filled rectangular waveguide are necessary to couple high-Q waveguide components to active planar ones. In addition, these transitions are necessary to be able to experimentally characterize the SIW components. They should have the following properties, namely broadband, low loss, and minimum VSWR.

Although the design of transitions is specific to a particular geometry, some basic guidelines can be appreciated. The design of broadband transitions, in general, requires that the level of the characteristic impedances be similar (impedance matching) and that the field distributions of the coupling structures be similarly oriented (field matching). The design of transitions between two dissimilar structures is complicated by the excitation of higher order modes at the discontinuity.

Different waveguide-to-SIW transitions are detailed in the literature. Most of these transitions are based on waveguide-to-microstrip or to-CPW (Cassivi et al. 2002; Deslandes 2010).

Microstrip/CPW-to-SIW Transition

The microstrip-to-SIW transition shown in Fig. 7 is based on the quarter-wave transformer principle. The main parameters of this transition are the width of the SIW (a_{eq}), the width of the microstrip taper (w) and the taper length (l). a_{eq} can be computed using Eq. 6. For each substrate having thickness h and permittivity ε_r, an optimum taper width is obtained by solving this equation (Deslandes 2010):

DOI 10.1007/978-981-4560-75-7_57-1

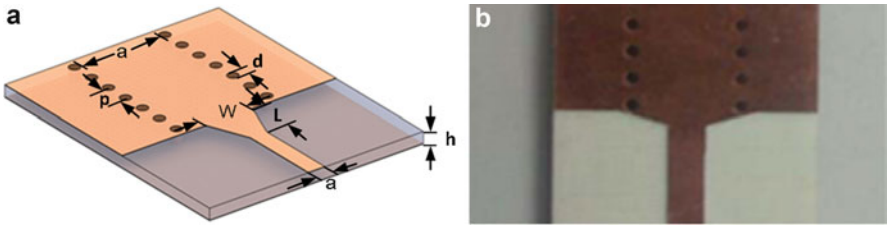


Fig. 7 Microstrip-to-SIW transition: (**a**) configuration and (**b**) example of a fabricated prototype

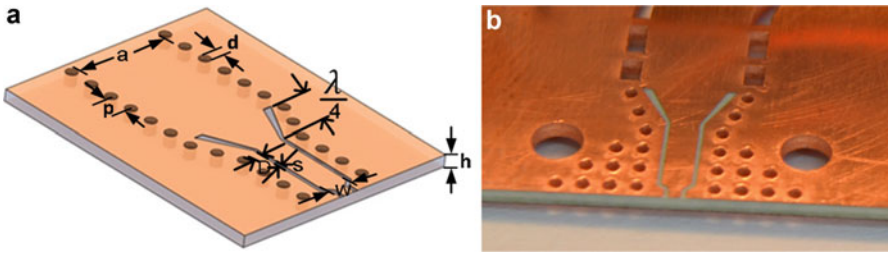


Fig. 8 GCPW-to-SIW transition: (**a**) configuration and (**b**) example of a fabricated prototype

$$\frac{1}{w_e} = \begin{cases} \dfrac{60}{\eta h}\ln\left(8\dfrac{h}{w} + 0.25\dfrac{w}{h}\right); & \dfrac{w}{h} < 1 \\ \dfrac{120\pi}{\eta h[w/h + 1.393 + 0.667\ln(w/h + 1.444)]}; & \dfrac{w}{h} > 1 \end{cases} \tag{24}$$

$$\frac{1}{w_e} = \frac{4.38}{a_{eq}} e^{-0.627\frac{\varepsilon_r}{\frac{\varepsilon_r+1}{2}+\frac{\varepsilon_r-1}{2}\frac{1}{\sqrt{1+12h/w}}}} \tag{25}$$

For a given substrate (h and ε_r) and SIW width (a_{eq}), to find w, Eqs. 24 and 25 are equalized, which yields the optimum taper width.

l is optimized to minimize the return loss over the full bandwidth, this length can be computed using some analytical equations described in Lu (1997).

Patrovsky et al. (2007) proposed a CPWG-to-SIW transition using an E-field coupling. The coupling slots are cut on the top surface of the SIW and placed next to the short circuit termination of the SIW. The coupling slots act like a magnetic dipole antenna with a strong E-field across the slot in the center but weaker at the end of the slot. In Li et al. (2009) Fig. 8, the sidewalls of the SIW are tapered along the triangle-shaped coupling slot in such a way that the direction of the electric field on the coupling slot is always perpendicular to the SIW sidewalls. This allows for a smooth transition. The tapered coupling slot also serves as an impedance transformer to transform any arbitrary impedance line in SIW to the CPWG port impedance exactly as the microstrip transition does.

To assure a single mode operation in the CPW, the following condition must be satisfied:

$$W + 2S + 2D < \frac{c}{2f_{\max}\sqrt{\varepsilon_r}} \tag{26}$$

W, S, and D are defined in Deslandes and Wu (2003), $f_{\max}$ is the maximum frequency of operation. The optimal taper length is $\frac{\lambda}{4}$ which gives a wider bandwidth.

DOI 10.1007/978-981-4560-75-7_57-1

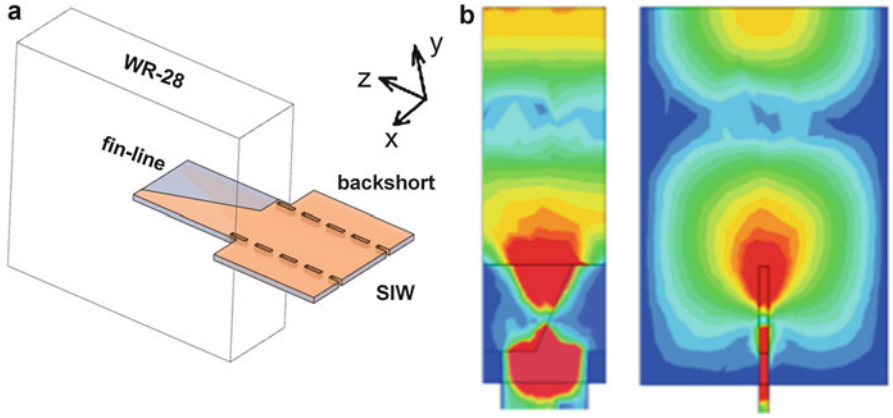


Fig. 9 Metallic waveguide-to-SIW transition using a fin line: (**a**) configuration and (**b**) simulated E-field magnitude distributions by HFSS at 34 GHz, along the transition XZ plan and YZ plan

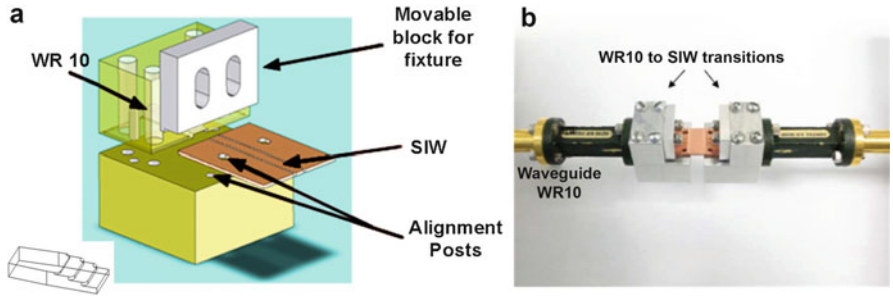


Fig. 10 WR10-to-SIW line transition using a stepped impedance: (**a**) configuration, (**b**) photograph of a back-to-back transition

Rectangular Waveguide-to-SIW Transition

In (Xia et al. 2006), a radial probe extended from an SIW inserted into a height-tapered waveguide is used, losses of about 2.5 dB are achieved at 33 GHz. In order to design the transition to a rectangular waveguide, two step transformers are used in Moldovan et al. (2006) with parallelepiped alumina probe to improve the scattering parameter, more than 3 dB of loss is measured in W band. In Li et al. (2009), a longitudinal slot window etched on the broad wall of SIW couples the energy between SIW and rectangular waveguide. This transition shows a return loss of less than 15 dB over a frequency range of 800 MHz at 35 GHz with loss of 2.5 dB in measurement and more than 1 dB in the simulation. In Li et al. (2010), a novel transition structure is realized by using a finline probe inserted into a WR-28 waveguide with an insertion loss of 1.4 dB. To measure this transition, two parts of the waveguide sandwich the substrate, which normally increases the width of the metallic waveguide. In Djerafi et al. (2012c), a WR-28-to-SIW transition based on ALTSA (Antipodal Linearly Tapered Slot Antenna) as illustrated in Fig. 9 is presented. A design constraint was to make the transition without any modification on the waveguide. A back-to-back transition has been fabricated, and measured results confirm performances of this transition. In Djerafi et al. (2013), a transition from the WR10 waveguide to SIW structure has been designed and fabricated. The proposed transition consists of a multistep impedance transformer milled into an aluminum block. The transition is illustrated in Fig. 10 where the parameters to be optimized in this transition are the aspect ratios and lengths of the steps.

DOI 10.1007/978-981-4560-75-7_57-1

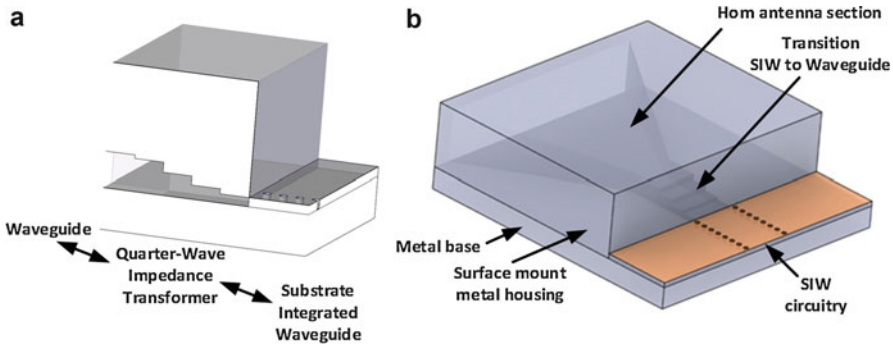


Fig. 11 (**a**) SIW-waveguide transition, (**b**) a surface mountable pyramidal horn antenna

Nonresonant SIW Antenna

The SICs and related antennas technologies converge toward the development of system-on-substrate (SoS) technology that allows for a complete and seamless integration of antennas and circuits. SIW supports the design of different antennas involving various elements and feeds as detailed in Wu et al. (2012). The more adaptable and popular antenna elements: slot (Wei et al. 2013) and patch backed by cavity. To miniaturize the size of antennas, different techniques can be used in the design of slot array antennas (Ding and Wu 2009, 2010). Printed antennas such as the microstrip patch, which might fulfill design requirements in the microwave range, do not provide sufficient performances in the millimeter-wave range. This is because of very high conductor losses arising from immense current densities (singularities) at the strip edges, especially in resonant structures. Disadvantages of resonant elements include also, the fact that it is usually narrowband, with bandwidths of a few percent being typical. Also, the radiation efficiency of the microstrip antenna tends to be lower than some other types of antennas, with efficiencies less than 70 % being typical.

An antenna operating in a traveling-wave configuration could provide the bandwidth and the efficiencies needed. Traveling-wave antennas are a class of antennas that use a traveling wave on a guiding structure as the main radiating mechanism. On the other hand, leakage can also be generated with enclosed waveguide structures with carefully designed geometrical asymmetry, open apertures, or slots. Therefore, leaky-wave or surface-wave antennas can be developed for millimeter-wave applications including tapered slot antenna (TSA), printed Yagi–Uda antenna or log-periodic dipole array, dielectric rod antenna, long-slot leaky-wave antenna. SIW supports the design of different antennas involving various elements and feeds as detailed in Wu et al. (2012). Over the last decade, various nonresonant antennas in SIW technique have been studied, which will be presented in the following section.

Horn Antenna

Horns provide high gain, relatively wide bandwidth, and they are not difficult to make. There are three basic types of rectangular horns. The rectangular horns are ideally suited for rectangular waveguide feeders. The horn acts as a gradual transition from a waveguide mode to a free-space mode of electromagnetic waves. The radiation fields from aperture antennas, such as slots, open-ended waveguides, horns, reflector and lens antennas, are determined from the knowledge of the fields over the aperture of the antenna. The theory of waveguide aperture antennas is generally used to study the horns. However, phase error occurs due to the difference between the lengths from the center of the feeder to the center of the horn aperture and the horn edge.

Horn antennas are now commonly used for many applications including microwave communications, feeds for reflector antennas, and radar elements. In Li and Wu (2004), an integrated *H*-plane horn antenna

DOI 10.1007/978-981-4560-75-7_57-1

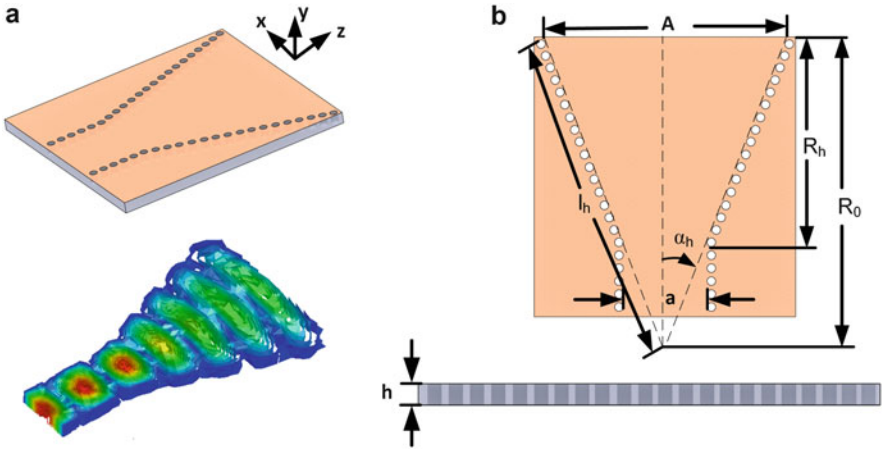


Fig. 12 (**a**) Geometry of H-plane SIW horn, (**b**) electric field distribution simulated at 35 GHz, (**c**) and (**d**) the parameters used in the analysis

was proposed. In Wang et al. (2010), a dielectric load was added to the SIW H-plane horn antenna in order to have a narrow beamwidth in both E-plane and H-plane. In Wong et al. (2008), a broadside horn antenna was implemented by the SIW technology.

E-Plane SIW Horn (Waveguide Transition)

The design of an appropriate SIW-waveguide transition passes through the design of a multiple stepped waveguide quarter-wave impedance transformer. The transformer makes use of a multiple reflection scheme. This transition can be used for designing a surface mountable horn antenna feed by SIW line. The waveguide port of the transition is flared into a large opening and forms an E-plane horn.

The proposed transition consists of a multistep impedance transformer milled into an aluminum block. The transition is illustrated in Fig. 11. The parameters to be optimized for this transition are the aspect ratios and lengths of the steps. The detail of the design has been well established (Pozar 2011).

The impedance in each section of the quarter-wave transformer is defined by:

$$Z_i = \frac{\pi^2 h_i \sqrt{\frac{\mu}{\varepsilon}}}{8a\sqrt{1-\left(\frac{\lambda_0}{2a}\right)^2}} \tag{27}$$

where h_i is the height of each section. The input impedance is defined by SIW line.

Since most of microwave integrated circuit modules need a metal housing, this structure can be installed as part of an external housing. A prototype presented by Li and Wu (2004) has an HPBW of 15° in H-plane and 27° in E-plane and a cross polarization below 30 dB.

H-Plane SIW Horns

The horn antenna integrated into substrate is a waveguide aperture flared into a large opening on the H-plane by keeping the perpendicular dimension constant. The radiation is focused in the H-plane. The geometry is shown in Fig. 12.

DOI 10.1007/978-981-4560-75-7_57-1

$$l_h{}^2 = R_0 + \left(\frac{A}{2}\right)^2, \tag{28}$$

$$\alpha_h = \arctan\left(\frac{A}{2R_0}\right), \tag{29}$$

$$R_h = \left(A - a_{eq}\right)\sqrt{\left(\frac{l_h}{A}\right) - \frac{1}{4}}, \tag{30}$$

To have a single mode excitation, the width and height of the SIW line must satisfy:

$$\frac{\lambda_0}{2\sqrt{\varepsilon_r}} < a_{eq} < \frac{\lambda_0}{2\sqrt{\varepsilon_r}} \tag{31}$$

and

$$h < \frac{\lambda_0}{2\sqrt{\varepsilon_r}} \tag{32}$$

where λ_0 is wavelength in free space.

The design rules follow the same principles of the conventional horn. The corresponding pattern can be computed using:

$$F_H(\theta, \varphi = 0^\circ) = \frac{1 + \cos(\theta)}{2} \frac{\mathrm{I}\left(\theta, \varphi = 0^\circ\right)}{\mathrm{I}(\theta = 0^\circ, \varphi = 0^\circ)} (\mathrm{H} - \text{plane pattern}) \tag{33}$$

$$F_E(\theta, \varphi = 90^\circ) = \frac{1 + \cos(\theta)}{2} \frac{\sin\left[\left(\frac{\beta b}{2}\right)\sin\theta\sin\varphi\right]}{\left(\frac{\beta b}{2}\right)\sin\theta\sin\varphi} (\mathrm{E} - \text{plane pattern}) \tag{34}$$

where

$$\begin{aligned} I(\theta, \varphi) = {} & e^{j\left(\frac{R_1}{2\beta}\right)\left(\beta\sin\theta\cos\varphi + \frac{\pi}{A}\right)^2}\left[C\left(S_2'\right) - jS\left(S_2'\right) - C\left(S_1'\right) - jS\left(S_1'\right)\right] \\ & + e^{j\left(\frac{R_1}{2\beta}\right)\left(\beta\sin\theta\cos\varphi - \frac{\pi}{A}\right)^2}\left[C\left(t_2'\right) - jS\left(t_2'\right) - C\left(t_1'\right) + jS\left(t_1'\right)\right] \end{aligned} \tag{35}$$

$$S_1' = \sqrt{\frac{1}{\pi\beta R_1}}\left(-\frac{\beta A}{2} - R_1\beta\sin\theta\cos\varphi - \frac{\pi R_1}{A}\right) \tag{36}$$

$$S_2' = \sqrt{\frac{1}{\pi\beta R_1}\left(\frac{\beta A}{2} - R_1\beta\sin\theta\cos\varphi - \frac{\pi R_1}{A}\right)} \tag{37}$$

DOI 10.1007/978-981-4560-75-7_57-1

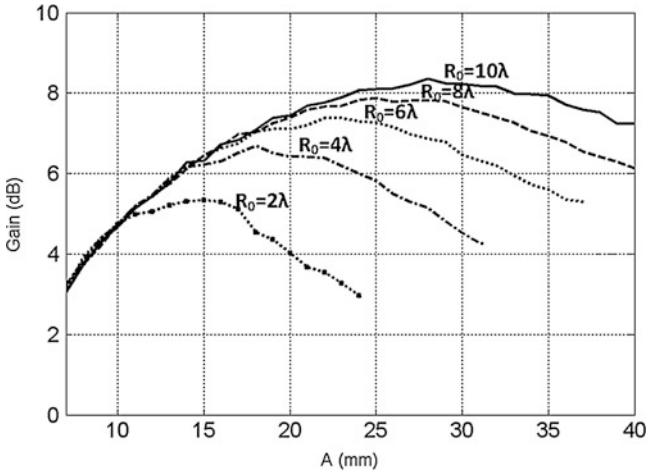


Fig. 13 Optimal gain for different lengths and apertures

$$t_1' = \sqrt{\frac{1}{\pi\beta R_1}\left(-\frac{\beta A}{2} - R_1\beta\sin\theta\cos\varphi + \frac{\pi R_1}{A}\right)} \tag{38}$$

$$t_2' = \sqrt{\frac{1}{\pi\beta R_1}\left(\frac{\beta A}{2} - R_1\beta\sin\theta\cos\varphi + \frac{\pi R_1}{A}\right)} \tag{39}$$

$C(x)$ and $S(x)$ are Fresnel integrals, which are defined as

$$C(x) = \int_0^x \cos\left(\tfrac{\pi}{2}\tau^2\right)d\tau \text{ and } S(x) = \int_0^x \sin\left(\tfrac{\pi}{2}\tau^2\right)d\tau$$

The directivity is calculated by

$$D = \frac{4\pi}{\lambda^2}\varepsilon_t\varepsilon_{ph}^H(Ah) \tag{40}$$

where

$$\varepsilon_t = \frac{8}{\pi^2}$$

$$\varepsilon_{ph}^H = \frac{\pi^2}{64t}\left((C(p_1) - C(p_2))^2 + (S(p_1) - S(p_1))^2\right)$$

$$p_1 = 2\sqrt{t}\left(1 + \frac{1}{8t}\right)$$

$$p_2 = 2\sqrt{t}\left(-1 + \frac{1}{8t}\right)$$

DOI 10.1007/978-981-4560-75-7_57-1

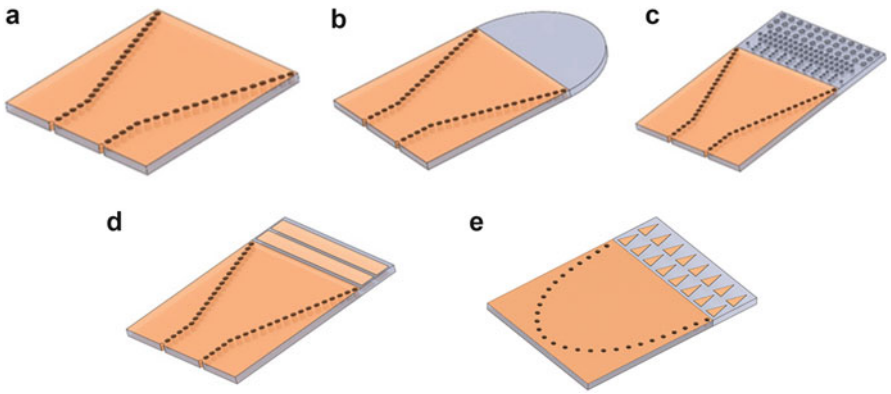


Fig. 14 SIW horn antennas (**a**) standard, (**b**) SIW horn with a dielectric lens (the lens has a constant permittivity, (**c**) SIW horn antenna with dielectric lens (the lens has different permittivity), (**d**) SIW horn with a 2 block printed transition to the air, (**e**) SIW horn antenna with a triangular grated transition and a superelliptical taper

$$t = \frac{1}{8}\left(\frac{A}{\lambda}\right)^2 \frac{\lambda}{R_0}$$

Figure 13 shows an example of the gain obtained using HFSS for different lengths (R_0) and apertures A. The substrate used has a dielectric constant of 2.94 and a thickness of 2 mm ($a_{eq}/2$)

For each length, there is an optimal aperture that gives the maximum gain. The figure demonstrates that by increasing the length, the gain enhancement is less and less significant.

For a very thin substrate thickness with respect to propagation wavelength, the electromagnetic waves will be confined within the substrate instead of radiating. The mismatch between the SIW horn and the air increases which results in a poor bandwidth.

Matching Bandwidth and Radiation Improvements

One of the solutions to improve the horn performance is the insertion of lens in front of the radiating aperture. Generally, dielectric lenses are constructed by extending the same dielectric used to build the horn. In the lens-corrected horn, the lens acts as a focusing element, which increases the front to back ratio and reduces the front phase error. When the substrate becomes thinner, generally below $\frac{\lambda_0}{6}$, the lens effects become less relevant. The different developed topologies in the literature are shown in Fig. 14 as well as the fabricated counterpart in Fig. 15.

SIW H-plane sectorial horn antenna is given in Fig. 15a. The rectangular waveguide and sectorial horn antenna are integrated by using the same single substrate based on the SIW technology in Wang et al. (2010). A reasonable length of 15 mm was determined to obtain an acceptable quadratic phase error and single mode on the aperture. The simulated gain and side lobe level are 6.73 dBi and −12.08, respectively; the simulated beamwidths of the E-plane and the H -plane are 180° and 48°, respectively. Efforts for improving the impedance matching of planar horn antenna have been reported in recent papers. Solutions based on integrated dielectric lens (Fig. 14b, c) or printed transition for SIW were proposed (Fig. 14d, e).

In Wang et al. (2010) a dielectric slab is placed in front of the aperture of the horn. This slab serves as the dielectric guiding structure in the H-plane. In the H-plane, for a horn with maximum gain, the aperture phase distribution along the H-plane is nearly uniform without the dielectric loading. This antenna is loaded by a dielectric elliptical slab with a length of 1.4 mm. In this case, the dielectric loading serves as the phase corrector in the H-plane to compensate the quadratic phase error caused from the shortened length. The gain is 5.75 dBi, the back sidelobe level is 11.78 dB and the beamwidths of the E- and H-plane

DOI 10.1007/978-981-4560-75-7_57-1

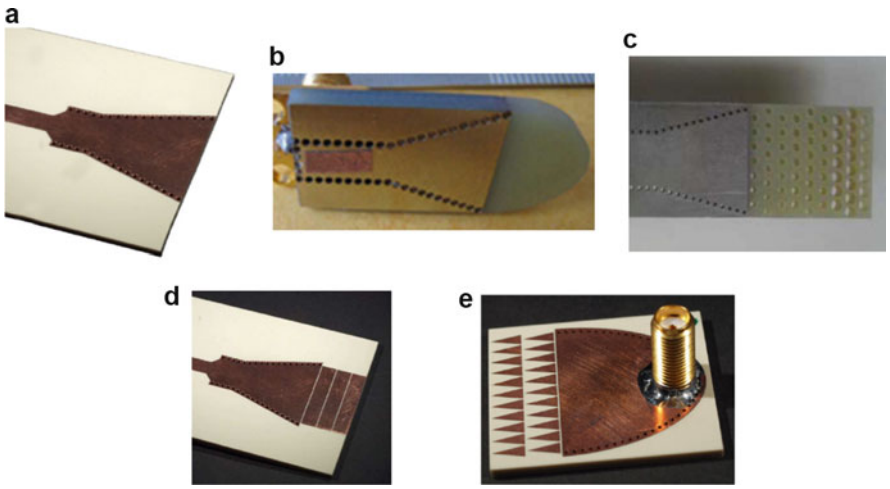


Fig. 15 Example of the fabricated horn antennas according to Fig. 14

are 178° and 63°, respectively. Different shapes of dielectric loading can be used (ellipse, rectangle, hyperbola, and parabola). In Cao et al. (2014) a planar SIW horn antenna with a modified dielectric lens was proposed, by drilling air vias in the substrate, different effective dielectric constants can be realized by adjusting via diameter. Therefore, the SIW horn antenna loaded by a dielectric slab with gradually decreasing dielectric constant is designed.

The presented solution in Esquius-Morote et al. (2013a) is focused on improving the reflection coefficient of H-plane SIW horn antennas with additional parallel plate block. The horn aperture practically behaves as a slot in an infinite ground plane with a front-to-back ratio (FTBR) close to unity. The proposed printed transition does not significantly modify the radiation characteristics of a conventional SIW horn. The main radiating element of an SIW horn with the proposed transition is indeed the open-ended termination of the last parallel plates waveguide. Therefore, the differences in radiation performances are mainly due to diffraction effects at the transition edges. Modifications of the transition geometry have been presented in Esquius-Morote et al. (2013b) to improve the radiation performances of SIW horn antennas (Figs. 14d and 15d).

The tapered ridge was proposed in Mallahzadeh and Esfandiarpour (2012), consisting of a simple arrangement of vias on the side flared wall within the multilayer substrate. VSWR below 2.5 for the whole frequency range of 18–40 GHz was realized by simulation.

Leaky-Wave Antenna

Leaky-wave antennas (LWA) are nonresonant, consisting of a series of perturbations on a metallic or dielectric waveguide in question. The perturbations produce radiation since they cause energy to leak away along the waveguide. Leaky-wave antennas have been studied for more than four decades, exhibiting interesting features due to their traveling-wave nonresonant nature. Leaky-wave antennas have the inherent property that the beam scans with frequency. For scanning applications, this is an important advantage. Most of the early leaky-wave antennas were based on enclosed rectangular or NRD waveguides, in which some kind of cuts or holes were introduced in order to produce the leakage of power along the length of the waveguide. In the last two decades, the research was focused in the millimeter-wave band, where new LWA designs have been proposed to overcome the technological problems associated to these higher frequencies.

Leaky-wave antennas support a fast wave on the guiding structure, where phase constant β is less than free-space wavenumber k_0. The leaky-wave is therefore fundamentally a radiating type of wave, which

DOI 10.1007/978-981-4560-75-7_57-1

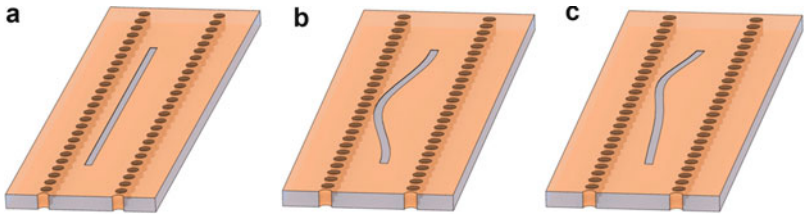


Fig. 16 Leaky-wave long slot antenna: (**a**) straight longitudinal slot, (**b**) sine squared meander pattern, and (**c**) Taylor power distribution meander pattern

radiates or "leaks" power continuously as it propagates on the guiding structure (and hence the name of the antenna). The operation is therefore quite different from a slow-wave or surface-wave type of antenna, where radiation mainly takes place at the end of the antenna.

Depending upon their architecture, leaky-wave antenna can be categorized as either uniform or periodic or quasi uniform periodic. Uniform antennas are homogeneous along the length of the guiding structure. Periodic structure has regular perturbations along their length. They both are derived from specific guided-wave counterparts and follow the same rules of formation of leaky wave. These two types can be constructed on the top wall (broad wall) or lateral wall (narrow wall). Another distinction that can be made is between a one-dimensional (1D) leaky-wave antenna and a two-dimensional (2D) leaky-wave antenna.

Uniform Leaky-Wave Structure

Long Slot Leaky-Wave Antenna

Figure 16 shows different configurations of SIW uniform leaky-wave structure. Usually, a fast wave ($b < k_0$) travels in enclosed guided-wave structures such as rectangular metallic waveguide and SIW. A physical cut that disturbs the guided wave along the longitudinal propagation direction will produce a leaky wave because the condition of forming a leaky wave has been fulfilled in this case. The long slot on the broadside of waveguide has to be precisely machined with a specific meandering profile in order to synthesize the desired radiation properties. Cheng et al. (2011) introduces a design process of SIW long slot leaky-wave antenna.

Using substrate with permittivity ε_r, the beam direction θ at the operation frequency is defined as:

$$\theta = \arcsin\frac{\lambda_0}{\lambda_g} = \arcsin\sqrt{\varepsilon_{r-}\left(\frac{c}{2a_{eq}f}\right)^2} \tag{41}$$

a_{eq}, the SIW equivalent width is generally selected to ensure the monomode operation.

As shown in Fig. 17, the scanning range of uniform leaky-wave antennas using a standard air-filled waveguide is about 12° with 4 GHz of frequency sweep around 35 GHz. The variation of pointing angle in the same bandwidth increases when the SIW line is used, with permittivity of 3. The scanning angle is three times in SIW compared to the hollow waveguide. In Cao et al. (2014) planar beam scanning substrate integrated waveguide (SIW) slot leaky-wave antenna (LWA) is proposed for enhancing scanning range and gain flatness using a modified composite right/left-handed transmission line (CRLH TL) structure.

It is assumed that the leaky-wave antenna is terminated with load which absorbs a fraction R of input power. Generally, the length L is considered equivalent to 20 wavelength to ensure efficiency of 95 % ($R = 0.05$). The slot width should be around 0.08 λ_g.

DOI 10.1007/978-981-4560-75-7_57-1

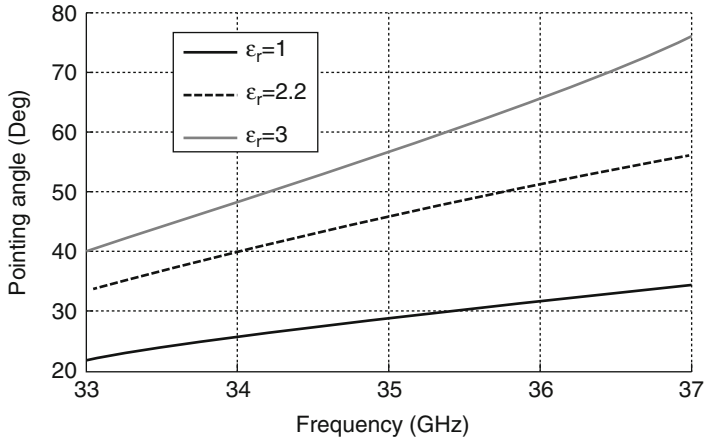


Fig. 17 Scanning range of uniform slot with different permittivity

Different meander contour can be considered other than the straight longitudinal slot; sine squared plotted in Fig.16b is defined by in Whetten and Balanis (1991)

$$x = \left(0.5 + 0.2sin^2\left(\frac{\pi}{L}z\right)\right)a \tag{42}$$

The slot starts at waveguide center ($x = 0.5\ a$ at $z = 0$) and returns at the center at the end of the slot. Different contours can be defined regarding the distribution of the radiated power. This type of meander contour provides the necessary aperture field taper for superior antenna far-field amplitude performance. Since the power within the waveguide is continuously lost along the slot, the ideal meander pattern must account for both the amount of power to be radiated at a given point along the slot, and the amount of power remaining in the guide. Thus, at the leading edge of the slot (nearest to the generator), the slot can remain near the centerline of the guide and still radiate sufficient power. Farther along the slot, as the power level decreases within the guide, the slot must meander closer to the sidewall in order to radiate the desired power level. The exact meander contour is usually very difficult to determine. Starting from the required amplitude distribution (Taylor (shown in Fig. 17c), Tchebechev, sinusoidal, etc.), the radiation per unit length α is defined by:

$$\alpha(z) = \frac{\frac{1}{2}|A(z)|^2}{\frac{1}{1-R}\int_0^L A(\zeta)^2 d\zeta - \int_0^z A(\zeta)^2 d\zeta} \tag{43}$$

Then the relation between offset x, and radiation per unit length α should be determined. Most of previous works determined the radiation properties of a long slot leaky-wave antenna by a numerical analysis of the fields in the slot. In Cheng et al. (2011), Ansoft HFSS was used to determine radiation per unit length, and normalized phase propagation per unit length, according to different offset. In Scharp (1982) the relation is defined as:

DOI 10.1007/978-981-4560-75-7_57-1

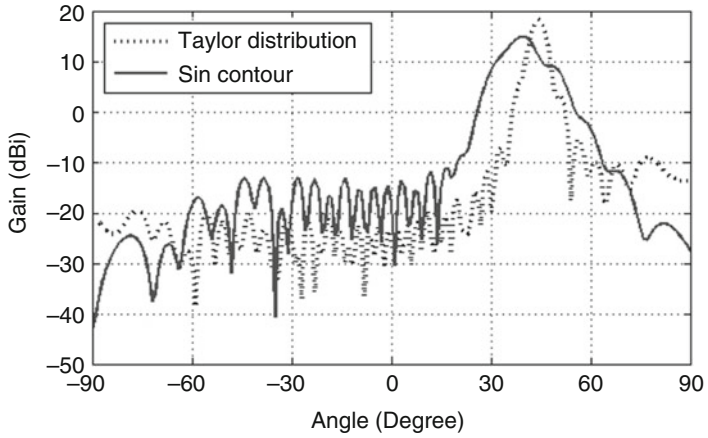


Fig. 18 Simulated patterns of leaky-wave antenna

$$x(z) = \frac{a}{\pi} arcsin \sqrt{\frac{h}{2d}\frac{\lambda_c^3}{\lambda\lambda_g L}\frac{\alpha\left(\frac{z}{L}\right)}{1-\int_0^{\frac{z}{L}}\alpha(y)dy}} \tag{44}$$

d for the width of the slot and h is the substrate thickness, the illumination being a function of the distance z divided by the length L.

Figure 18 gives simulated radiation patterns for an LWA antenna with Sin contour and Tayler distribution. Dielectric with 2.2 of permittivity and 0.5 mm of thickness with length of 115 mm (20 λ_g at 35 GHz) is used in the two designs. The antenna with Taylor distribution shows higher gain, narrower 3-dB beam-width and better sidelobe levels. However, the sin contour slot is very simple to design and in general its performance is still acceptable.

Lateral Slot Leaky-Wave Antenna

Using its half-mode (HM) version, the SIW structure offers another advantage when used in the design of leaky wave structures. When only the mode TE10 is propagating in SIW waveguide, the maximum value of E-field is located at the vertical center plane along the propagating direction and the center plane can be considered as an equivalent magnetic wall. As a result, the half-mode substrate integrated waveguide (HMSIW) can be simply achieved by dividing the SIW into two equal parts. HMSIW was proposed by Hong et al. (2006). In an SIW structure, the ratio of width a to height b is usually very large when an SIW is cut into two half waveguides along the center plane, the resulting open aperture between the top and bottom planes is thus very close to a magnetic wall. The dominant mode in such a structure is just like one half of the dominant mode TE_{10} in a complete SIW. In the area close to the aperture, the field is out-diffused a little bit to satisfy its boundary conditions. On the other hand, an HMSIW can also be regarded as a half of microstrip line, which is terminated by a short-circuited plane in the middle. The first higher order mode of the microstrip line has been demonstrated to be an efficient leaky-wave line source. Usually an electric wall is inserted in the middle plane of a microstrip line to suppress the dominate mode (quasi-TEM mode). Interestingly, the working mechanism of a half microstrip line excited by its first high-order mode is exactly the same as that of an HMSIW. Since the leaky-mode is the fundamental mode in the HMLWA, the excitation becomes easier, and complicated feeding structures are not needed anymore. A leaky-wave antenna based on the HMSIW technique was proposed and investigated in Xu

DOI 10.1007/978-981-4560-75-7_57-1

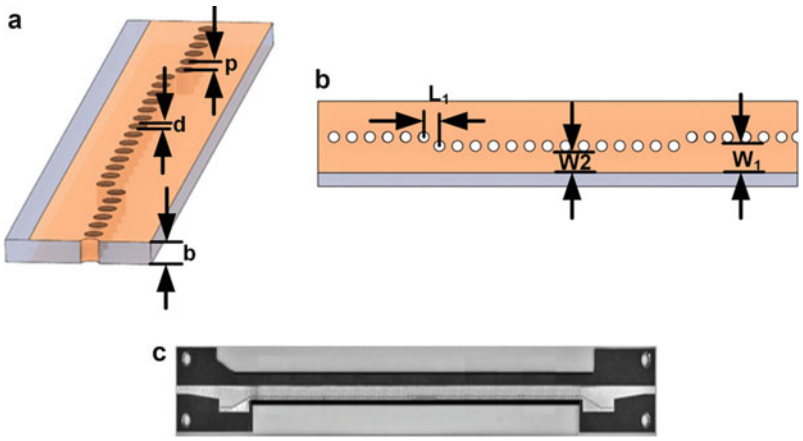


Fig. 19 (**a**, **b**) Configuration of HMSIW LWA, (**c**) a fabricated prototype

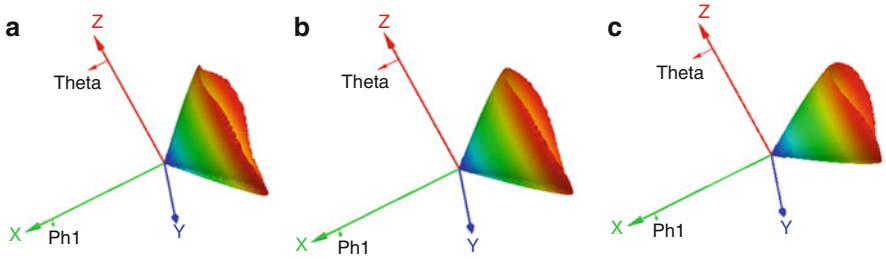


Fig. 20 Radiation pattern of HMLWA with different thickness at 35 GHz: (**a**) 5 mil, (**b**) 10 mil, (**c**) 20 mil

et al. (2008). The antenna is characterized by many attractive advantages, including compact size, easy fabrication, low cost, low loss, wide bandwidth, and direct integration with planar circuits.

The proposed LWA topology is shown in Fig. 19, to improve the return loss, a section of HMSIW with gradually tapered width from W_2 to W_1 is adopted. L is the antenna length, b is the thickness of substrate and the dielectric constant of the used substrate is 2.2, W_2 is the width between the metal via holes and leakage aperture (also called aperture depth when seen from antenna angle), and W_0 is the width between the metal via holes and the other edge of waveguide (input HMSIW lines). The width of additional dielectric is designed less than $\lambda/4$ to avoid the undesired channel modes.

The far-field radiation pattern of this antenna indicates a quasi omnidirectional characteristic, which is a hollow cone centered on the axis. The beam will then be in the form of a conical beam that opens along the negative z axis at an angle θ that increases with frequency, approaching a donut-shaped pattern at broadside. The angle of the cone can be controlled using the thickness as shown in Fig. 20. As the used substrate is thicker as the aperture of the cone is smaller.

Periodic Structures

This type of leaky-wave antenna consists of a slow-wave structure that has been modified by periodically modulating the structure in some fashion. A typical example is a rectangular waveguide that is loaded with a dielectric material (so that the TE_{10} mode is a slow wave) and then modulated with a periodic set of holes or slots, as shown in Fig. 21. Many of the features common to 1D periodic leaky-wave antennas may be appreciated by consideration of this simple (but practical) structure. The antenna can be designed to radiate from backward to forward. In rectangular waveguide, (Hyneman 1959) shows a closely spaced slotted traveling wave array. All slots have the same length, the waveguide does not taper, and the

DOI 10.1007/978-981-4560-75-7_57-1

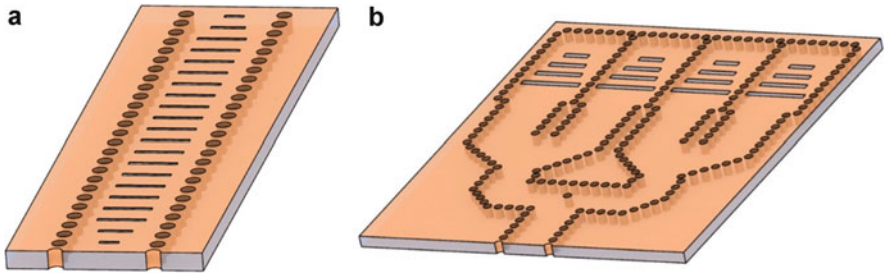


Fig. 21 (**a**) Configuration of periodic slot LWA, (**b**) example of array architecture

waveguide is filled with air. Dion (1958) describes relationships among parameters within nonresonant slotted arrays.

It is assumed that the relative permittivity of the filling material is sufficiently high so that the TE_{10} mode is a slow wave over the frequency region of interest. This will be the case provided

$$\varepsilon_r > 1 + \left(\frac{\pi}{k_0 a_{eq}}\right)^2 \tag{45}$$

According to Floquet's theorem, the infinite numbers of space harmonics of guided mode inside the periodic LWA are related to,

$$\beta_n = \beta_0 + 2n\pi/p \tag{46}$$

The first kind of leaky-wave antennas provides radiation into the forward quadrant and can yield scanning from broadside to forward end-fire directions. The scanning range for periodic leaky-wave antennas reaches from backward end-fire through broadside directions into a part of the forward quadrant. The dominant mode on the former type represents a fast wave, while the latter type is a slow wave structure. As a result, the dominant mode on periodic leaky-wave antennas does not radiate and radiation is achieved by using one of its space harmonics.

The first higher-order radiating harmonic β_{-1} inside the dielectric-filled rectangular waveguide for which $(\beta_n/k_0) \leq 1$ is related as:

$$\frac{\beta_{-1}}{k_0} = \sqrt{\varepsilon_r}\sqrt{1 - \left(\frac{c}{2af\sqrt{\varepsilon_r}}\right)^2} - \frac{n\lambda_0}{d} \tag{47}$$

Dispersion behavior inside the LWA is explained through the attenuation constant α and the propagation constant β. An angle of radiation ϕ_0 is a function of space harmonic β_n. If $\beta_{-1} < 0$, antenna radiates in the backward direction, as frequency is increasing $\beta_{-1} > 0$, antenna radiates in the forward direction.

$$\phi_0 = sin^{-1}\left(\frac{\sqrt{\varepsilon_r}\beta_{-1}}{k_0}\right) \tag{48}$$

The distance d_0 corresponding to this condition is

$$d_0 = \frac{\lambda}{1 + |cos\,\phi_0|}\pi r^2 \tag{49}$$

DOI 10.1007/978-981-4560-75-7_57-1

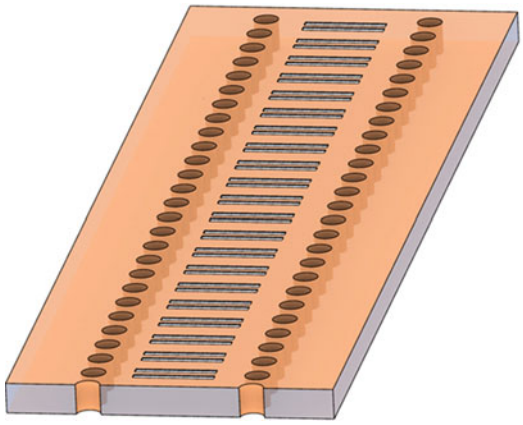


Fig. 22 The reflection cancelation forward-wave LWA

Since the slots cause a perturbation to the TE mode of the SIW, to suppress the reflection, the slot lengths are tapered linearly at both ends as shown in Fig. 21. The slot length is transformed from zero to slot length and keep the other parameters of the slots constant. To enhance the bandwidth, the SIW width can be tapered as a cosine function at both ends.

The periodic frequency scanning antenna generally has a poor radiation efficiency around broad side due to the existence of open stop band in the narrow frequency band. The reason is that the constructive addition of all series or shunt admittances along the transmission line forces input impedance to become zero. The addition of reflection cancelation stub at the distance of $\lambda_g/4$ from excitation slot avoids the undesired open stopband. The dual slot structure was proposed in Yamamoto et al. (2002), but the structure is of broadside nature only. The periodic array of eight radiating transverse slots is embedded on the broad wall of the guiding structure. The distance (S_1) between the excitation slot and reflection cancelation slot is approximately one quarter of a guided wavelength (λ_g) as shown in Fig. 22. The total phase shift between the two slots is approximately 180°. Hence, the reflected waves from the two slots are added in out-of-phase resulting in a very low reflection coefficient at the input. The slot pairs are periodically spaced at 1 λ_g, hence progressive phase is multiple of 2π, resulting in the fact that all the slot pairs are added in phase.

One-port LWA, steering from backward to forward directions including the broad side, is studied and implemented in the SIW technology. The azimuth plane scanning of multibeam antenna (MBA) is obtained through the phase control of array of nine 1-D elevation plane frequency scanning LWAs, each fed through the vertical corner on the array port contour of a Rotman lens. The complete antenna prototype fabricated on single Rogers RT/Duriod 6002 substrate with $\varepsilon_r = 2.94$ and 0.508 mm thickness is simulated through Ansoft HFSS.

The periodic frequency scanning antenna has poor radiation efficiency because of the existence of an open stopband around broad side in the narrow frequency band. In Guntupalli (2014), SIW LWA using the concept of reflection cancelation was developed. Directional pattern characteristics were studied and demonstrated from −37° to +18° by changing frequency from 58 to 78 GHz at an interval of 1 GHz. A matched termination was designed, validated and integrated to realize single port LWA. A 2-D scanning multibeam conformal array was realized by utilizing a Rotman lens BFN feeding nine LWAs in the azimuth plane for E-plane steering and natural frequency scanning property of 1 × 8 LWA for H-plane steering. The whole 3-D circuit is constructed after fabricating nine LWAs and then inserted into the slots made on the Rotman lens BFN (Fig. 23). The antenna array has seven inputs for beam switching, four dummy ports for sidewall absorption and nine LWAs on the contour of the Rotman lens. The lens is

DOI 10.1007/978-981-4560-75-7_57-1

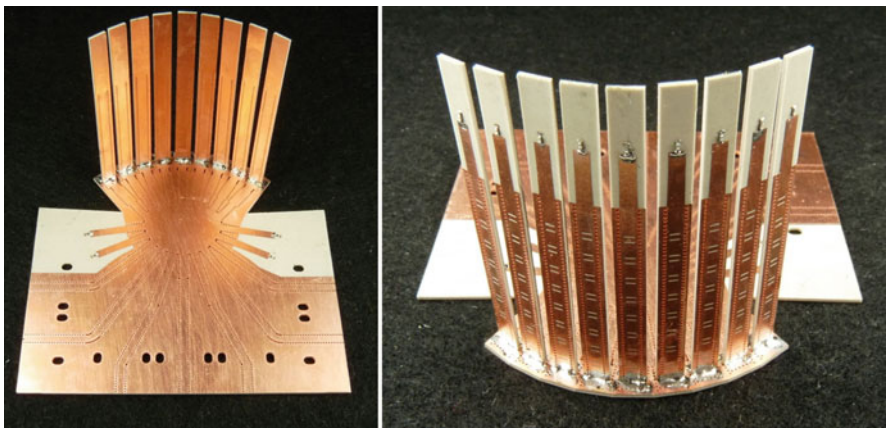


Fig. 23 Two-dimensional scan conformal multibeam array antenna (Figure from http://creer.grames.polymtl.ca/ressources/realisations/)

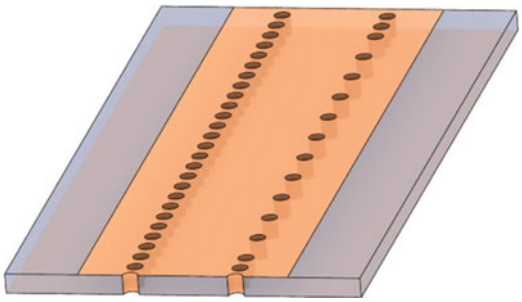


Fig. 24 Open periodic waveguides

synthesized to distribute the desired amplitude to the radiating elements located on the output ports of the Rotman lens contour.

Laterally Open Periodic Waveguides

The scheme of a lateral periodic open SIW LWA is shown in Fig. 24. This configuration is similar to the HMLWA with the addition of a second row of conducting vias separated a distance P, which behaves as an inductive partially reflective surface (PRS) (Martinez-Ros et al. 2010) and allows the propagation of the fundamental mode in an SIW. In the SIW LWA introduced by Deslandes and Wu (2005), there is a lateral wave that radiates from the open end of the truncated substrate. The radiation mechanism was the basis for identifying additional limitations originating from Snell's law at the dielectric-air interface. In general case, the periodic modulation of the conductors that gives the periodic LWA a larger scanning range and a more flexible design scheme than the uniform leaky-wave antenna. The lateral LWA radiates in the form of a space leaky-wave created in the discontinuity between the grounded microstrip lateral edge and the extended dielectric-filled ground plane as any periodic LWA.

Antipodal Tapered TSA

TSA

The Tapered Slot Antenna (TSA) consists of a tapered slot cut in a thin film of metal with or without an electrically thin substrate on one side of the film as shown in Fig. 25. The slot is narrow toward one end (s) for efficient coupling to devices such as mixer diodes. Away from this region (L_{ant}), the slot is tapered and

DOI 10.1007/978-981-4560-75-7_57-1

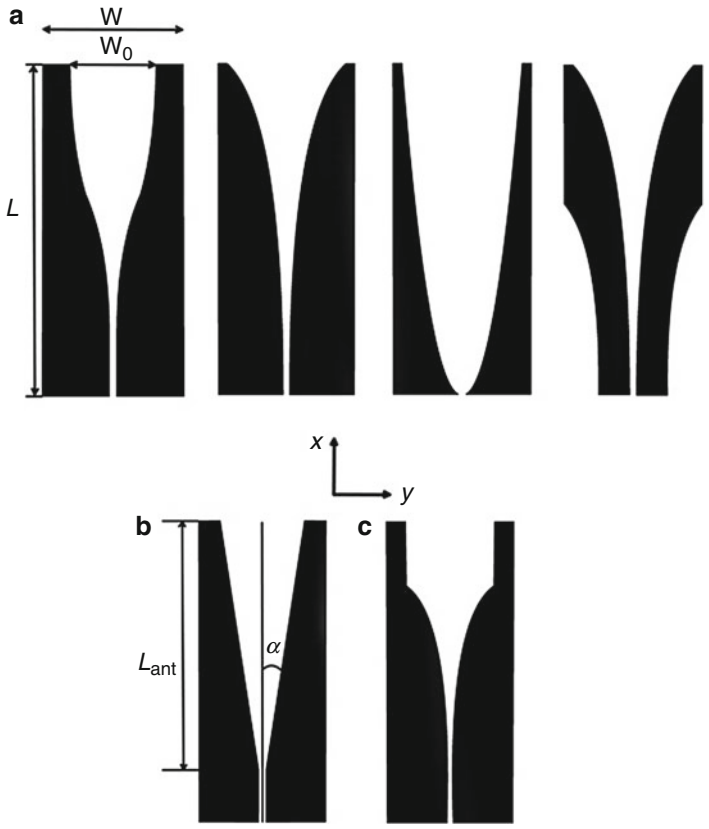


Fig. 25 Differently shaped tapered-slot antennas: (**a**) Nonlinear: Exponential (Vivaldi, *ETSA*), Fermi, Parabolic (*PTSA*), Dual Exponentially (*DETSA*), (**b**) Linear (*LTSA*) and (**c**) Constant Width Slot

a traveling wave propagating along the slot radiates in the end-fire direction (w_0). Antenna is in principle a transition from a guided wave to an unbounded wave or vice versa. Looking at typical TSA, this definition is quite intuitive. As the width of the slot increases, the characteristic impedance increases as well providing smooth transition to the free space characteristic impedance of $120\pi\ \Omega$. The antenna was first proposed by Lewis et al. (1974). Prasad and Mahapatra (1979) first introduced the linearly tapered slot antenna (LTSA) in 1979. Their antenna of about one free space wavelength long was etched on a piece of 25-mil alumina substrate. In the same year, Gibson (1979) reported a TSA with exponentially tapered profile, also known as – Vivaldi antenna, which demonstrated a bandwidth of 8–40 GHz.

An SIW-based antipodal LTSA was demonstrated in Hao et al. (2005) and width corrugated version in Djerafi and Wu (2012b). In the antipodal version of the TSA, the metallization on either side of the substrate is flared in opposite directions to form the tapered slot. When the SIW waveguide is used to feed the antipodal TSA, the bandwidth limitation caused by balun can be removed and, thus, wideband characteristics are indeed obtainable. The SIW TSAs have many advantages such as low profile, low weight, wideband characteristics, and they are easy to integrate with other planar devices. The TSA's antennas are very sensitive to the thickness and the permittivity of their supporting substrate. They can exhibit higher cross-polarized lobes in the (E or H)-plane and sidelobe in the H-plane. This is not only very critical in element design but affects considerably the array performances. The different material, antenna shape, as well as mutual coupling have been widely studied for one side TSAs unlike the SIW antipodal version (Fig. 26).

In the antipodal version, a balanced microstrip feed is used in a double-sided arrangement for the antenna. When SIW is used to feed the antipodal TSA (Fig. 27), the bandwidth limitation caused by balun

DOI 10.1007/978-981-4560-75-7_57-1

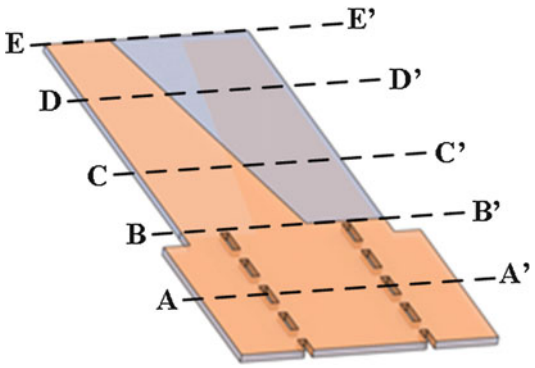


Fig. 26 SIW TSA

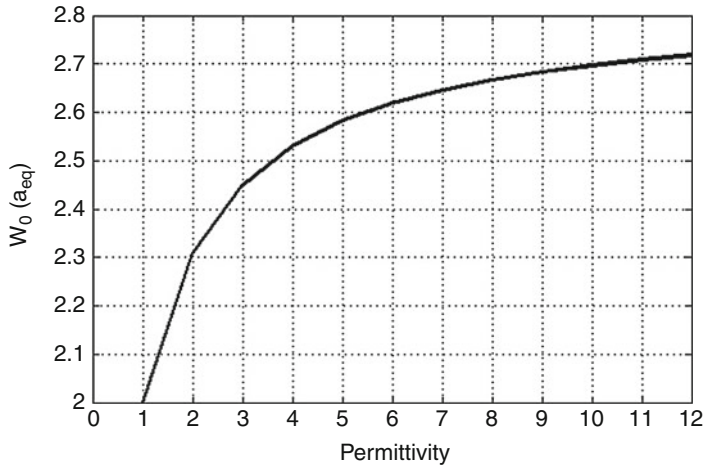


Fig. 27 SIW TSA end-fire slot width (w_0) versus the used material permittivity

can be removed. The performance of the antenna is determined both by its geometry and by the thickness and permittivity of the supporting dielectric substrate.

Slot-line field components are not confined to the substrate alone. They extend into the air regions above the slot and below the substrate also. Therefore, the energy is distributed between the substrate and the air regions. Consequently, the effective dielectric constant for a slot-line (ε_{re}) is less than the substrate permittivity ε_r. Galejs (1962) has shown that the zeroth-order value of ε_r for a slot on an infinitely thick substrate is the average dielectric constant of the two media, that is, the aperture can be calculated using the following equation

$$w_0 = \frac{c}{f_{low}\sqrt{(\varepsilon_r + 1)/2}} \tag{50}$$

To cover the entire monomode bandwidth, the lower frequency is equal to the SIW first mode cut-off, which gives

$$w_0 = \frac{2a_{eq}\sqrt{\varepsilon_r}}{\sqrt{(\varepsilon_r + 1)/2}} \tag{51}$$

DOI 10.1007/978-981-4560-75-7_57-1

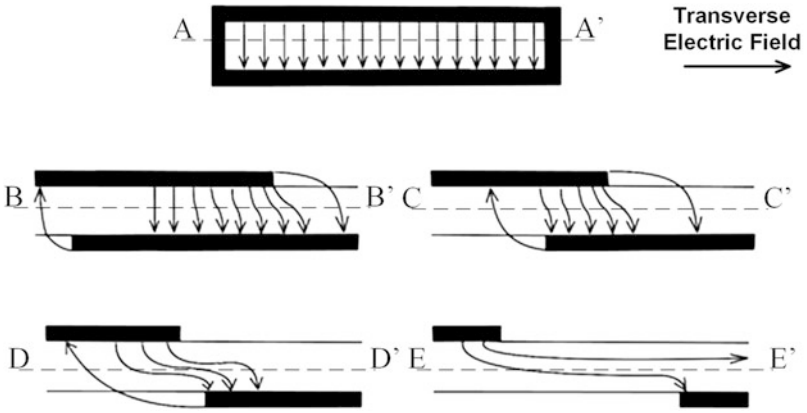


Fig. 28 Electric field at various cross sections along the transition from SIW waveguide to end-edge of the antenna

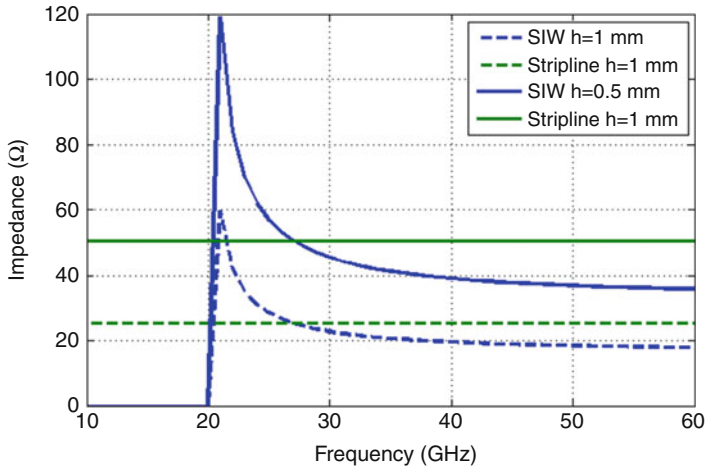


Fig. 29 SIW and stripline impedance

Figure 27 shows that it is possible to reduce the width of the antenna efficiently using dielectric with permittivity lower than 4.

Balun

The tapered antipodal finline smoothly transforms the incident TE_{10} (Fig. 26 cut A) waveguide mode to a quasi microstrip mode which is propagating in the area of overlapping antipodal fins of the transition (Fig. 26 cuts B and C). E field in the transition area concentrates and rotates by almost 90° until waveguide mode transforms to quasi slot mode (Fig. 26 cut D). The E field transformation process in different cut planes of the transition is shown in Fig. 28. Smooth transformation of the field mode allows the transition to operate in a wide frequency band. The region II (plane BB' to CC') act as balun and impedance transformer. In addition, region II transforms the impedance of the TE_{10} waveguide mode to high impedance of slot line component.

Figure 29 shows the SIW waveguide impedance according to the Eq. 21 and the impedance of the overlapping fins at Cut BB' considered as parallel plat line. When the thickness of SIW changes, the cut-off frequency remains the same and the impedance changes. The parallel plat impedance is higher than its SIW counterpart, when the overlapping section diminishes, the impedance increases ensuring a wideband operation. The limit of width of the overlapping section is the half of the SIW width..

DOI 10.1007/978-981-4560-75-7_57-1

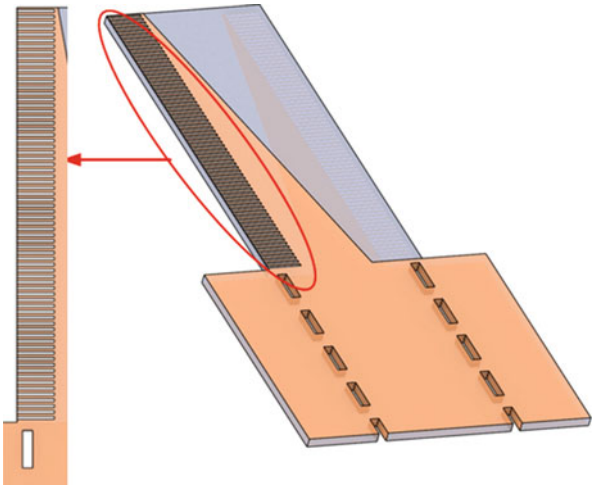


Fig. 30 TSA antenna with corrugation

Corrugation

TSA width w should be larger than 2λ. A degradation of radiation pattern has been observed for a narrower TSA width (Schaubert 1993). This degradation in the radiation pattern is a significant problem for the effective design of compact TSAs. Pattern improvement for a narrow width antenna can be achieved by using a corrugation structure, which makes this technique suitable for the formation of antenna arrays where a small spacing between antenna elements is needed. A Fermi TSA antenna with a corrugation was proposed in Sugawara et al. (1998) and in Sato et al. (2003). This technique was verified with SIW in Djerafi and Wu (2012b). It has been successfully demonstrated that the use of corrugations along the sides as shown in Fig. 30 can reduce the antenna width and improve the VSWR over a wide frequency range, as well as suppress the side lobe levels. The corrugation structure is used to alter the phase of currents flowing along the outer edge of the substrate, changing the orientation of the electric field at the edge of the antenna substrate.

To explain this improvement, simulated E-field distributions along the antipodal LTSA structures with and without corrugation are shown in Fig. 31. The E-field counterparts in the transverse plane over the radiation region are described. In the case without corrugation, the electrical field orientation at the edge of the antenna substrate is opposite to that in the antenna aperture. Some reference claims that the alternation of strips and slots behaves like a slow-wave structure that increases effective aperture size resulting in lower reflection of the radiated wave. As observed in Fig. 31 this phenomenon is not observed. The current distribution illustrated in Fig. 32 clearly shows that the energy is more confined in the central region and radiation from the side is limited.

The effective antenna aperture is thus increased by the presence of the corrugation, which ensures a plane wave phase front over the antenna aperture. In standard antenna, the antipodal nature of the antenna gives rise to very high levels of cross-polarization, particularly at high frequencies because the skew in the slot fields is close to the throat of the flare. As shown in the field distribution in the transverse view in Fig. 31, the field in the center region is more highlighted in the structure without corrugation. Out of this perturbed region, the field is parallel to the two sides of the substrate compared to the one in the antenna without corrugation which is arranged in an arc.

Physical dimension of the slot corrugation is constrained by the fabrication process. In fact, the smallest possible slot width, feasible under this process, is 236 um in this work. The slot lengths are selected to be around $\lambda_r/4$, where λ_r is the wavelength in the dielectric substrate. The SIW antipodal LTSA is optimized

DOI 10.1007/978-981-4560-75-7_57-1

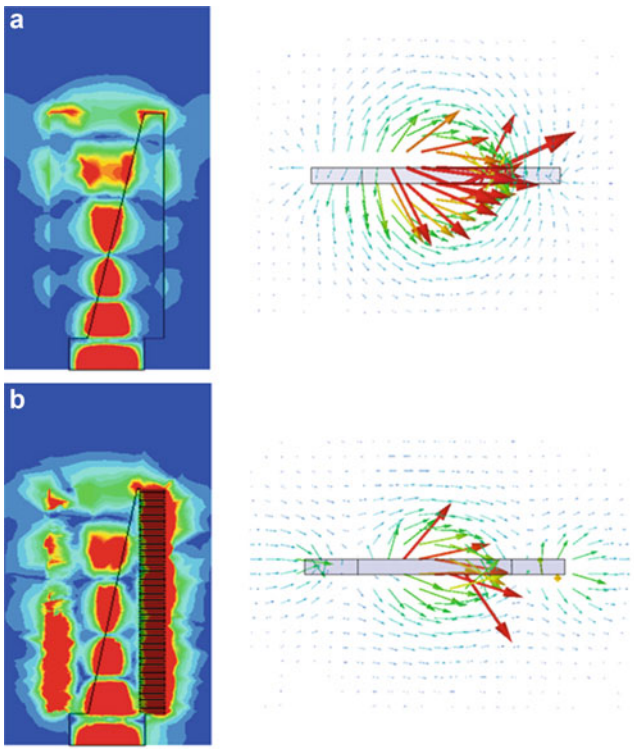


Fig. 31 Simulated E-field magnitude distributions obtained by HFSS at 28 GHz, along the antenna and in transverse cut (**a**) antipodal LTSA without corrugation, (**b**) antipodal LTSA with corrugation

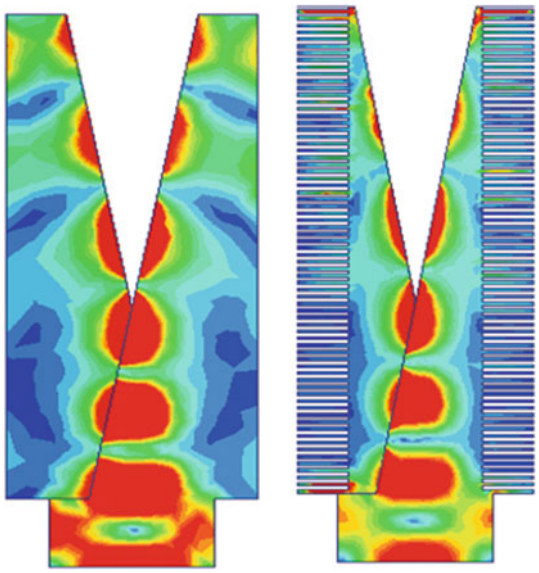


Fig. 32 Simulated current distribution comparison between corrugation and no-corrugation structures

around 28 GHz. Substrate RT/Duroid 6002 with thickness h of 508 um and dielectric constant of $\varepsilon_r = 2.94$ is used.

Figure 33 shows simulated pattern results of antennas with and without corrugation. Without corrugation, the E-plane pattern beamwidth is excessively large and the SLL in H-plane is 12 dB. Once the corrugation is used, a noticeable decreasing of 3 dB beamwidth can be observed in the E-plane, while the

DOI 10.1007/978-981-4560-75-7_57-1

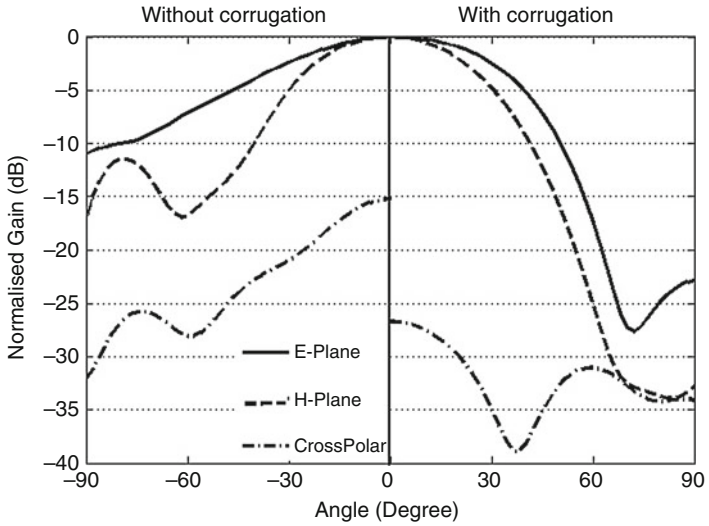


Fig. 33 Simulated radiation patterns with and without corrugation obtained by HFSS at 28GHz

Table 2 Typical properties of end fire traveling antenna

Type	Maximum gain	Low sidelobes
Directivity (dB)	$10\ \log\left(10\frac{L}{\lambda_0}\right)$	$10\ \log\left(4\frac{L}{\lambda_0}\right)$
Beamwidth	$55/\sqrt{\frac{L}{\lambda_0}}$	$77/\sqrt{\frac{L}{\lambda_0}}$

beamwidth in the H-plane increases. The sidelobe in the H-plane for this antenna is 23 and 33 dB in the E-plane. The cross-polarization level with corrugation is found to be better than 27 dB at the frequency of design and without corrugation is 15 dB.

The impact of the slot width is verified for its dimensional variation from 6 to 12 mil. A small effect on the SLL level is observed without gain degradation. For the slot depth, the variation affects the gain of the antenna by changing the effective aperture. It is possible to optimize the corrugation slot to achieve the same beamwidth in both E- and H- planes.

Maximal Gain Versus Length

Generally, directivity increases as the length, L, of a TSA is increased. For lengths between three and eight wavelengths, the increase is linear according to Formula in Table 2 defined in Yngvesson et al. (1985). In this table, the tapered dielectric rod antennas are classified by Zucker as low sidelobes traveling wave.

Figure 34 shows the studied linear and constant with TSA antennas. The CWTSA is a kind of tapered slot antenna for which the width of the radiating slot is constant over the antenna length. Since the slot line that feeds the antenna is generally narrow, there should be a tapering section between the feeding slot and the radiating slot in order to achieve a better impedance matching. If such a transition is not used, there could be a high return loss due to the abrupt change in the characteristic impedance.

Figures 35 and 36 show the gain variation with antenna length. In the LTSA version the gain increases linearly between 1 λ_0 and 3 λ_0 and saturates after 14 dBi. The real constraint with this version is the angle α (according to Fig. 25). The taper angle in the case of the LTSA version is: $5° \leq 2\alpha \leq 15°$. If the length increases, the antenna will be out of the acceptable angle range after a specific value. The linear dependence of the gain on L/λ_0 is an important characteristic of all end-fire traveling-wave antenna. For

DOI 10.1007/978-981-4560-75-7_57-1

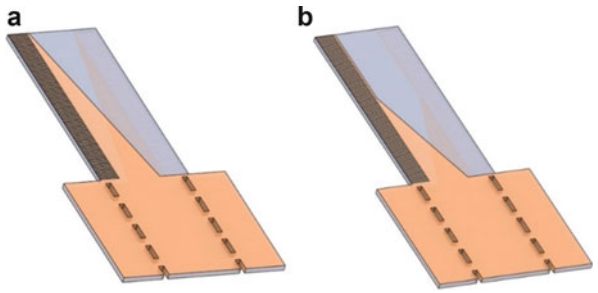


Fig. 34 (**a**) Linear taper, (**b**) CWTSA with linear taper

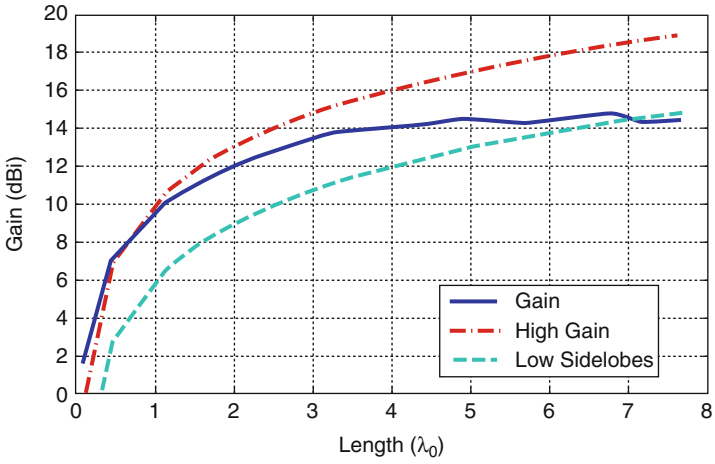


Fig. 35 Effect of antenna length variations on gain of LTSA

CWTSA, the typical numbers are gain of 12–18 dBi with length of the antenna L in the range of 3 λ_0 to 8 λ_0. For longer antennas, the coefficient multiplying L/λ_0 decreases somewhat.

Figures 37 and 38 show the effect of length variations on E- and H-plane patterns for LTSA and CWTSA. The CTWSA shows that the beam width and gain essentially following the high gain standard curve for length below 5 λ_0. In the TSA, the E-plane beam with curve has a very steep slope for length of 4–5 λ_0, exactly as standard curves. After 3 λ_0, the beamwidth in the E- and H-plane follow the low side lobe curve. While TSA elements have many features in common with the general class of traveling-wave antenna, there are some differences, for example, the beamwidth often follows $(1/\sqrt{L})$ while the E-plane varies more slowly with L. The E- and H- bandwidths follow the low side lobe type curve in the case of LTSA and high gain curve for the CWTSA version.

Substrate Thickness

There is a trade-off in substrate thickness; a too "thick" substrate launches surface modes, which degrade the antenna performance, whereas a too "thin" substrate yields higher cross-polarizations and less symmetrical beam patterns. The optimum substrate thickness h has experimentally been defined by Yngvesson et al. (1985) as a function of factor, t_{eff}:

$$t_{eff} = b(\sqrt{\varepsilon_r} - 1)/\lambda_0 \tag{52}$$

DOI 10.1007/978-981-4560-75-7_57-1

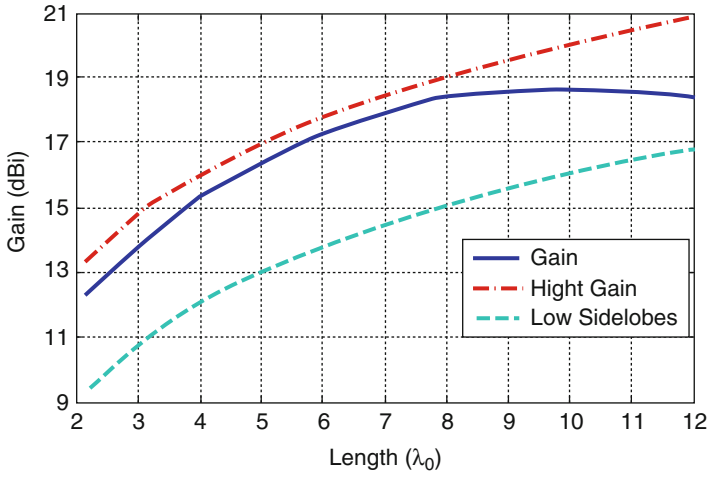


Fig. 36 Effect of antenna length variations on gain of CWTSA

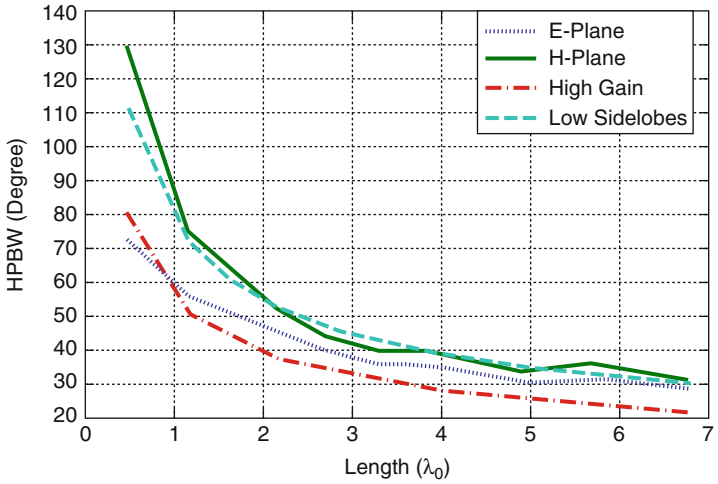


Fig. 37 Effect of antenna length variations on E- and H-plane patterns for LTSA

An acceptable range of t_{eff} for a good TSA opeacceptable range of t_{eff} for a good TSA operation is ration is determined to be between 0.005 and 0.03.

Considering the criteria of 3 dB gain bandwidth, the CWTSA gain is still in this bandwidth from t_{eff} of 0.015–0.053 as shown in Fig. 39. The width of the bandwidth is still in the same order of Yngvesson range (in this simulation ε_r is fixed and thickness modified). With a smaller length, the LTSA gain variation shows a larger t_{eff} range (Fig. 40). This condition is not applicable to antipodal SIW TSAs, which is less subject to substrate mode. LTSA is designed using substrate with 2.94 of relative permittivity and 0.508 mm of thickness, which is equivalent to $t_{eff} = 0.034$ at 28 GHz. Obviously, this value is out of the Yngvesson range. The designed antenna still works adequately in both simulation and measurement. To examine the possibility of using material out of the defined range, two different thicknesses are tested here: $t = 0.05$ mm and $t = 1.25$ mm. This is equivalent to t_{eff} of 0.0034 and 0.101, respectively. Figure 41 illustrates E-plane radiation patterns of the two selected structures. For the two substrate, SLL is still low and an acceptable gain is achieved. These results demonstrate the possibility to design the antipodal

DOI 10.1007/978-981-4560-75-7_57-1

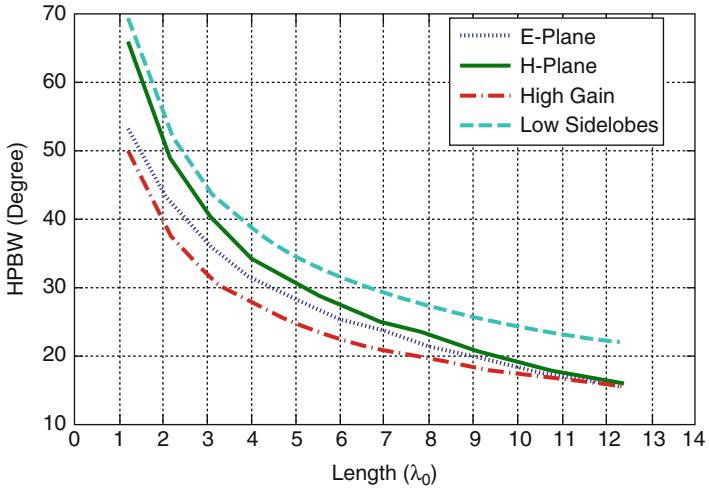


Fig. 38 Effect of antenna length variations on E- and H-plane pattern for CWSA

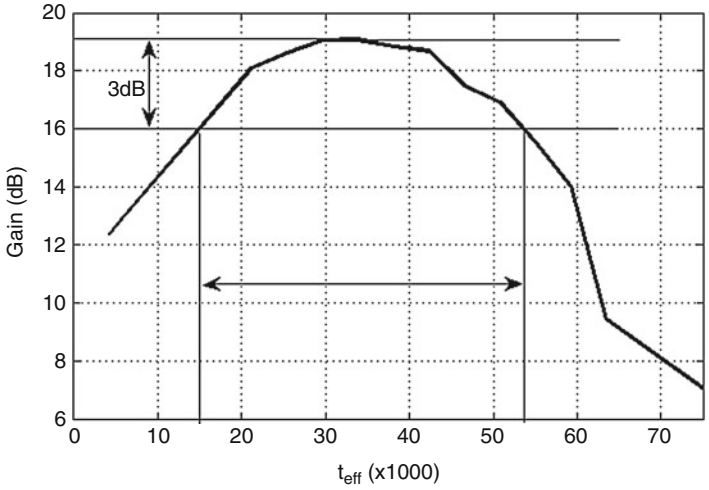


Fig. 39 Simulated gain obtained by HFSS for CWTSA at 28 GHz versus substrate thickness

corrugated TSA with materials out of Yngvesson's range. In this chapter, new criteria will be established for the SIW antipodal version.

Prototypes

Several TSA designs have been reported; as shown in Fig. 25. The tapered slot section which constitutes the radiating region can take on any of three geometric profiles:

- Nonlinear taper (Vivaldi, tangential or parabolic),
- Linear taper (LTSA),
- Constant width (CWTSA).

Vivaldi

The Vivaldi antenna has been studied since its first appearance. It undergoes many improvements in the last three decades, including introducing corrugated edges to form a better radiation pattern (Sugawara

DOI 10.1007/978-981-4560-75-7_57-1

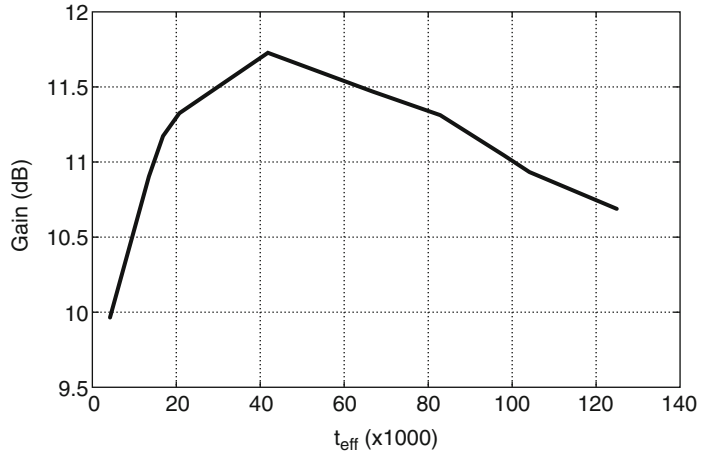


Fig. 40 Simulated gain obtained by HFSS for LTSA at 28 GHz as a function of substrate thickness

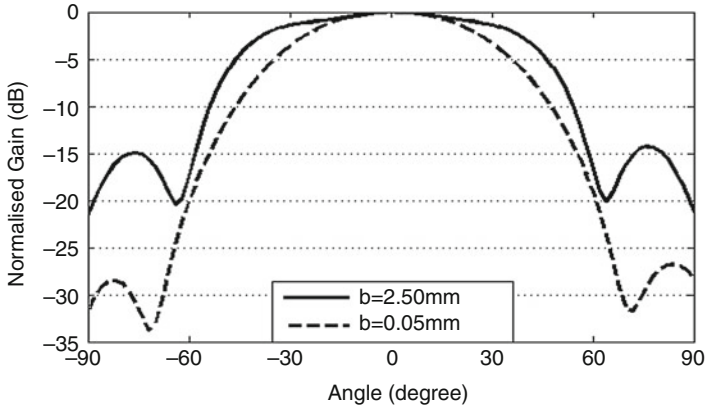


Fig. 41 Simulated radiation pattern obtained by HFSS at 28 GHz in connection with substrate thickness of 0.05 and 2.5 mm

et al. 1998), or changing its beam width by changing the shape, length, dielectric constant, and dielectric thickness. As one of the improvements, antipode Vivaldi antenna was proposed (Kim and Chang 2004).

The exponential taper profile is defined by the following equation:

$$y(x) = Ae^{Rx} + y_0 \tag{53}$$

where R specifies how the slot is widening, parameter A is determined by the length L and also the width w_0 of the taper and parameters y_0 is equal to L-L_{ant}. Very wide band performance can be acquired using an antipodal Vivaldi antenna with its easy wideband transition from microstrip line to radiation part of the antenna (Yang 2008).

Dual Exponentially Tapered Slot Antenna (DETSA), also known as Bunny-Ear Antenna (BEA), is an end-fire traveling wave antenna used for UWB applications. This antenna is a special form of Tapered Slot Antenna (TSA). It is created by tapering the outside edge of the slotline conductor of a Vivaldi antenna. The outer slot taper of the DETSA is described by an exponential function. This adds additional antenna design degrees of freedom.

DOI 10.1007/978-981-4560-75-7_57-1

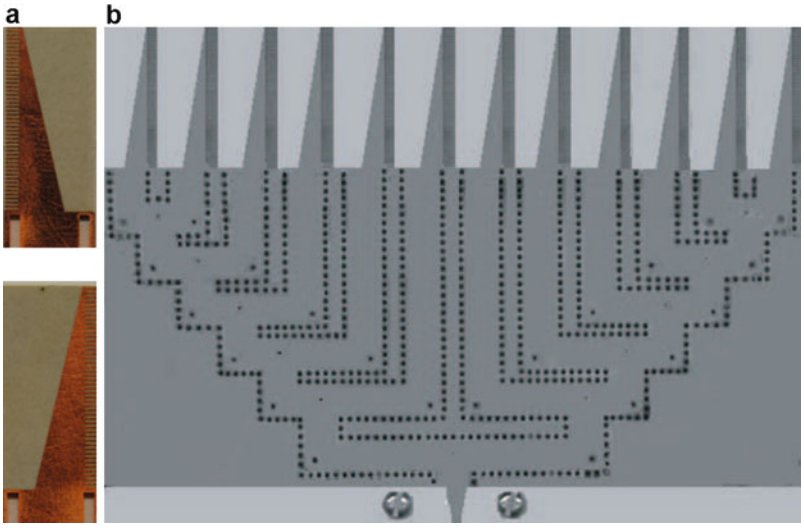


Fig. 42 (**a**) *Top* and *bottom* photograph of the fabricated SIW corrugated antipodal LTSA ($L = 25.6$ mm, $a_{eq} = 5$ mm, $w = 8.44$ mm), (**b**) example of array using LTSA

Fermi

The designed feed antennas are based on the TSA concept of Sugawara et al. with a Fermi taper. It provides symmetric radiation patterns and low side lobe level in both E-and H plane, The Fermi-TSA opens out the shortest while the Vivaldi opens out with the longest extension and the linear TSA is in the middle. The Fermi-TSA profile is defined by the following:

$$y(x) = \frac{w_0}{1 + e^{-R(x-b)}} \tag{54}$$

where R and b are variable. The antipodal version was suggested first in Gazit (1988).

The Parabolic

The inner side is shaped as a parabola

$$y(x) = Ax^2 + y_0 \tag{55}$$

The parabolic type opens out the shortest while the Vivaldi opens out with the longest extension and the Fermi-TSA linear is in the middle. Consequently, the beamwidth is narrowest for the PTSA and wide for the FTSA and widest for Vivaldi (Wang et al. 2009).

LTSA

The linearly tapered slot antennas (LTSA) have been reported for years (Simons et al. 1992; Yngvesson et al. 1985, 1989; Nesic 1991; Wu et al. 1998). Due to their salient features such as narrow beam width, high element gain, wide bandwidth, and small transverse spacing between elements in arrays, it attracts lots of interests in many applications. Practically, the performance of the LTSA is mainly determined by the feeding system. The angle of the aperture should be lower than 15°.

In this section, different types of the SIW TSA antenna are fabricated and tested.

DOI 10.1007/978-981-4560-75-7_57-1

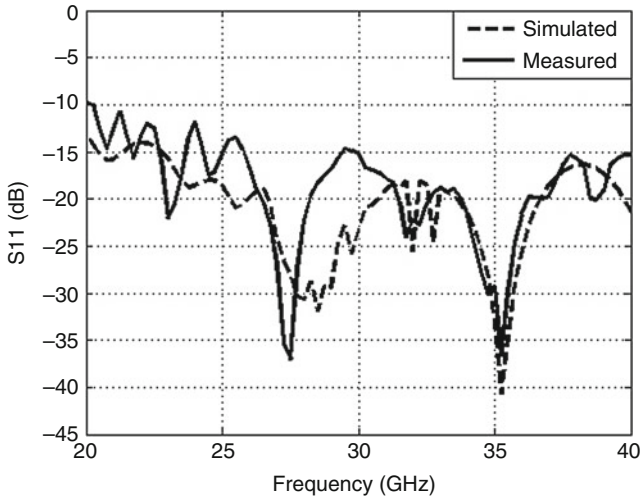


Fig. 43 Measured and simulated return losses

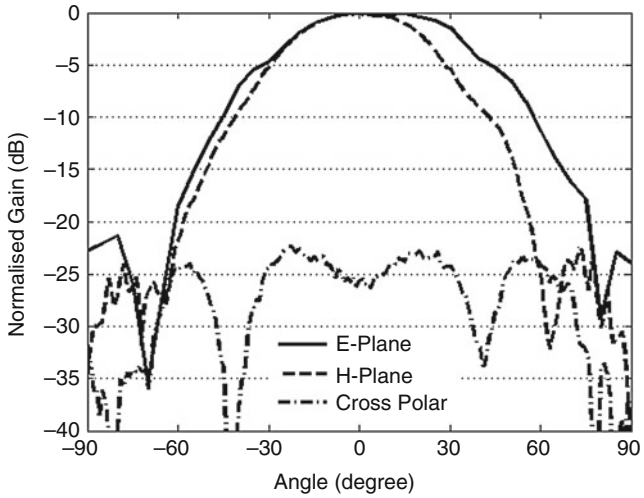


Fig. 44 Measured radiation patterns at 28 GHz

LTSA

SIW antipodal LTSA is optimized to work around 28 GHz for LMDS or future 5G systems. Substrate RT/Duroid 6002 with thickness b of 0.508 mm and dielectric constant of $\varepsilon_r = 2.94$ is used. Parameters of the optimized structure are shown in Fig. 42. Simulated and measured return losses of the proposed antenna are shown in Fig. 43. The return losses are greater than 15 and 10 dB in the simulation and measurement, respectively. Such an observed degradation is probably caused by the mismatch effect of transitions.

E-plane and H-plane patterns of the antenna as well as the cross-polarizations at 28GHz are measured and shown in Fig. 44. The achieved gain is 12.2 dBi against the simulated 11.9 dBi. This difference is within the calibration-related tolerance range of the antenna reference in used anechoic chamber. The 3 dB beamwidth is 59° in the E-plane and 48° in the H-plane and the first primary side lobe is located at −22 dB. It can be seen that the measured results are in good agreement with the simulated ones. The cross-polarization level shows a 4 dB degradation compared to the simulated counterpart.

DOI 10.1007/978-981-4560-75-7_57-1

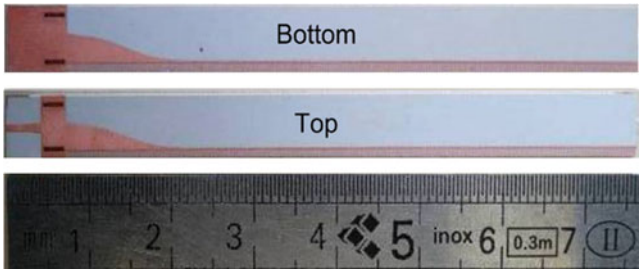


Fig. 45 *Top* and *bottom* photograph of the fabricated SIW corrugated antipodal Fermi-TSA ($L = 65.61$ mm, $w = 6.8$ mm, $a_{eq} = 3.8$ mm)

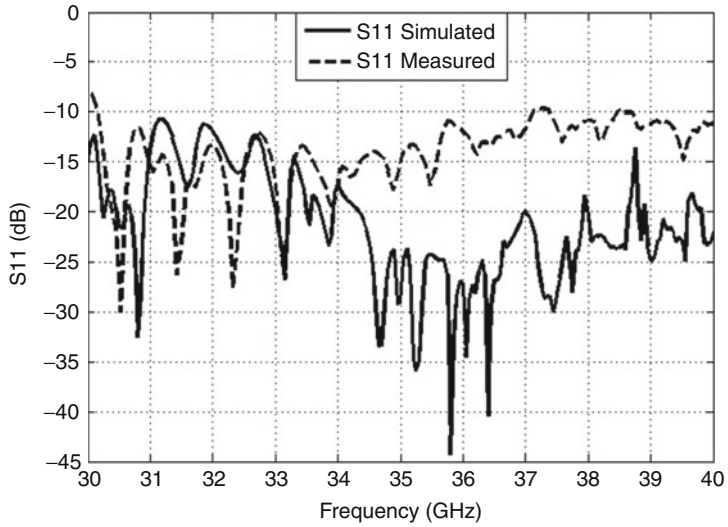


Fig. 46 Simulated and measured S-parameter characteristics of the SIW corrugated Fermi-TSA

Corrugated Fermi

A single corrugated antipodal Fermi-TSA has been designed to cover the band of interest around 35 GHz for imaging array applications. Substrate RT/Duroid 6002 with thickness of 0.508 mm and dielectric constant of $\varepsilon_r = 2.94$ is used. To realize the desired gain, antenna length L_1 has been increased to reach the value of 8 λ_0. The antenna width has been slightly modified due to the restrictions related to the design of antenna arrays where the width of the antenna is 0.79 λ_0. The photograph of the fabricated SIW corrugated Fermi-TSA (top and bottom) is shown in Fig. 45.

Simulated and measured reflection coefficients versus frequency are presented in Fig. 46. It can be noticed that the antenna bandwidth covers the entire considered frequency band of 28.57 % at 35 GHz for both measured and simulated results that are obtained by the use of HFSS.

Figure 47 presents simulated and measured copolar E-plane and H-plane radiation patterns. From these results, it can be seen that there is a good agreement between the measured and simulated radiation patterns. The 3 dB beam widths are approximately 25.33° and 28° for both E-plane and H-plane patterns, respectively. The side lobes are lower in the E-plane compared to the H-plane. At 35 GHz, the antenna gain is about 19.4 dBi based on simulations and almost 18.4 dBi is obtained in practice according to measured results. The cross-polarization level is found to be better than 20 dB thanks to the overlapped flaring metal in the SIW version. The excellent radiation properties and compact size of this antenna make it ideal as either a standalone antenna with a broad pattern or as an array element.

DOI 10.1007/978-981-4560-75-7_57-1

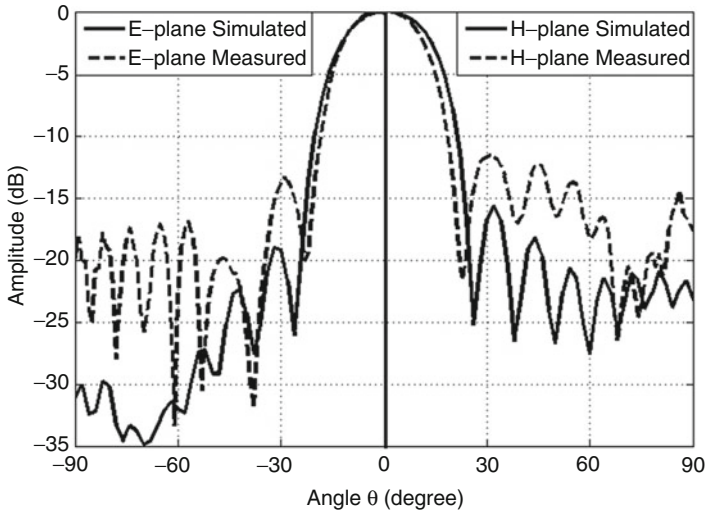


Fig. 47 E-plane and H-plane radiation pattern performances of the SIW corrugated Fermi-TSA at 35 GHz

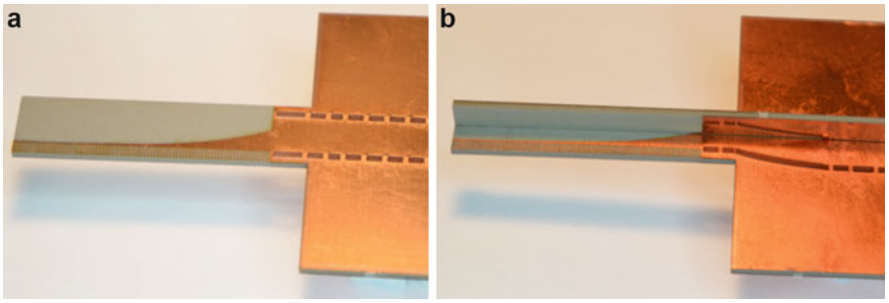


Fig. 48 Photography of the fabricated: (**a**) planar corrugated PTSA and (**b**) corrugated L-Folded PTSA

Corrugated SIW Planar and Folded Antipodal Parabolic Tapered Slot Antenna

A folded tapered slot antenna is presented in Doghri et al. (2013), this antenna represents a three dimensional version of Parabolic Tapered Slot Antenna. The PTSA is folded along its longitudinal center line which gives the possibility to have 45° and −45° polarizations.

Planar PTSA

To design an L-folded TSA, the first step is to design a planar TSA with respect to the required half power beamwidth, side lobe level, gain, etc. over all the desired operating frequency bandwidth. The geometrical parameters such as width, dielectric thickness, taper profile, length, and so on are optimized to meet requirements. Several TSA designs have been reported. For the demonstration purpose of the L-folded TSA PTSA (Parabolic Tapered Slot Antenna) is used in this design to obtain an optimal matching and bandwidth. The corrugation is also used for effectively reducing the width of the TSAs without degrading the antenna performances. The fabricated planar prototype is shown in Fig. 48, radiation and matching performances in Figs. 49 and 50 successively.

Folded SIW Line

An SIW line folded along its longitudinal center line is used to properly feed the L-folded TSA (LFSIW). Figure 51 illustrates a cross-sectorial view of SIW and LFSIW with the transverse electric and magnetic field lines of the fundamental TE_{10} mode. The junction acts as a tapered matching section between the two

DOI 10.1007/978-981-4560-75-7_57-1

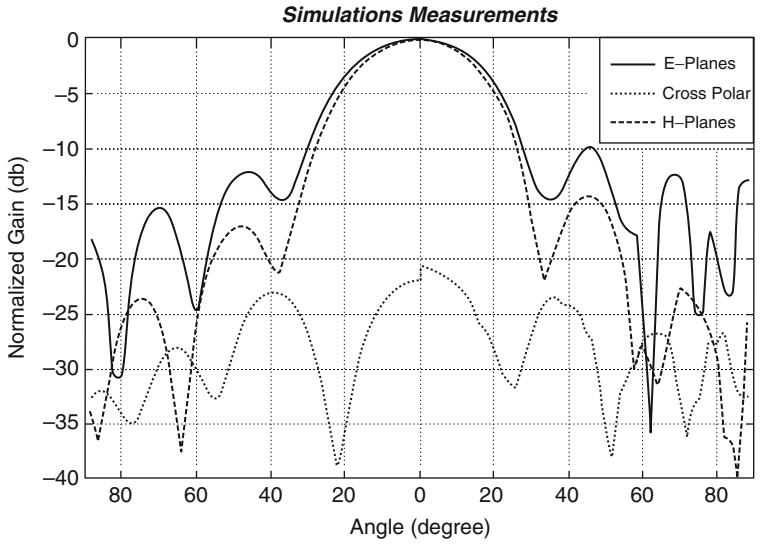


Fig. 49 Simulated and measured radiation patterns of the corrugated PTSA at 35 GHz

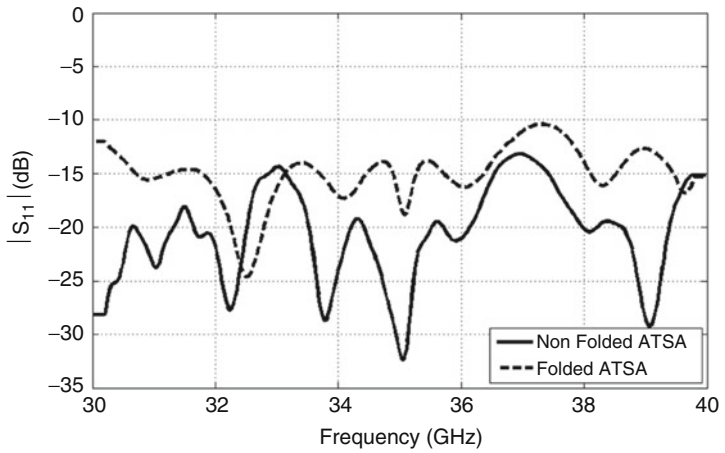


Fig. 50 Measured return loss of the corrugated APTSA and corrugated L-Folded APTSA at 35 GHz

waveguide types. The taper is optimized to achieve the minimum reflection coefficient. The optimized length of the L-Junction is about 1.41 λ_g at the center frequency. The back-to-back structure layout is illustrated in Fig. 52 with the taper shape. The back-to-back configuration is used to measure performances of the junction. The fabricated prototype is shown in Fig. 52b. The fabrication processes involve laser micromachining permitting arbitrarily shaped perforations, long and large rectangular vias with rounded corners that define SIW walls. After machining the structure, vias are electroplated. A back-to-back configuration is used to measure the performance of the junction because it was not possible to measure directly the L-shaped SIW line.

Figure 52a shows the measured reflection and transmission coefficients. The back-to-back junction achieves a return loss of better than 28 dB and an insertion loss of better than −0.9 dB over the 30–40 GHz bandwidth. The quality of the design is demonstrated by the E-field distribution inset in Fig. 52.

DOI 10.1007/978-981-4560-75-7_57-1

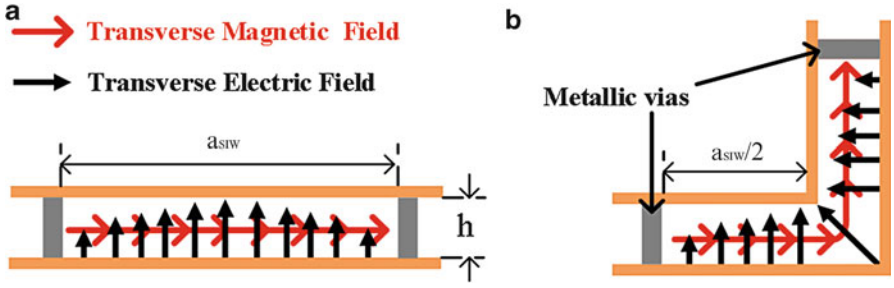


Fig. 51 Geometry of (**a**) SIW and (**b**) LFSIW

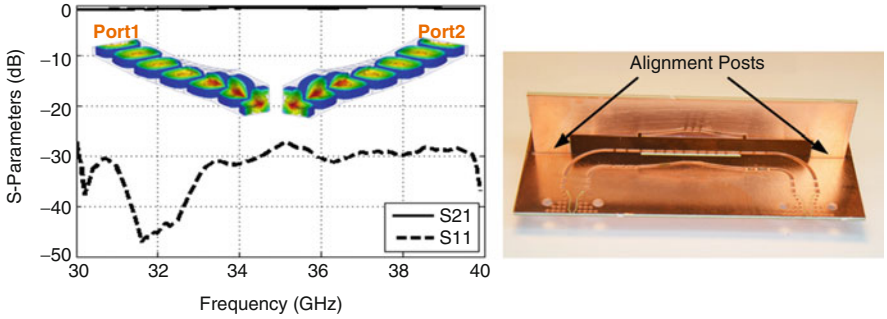


Fig. 52 (**a**) Back-to-back L-junction with measured results and simulated E-field magnitude distributions obtained by HFSS at 35 GHz, along the junction. (**b**) Assembled back-to-back L-junction

Folded PTSA

The third design step consists in folding the planar TSA. The previously introduced L-junction is used to feed the L folded PTSA. The main geometrical parameters are not altered by the folding step. The fabrication of this antenna involves two PCBs that are illustrated with dimensions in Fig. 53. The fabricated prototype is shown in Fig. 48b. Two posts are used for alignment purposes. When assembled, the two corners are metalized for a proper electrical connection between the two PCBs.

Similar performances are observed when the proposed L-folded antenna is compared with the nonfolded counterpart. Figure 54 shows the measured and simulated radiation patterns in the E and H-planes, corresponding to the 45° and −45° planes. The measured gain is 11 dBi while the simulated gain is 13 dBi. The 3 dB beamwidth is 38° in the E and H-planes. The cross polarization is 10.7 dB in simulation and 8.1 dB in measurement. The folded version has a better symmetry in the E and H-planes. However, the gain and the cross polarization are reduced.

Coupling

Mutual coupling between elements is an important issue in designing antenna arrays. It modifies impedance, radiation pattern, beamwidth, and directivity of an array, and even degrades the performance. Therefore, it is a parameter that must be taken into account during the design process. The previously described antenna topology has been studied.

Using HFSS simulations to evaluate the isolation levels between two adjacent elements (Fig. 55). The isolation is about 13 dB between the elements aligned along their *E*-plane and separated by 0.68 λ_0. The dielectric is subtracted in the central region and the isolation between the two elements is 3 dB lower. To explain the coupling effect and a possible improvement, simulated E-field distributions are shown in Fig. 56 along the antipodal Fermi-TSA structures with and without the substrate removal. The effective antenna aperture is thus increased by such a removal, which ensures a plane wave phase front over the

DOI 10.1007/978-981-4560-75-7_57-1

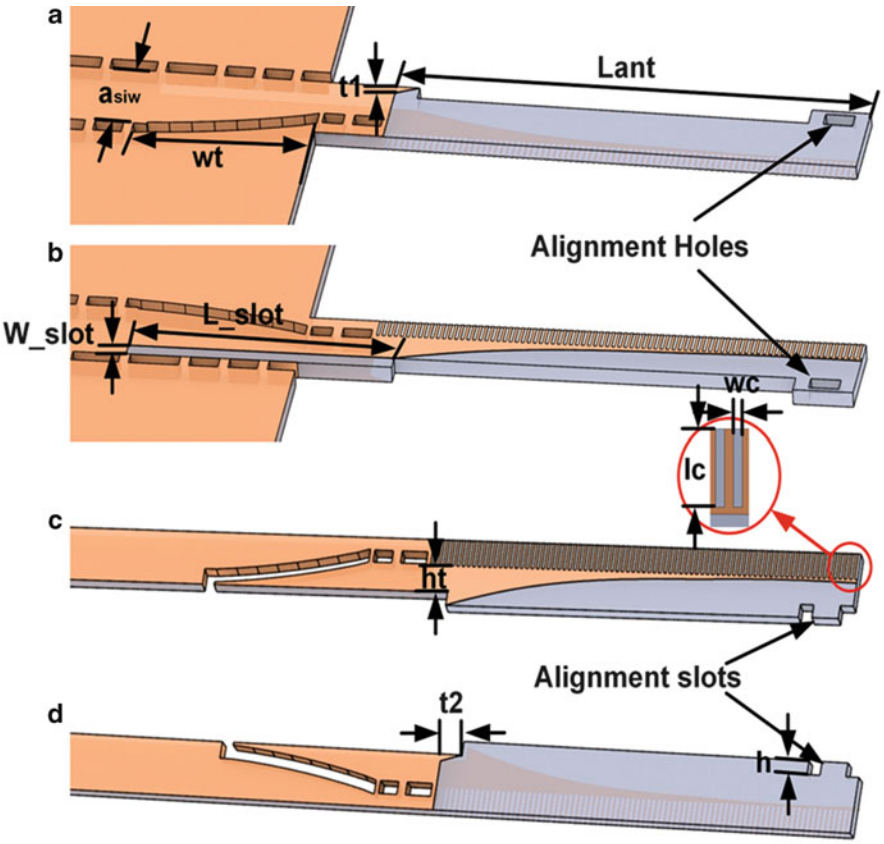


Fig. 53 Geometry of the corrugated folded TSA with Lant = 24.5 mm, t1 = 0.76 mm, a_{SIW} = 4 mm, L_slot = 13.4 mm, W_slot = 0.76 mm, wt = 9.3 mm, ht = 1.6 mm, t2 = 1.2 mm, lc = 1.23 mm, wc = 0.13 mm, h = 0.76 mm. (**a**) *top* and (**b**) *bottom* view of main PCB, (**c**) *top* and (**d**) *bottom* view of inserted PCB

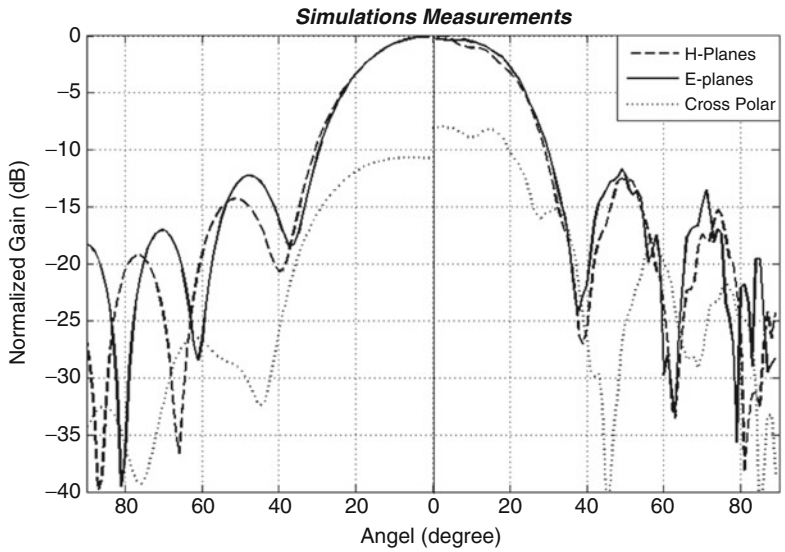


Fig. 54 Simulated and measured radiation patterns of the corrugated L-folded PTSA at 35 GHz

DOI 10.1007/978-981-4560-75-7_57-1

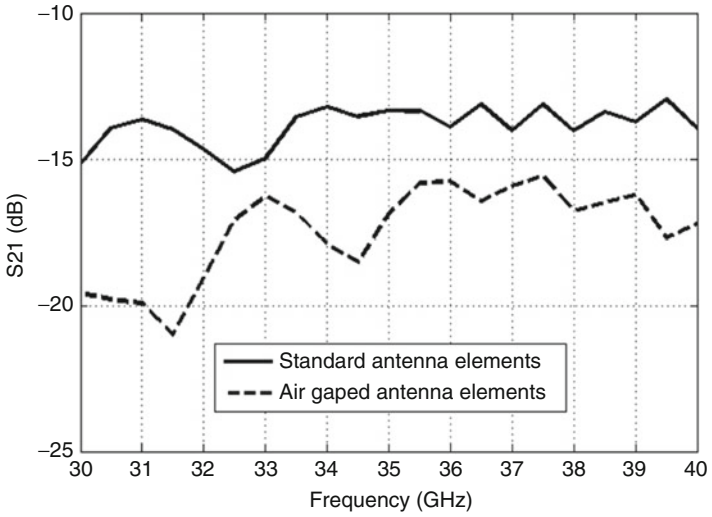


Fig. 55 Simulated isolation between the antennas

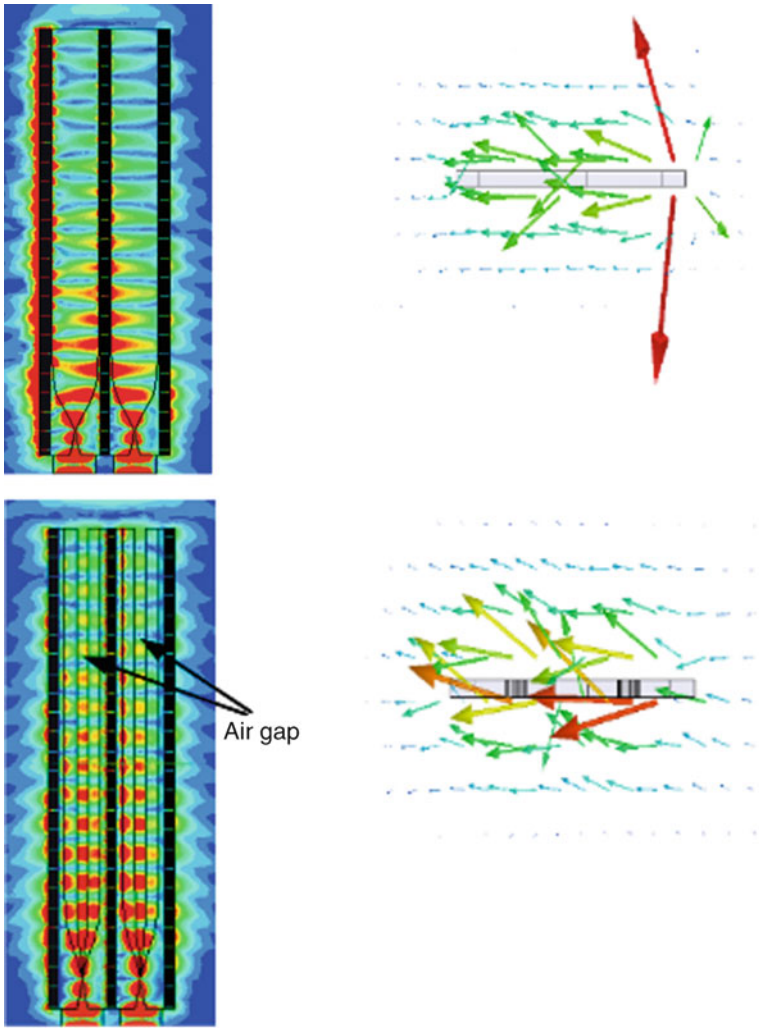


Fig. 56 Simulated mutual coupling for standard two element Fermi-TSA and for air gaped two element Fermi-TSA

DOI 10.1007/978-981-4560-75-7_57-1

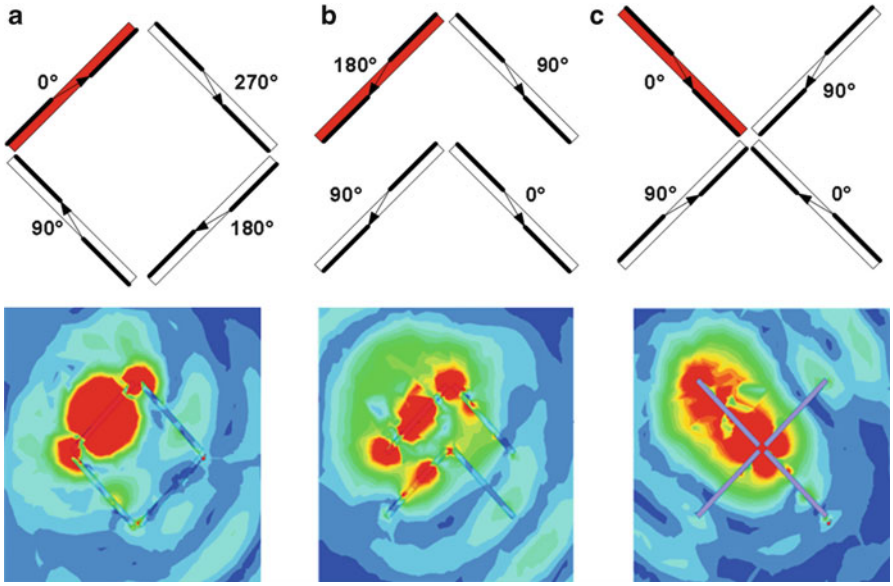


Fig. 57 Simulated E-field magnitude distributions obtained by HFSS in transverse cut (**a**) case1: square configuration, (**b**) case2: V configuration; (**c**) case3: X configuration

antenna aperture. As shown in the transverse view of the field distribution in Fig. 56, the field in the center region is more highlighted in the structure without the subtraction. In the modified antenna, the field is more spread along the transverse section. In fact, the air gap creates discontinuities and the field is confined along the air gap. When used in the two element array, the air slotted version gives more than 1 dB of improvement. The gain of one simulated Fermi-TSA element in the presence of a second passive element is 2 dB lower than that of an isolated element. In the H-plane, as the field is confined around the material, the coupling is very small between two elements disposed in the parallel arrangement.

Circular polarization can be ensured in wide bandwidth with high gain performances based on LTSA and sequential rotation. This element can then be integrated into a subarray cell to determine the optimal geometry. The four elements are physically rotated relative to each other and the feed phase is individually adjusted for each arrangement as shown in Fig. 57. In the figure, the field distribution in the transverse view when just one port is excited is illustrated. The mutual coupling effect changes with configuration, the first cell shows a better gain performance. The coupling is very high between two elements disposed in the parallel arrangement as in V configuration and directly affect the center of the parallel element. The total gain is in the same level of one element. The 3rd case is the complement of the first one to build a larger antenna array. The field distribution in the three cases gives some idea about the expected asymmetrical radiation pattern.

Application to Imaging

Passive millimeter-wave radiometer for imaging has found applications primarily in the security and defense industries. It is desired because of its ability to image through typical obscurants such as smoke, dust, and materials like plastic and textile, in certain frequency ranges. Certain frequency windows present low atmospheric losses (centered at 35, 94, 135, and 220 GHz). Many designers typically build passive sensors to operate within these windows in order to exploit low atmospheric attenuation and increased image contrast compared to lower frequencies.

DOI 10.1007/978-981-4560-75-7_57-1

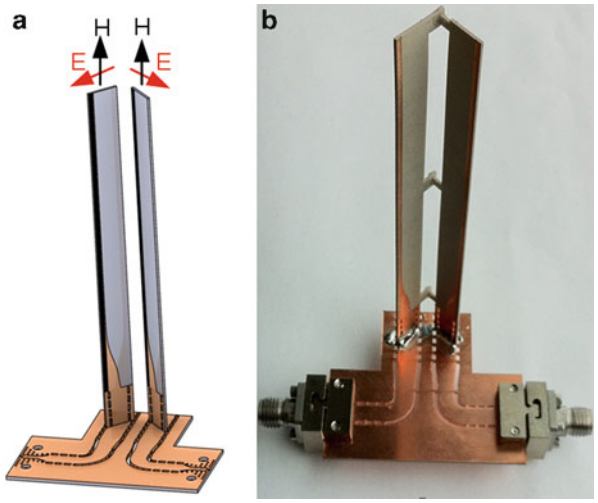


Fig. 58 (**a**) Topology of the SIW dual polarization system, (**b**) photograph of the fabricated SIW dual polarization system

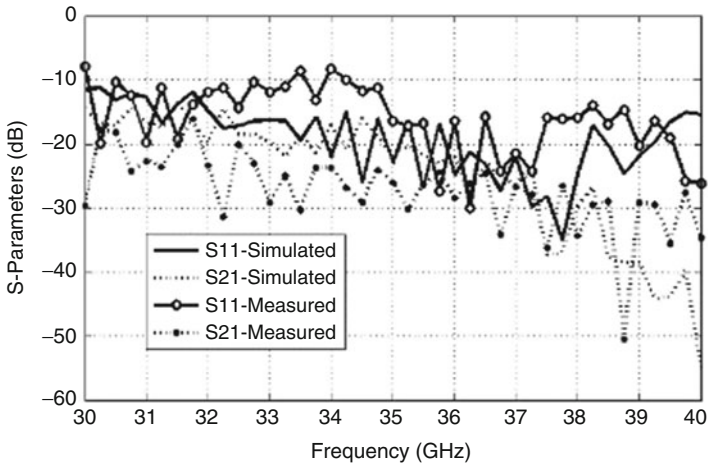


Fig. 59 S-parameter characteristics of the prototyped antennas

The antenna architecture for imaging application should satisfy defined conditions in term of high gain and low side lobe level (SLL). In addition, low cost and light weight characteristics are always required for highly integrated, reliable and portable systems. In fact those systems are generally duplicated to achieve a better resolution, which increase the weight and total volume.

Thanks to the intrinsic shielding of the proposed SIW feeding, a compact feed can be obtained without resorting to any coupling in the feeding network that would deteriorate isolation and therefore increase cross-polarization. Different antennas based on TSA element with orthogonal polarization are presented in the following. An array of 128 elements shows low SLL and height gain with just 200 g of the total weight is used to build image.

Dual-Polarized end-Fire Antenna

Progressive plane wave can be decomposed into two linearly polarized waves formed orthogonal vectors; which are also orthogonal to the direction of propagation. In the case of the radiometry these orthogonal components are referred to as the vertical and horizontal polarization. The natural radiation emitted by the human is unpolarised, but after interaction with certain objects around it becomes polarized. One of the

DOI 10.1007/978-981-4560-75-7_57-1

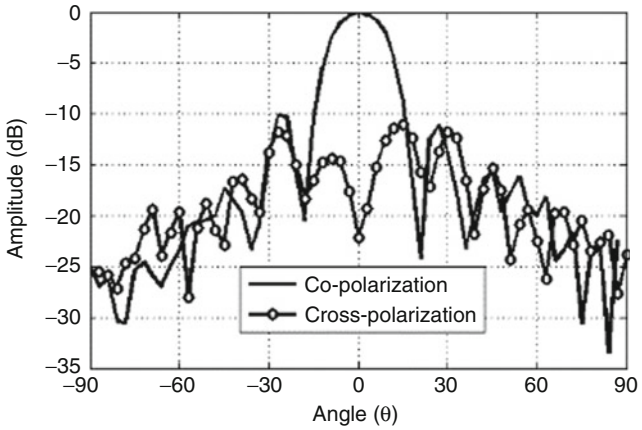


Fig. 60 Radiation pattern performances at 35 GHz

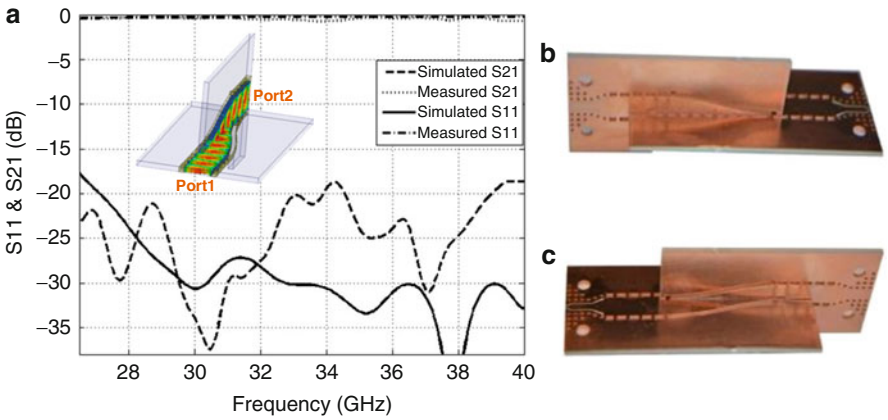


Fig. 61 (**a**) Measured and simulated S-parameters of the proposed twist, with E-field magnitude distribution simulated at 35 GHz, (**b**, **c**) shows the fabricated SIW 90° twist *front* and *back* view respectively

challenging issues in utilizing such full polarimetry is the need of a good discrimination between the H- and V- polarization, which finally leads to an accurate imaging synthesis. Therefore, achieving good isolation between the H- and V- polarization port, along with the important antenna parameters including antenna gain, impedance matching, co- and cross- polarization discrimination, etc., is crucial. Three different configurations for building dual polarization TSA are presented in this section.

A Dual Polarization Fermi Tapered Slot Antenna Setup Using E-Plane Corner

Compared to other types of tapered slot antennas, the corrugated Fermi-ATSA provides almost the same E-plane and H-plane patterns, presenting at the same time low side lobe levels. Two 45° and −45° vertical arm rotated E-plane bends are arranged to feed two Fermi antennas in posture of 90°, as illustrated in Fig. 58a. The photograph of the fabricated prototype is shown in Fig. 58b.

Measured and simulated S-parameters of the dual polarization setup are shown in Fig. 59 (El Khatib et al. 2012a). The isolation between the two ports is almost better than 20 dB in simulation over the entire 32–40 GHz band and in the measured one. The measured performance matches well with the simulation. Figure 60 presents measured copolar (E-plane) radiation patterns of the design at the 0°, 90° planes angle, the two ports are fed in phase simultaneously. The E field amplitude in the 0° plane is the geometrical

DOI 10.1007/978-981-4560-75-7_57-1

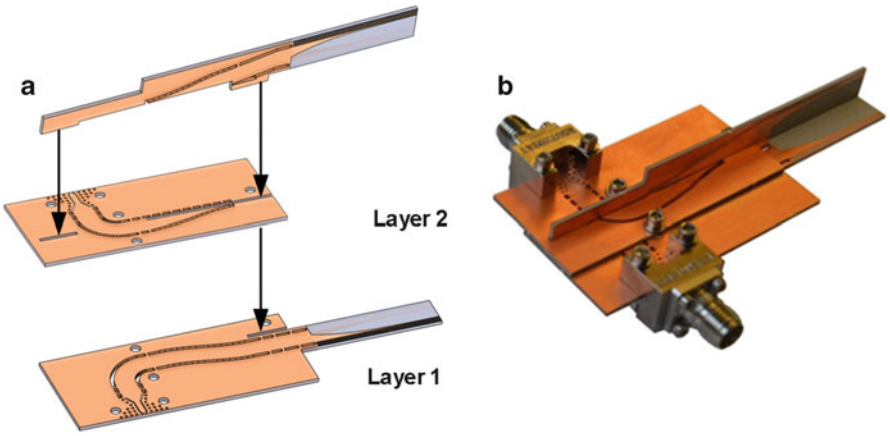


Fig. 62 SIW dual-polarized end-fire antenna (**a**) assembling scheme (**b**) fabricated prototype

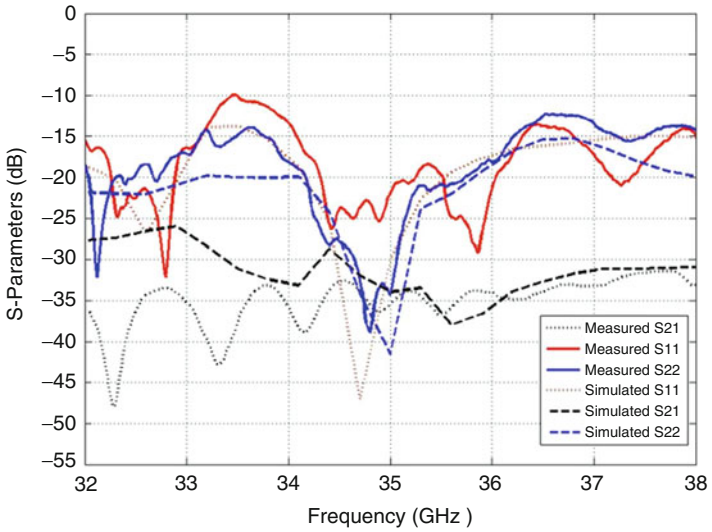


Fig. 63 Measured and simulated return losses and isolations

addition of the E field in the two planes. In counterpart, the 90° plane is the subtraction of the two. The radiation curves show some little dissymmetry due to the presence of the other element. Difference between the E field at 0° and 90° plane is 22 dB in the broadside direction.

A Dual Polarization Fermi Tapered Slot Antenna Setup Using Twist

In the second configuration, two PTSAs are disposed perpendicular to each other which give us the vertical and horizontal polarization. To bring the signal of the vertical antenna to the horizontal an SIW 90° twist is used (Doghri et al. 2013).

The SIW 90° twist consists of a progressive deformation of an L-Shaped SIW line shown in Fig. 61. The width is progressively decreasing in one plane and increasing in the other. The signal from the horizontal plane will move to the vertical one and vice versa. To validate the proposed device, a prototype operating over the *Ka*-band (26.5–40 GHz) was designed based on RT/Duroid 6002 substrate with dielectric constant of $\varepsilon_r = 2.94$ and substrate thickness of 0.76 mm. The fabricated prototype involving two LEGO-like interconnected PCBs is shown in Fig. 61b, c. Figure 61a illustrates the measured

DOI 10.1007/978-981-4560-75-7_57-1

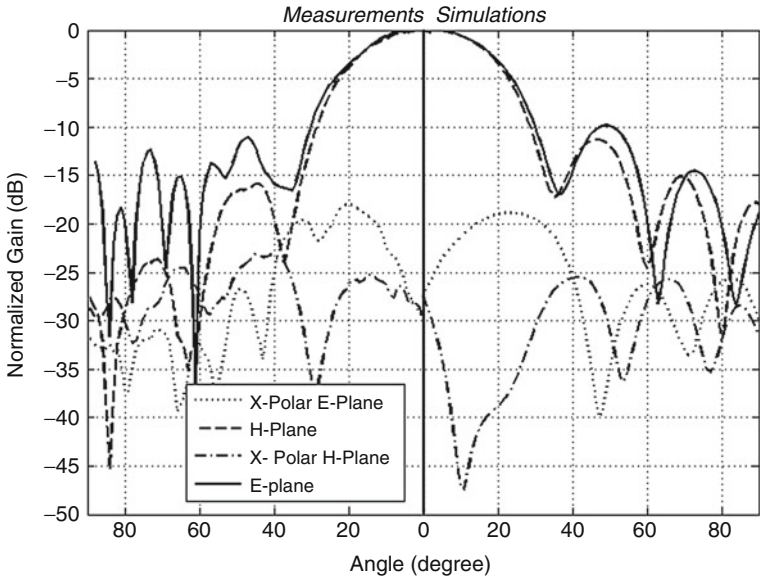


Fig. 64 Simulated and measured radiation patterns corresponding to the vertically polarized antenna (fed by the twist) at 35 GHz

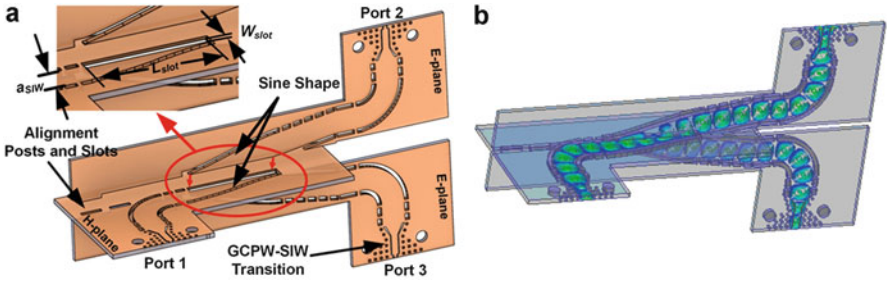


Fig. 65 (**a**) Geometry of the E-plane junction with $a_{SIW} = 4$ mm, $W_{slot} = 0.762$ mm and $L_{slot} = 15.2$ mm, (**b**) E-field magnitude distribution simulated at 35 GHz

performance of the proposed twist, a reflection coefficient of less than −18 dB and an insertion loss better than 1 dB over the *Ka*-band is achieved.

Two high gain and broadband corrugated PTSAs are used as radiating elements. Some slots and holes are used for alignment purposes. The fabricated prototype is shown in Fig. 62, two K connectors are added to feed the access. The measured and simulated return losses and isolations are shows in Fig. 63. The structure achieves a return loss better than −10 dB and an isolation higher than 32 dB from 32 to 38 GHz (22 % bandwidth centered at 35 GHz). Due to the shielding of the structure, the interference in the feeder is minimized to a negligible level.

Simulated and measured radiation characteristics of this dual-polarized antenna when one port is excited while the other port is matched are shown in Fig. 64. The vertically polarized antenna fed by the twist achieves a gain of 14.44 dBi at 35 GHz while the horizontally polarized antenna achieves a gain of 14.28 dBi. The two patterns have a cross-polarization of better than 27 dB in the 0° direction. A small asymmetry is observed in the co- and cross-polarization patterns because of the presence of the other elements. This solution shows a better isolation than the first solution with corner.

DOI 10.1007/978-981-4560-75-7_57-1

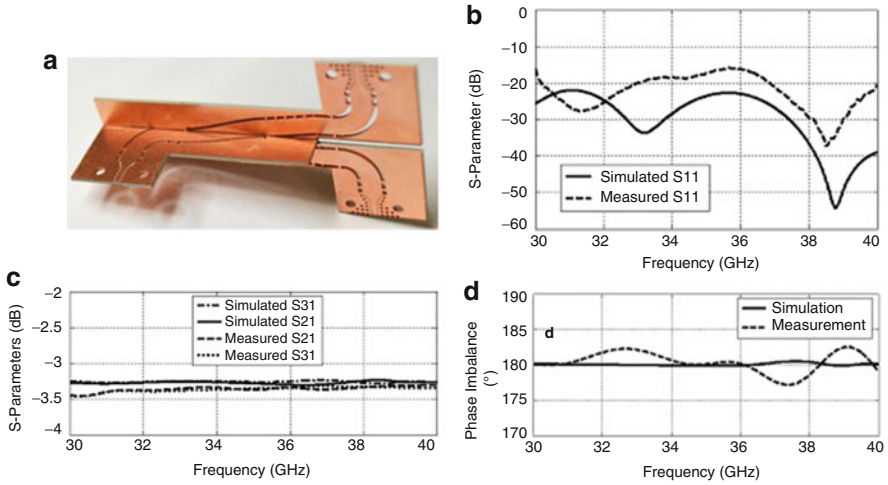


Fig. 66 Measured and simulated: (**a**) photograph of the fabricated prototype, (**b**) input return losses, (**c**) transmission coefficients (**d**) output phase differences

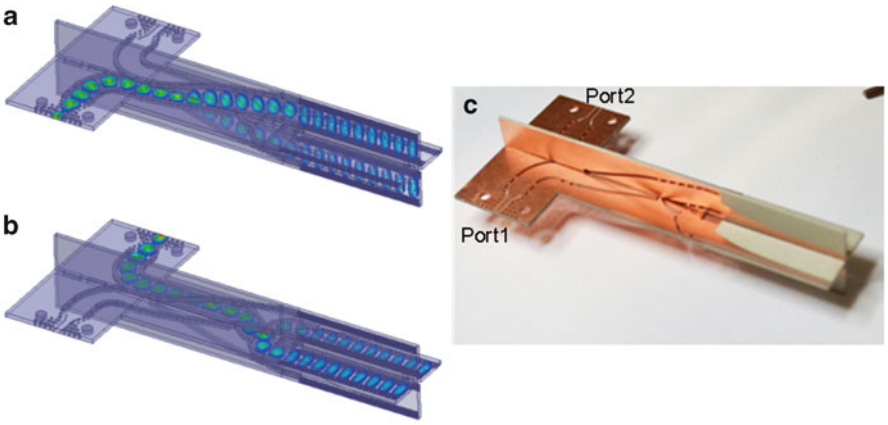


Fig. 67 (**a**) and (**b**) E-field magnitude distribution when feeding the horizontal and vertical antennas, (**c**) fabricated prototype

Four Elements Dual-Polarized end-Fire Antenna

Compared to two-element configuration, four elements will give a better symmetry between the two vertical and horizontal planes. The horizontal polarization is obtained thanks to the use of 3-port E-plane junction. Figure 65 illustrates the proposed E-plane junction. The proposed E-plane junction consists of a progressive deformation of a T-shaped junction to rectangular waveguide with TE_{20}. The resulting mode is then divided in two arms to feed two TE_{10} mode SIW lines with equal amplitude and 180 phase difference.

As depicted in Fig. 65, the E-plane junction divides and rotates the mode signal of the input SIW line in H-plane to the mode signal of the two output SIW lines in E-plane. An experimental prototype operating over the frequency range of 30–40 GHz was designed using RT/Duroid 6002 substrate from Rogers Corporation, having substrate thickness of 0.76 mm. The equivalent synthesized bilateral walls of the SIW are rectangular metalized slot arrays with width 0.762 mm (the minimum is defined by the metallization process, in presented case 0.254 mm).

The photograph of the fabricated prototype is shown in Fig. 66a. The prototype was measured using a vector network analyzer and a Line-Reflect-Line (LRL) calibration to de-embed the connectors, transitions and 90 bend effects. As shown in Fig. 66b, c, return loss and insertion loss of, respectively, less than

DOI 10.1007/978-981-4560-75-7_57-1

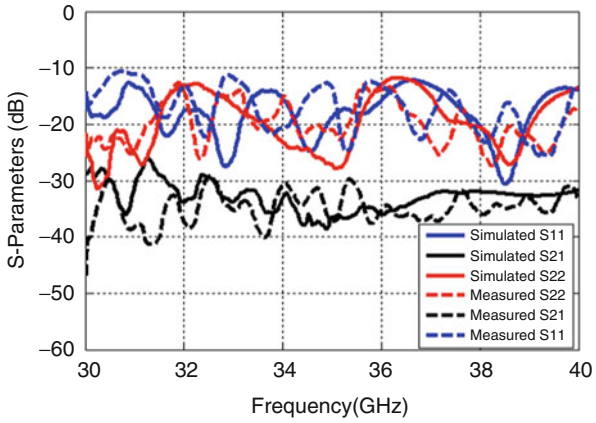


Fig. 68 Measured and simulated return losses and isolations

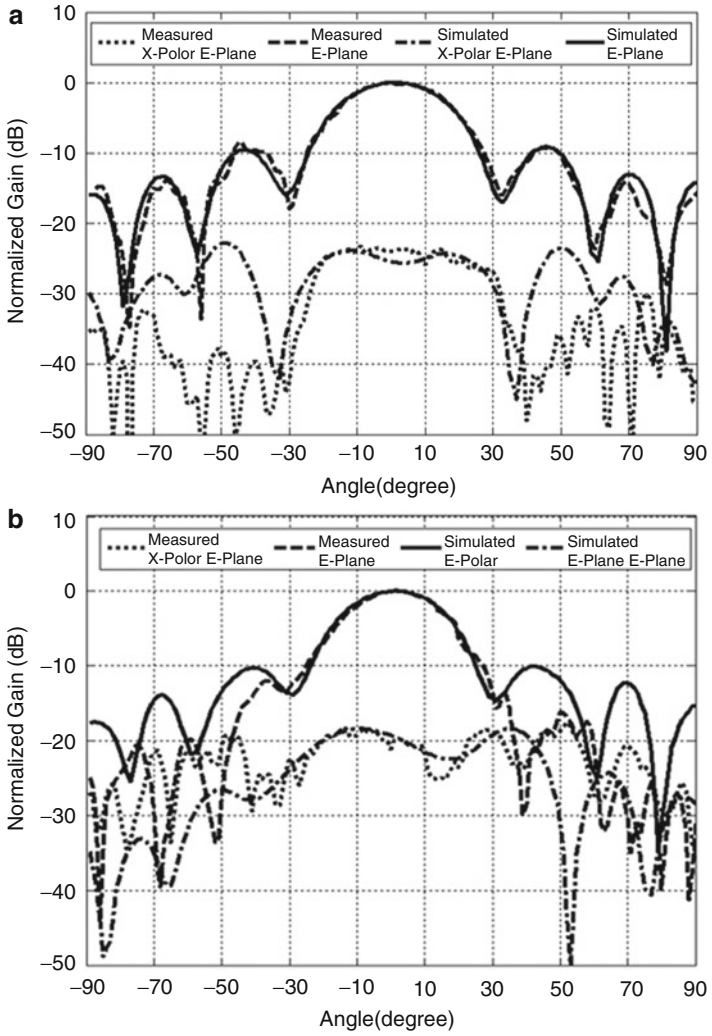


Fig. 69 35 GHz simulated and measured E-plane radiation patterns (**a**) the horizontally polarized array and (**b**) vertically polarized array

DOI 10.1007/978-981-4560-75-7_57-1

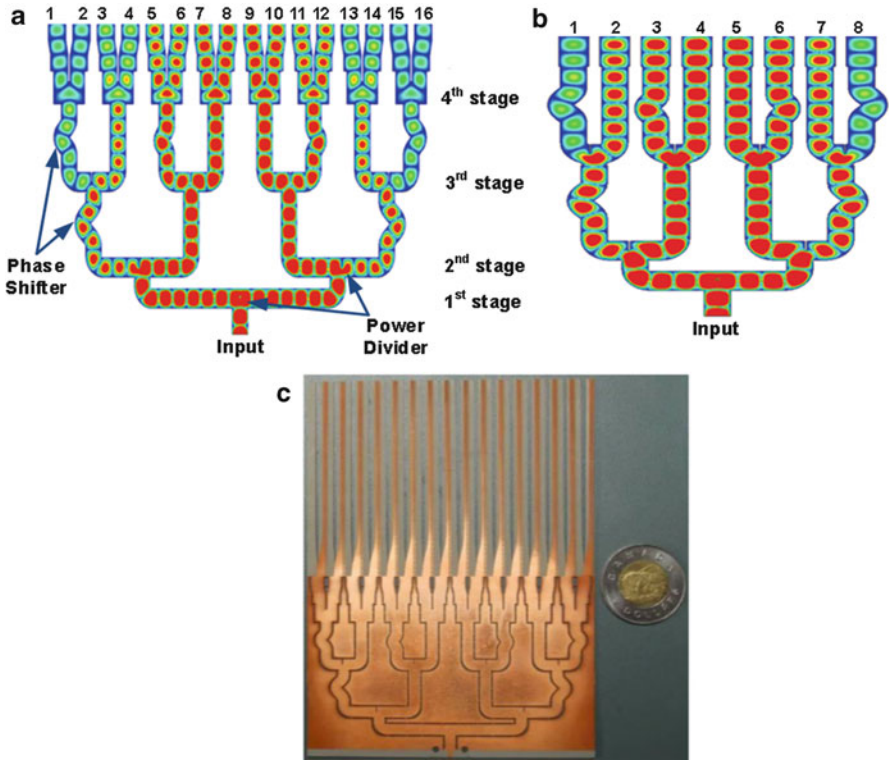


Fig. 70 (**a**) and (**b**) Simulated E-field magnitude distribution obtained by HFSS at 35 GHz, along the unequally distributed 1/16 and 1/8 principal power dividers, (**c**) fabricated 1/16 power dividers

16 and 0.46 dB, are observed from 30 to 40 GHz. In addition, as shown in Fig. 66d, the E-plane junction provides a 180 phase difference with phase imbalance of less than $\pm 3^\circ$ over all the frequency range.

The proposed junction is used to feed four PTSA elements. The prototype shown in Fig. 67 involves three Lego-style PCB blocks accurately aligned using posts and slots.

Figure 68 shows the measured and simulated return losses and isolations. In this case, ports 1 and 2 correspond to the vertical and horizontal ports, respectively. The structure achieves a return loss better than 10.6 dB and isolation better than −26 dB from 30 to 40 GHz (28 % bandwidth). Figure 69 compares simulated and measured radiation patterns of the horizontally and vertically arranged arrays, respectively. The horizontally polarized array achieves a gain of 14.6 dBi at 35 GHz while the vertically polarized array fed by the E-plane junction achieves a gain of 14 dBi. Thanks to the symmetry of the array arrangement, a good symmetry of the radiation patterns is observed.

Fermi Tapered Slot Antenna Array for Imaging Applications

A Villeneuve distribution is used for sidelobe control in the proposed 8 by 16 array (El Khatib et al. 2012b). Figure 70a shows the 1/16 power divider, the 4-stage H-plane SIW power divider makes use of an equal and unequal-split T-junction power divider to achieve the required aperture distribution. H-plane bends are used to compensate different delays generated by the imbalanced power splitting. The equal power division in the fourth stage of the power divider represents a compromise that is needed due to the tight dimensions that are dictated by the element spacing. All transmission coefficients fluctuate around the ideal coupling factor in the band of 30–40 GHz. Similar to the previous 1/16 power divider, an 8-way T-junction power divider is designed in order to accommodate the feeding for eight samples of the above mentioned 1×16 linear array, thus forming the proposed integrated 8×16 planar antenna array.

DOI 10.1007/978-981-4560-75-7_57-1

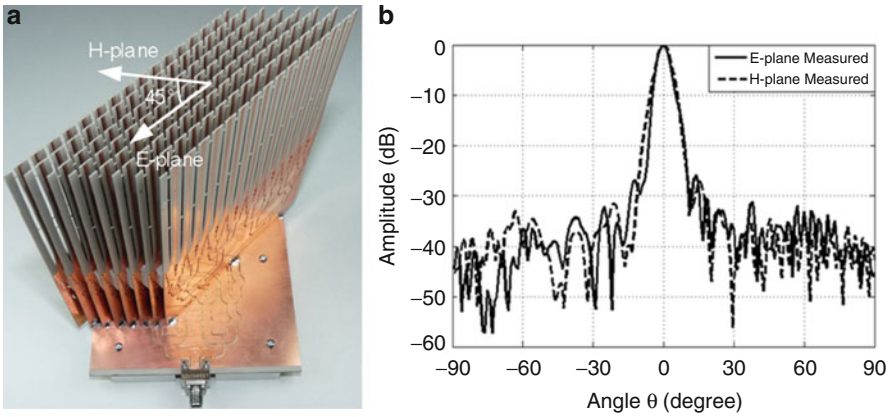


Fig. 71 (**a**) Photograph of the proposed three-dimensional architecture of 128 element antenna array, (**b**) E-plane and H-plane radiation pattern performances of the 8 × 16 SIW corrugated Fermi-TSA planar array at 35 GHz

Fermi element described in the SIW TSA section is selected to ensure initial high gain and broadband characteristics for the array. At 35 GHz, the simulated gain is about 19.4 and 18.4 dBi is obtained in measurement. The mutual coupling between the elements of antenna arrays must be taken into account during the design process. The previously described antenna topology has been studied to evaluate the isolation levels between two adjacent elements along their *E*-plane.

The isolation is about 13 dB between the elements aligned and separated by 0.68 λ_0. The dielectric is subtracted in the central region and the isolation between the two elements is 3 dB lower. In the modified antenna, the field is more spread along the transverse section. In fact, the air gap creates discontinuities and the field is confined along the air gap. When used in the two element array, the air slotted version gives more than 1 dB of improvement. The gain of one simulated Fermi-TSA element in the presence of a second passive element is 2 dB lower than that of a single element. The coupling is very small between two elements disposed in the parallel arrangement.

Figure 71 illustrates this novel planar array composed of 8 × 16 antenna elements of corrugated antipodal Fermi-TSA, and fed with an SIW power divider. A 45° rotated E-to-H-plane interconnect ensures, with eight 1 × 16 and one 1 × 8 power dividers, the construction of one block feeding network for 128 antenna element array. The whole antenna is fabricated with standard PCB process (7 PCB plaques). The gain of the planar array is 27 dBi, and the SLL is better than 26 dB in both E-plane and H-plane as shown in Fig. 71b, the beamwidth is about 5.15° in the E-plane and 6.20° in the H-plane. The total weight of the array is less than 200 g, showing an important advantage of SIW technology in payload's design efficiency.

Imaging System

There is a number of different ways to obtain a passive radiometric image from a scene (Goldsmith et al. 1993). Phased arrays have been limited in their success at millimeter wavelengths due to difficulties designing efficient phase shifter as well as relatively large feed system. Mechanically scanned systems have been employed for different imaging applications in the millimeter range, but suffer the very major handicaps of expensive, relatively unreliable mechanical systems together with low data rate from the single receivers utilized. Another way of acquiring an image is to implement a FPA (Focal Plane Array), which uses an array of independent detector elements in the focal surface of an imaging lens or reflector. Recently, moderate-cost systems using small FPAs combined to mechanical scanning have been built allowing real-time imaging (Doghri et al. 2011).

DOI 10.1007/978-981-4560-75-7_57-1

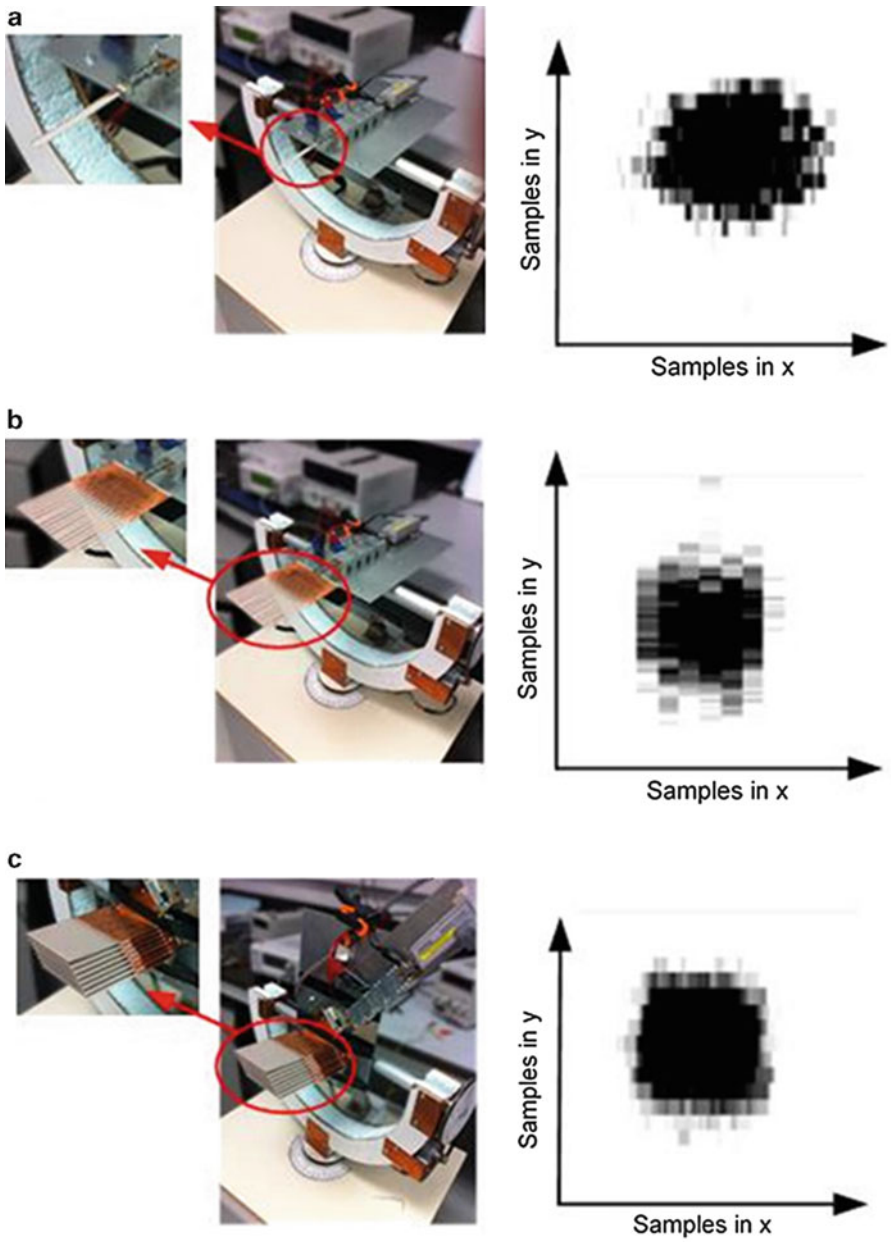


Fig. 72 Heated square object imaged by three different antenna configurations with: (**a**) single element, (**b**) linear array, (**c**) planar array

For passive imaging system, the antenna is used together with a very sensitive receiver to detect and provides measurements of the electromagnetic radiation intensity received within a solid angle subtended by the antenna beam. It is necessary to accommodate the antenna in this system design with high gain, low side lobe level and low cross-polarization. Array antenna schemes can be used to accomplish these goals. However, they present a number of problems in connection with cost, weight, and fabrication complexity of the feed networks. Various discontinuities within a feed network should be avoided or used with caution since noise emitted or interference created by such components into the receiver could desensitize and confound the radiometric measurement. Images issued from a simple receiver composed of a front-end circuit, which has two prime tasks, namely, the input frequency-band selection and the amplification of the received signal into a proper level for the detector. A novel SIW prototyping approach was used to

DOI 10.1007/978-981-4560-75-7_57-1

build up the receiver. This approach consists of SIW subcircuits realizing simple functions and interconnects as LEGO building blocks to build up a complex function out of basic elements. Putting the SIW transmission lines in front of each other with sufficient pressure passes the electromagnetic wave through with a negligible loss.

The signal from the detector is fed to the analog to digital (A/D) input of the microcontroller and then sent to the computer through a USB port and processed by the data acquisition software developed on the Matlab platform.

The tests were done using respectively one element, a planar array composed by 16 elements and a 2D array composed by 128 elements shown in Fig. 71 shows the three images corresponding to the three scenarios. By increasing the antenna gain, the resolution is increased as shown in Fig. 72. The 2D array gives a closer image of the square as it presents a higher gain and a better beam symmetry in the two planes compared to one element and the planar array. The image resolution can be further improved by increasing the number of array elements or by increasing the gain of each element.

Conclusion

It is well recognized that the antenna is one of the most important system components that limit or enhance system performance, depending on the design of such a component. Generally, antenna elements cannot be conveniently integrated in chip-set because either they are too large or the required performance such as efficiency cannot be achieved by integrated components. In some cases, they could simply be considered as part of the package that embeds the chip-set, which may be of importance for millimeter-wave system design. In most high-gain antenna applications, array geometries are always required that involve both radiating elements and feed network. Special feed networks in the form of beam-forming networks present the engine for the development of multibeam, beam-scanning and beam-agile systems. In this chapter, nonresonant SIW antenna structures well suited for a wide range of applications over millimeter-wave and terahertz frequency bands are presented and discussed.

Different dielectric loaded H–plane SIW horn antennas are presented. Through a proper choice of the structure and parameters, the maximum gain for the given size or the minimum size for a given gain may be achieved.

Two different kinds of leaky-wave structures have been discussed in this chapter. One is related to uniform guided-wave structures while the other consists of an array of periodic guided-wave structures. They share common features but also exhibit many interesting differences in connection with electrical and mechanical aspects. They exhibit low losses, very flexible radiation characteristics, and excellent mechanical simplicity. Tapered slot antennas based on SIW feed with different shapes are presented with simulation and measurement results. Configurations which ensure dual polarization and applications of an array on imaging system are detailed. Other type of SIW-based nonresonant antenna can be fed by SIW or integrated to the SIW antenna like Yagi, Lenses, Rod.

Advances in thick film technology have been made such as photoimageable and photoetchable conductor pastes, which have enabled system-in-package (SiP) techniques to be demonstrated well into the up-millimeter-wave frequency range (500 GHz, for example). LTCC technology has been known to well synergize with SIW techniques since it offers multilayered 3D passive integration. Proposed LEGO-style blocks fabricated and assembled with easy-to-connect and manipulate PCB pieces have been studied and demonstrated in different components as well as for imaging system. This technique will play a critical role with LTCC and PCB fabrication processes.

DOI 10.1007/978-981-4560-75-7_57-1

Cross-References

- ▶ Beam-scanning Leaky Wave Antennas
- ▶ Broadband and Multiband Planar Antennas
- ▶ Millimeter-Wave Antennas and Arrays
- ▶ Phased Arrays
- ▶ Transmission Lines
- ▶ Waveguide Slot Array Antennas

References

Bozzi M, Georgiadis A, Wu K (2011) Review of substrate-integrated waveguide circuits and antennas. IET Microwaves Antennas Propag 5:909–920

Bozzi M, Pasian M, Perregrini L (2014) Modeling of losses in substrate integrated waveguide components. In: International conference on numerical electromagnetic modeling and optimization for RF, microwave, and terahertz applications (NEMO), Pavia, 14–16 May 2014, pp 1–4

Cai Y, Qian Z-P, Zhang Y-S, Jin J, Cao W-Q (2002) Bandwidth enhancement of SIW horn antenna loaded with air-via perforated dielectric slab. IEEE Antennas Wirel Propag Lett 13:571–574

Caloz C, Jackson DR, Itoh T (2011) Leaky-wave antennas. In: Frontiers in antennas: next generation design & engineering. McGraw-Hill, New York

Cao W, Chen ZN, Hong W, Zhang B, Liu A (2014) A beam scanning leaky-wave slot antenna with enhanced scanning angle range and flat gain characteristic using composite phase-shifting transmission line. IEEE Trans Antennas Propag 62(11):5871–5875

Cassivi Y, Perregrini L, Arcioni P, Bressan M, Wu K, Conciauro G (2002) Dispersion characteristics of substrate integrated rectangular waveguide. IEEE Microwave Wirel Compon Lett 12(9):333–335

Cheng YJ, Wu K, Hong W (2008) Power handling capability of substrate integrated waveguide interconnects and related transmission line systems. IEEE Trans Adv Packag 31(4):900–909

Cheng YJ, Hong W, Wu K, Fan Y (2011) Millimeter-wave substrate integrated waveguide long slot leaky-wave antennas and two-dimensional multibeam applications. IEEE Trans Antennas Propag 59(1):40–47

D'Orazio William (2004) Study of substrate integrated waveguide circulators for millimeter wave applications. In: Mémoire présenté en vue de l'obtention du diplôme de maîtrise ès sciences appliquées (génie électrique)'Thèse – École polytechnique

Deslandes D (2010) Design equations for tapered microstrip-to-substrate integrated waveguide transitions. In: International microwave symposium, Anaheim, 23–28 May 2010, pp 704–707

Deslandes D, Wu K (2003) Single-substrate integration technique of planar circuits and waveguide components. IEEE Trans Microwave Theory Tech 51:593–596

Deslandes D, Wu K (2005) Substrate integrated waveguide leaky-wave antenna: concept and design considerations. In: Microwave conference proceedings, APMC 2005. Asia-pacific conference proceedings, Suzhou, pp 346–349

Deslandes D, Wu K (2006) Accurate modeling, wave mechanisms, and design considerations of a substrate integrated waveguide. IEEE Trans Microwave Theory Tech 54(6):2516–2526

Ding Y, Wu K (2008) Miniaturization techniques of substrate integrated waveguide circuits. In: IEEE MTT-S inter microwave workshop series (IMWS) on art of miniaturizing RF and microwave passive components, Chengdu, pp 63–66

Ding Y, Wu K (2009) A 4 x 4 ridge substrate integrated waveguide (RSIW) slot array antenna. IEEE Antennas Wirel Propag Lett 8:561–564
Ding Y, Wu K (2010) T-type folded substrate integrated waveguide (TFSIW) slot array antenna. IEEE Trans Antennas Propag 58(5):1792–1795
Dion A (1958) Nonresonant slotted arrays. IRE Trans Antennas Propag 1958:360–365
Djerafi T, Wu K (2012a) 77 GHz planar Butler matrix without crossover. IEEE Trans Antennas Propag 60(10):4949–4954
Djerafi T, Wu K (2012b) Corrugated substrate integrated waveguide (SIW) antipodal linearly tapered slot antenna array fed by quasi-triangular power divider. Prog Electromagn Res C 26:139–151
Djerafi T, Wu K (2013) Substrate integrated waveguide (SIW) techniques: the state-of-the-art developments and future trends. J Univ Electron Sci Technol China 42(2):171–192
Djerafi T, Ghassemi N, Kramer O, Youzkatli-el-Khatib B, Guntupalli AB, Wu K (2012a) Small footprint multilayered millimeter-wave antennas and feeding networks for multi-dimensional scanning and high-density integrated systems. Radioengineering 21(4):935–945
Djerafi T, Aubert H, Wu K (2012b) Ridge substrate integrated waveguide (RSIW) dual-band hybrid ring coupler. IEEE Microwave Wirel Compon Lett 22(2):70–72
Djerafi T, Ghiotto A, Wu K (2012c) Antipodal fin-line waveguide to substrate integrated waveguide transition. In: IEEE MTT-S international microwave symposium digest, Montreal, pp 1–3
Djerafi T, Patrovsky A, Wu K, Tatu SO (2013) Recombinant waveguide power divider. IEEE Trans Microwave Theory Tech 61(11):3884–3891
Doghri A, Djerafi T, Ghiotto A, Wu K (2011) Early demonstration of a passive millimeter-wave imaging system using substrate integrated waveguide technology. In: Mediterranean microwave symposium (MMS), 2011 11th, Hammamet, pp 215–218
Doghri A, Djerafi T, Ghiotto A, Wu K (2013) SIW 90-degree twist for substrate integrated circuits and systems. In: IEEE MTT-S international microwave symposium digest, Seattle, pp 1–3
Doghri A, Ghiotto A, Djerafi T, Wu K (2014) Broadband E-plane junction for three-dimensional substrate integrated waveguide circuits and systems. IEEE Microwave Wirel Compon Lett 24(11):739–741
El Khatib BY, Djerafi T, Wu K (2012a) Substrate integrated waveguide vertical interconnects for three-dimensional integrated circuits. IEEE Trans Compon Packag Manuf Technol 2(9):1526–1535
El Khatib BY, Djerafi T, Wu K (2012b) Three-dimensional architecture of substrate integrated waveguide feeder for fermi tapered slot antenna array applications. IEEE Trans Antennas Propag 60(10):4610–4618
Esquius-Morote M, Fuchs B, Zürcher J-F, Mosig JR (2013a) A printed transition for matching improvement of SIW horn antenna. IEEE Trans Antennas Propag 61(4):1923–1930
Esquius-Morote M, Fuchs B, Zurcher J, Mosig JR (2013b) Novel thin and compact H-plane SIW horn antenna. IEEE Trans Antennas Propag 61(6):2911–2920
Galejs J (1962) Excitation of slots in a conducting screen above a lossy dielectric half space. IRE Trans AP-10:436–443
Gazit E (1988) Improved design of the Vivaldi antenna. Proc Inst Elect Eng pt H 135:89–92
Gibson PJ (1979) The Vivaldi aerial. In: Proceedings of the 9th European microwave conference, Brighton, pp 101–105
Goldsmith PF, Hsieh C-T, Huguenin GR, Kapitzky J, Moore EL (1993) Focal plane imaging systems for millimeter wavelengths. IEEE Trans Microwave Theory Tech 41:1664–1675
Grigoropoulos N, Young PR (2004) Compact folded waveguides. In: 34th European microwave conference, Amsterdam, pp 973–976
Grigoropoulos N, Izquierdo BS, Young PR (2005) Substrate integrated folded waveguides (SIFW) and filters. IEEE Microwave Wirel Compon Lett 15(12):829–831

DOI 10.1007/978-981-4560-75-7_57-1

Guntupalli AB (2014) Low-cost integrated waveguide antenna front-end solutions for fifth generation cellular systems and beyond. Dissertation thesis, Ecole polytechnique de Montreal

Hao Z-C, Hong W, Chen J-X, Chen X-P, Wu K (2005) A novel feeding technique for antipodal linearly tapered slot antenna array. IEEE MTT-S Int Microw Symp Dig 12–17:1641–1643

Hong W, Wang Y, Lai QH, Liu B (2006) Half mode substrate integrated waveguide: a new guided wave structure for microwave and millimeter wave application. In: Proceedings of the joint 31st international conference on infrared millimeter waves and 14th international conference on terahertz electron, Shanghai

Hyneman RF (1959) Closely-spaced transverse slots in rectangular waveguide. IRE Trans 1959:335–342

Johansson JF (1989) A moment method analysis of linearly tapered slot antennas. In: IEEE antennas and propagation society international symposium, San Jose, pp 383–386

Kim SG, Chang K (2004) Ultra wideband exponentially-tapered antipodal Vivaldi antennas. In: IEEE AP-S, Monterey, pp 2273–2276

Lewis LR, Fasset M, Hunt J (1974) A broadband stripline array element. In: IEEE antennas and propagation society international symposium (digest), Atlanta, pp 335–337

Li ZL, Wu K (2004) A new approach to integrated horn antennas. In: Proceedings of international of symposium on antenna technology and applied electromagnetics, Montreal, pp 535–538

Li L, Chen X, Khazaka R, Wu K (2009) A transition from substrate integrated waveguide (SIW) to rectangular waveguide. In: Asia pacific microwave conference, 2009, Singapore, pp 2605–2608

Li J, Wen G, Xiao F (2010) Broadband transition between rectangular waveguide and substrate integrated waveguide. Electron Lett 46(3):223–224

Lu K (1997) An efficient method for analysis of arbitrary nonuniform transmission lines. IEE Trans Microwave Theory Tech 45(1):9, 14

Mallahzadeh AR, Esfandiarpour S (2012) Wideband H-plane horn antenna based on ridge substrate integrated waveguide (RSIW). IEEE Antennas Wirel Propag Lett 11:85–88

Martinez-Ros AJ, Gomez-Tornero JL, Goussetis G (2010) Planar leaky-wave antenna with flexible control of the complex propagation constant. IEEE Trans Antennas Propag 60(3):1625–1630

Meinel H (1995) Commercial applications of millimeterwaves – history, present status and future trends. IEEE Trans Microwave Theory Tech 43(7):1639–1653

Moldovan E, Bosisio RG, Wu K (2006) W-band multiport substrate integrated waveguide circuits. IEEE Trans Microwave Theory Tech 54:625–632

Nanzer J (2012) Microwave and millimeter-wave remote sensing for security applications. Artech House, Norwood

Nesic A (1991) Endfire slot line antennas excited by a Coplanar waveguide. In: IEEE AP-S international symposium, digest, London, pp 700–702

Oliner AA, Jackson DR, Volakis JL (2007) Leaky-wave antennas, Chapter 11. In: Antenna engineering handbook, 4th edn. McGraw-Hill, New York

Patrovsky A, Daigle M, Wu K (2007) Millimeter-wave wideband transition from CPW to substrate integrated waveguide on electrically thick high-permittivity substrates. In: Proceedings of 37th European Microwave Conference (EuMC37), Munich, pp 138–141

Pozar DM (1983) Surface wave effects for millimeter wave printed antennas. In: Proceedings of the antennas and propagation society international symposium, Houston, vol 21, pp 692–695

Pozar DM (2011) Microwave engineering, 4th edn. Wiley, New York

Prasad SN, Mahapatra S (1979) A novel MIC slot line aerial. In: Proceedings 9th European microwave conference, Brighton, pp 120–124

Salmon NA (2004) Scene simulation for passive and active millimetre and sub-millimetre wave imaging for security scanning and medical applications. SPIE 5619:129–135

DOI 10.1007/978-981-4560-75-7_57-1

Sato H, Sawaya K, Arai N, Wagatsuma Y, Mizuno K (2003) Broadband FDTD analysis of fermi antenna with narrow width substrate. In: IEEE antennas and propagation society international symposium, 2003, Columbus, vol 1, pp 261–264

Scharp GS (1982) Continuous slot antennas. US Patent 4,328,502, 4 May 1982

Schaubert DH (1993) Scanning characteristics of stripline- fed tapered slot antennas on dielectric substrates. In: Proceedings of the 1993 antenna applications symposium, Monticello

Schoebel J, Herrero P (2010) Planar antenna technology for mm-wave automotive radar, sensing, and communications. In: Kouemou G (ed) Radar technology InTech. ISBN:978-953-307-029-2, doi:10.5772/7185. http://www.intechopen.com/books/radar-technology/planar-antenna-technology-for-mm-wave-automotive-radar-sensing-and-communications

Simons RN, Lee RQ, Perl TD (1992) Non-planar linearly tapered slot antenna with balanced microstrip feed. In: IEEE AP-S international symposium, Chicago, vol 4, pp 2109–2112

Solbach K, Schneider R (1999) Review of antenna technology for millimeter wave automotive sensors. In: Proceedings of the European microwave conference, Munich, pp 139–142

Sugawara S, Maita Y, Adachi K, Mori K, Mizuno K (1998) Characteristics of a MM-wave tapered slot antenna with corrugated edges. IEEE Trans Microwave Theory Tech 2:533–536

Sutinjo A, Okoniewski M, Johnston RH (2008) Radiation from fast and slow traveling waves. IEEE Antennas Propag. Mag. 50(4):175–181

Uchimura H, Takenoshita T, Fujii M (1998) Development of `a laminated waveguide. IEEE Trans Microwave Theory Tech 46(12):2438–2443

Vye D (2011) Divine innovation: 10 technologies changing the future of passive and control components. Microw J 54(11):22–42

Walter CH (1965) Traveling wave antennas. McGraw-Hill, New York

Wang N-B, Jiao Y-C, Song Y, Zhang L, Zhang FS (2009) A microstrip-fed logarithmically tapered slot antenna for wideband applications. J Electromagn Waves Appl 23(10):1335–1344

Wang H, Fang D-G, Zhang B, Che W-Q (2010) Dielectric loaded substrate integrated waveguide (SIW) H-plane horn antennas. IEEE Trans Antennas Propag 58(3):640–647

Wei J, Chen ZN, Qing X, Shi J, Xu J (2013) Compact substrate integrated waveguide slot antenna array with low back lobe. IEEE Antennas Wirel Propag Lett 12:999–1002

Whetten F, Balanis CA (1991) Meandering long slot leaky-wave antennas. IEEE Trans Antennas Propag 39(11):1553–1559

Wong M, Sebak AR, Denidni TA (2008) A broadside substrate integrated horn antenna. In: IEEE AP-S international symposium 2008, Atlanta, pp 1–4

Wu K (2001) Integration and interconnect techniques of planar and nonplanar structures for microwave and millimeter-wave circuits-current status and future trend, In: Asia–pacific microwave conference 2001, Taipei, Taiwan, pp 411–416

Wu C, Shen L, Deng GY, Shen Y, Litva J (1998) Experimental study of a wide band LTSA which is fed by an inverted microstrip line (IML). In: IEEE AP-S international symposium, San Diego, vol 4, pp 2328–2331

Wu K, Deslandes D, Cassivi Y (2003) The substrate integrated circuits-a new concept for high-frequency electronics and optoelectronics. In: Telecommunications in modern satellite, cable and broadcasting service, TELSIKS 2003, 6th international conference, vol 1, pp P-III

Wu K, Cheng YJ, Djerafi T, Hong W (2012) Substrate integrated millimeter-wave and terahertz antenna technology. Proc IEEE 100(7):2219–2232

Xia L, Xu R, Yan B, Li J, Guo Y, Wang J (2006) Broadband transition between air-filled waveguide and substrate integrated waveguide. Electron Lett 42(24):1403–1405

DOI 10.1007/978-981-4560-75-7_57-1

Xu F, Zhang Y, Hong W, Wu K, Cui TJ (2003) Finite difference frequency-domain algorithm for modeling guidedwave properties of substrate integrated waveguide. IEEE Trans Microwave Theory Tech MTT-51(11):2221–2227

Xu J, Hong W, Hongjun H, Kuai Z, Wu K (2008) Half-mode substrate integrated waveguide (HMSIW) leaky-wave antenna for millimeter-wave applications. IEEE Antennas Wirel Propag Lett 7:85–88

Yamamoto S, Hirokawa J, Ando M (2002) A beam switching slot array with a 4-way Butler matrix installed in a single layer post-wall waveguide. In: Proceedings of the IEEE AP-S international 1241 symposium, San Antonio, vol 1, pp138–140

Yan L, Hong W, Wu K, Cui TJ (2005) Investigations on the propagation characteristics of the substrate integrated waveguide based on the method of lines. Proc IEE Microwave Antennas Propag 152:35–42

Yang Y, Wang Y, Fathy AE (2008) Design of compact Vivaldi antenna arrays for UWB see through wall applications, Progress In Electromagnetics Research, PIER vol 82, pp 401–418

Yngvesson SK, Schaubert DH, Korzienowski TL et al (1985) Endfire tapered slot antennas on dielectric substrates. IEEE Trans Antennas Propag AP-33(12):1392–1400

Yngvesson KS, Korzeniowski TL, Kim Y-S, Kollberg EL, Johansson JF (1989) The tapered slot antenna – a new integrated element for millimeter-wave applications. IEEE Trans Microwave Theory Tech 37(2):365–374

Zhu L, Wu K (2002) Field-extracted lumped-element models of coplanar stripline circuits and discontinuities for accurate radio-frequency design and optimization. IEEE Trans Microwave Theory Tech 50(4):1207–1215

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_59-1

Ultra-Wideband Antennas

Max J. Ammann[a]*, Matthias John[b] and Giuseppe Ruvio[c]
[a]School of Electrical and Electronic Engineering, Dublin Institute of Technology, Dublin, Ireland
[b]The Telecommunications Research Centre, (CTVR), Trinity College, University of Dublin, Dublin, Ireland
[c]Antenna & High Frequency Research Centre, Dublin Institute of Technology, Dublin, Ireland

Abstract

The focus of UWB antenna research activity has matured in recent years and currently mainly concentrates on applications such as biomedicine and security. Early UWB antenna designs were driven by the FCC allocation of spectrum in 2002 and focussed on obtaining wide impedance bandwidths with reasonable group delay characteristics. Many of these were simple planar monopoles antennas with canonical geometries. The emergence of new applications channelled the emphasis towards miniaturisation and integration into devices. This required optimisation of the antenna geometries to ensure that good system performance is achieved from the integrated antenna. Many optimisation techniques are available including the spline technique to generate the outline of the antenna element and ground plane. Simple methods based on genetic algorithms are employed and evolutionary algorithms which are capable of optimising for multiple goals are beneficial when multiple antenna parameters are simultaneously investigated. These techniques have proven advantageous especially when time-domain performance is critical and provide solutions for both single-ended and differential feed arrangements. The main applications using UWB channels in the 3.1 GHz −10.6 GHz spectrum are localization and tracking applications, mainly employing impulse radio UWB imaging, and generally using linear polarization. However circularly-polarized UWB antennas have been developed, both directional and omnidirectional and are being investigated across various systems.

Keywords

Planar monopole antennas; Splines; Optimisation; Time domain performance; Impulse response; Asset tracking antennas; Imaging antennas; Vivaldi antennas; Circular polarisation; Dipole antennas; Miniaturisation; Slot antennas; Radio tags

Introduction

In 1897, Guglielmo Marconi sent the first wireless communication over open sea using a spark-gap transmitter and in 1901, performed the first transatlantic communication from Poldhu, England to St. John's, Newfoundland (Balanis 2005). Because of the spark-gap transmitter, the signal used by Marconi was inherently wideband. Since then the evolution of radio technology has greatly increased. For many of the following decades, radio technology advancements focussed mainly on narrow band transmission, due to the congestion in the electromagnetic spectrum. However in the late 1960's, researchers started to study Ultra Wide Band (UWB) technology. Between 1977 and 1989 the United State Air Force (USAF) had a program on the UWB system (Barret 2001) and some universities focused

*Email: max.ammann@dit.ie

DOI 10.1007/978-981-4560-75-7_59-1

on the interaction of short pulses with matter. In 1994 T.E. McEwan invented the Micropower Impulse Radar (MIR) which proved to be the first compact, inexpensive and low power radar, which consumes only microwatts from batteries (Azevedo and McEwan 1996).

In 2002, the Federal Communication Commission (FCC) allocated the 3.1–10.6 GHz for UWB unlicensed use and standardisation began. It limited the maximum Equivalent Isotropic Radiated Power (EIRP) for UWB radio to −41.3 dBm/MHz (74 nW/MHz) and specified that a UWB signal must have, at any point in time, a fractional bandwidth equal to or greater than 20 % or have a 500 MHz bandwidth regardless of the fractional bandwidth (Allen et al. 2007). The high-data rate UWB standard was prepared by the IEEE 802.15.3a task group (Mandke et al. 2003), but ceased in early 2006. However, a high data rate Multi-Band OFDM scheme was standardized by the ECMA-368 standard. In 2004, the IEEE 802.15.4a task group was established for low rate UWB systems which chose to use two approaches for the physical layer, the Direct Sequence UWB (DS-UWB) and the Chirp Spread Spectrum (CSS) (IEEE 2011). The IEEE since published the IEEE 802.15.6 standard for Body-Area-Networks, which provides for UWB channels. Because of their very wide bandwidth, UWB signals have a high temporal resolution (typically less than 1 ns), making this technology particularly suitable for Wireless Sensor Network (WSN) and Real Time Location System (RTLS).

Some of the most popular antennas employed in modern wideband wireless communications systems are the dipole and monopole family. These elements exhibit broad impedance bandwidths which can be extended by increasing the radius of the cylindrical element. This is true up to a point where the stepped radius from the feed probe to the cylindrical element becomes abrupt. Tapering this transition is beneficial and used in wideband elements such as biconical dipoles, but cost limits these to laboratory applications. However, it has been shown that replacing the cylindrical element of a monopole or dipole with planar elements provide a wider impedance bandwidth, and with some simple adjustment, UWB bandwidth. The first mention of the planar monopole antenna was in a textbook in 1968 by Meinke and Gundlach (1968), who describe it as a variant of the cylindrical and conical monopole. Dubost and Zisler provided more detail about the wide impedance characteristics of this antenna in 1976 (Dubost and Zisler 1976). A disc-shaped planar monopole antenna was subsequently reported in 1991 by Honda et al. (1992), who proposed this antenna for the Japanese television band (90–770 MHz). In 1992 he reported one which had a 10 dB impedance bandwidth ratio in excess of 10:1 (Honda et al. 1991). A model for determining the value of the input impedance of the circular disc-shaped monopole was reported in 1993 by Hammoud et al. (1993) based on the Method of Moments. Circular and elliptical disc monopoles were also studied in 1998 by Agrawall et al. (1998), who proposed a formula for predicting the frequency corresponding to the lower edge of the impedance bandwidth. Since then a plethora of canonical shapes were reported to provide UWB impedance bandwidth. These include square, rectangular, elliptical, circular, trapezoidal and triangular (Ammann 1999; Evans and Ammann 1999). The use of triangular and bow-tie geometries for the planar element were also investigated by Chen (2000). Furthermore, bevelling of the shapes and feed asymmetry was introduced to further widen the impedance bandwidths and very wide impedance bandwidth ratios were realised (Ammann 2001; Ammann and Chen 2003a, b).

At the time, these elements were studied when placed above a classical ground plane and in many cases, only the impedance properties were reported. As time moved on, many of the geometries migrated to printed antennas geometries where the ground planes are printed on the same PCB (Kuo and Wong 2002; Go and Jung 2004; Liang et al. 2004). The ground planes, which were often excluded from any design rules, were populated with components and modules, which is particularly attractive for portable terminal devices (Fig. 1).

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_59-1

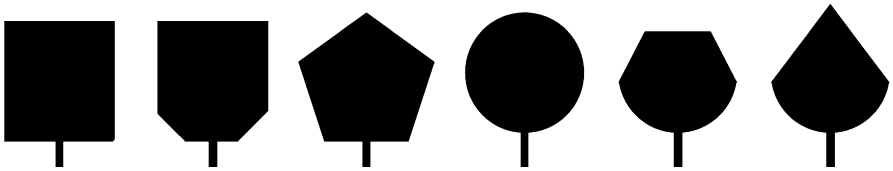


Fig. 1 A selection of planar monopole shapes including square, tapered square, pentagonal, circular and combinations of these

Optimised UWB Antennas

Geometries of recent UWB antennas are based on simple geometric elements, such as rectangles (Chen et al. 2006a), circles (Liang et al. 2005) or ellipsis (Angelopoulos et al. 2006), or a combination of these (Griffiths et al. 2005). Examples for combining multiple elements are a combination of rectangles, triangles and circles in (Karacolak and Topsakal 2006) or polygons, circles and exponential curves in (Chen et al. 2006b).

To optimise a specific antenna design, a number of optimisation methods are available ranging from simple interpolation methods to evolutionary algorithms capable of optimising for multiple goals. There is a large number of publications available dealing with electromagnetic optimisation. Evolutionary methods include genetic algorithm (GA) (Johnson and Rahmat-Samii 1997; Haupt 2007; John and Ammann 2007) and particle swarm optimisation (PSO) (Jin and Rahmat-Samii 2005; Donelli and Massa 2005). Today, a variety of methods are available in commercial EM software such as CST, Feko or HFSS. In the following example the GA in CST Microwave Studio is used.

This section analyses antenna geometries based on quadratic Bézier spline curves (Foley et al. 1993; Bartels et al. 1996). Splines are curves generated by quadratic interpolation between control points. This spline is used to outline the radiating element and/or ground plane of an antenna. The benefit of this technique is an inherently rounded shape of the radiating element and a straight forward way of assigning parameters to the control points. The design methodology overcomes the geometrical constraints of a circular or elliptical disk and the need of combining simple geometric elements.

Spline Theory

A spline is a function defined by polynomials. In mathematics, splines are used for interpolation and smoothing problems. In computer graphics, the term spline refers to a piecewise polynomial parametric curve used to illustrate complex shapes (Bartels et al. 1996). Splines are classified by the degree of the polynomial, such as linear, quadratic, cubic or higher order splines. In computer graphics and CAD programs, quadratic and cubic splines are most common (Foley et al. 1993). Bézier curves are named after Dr. Pierre Bézier who used these curves to design automobile bodies for Renault in the 1960s.

The electromagnetic simulator CST Microwave Studio (MWS) uses the quadratic form of these curves. One curve segment of a quadratic Bézier spline is the path traced by the function B(t)

$$B(t) = (1-t)^2 P_0 + 2t(1-t)P_1 + t^2 P_2; t \in [0,1], \quad (1)$$

when t is swept from 0 to 1. The points P_n are defined by their x- and y-coordinates in a two dimensional plane.

Figure 2 shows this path for various *t*. It can be seen that, as *t* is varied in five steps from 0 to 1, the '×' (green) traces a quadratic curve (red). To create more complex shapes, an arbitrary number of Bézier curves can be attached together in a chain.

DOI 10.1007/978-981-4560-75-7_59-1

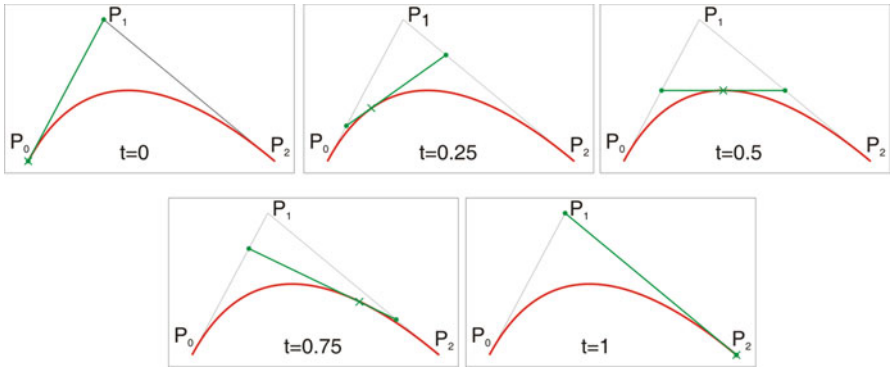


Fig. 2 Quadratic Bézier curve for various t

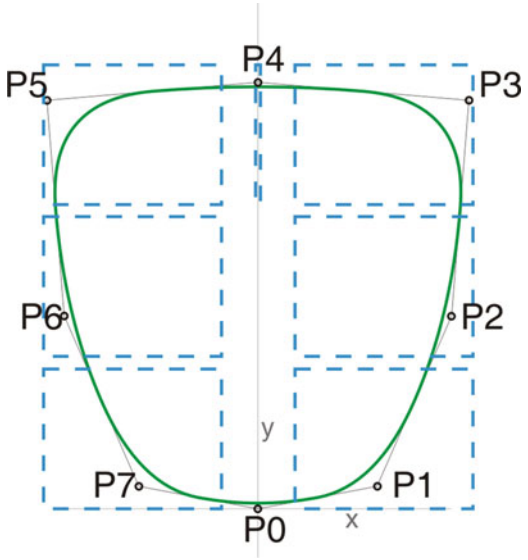


Fig. 3 Control points and bounds for spline geometry

Example of a Spline Based Antenna Geometry

As an example a quadratic Bézier spline is used to describe the outline of the radiating element of a printed monopole antenna in the following way. The spline curve is defined by 8 control points P0–P7. Figure 3 shows the outline curve and its control points. Point P0 is fixed at (0,0), this is where the antenna is fed by a 50 Ω microstrip line. Points P1–P3 are defined their by x and y-coordinates, which are variable parameters. Point P4 is fixed at $y = 0$ and can only be moved along the x-axis. Points P5–P7 are derived from points P1–P3 by mirroring them along the x-axis. In order to achieve a closed curve, the curve ends again in point P0. This creates a symmetrical radiator with the intention of achieving an omnidirectional radiation pattern. The resulting curve can be used as the outline of a radiating element of a printed monopole antenna. It is placed on the substrate with a partial ground plane on the rear, the enclosed area filled with copper and connected to a microstrip feedline.

The coordinates of the control points are used as parameters for optimisation by the genetic algorithm in the CST Studio Suite. Bounds are defined for the coordinates of each point, they are shown in Fig. 3 by blue rectangles. These bounds serve two purposes. Firstly, they constrain the size of the design to be

DOI 10.1007/978-981-4560-75-7_59-1

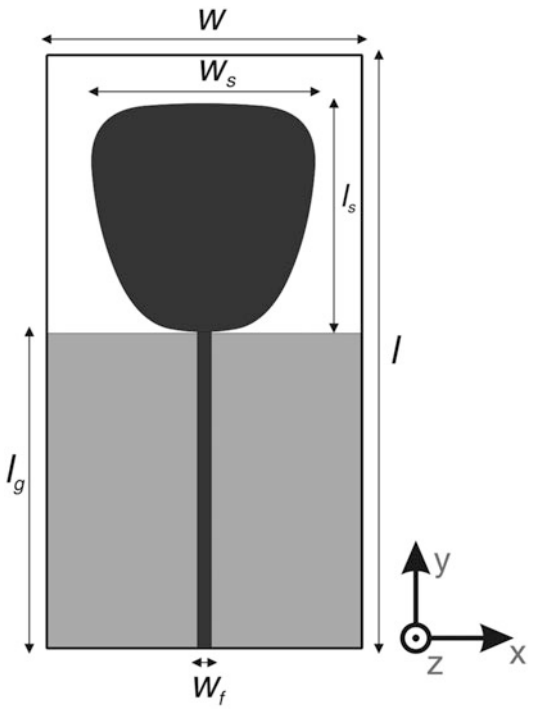


Fig. 4 Geometry of the optimised spline based monopole

within the limits defined by the substrate size. This ensures that the radiating patch does not extend over the side of the substrate. Secondly, the bounds ensure that the resulting spline does not form a loop.

Due to the wide operating bandwidth anticipated from this design, Rogers RO4350B was used as the RF substrate. The substrate size was w = 45 mm and l = 85 mm with the 45 mm square ground plane located on the rear side. A 2.4 mm wide 50 Ω microstrip feedline is used to feed the antenna. The complete geometry is shown in Fig. 4. It is the result of a GA optimisation with a population size of 30 which was evolved over 20 generations. The fitness function was set to maximise the −10 dB S11 bandwidth within the 0–20 GHz simulation bandwidth. In the resulting geometry, it can be seen that the radiating element curves smoothly away from the feedpoint. The maximal possible height is exploited as point P4 is placed at the maximum of its bound. The dimensions of the spline based radiating element are l_s = 33 mm by w_s = 32 mm.

Bandwidth

The simulated and measured S11 are shown in Fig. 5. It can be seen that the measured S11 is better than −10 dB from 1.44 GHz to 14.7 GHz. This corresponds to an impedance bandwidth ratio of 10.2:1, which is very wide for a printed monopole.

Surface Currents

The surface currents in the metallised parts of the antenna have been computed for various frequencies. They are show in Fig. 6. For the lowest frequency of 2 GHz (Fig. 6a), the plot indicates a first order resonance. Animation of a series of plots at this frequency for different phases shows that this mode has a standing wave current pattern. Higher modes at 8 GHz or 16 GHz (Fig. 6c and d) show more complex current patterns. An animation of these plots in a phase sweep shows that the current is travelling along the edge of the radiating element. It is concluded that this antenna is working in a hybrid mode (Chen 2007) of operation between a standing wave and a travelling wave antenna.

DOI 10.1007/978-981-4560-75-7_59-1

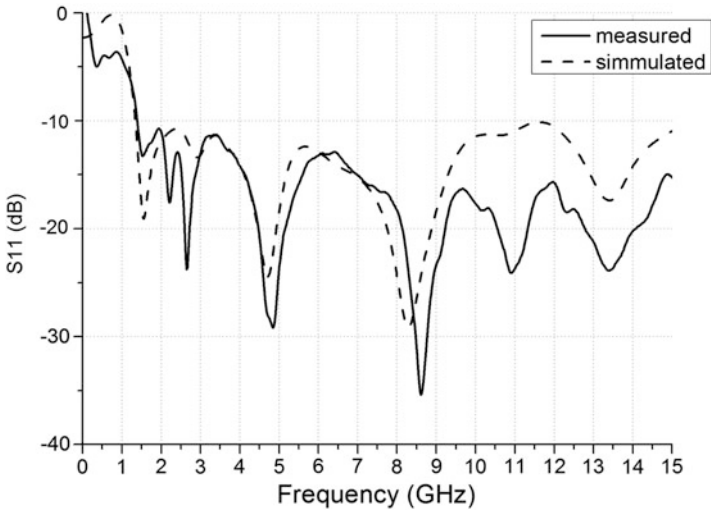


Fig. 5 Simulated and measured S11 of the spline based monopole

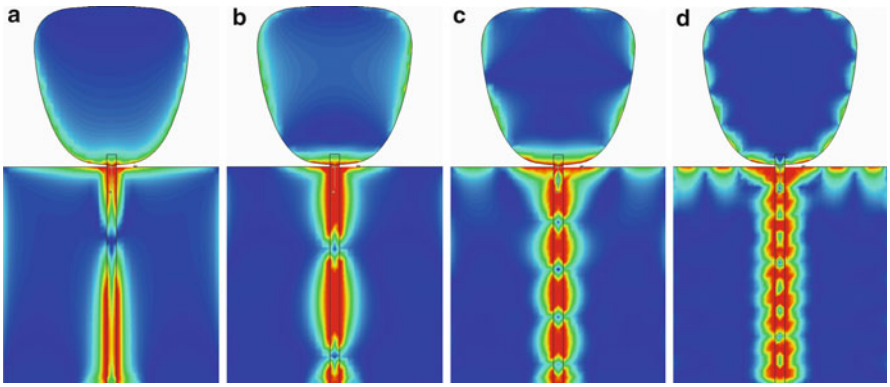


Fig. 6 surface current in the spline based monopole at 2, 4, 8, and 16 GHz

Radiation Pattern

The measured radiation patterns are shown in Fig. 7. The H-plane patterns are found to be omnidirectional up to about 8 GHz. The measured gain is 2.8 dBi at 2 GHz, 4.3 dBi at 6 GHz, 4.8 dBi at 10 GHz and 5.3 dBi at 14 GHz. The simulated radiation efficiency at these frequencies is 91, 96, 92 and 89 % respectively.

Group Delay

Group delay was measured in three orientations of the antennas, front-front, front-side and side-side. Measurement results are shown in Fig. 8. In the UWB band from 3.1 GHz to 10.6 GHz, the group delay variation is below 0.5 ns. Group delay is the derivative of the pulse response of the S21 between the two antennas. A linear phase response is required for low group delay variation. There was no explicit optimisation goal for a linear phase response in this example. A linear phase response is inherent in the spline geometry as the current path is very smooth around the outside of the radiating element. Hence the spline based design methodology is suitable for designing antennas for communication applications where low pulse distortion is required.

DOI 10.1007/978-981-4560-75-7_59-1

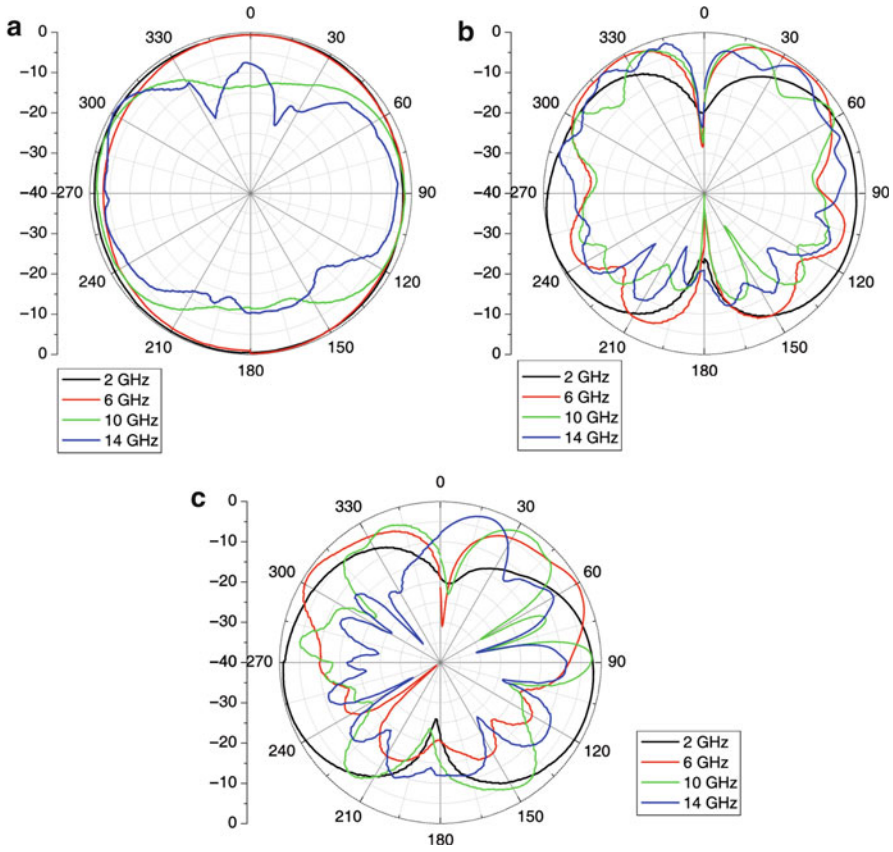


Fig. 7 Measured radiation patterns for the spline based monopole antenna in x-z, x-y and y-z planes

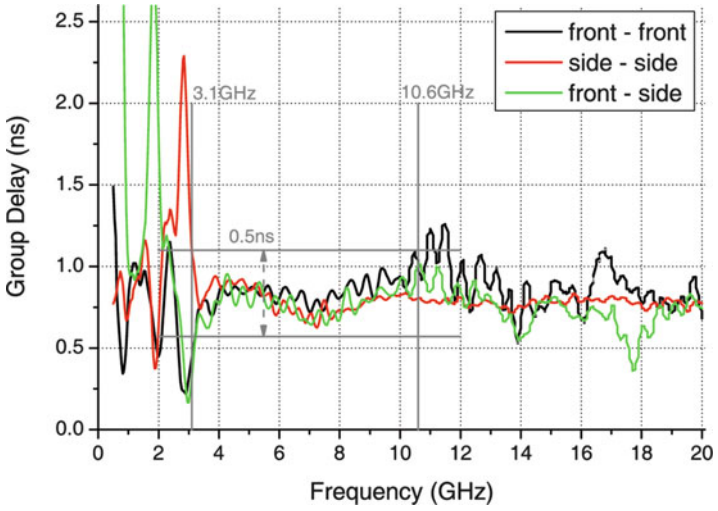


Fig. 8 Measured group delay for a pair of spline based monopole antennas in three orientations

This spline based optimisation technique can be easily adopted to design antennas of constrained size. Figure 9 shows a 40 × 30 mm monopole operating in the 3.1–10.6 GHz band with the ground plane shape also optimised (John and Ammann 2009).

DOI 10.1007/978-981-4560-75-7_59-1

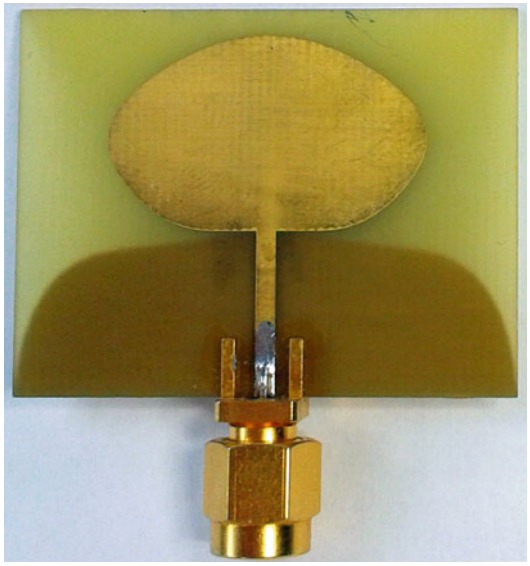


Fig. 9 Small UWB antenna developed at the Dublin Institute of Technology

UWB Antennas Optimised for Time Domain Applications

Impulse radio Ultra Wide Band (IR-UWB) systems use very short pulses (typically less than 1 ns) to achieve very broad bandwidth for low powered communications. The short pulses can be used for UWB Real Time Location System (RTLS) for proximity ranging or localization with centimeter precision. Earlier publications consider and analyse time domain performance of UWB antennas (Schantz and Fullerton 2001; Kwon 2006), but do not optimize for it. Schantz, in (2004) gives general rules on how to minimize dispersion in UWB antennas. Recent publications synthesize UWB antennas from a time-domain perspective by optimizing antenna geometries depending on the pulse shape they radiate (Telzhensky and Leviatan 2006; Dumoulin et al. 2012; Lizzi et al. 2012).

UWB Pulse Shapes

UWB antenna literature predominantly reports Gaussian and Rayleigh pulses, and their derivatives, for antenna excitation (Chen et al. 2004; Lizzi et al. 2009; Chen et al. 2005; Guo et al. 2007). Their characteristics include very low level side lobes in the frequency-domain and fast energy damping in the time-domain. However, the energy distribution across the frequency range is typically bell-shaped and antennas that are optimized with these pulse profiles do not have the bandwidth for an efficient fit to the 802.15.4a spectrum mask (Chen et al. 2005), (Chamaani et al. 2011), (Cerny and Mazanek 2007). While a series of modulated Gaussian pulses can be sequenced (IEEE 2007) to cover the entire UWB mask, the approach increases the duration of the excitation pulse. It extends antenna simulation run-times and adds complexity to the time-domain analysis of the models.

Alternatively, an SRRC pulse overcomes the limitations. The analytical form of the SRRC pulse is shown in Eq.2 (Zhang et al. 2003).

DOI 10.1007/978-981-4560-75-7_59-1

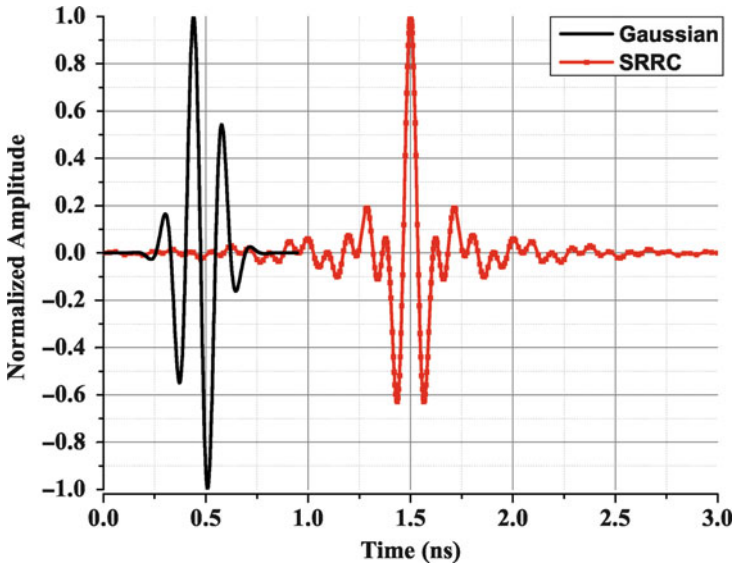


Fig. 10 Modulated Gaussian and Square Root Raised Cosine pulses with respective peak values normalized to unity

$$F_{SRRC}(t) = \left[\frac{2\beta}{\pi\sqrt{T_s}} \times \frac{\cos\left[(1+\beta)\pi\frac{t}{T_s}\right] + \frac{\sin\left[(1-\beta)\pi\frac{t}{T_s}\right]}{4\beta\frac{t}{T_s}}}{\left[1-\left(4\beta\frac{t}{T_s}\right)^2\right]} \right] \tag{2}$$

where t is the time, $T_S = 1/R_S$, R_S being the symbol rate and β is a dimensionless roll-off factor for bandwidth control.

Figure 10 shows an amplitude modulated Gaussian pulse and an amplitude modulated SRRC pulse (roll-off factor of 0.1 and symbol rate of 7.1 GHz); both centred at 6.85 GHz and with the respective peak values normalized to unity. Figure 11 shows the corresponding Power Spectrum Density (PSD) plots with the 802.11.4a UWB indoor use spectrum mask.

The evenly contoured power spectrum of the SRRC pulse, compared to that of the Gaussian pulse, conveys a broader bandwidth at a higher energy level. Although many UWB systems only operate using channels with bandwidths in range 500–1300 MHz, different channels can be used in different jurisdictions and hence it is beneficial to optimise over the full UWB range. The design goal is to create antennas that can support pulses of increased bandwidths across the spectrum profile of the SRRC.

Differentially Fed UWB Antennas

While monopole antennas are suitable for most applications, they need a balanced to unbalanced transformer (BALUN), when connected to a differential IC output. The introduction of the balun incurs losses and dispersion in both transmission and reception mode, impairing the accuracy and range of the overall system. Balanced antennas, such as dipoles or slot antennas can be differentially fed making their integration into the system more straightforward.

DOI 10.1007/978-981-4560-75-7_59-1

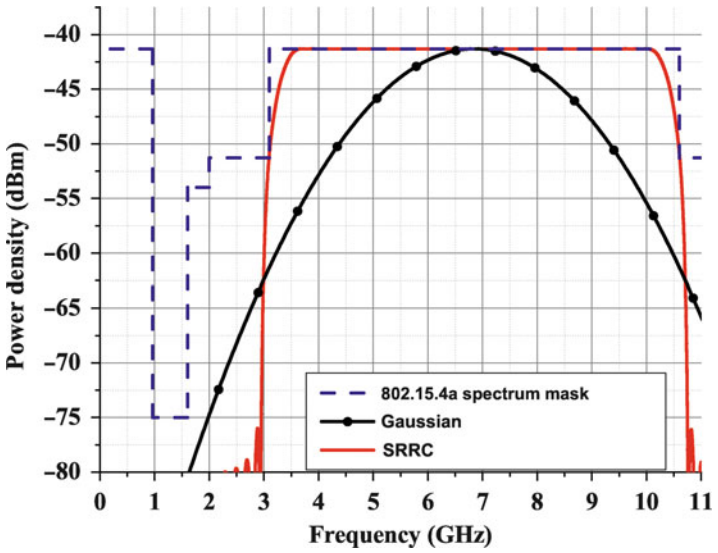


Fig. 11 802.15.4a indoor use spectrum mask with power spectrum density of Gaussian and SRRC pulses; respective peak values normalized to −41.3 dBm

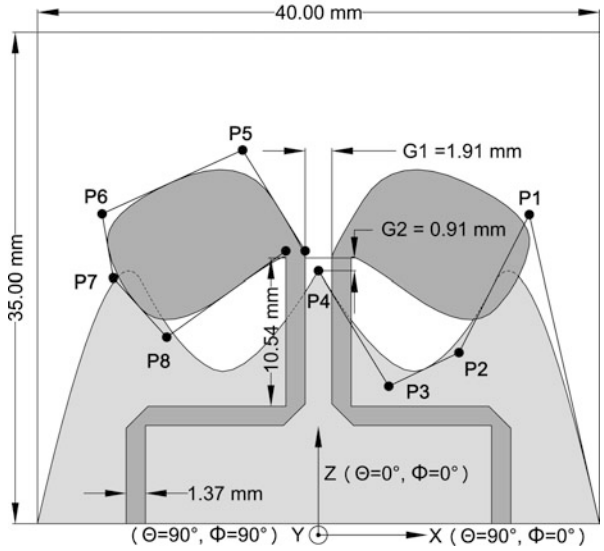


Fig. 12 Gaussian Dipole (GD) geometry

In this example a differentially fed balanced dipole is examined. The dipole geometry is optimised using both a SRRC excitation and a modulated Gaussian like excitation signal. Spectral and time-domain performances of the resulting antennas are compared.

The dipole antennas were optimised for wide-band performance and pulse fidelity, using a SRRC excitation signal for the SRRC dipole (SD) and a modulated Gaussian like excitation signal for the Gaussian dipole (GD) shown in Figs. 10 and 11. Antennas GD and SD, shown in Figs. 12 and 13, are planar dipole-like with a balanced microstrip feed arrangement which requires a small ground plane. They consist of two identical dipole arms (defined by 4 spline points) fed with 50 Ω dual mitred microstrip lines and a mirrored splined ground plane (4 points). The spline points for the radiators have two degrees of

DOI 10.1007/978-981-4560-75-7_59-1

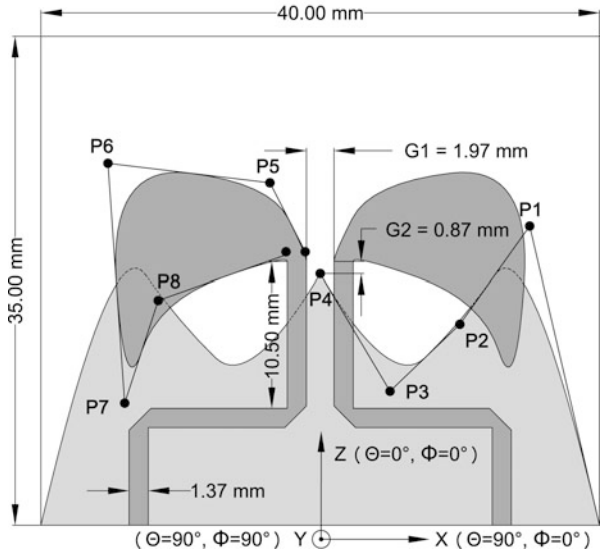


Fig. 13 SRRC Dipole (SD) geometry

freedom (X and Z-axis), while the spline points for the ground plane are restricted to one degree of freedom (Z-axis) in order to limit the complexity of the problem.

Both antennas were fabricated on a 40 × 40 × 0.7 mm FR4 double-sided laminate with a dielectric constant $\varepsilon_r = 4.3$ and loss tangent = 0.02. The antennas were initially performance-optimized without SMA connectors but the simulation model included the connectors to allow a fair comparison between measurement and simulation. The dimensional constraints of the antenna, defined in millimetres, are shown below.

- $P_{1,2,3,4}\ (z) = 5 < z < 26$
- $P_5\ (x) = 0 < x < 4$
- $P_5\ (z) = P_4 < z < 40$
- $P_{6,7,8}\ (x) = 10 < x < 20$
- $P_{6,7,8}\ (z) = P_5 < z < 40$
- $G_1 = 1 < x < 5$
- $G_2 = 0 < z < 5$

Optimisation

The antennas were optimised for best fidelity factor between the excitation signal derivative and the radiated signal. For time-domain optimization, E-field-probes were located 30 cm from the dipoles in the azimuth plane. Simulation time was halved by using field symmetry in the H-plane and by limiting probes to [$\varphi = 90°$, $-90° \leq \theta \leq 90°$ in 5° steps].

A weighted cost function was post-processed for the design iterations using Eq.3.

$$Cost = -0.3 \times \frac{\Gamma}{\alpha} - 0.7 \times \frac{\sum FF}{\gamma}, \text{for } \Gamma < \alpha \tag{3}$$

$$Cost = -0.3 - \left(0.7 \times \frac{\sum FF}{\gamma}\right), \text{for } \Gamma \geq \alpha$$

DOI 10.1007/978-981-4560-75-7_59-1

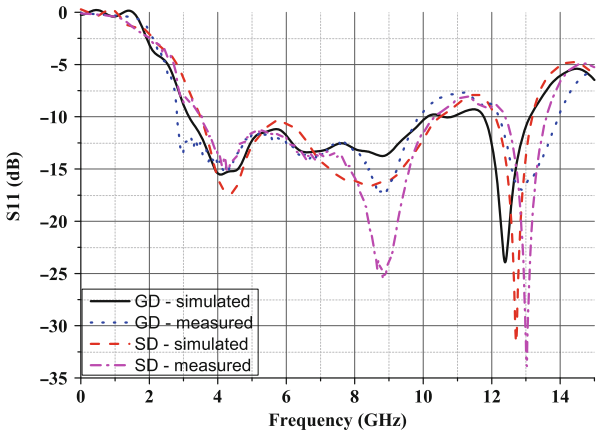


Fig. 14 Simulated and measure S11 for the GD and SD dipoles

where Γ is the least matched magnitude in the S11, α is the S11 match target, *FF* is the fidelity factor (Chen et al. 2005) (field probes relative to the derivative of the excitation pulse) at each angle and γ is the number of FF values. The empirically selected coefficient weightings were chosen to provide a balance between optimization time and accuracy.

Additionally, a subroutine was used to skip simulations of non-realistic structures or re-simulation of identical structures (Guo et al. 2007), more of which occur as the algorithm converges on the optimum range of possible solutions. These significantly reduce the optimization time with minimum effect on the GA behavior. The algorithm also uses an intermediate crossover, the roulette wheel selection, and stops after 30 population iterations (maximum antenna simulations $\leq$ 8085).

Frequency-Domain Results

The simulated and measured S11 for the two Dipoles *GD* and *SD* are illustrated in Fig. 14. The dipoles exhibit an 8 dB return loss across 2.75–10.6 GHz and 2.9–13.8 GHz, respectively. The antennas have good impedance matching for the 3.1–10.7 GHz band, with good agreement between simulations and measurements.

The dipole radiation patterns were similarly measured in the H-plane (Z-Y plane). Figures 15 and 16 show good stability in the radiation pattern across 3.1–10.6 GHz. The dipoles achieve an average realized gain of -0.27 dBi (std. dev. = 2.45) and 0.8 dBi (std. dev. = 2.47) respectively. *Dipole GD* has a maximum gain of 6.4 dBi (θ, $\varphi = 90°$, 55°) at 8.88 GHz, while *Dipole SD* has a maximum gain of 6.87 dBi (θ, $\varphi = 90°$, 60°) at 9.4 GHz.

Time Domain Measurement and Results

To quantify the time-domain performance of an antenna, or a system of two antennas, the most common metric is the Fidelity Factor (FF). The FF is established by the maximum absolute value of the cross-correlation coefficient of the two normalized pulses. It is a useful measure of how much time-domain dispersion an antenna will add to a pulse being transmitted through it.

The antenna and its operating mode determine if the radiated pulse will be radically different to the source pulse. It has been shown that the transient response of an antenna in transmitting mode is proportional to the time derivative of the impulse response of the same antenna in receiving mode (Kanda 1986; Ghosh et al. 2006; Rambabu et al. 2009). Researchers have demonstrated that radiation of UWB pulses involve fields that are time-delayed time-derivatives of the signal currents from the various parts of the transmitting antenna (Siwiak and McKeown 2004). Hence, the simulated radiated pulses have to be cross-correlated with the 1st-order derivative of the excitation pulse.

DOI 10.1007/978-981-4560-75-7_59-1

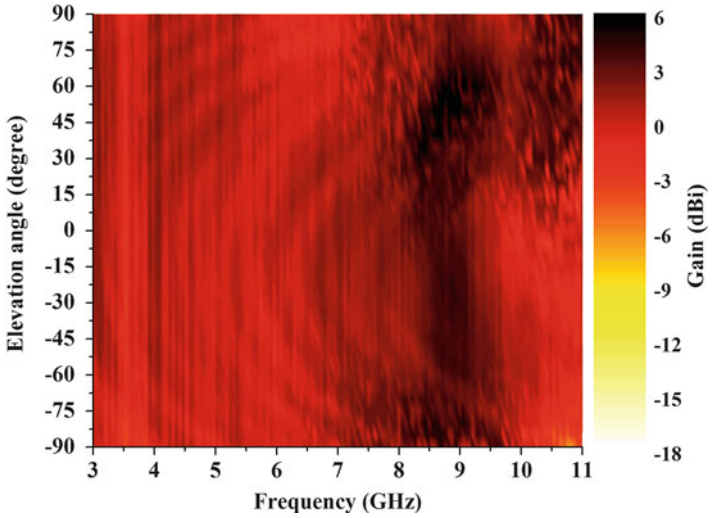


Fig. 15 Measured Dipole GD radiation pattern in the $\varphi = 90^\circ$ plane

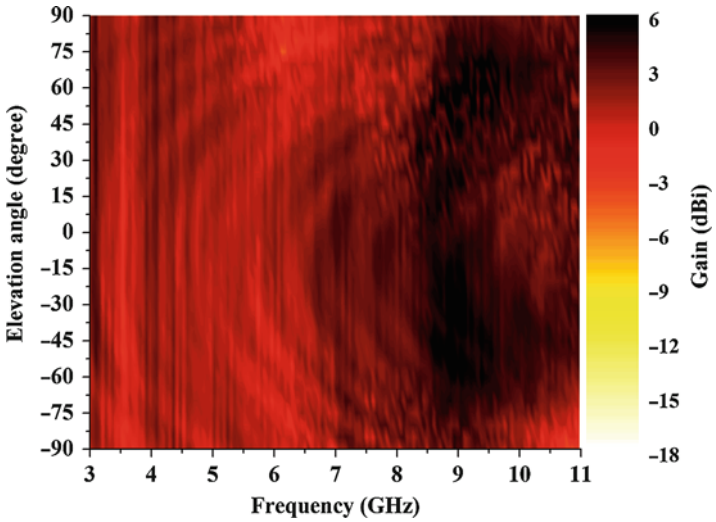


Fig. 16 Measured Dipole SD radiation pattern in the $\varphi = 90^\circ$ plane

Measurement Setup To generate the SRRC UWB pulse, a digital sampled waveform was exported to a Tektronix AWG 7122C Arbitrary Waveform Generator (AWG). Due to its output voltage limitation, the AWG output signal was amplified with a Picosecond Pulse Labs wideband amplifier (Model 5865) and fed to the antenna under test. Also, in order to accurately measure the radiated pulse, the receiving antenna should not be dispersive. Consequently a directional tapered slot antenna was designed and its geometry is shown in Fig. 17. Using the SRRC pulse (see Fig. 10) as an incident signal instead of excitation signal, it was optimised for optimal time-domain performance on bore-sight in receiving mode. In this operating mode, the antenna has a simulated pulse fidelity factor of 98.6 % at the bore-sight. This result indicates that the received signal at the antenna's port is nearly identical to the replica of the incident signal.

A distance of 20 cm was chosen between the transmitting and receiving antenna to maintain the farfield measurement while achieving an adequate signal to noise ratio (SNR) because from a distance of 50 cm, the SNR dropped to an unacceptable level, impairing the time-domain measurement. Figure 18, depicts the time-domain measurement setup.

DOI 10.1007/978-981-4560-75-7_59-1

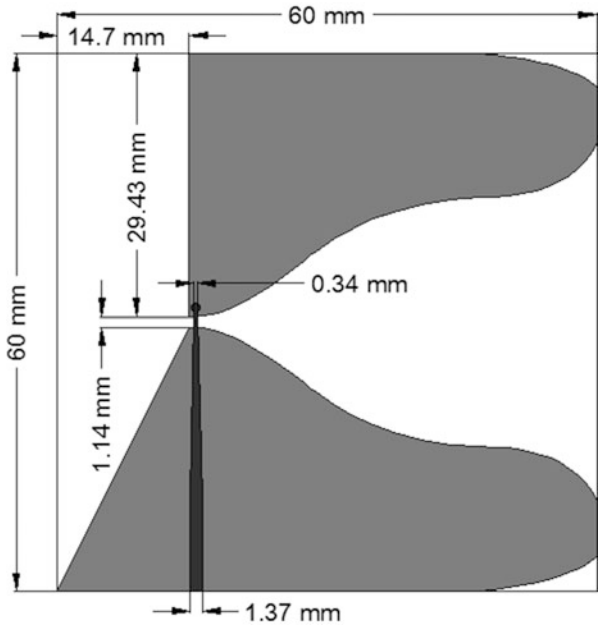


Fig. 17 Tapered slot antenna optimised for time-domain performance in receiving mode

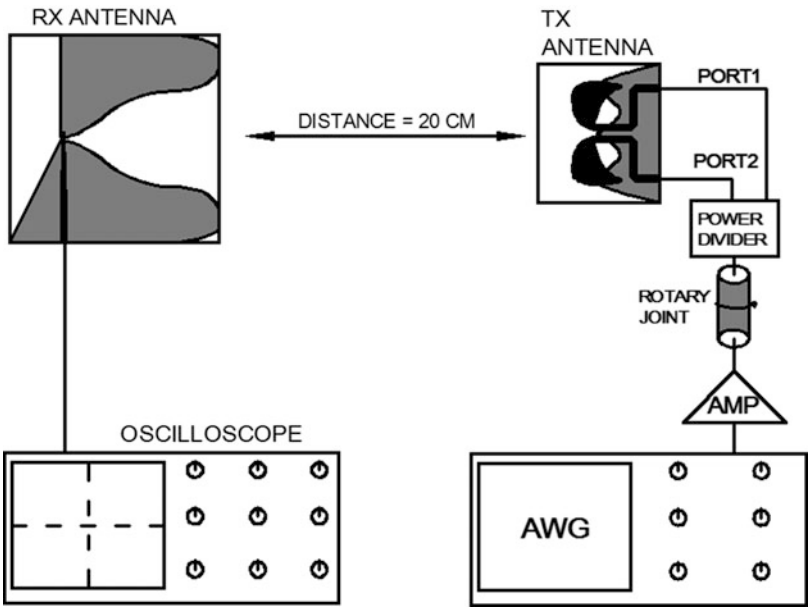


Fig. 18 Time-domain measurement setup

Pulse Equalisation

Because of the imperfect impulse response of the transmit network components, the AWG generated SRRC pulse fed at the AUT's input port suffers from impaired frequency- and time-domain characteristics. If such a signal is fed to the transmit antenna, the measured pulse fidelity will not represent the full transient performance of the AUT.

Figure 19 illustrates the AWG input waveform *in(t)* with the pulse *out(t)* offered to the transmit antenna. To subdue the component losses, the input signal was compensated using Eq.4, where the Inverse Fast

DOI 10.1007/978-981-4560-75-7_59-1

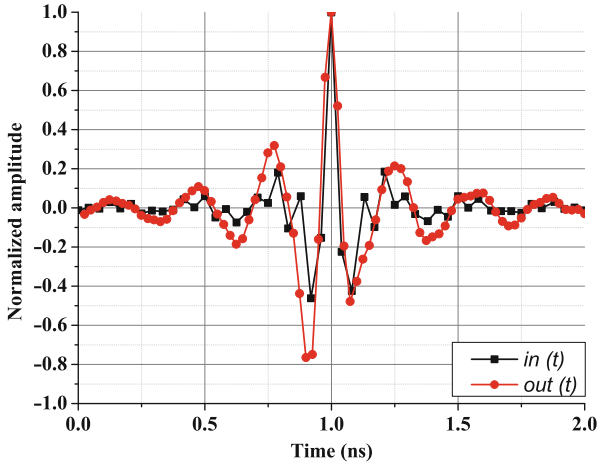


Fig. 19 *in(t)* waveform to AWG and *out(t)* pulse

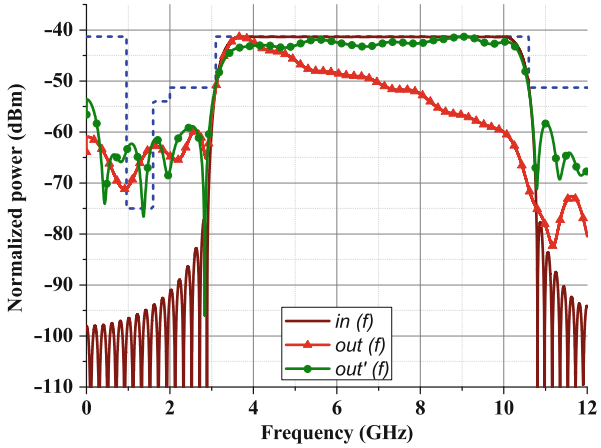


Fig. 20 Input, output and output equalized pulse Power Spectrum

Fourier Transform (IFFT) and the Fast Fourier Transform (FFT) are used to manipulate the pulse in frequency- and time-domain. In Fig. 20, the power spectrum density of the compensated signal *out'(t)*, the input signal *in(t)* and the initial output signal *out(t)* are shown.

$$in'(t) = IFFT\left[FFT(in(t)) / FFT(out(t))\right] \tag{4}$$

Compared to the original SRRC signal *in(t)*, the compensated signal *out'(t)* energy distribution level share many characteristics in the UWB band, demonstrating that the transmit network impulse response was equalised. Hence, the *out'(t)* signal was fed at the transmitting antenna port allowing an accurate time-domain characterisation of the AUT.

Antenna Impulse Response De-embedding

In the last section, an equalisation method was presented, suppressing the transmit network impulse response from the measurement. However, in order to get an accurate measure of the radiated signal, the receive network impulse response, which includes the one from the receiving antenna, should also be removed.

DOI 10.1007/978-981-4560-75-7_59-1

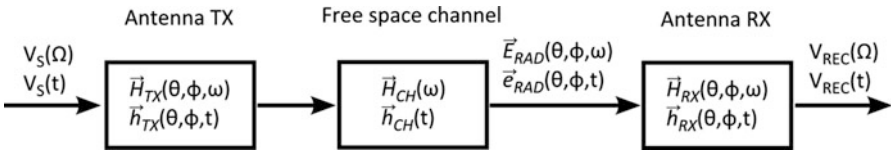


Fig. 21 UWB channel model

Using two identical antennas, in an antenna system configuration, and by using the method described by (Duroc et al. 2007), it is possible to estimate the transient and frequency response of the AUT. Indeed, the UWB channel model can be decomposed in three different blocks, consisting of the transmitting antenna, the free space channel, and the receiving antenna. Each block can be described by a transfer function such as, $\vec{H}_{TX}(\theta,\varphi,\omega), H_{CH}(\ \Omega)$and $\vec{H}_{RX}(\theta,\varphi,\omega)$, as shown on Fig. 21.

Hence the UWB channel can be completely described in both frequency- and time-domain as shown in Eqs.5 and 6 (Kanda 1986).

$$\frac{V_{REC}\ (\omega)}{V_S\ (\omega)} = \vec{H}_{TX}(\theta,\varphi,\omega) \times H_{CH}(\omega)\ \times\ \vec{H}_{RX}(\theta,\varphi,\omega) \tag{5}$$

$$V_{REC}(t) = \vec{h}_{TX}(\theta,\varphi,t) * h_{CH}(t)\ *\ \vec{h}_{RX}(\theta,\varphi,t) * V_S(t) \tag{6}$$

where Ω is the angular frequency, and the symbol $*$ is the convolution symbol. The free space channel transfer function can be defined using Eq.7, where c is the speed of light in free space, d is the antenna separation and λ is the wavelength.

$$H_{CH}(\omega) = \frac{\lambda}{2d}\exp\left(-j\omega\frac{d}{c}\right) \tag{7}$$

Then, by measuring the vectorial S21 of two identical antennas with same orientation, it is possible to deduce the transfer function of the antenna in transmitting and receiving mode as shown in Eqs. 8 and 9.

$$\vec{H}_{TX}(\theta,\varphi,\omega) = \sqrt{\left(\frac{\vec{S}_{21}\ (\theta,\varphi,\omega)}{H_{CH}(\omega)}\ \frac{j}{\lambda}\right)} \tag{8}$$

$$\vec{H}_{RX}(\theta,\varphi,\omega) = \sqrt{\left(\frac{\vec{S}_{21}\ (\theta,\varphi,\omega)}{H_{CH}(\omega)}\ \frac{\lambda}{j}\right)} \tag{9}$$

By knowing the transfer function of the receiving antenna, the free space channel and the antenna system's S_{21}, the transfer function of the unknown transmitting antenna can be calculated has shown on Eq.10.

$$\vec{H}_{TX}\ (\theta,\varphi,\omega) = \frac{\vec{S}_{21}(\theta,\varphi,\omega)}{H_{CH}(\omega)\ \times\ \vec{H}_{RX}(\theta,\varphi,\omega)} \tag{10}$$

DOI 10.1007/978-981-4560-75-7_59-1

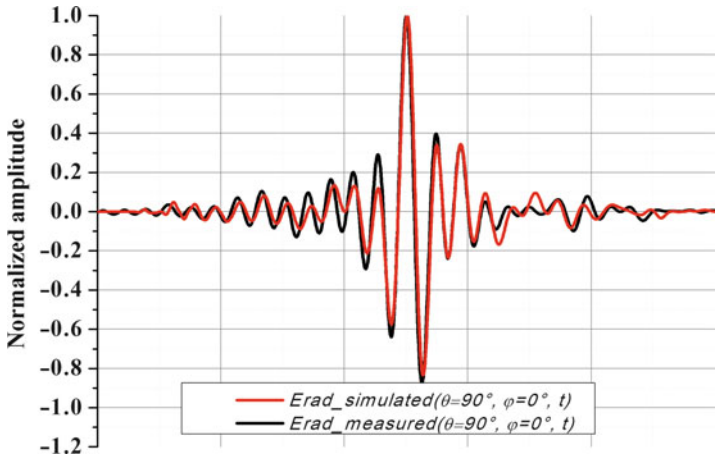


Fig. 22 Simulated and measured radiated pulse

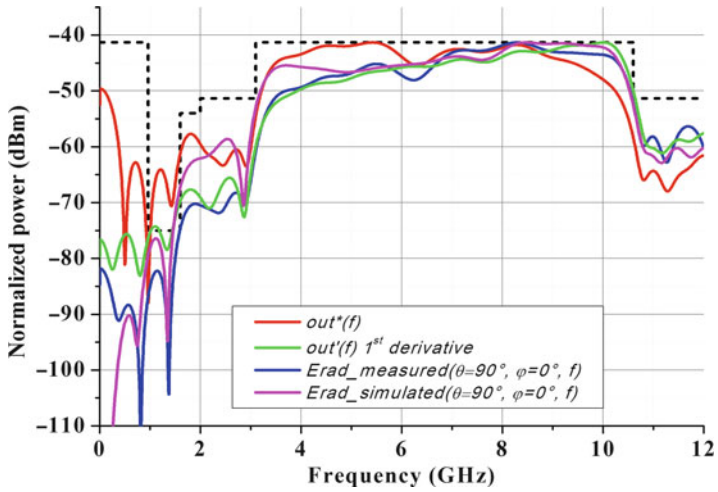


Fig. 23 Received output signal *out*(t)*, 1st derivative pulse *out'(t)*, simulated radiated signal $\vec{e}_{RAD\_simulated}(\theta = 90°, \varphi = 0°, t)$, measured radiated signal $\vec{e}_{RAD\_measured}(\theta = 90°, \varphi = 0°, t)$, normalised power spectrum density

Hence the incident radiated pulse $\vec{e}_{RAD}(\theta, \varphi, t)$ at the receiving antenna can be deduced by using Eq.11

$$\vec{e}_{RAD}(\theta, \varphi, t) = \vec{h}_{RX}(t) \overline{*} \left[IFFT \left(\vec{S}_{21}(\theta, \varphi, \omega) \right) \right] \tag{11}$$

where the symbol $\overline{*}$ stands for the deconvolution. Using this method, the measured radiated pulse from the transmitting antenna can be compared with the simulated radiated field captured from the virtual probe in the simulation model. It is seen that the measured radiated signal $\vec{e}_{RAD\_measured}(\theta = 90°, \varphi = 0°, t)$ is in good agreement with the simulated radiated signal $\vec{e}_{RAD\_simulated}(\theta = 90°, \varphi = 0°, t)$ as seen in Fig. 22. Figure 23 demonstrates that the signal $\vec{E}_{RAD\_measured}(\theta = 90°, \varphi = 0°, f)$ energy distribution shares similarities, in the UWB frequency range, with the energy spectrum of the $\vec{E}_{RAD\_simulated}(\theta = 90°, \varphi = 0°, f)$ and the 1st order derivative of the reference signal *out'(f)*,

DOI 10.1007/978-981-4560-75-7_59-1

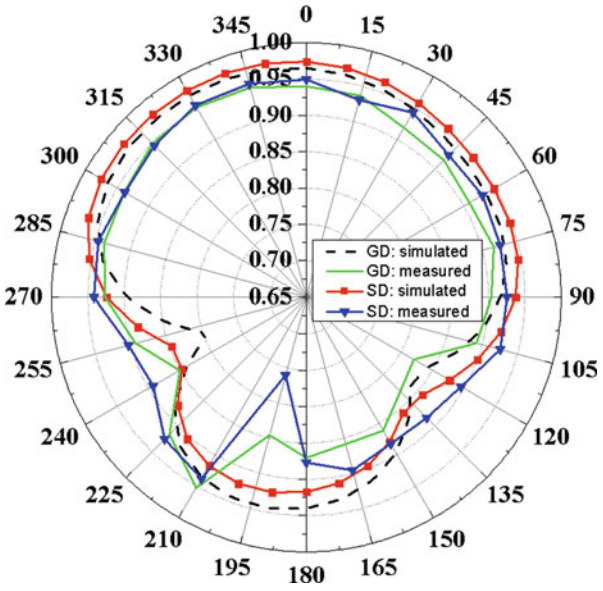


Fig. 24 Measured and simulated FF for an antenna system for Dipole GD and Dipole SD with an equalized modulated SRRC excitation pulse

demonstrating that the transmitting antenna tends to radiate a signal that is approximately proportional to the reference signal derivative at $\theta = 90°$, $\varphi = 0°$.

The transmitting antenna was rotated through 15° angle steps in the radiation plane. The oscilloscope was set to 64-sweep averaging to mitigate spurious sampling artefacts in the measured waveforms. Separate equalization eliminated the dispersive effects due to the cables between the reference antenna and the oscilloscope. The resultant measurement data for the antenna under test was then equivalent to the simulated field-probe data.

The FF was computed by correlating the post-processed received pulse waveforms with the 1st order derivative of the AWG output pulse out'(t). Figure 24 shows the measured FF polar plot for Dipole GD and Dipole SD alongside the simulated results from a modulated SRRC excitation pulse. Good agreement between simulation and measurement is achieved. Furthermore, the measurements validate that an antenna optimized for the full SRRC UWB pulse (Dipole SD) has better time-domain performance for the full UWB range than an antenna optimized with a pulse that does not fit to the FCC UWB spectral mask (Dipole GD).

UWB Localization and Tracking Applications

Impulse Radio UWB (IR-UWB) technology addresses one of the most topical frontiers of emerging wireless communication: precise indoor location and tracking. It allows both accurate measurement of time and data communications to occur simultaneously and so enable a wide variety of applications by developers of real-time location and indoor positioning systems, location-based services, Internet of Things and Wireless Sensor Networks. In healthcare applications doctors and equipment can be located to within centimetres, or children and elderly people can be monitored for their safety and security. In the automotive sector motorists can open and start their cars by simply approaching them. Defence and security can take advantage of weapons that are programmed to respond only to the licensed holder, and military personnel in the field can be tracked. IR-UWB radio devices are currently commercialised (DecaWave, TimeDomain, Zebra Technologies to cite a few) and overcome the intrinsic problems of

DOI 10.1007/978-981-4560-75-7_59-1

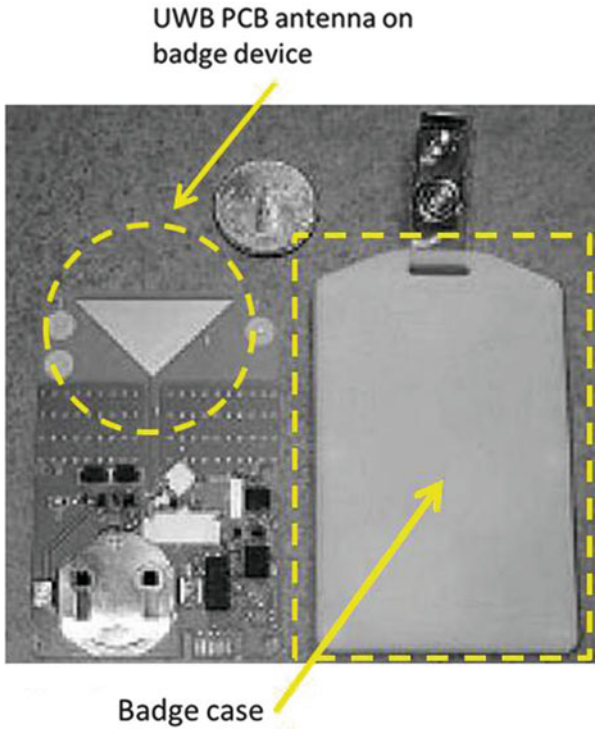


Fig. 25 Tag for IR-UWB personnel tracking (Zebra Technologies DartTag Badge device)

narrowband radio schemes such as high power consumption and poor performance in multipath environments. IR-UWB technology allows longer battery operation and it also enables higher degrees of flexibility in the deployment of wireless networks and sensor positioning, together with increased communication reliability.

Real Time Location Systems are used to track and identify asset tags in real time generally using simple, inexpensive nodes (i.e., badges or tags) attached to or embedded in assets. These tags transmit wireless signals to devices called anchors or readers which employ the wireless signals to determine their location. Current commercial products provide a precision of 10 cm at a distance of up to 300 m in Line-of-Sight (LOS) mode even when devices are moving at speed. High node density is also supported with the accurate location of up to 11,000 items to such precision in a 20 m radius.

Dedicated antenna design is required to deliver this performance. On the tag side, omni-directionality and high compactness are desirable together with pulse fidelity and performance resilience in diverse operating configurations (i.e., close to human body or attached to different materials). But also easy antenna integration on the tag PCB is a required characteristic enabling further miniaturization. Monopole and dipole antennas are a convenient compromise in terms of operating bandwidth, omnidirectional radiation pattern and compactness when integrated on PCB.

In Fig. 25 a commercial product from Zebra Technologies Ltd. for personnel tracking is shown (Online http://www.zebra.com/us/en/products-services/location-solutions/dart-uwb/dart-tag.html). By printing the triangular monopole on the PCB, good compactness and cost reduction are achieved. The tag is approximately the size of a credit card (ID-1 format, 85.60 × 53.98 mm).

Figure 26 shows a tag where a ceramic UWB antenna is integrated on the PCB with the chip. The chip itself together with the metallized board acts as ground for the bevelled monopole (Online http://www.decawave.com/products/dwm1000-module).

While larger antenna solutions are afforded on the anchor device, the requirement for high gain and minimum group delay variation across frequency and angular beamwidth makes the design very

DOI 10.1007/978-981-4560-75-7_59-1

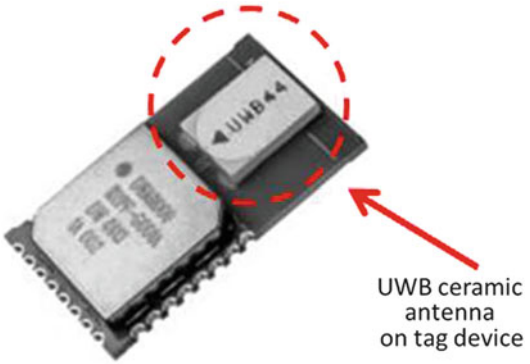


Fig. 26 Tag for IR-UWB radio localization and tracking (DecaWave DWM1000 Module)

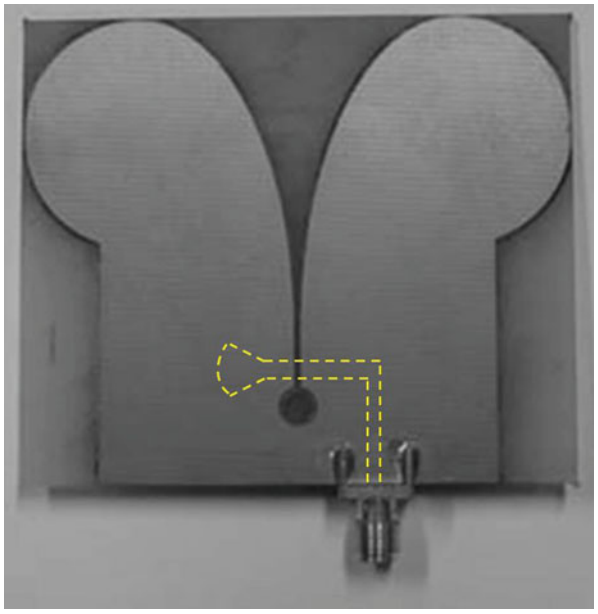


Fig. 27 Modified Vivaldi antenna developed at Queen Mary, University of London

challenging. More directive antennas at the anchor station enable the coverage of larger operating areas. Printed Vivaldi antennas are good candidate for this application due to their large operating bandwidth and directivity.

This antenna with an exponentially tapered slot enables a good trade-off between bandwidth and minimal length. Figure 27 shows a variation of a printed Vivaldi antenna, developed at Queen Mary College, University of London (Valderas et al. 2011). The edge of both arms was rounded to further improve the bandwidth and minimize distortion from the arm-extremities. The antenna is fed through a microstrip to a slot line transformer. The direction of maximum radiation in the Vivaldi antenna corresponds to the aperture between the two arms. In order to reduce the total length of printed tapered slot antennas, variations in the arm profiles were proposed. For instance, the bunny ear antenna (Fig. 28) combines an exponentially tapered aperture with a slot line which is designed to match the 50 Ω input impedance. The element is symmetrically printed on both sides of the substrate and can be easily integrated into a dual-polarized antenna with two centrally intersected PCB slabs (Lee and Livingston 1993).

DOI 10.1007/978-981-4560-75-7_59-1

Fig. 28 Bunny-ear antenna

In their stand-alone conventional configuration, dipole antennas have omnidirectional radiation patterns in the H-plane. The presence of a reflector restricts the operating frequency range of the antenna as it must be placed at a quarter-wavelength distance from the dipole. However, several reported dipole-based antennas have been recently proposed to combine UWB performance, low-profile design and directional radiation patterns (Fig. 29). Examples of band enhancement techniques include the integration of composite reflectors such as E- and H-type corrugated reflectors to support surface currents across a large frequency range (Wu et al. 2010) and minimize destructive current images. In (Wang et al. 2012) a reversed T-match dipole with a flat ground plane was introduced achieving a 77 % operating bandwidth. Small variation input impedance, large bandwidth and stable radiation patterns were reported in (Qu et al. 2009) using a cavity-backed folded triangular bowtie antenna with a rimmed circular reflector.

Circularly Polarized (CP) antennas have been used in a number of radar, communication and navigation systems for many years to mitigate polarization mismatch and reduce orientation constraints between the transmitting and receiving antennas. Recently, systems capable of transmitting and receiving circularly polarized UWB waves were proposed for accurate radio-localization. The objective is to combine the natural immunity of CP transmit/receive systems to arbitrary polarisation alignment with high spatial and temporal resolutions provided by UWB short pulse signals (Shlivinski 2009). But this creates two significant challenges for the antenna engineer: increasing circular polarization across a large operating bandwidth while satisfactory time-domain performance is provided. In Qing et al. (2007) use a three-stub hybrid coupler to provide controlled power division to two orthogonal UWB slot antennas over the frequency range 2.8–5.2 GHz, shown in Fig 30.

The investigation is carried out in terms of typical frequency-domain parameters to characterize CP antennas (i.e., impedance matching, gain, axial ratio, radiation patterns) with further analysis in the time-domain in terms of pulse-distortion, group delay variation and fidelity.

The same operating principle of directing equal signals to two orthogonal UWB antennas is also adopted by Narbudowicz et al. in (2012) but applied to directive antipodal Vivaldi antennas. The input signal is split using a 3-stage Wilkinson divider to provide smooth transmission with minimal variation between antenna feeds. The desired 90° phase shift between the antenna feeds is obtained in two stages by

DOI 10.1007/978-981-4560-75-7_59-1

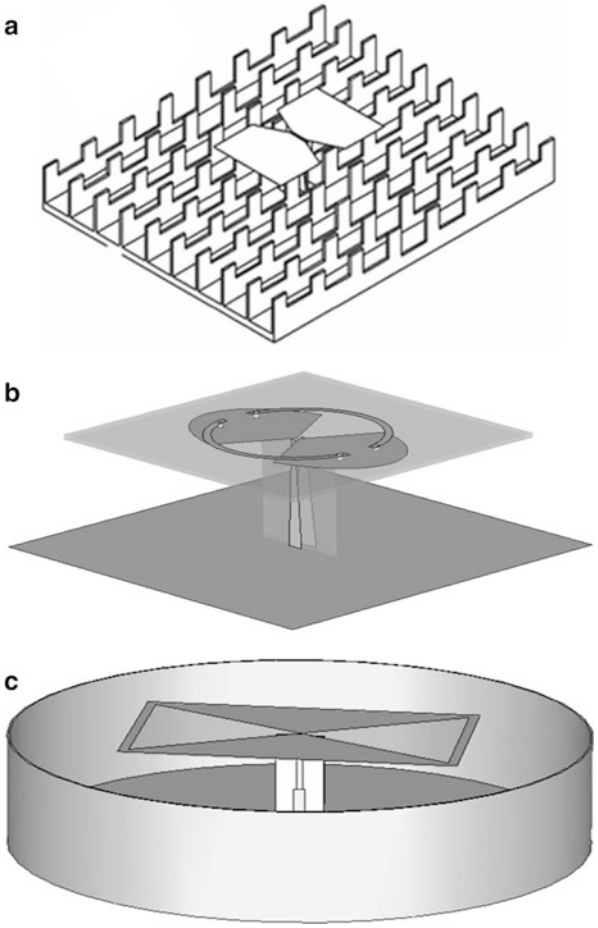


Fig. 29 Bandwidth enhancement techniques for directive dipole antennas. (a) E- and H-type corrugated reflector; (b) Reversed T-match dipole; (c) cavity-backed folded triangular bowtie antenna

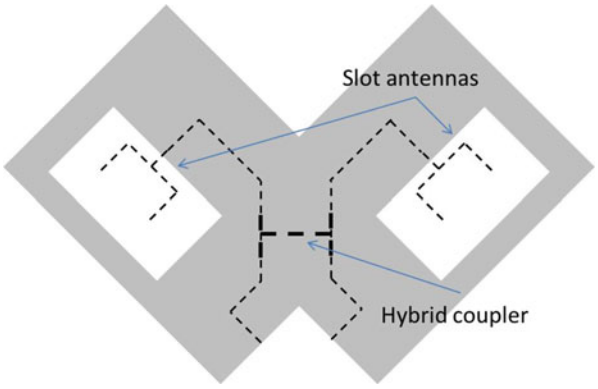


Fig. 30 Ultra-wideband Circularly Polarized Wide-slot Antenna Fed by Three-stub Hybrid Coupler

coupling the signal through the elliptical slot in the ground plane to a microstip line located on the rear side of the dielectric (Fig. 31). Although the mismatch of the two phase centers generates a small tilt in the main beam, this can be mitigated by placing the antennas as close as possible.

DOI 10.1007/978-981-4560-75-7_59-1

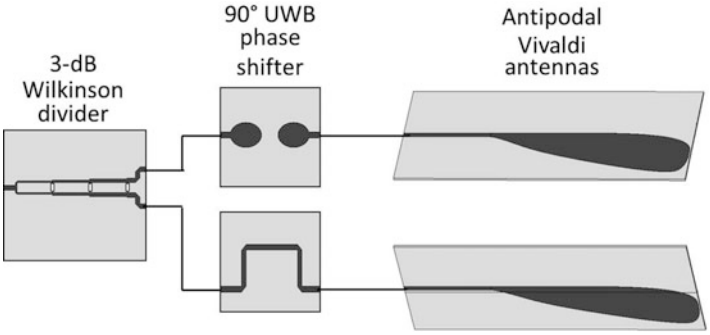


Fig. 31 Vivaldi array for generation of directional UWB circular polarization

Biomedical UWB Imaging

Microwaves and millimeter waves have been widely used as a successful and low-cost imaging method for dielectric structures. In this spectrum, electromagnetic waves have wavelengths ranging from a millimeter to a meter and can penetrate into several optically opaque media. In particular, when UWB radar technology is used, target detection can be achieved with high accuracy. Compared to narrowband systems, UWB radar offers greater imaging resolution and improved target characterization. Recently, UWB radar technology has been successfully proposed in biomedical diagnostics.

UWB microwave imaging techniques for biomedical scenarios are based on the transmission of low-power short microwave pulses (tens picoseconds to nanoseconds) into the body and the consequent detection of the differentiated scattering due to the dielectric contrast between the target and healthy tissues. The advantage of using short pulses is due to easier identification of the target contribution in the received scattered signal. This consists of the incident signal together with a large number of delayed and attenuated replicas from dielectric discontinuities inside human tissue. Narrowband antennas would generate larger pulses and overlapping signal replicas which can hide the informative signals coming from the target. The short pulses are advantageous in detecting a target in human tissue, but the UWB antenna design presents several challenges.

Human tissue such as skin, muscle, fat, brain, and many others present very different electrical properties from those of air. Starting with the skin layer with a relative permittivity around 30 and a specific conductivity of about 3 S/m in the low GHz range, deeper layers of fat, muscle, fibro-glandular and connective tissues contribute to the rejection of the microwave power due to reflection and dissipation. Moreover, the electrical properties of tissue exhibit strong dispersion with frequency in the RF and microwave spectra. In this difficult propagation scenario the electromagnetic signal is subjected to strong attenuation and distortion. Moreover, the attenuation between incident and penetrated signal challenges the dynamic range of the measurement setup and the UWB signal distortion can make the imaging formation process extremely difficult, resulting in clutter which masks the signal scattered by the target. The attenuation and distortion limitations tolerable by the imaging system translate into the following requirements for a successful diagnostic procedure:

- The system must provide appropriate penetration into the body area under assessment;
- The coupling mechanism between sources, sensors and the body must be stable across the operating bandwidth and screening configurations.

The design of the UWB antenna plays a vital role. The first consideration comes from the strong rejection of electromagnetic signals by skin and deeper layers for frequencies above 10 GHz. A few

DOI 10.1007/978-981-4560-75-7_59-1

centimeters of tissue along the direction of propagation can attenuate a 10 GHz signal by more than 80 dB. Good signal penetration can be achieved at lower frequencies (i.e., below 1 GHz). However, the allocation of the operating bandwidth is compromised by the requirement for compact antennas. The diagnostic device design involves the integration of small antennas to facilitate manoeuvrability or array integration. Size reduction is often achieved by immersing the antenna into a liquid solution with a dielectric constant similar to that of tissue. This medium provides an impedance transformation between the antenna and the body, improving tissue coupling.

Another important constraint for the designer is to provide resilience of antenna performance under all operating configurations. In particular, adequate impedance matching must be maintained with the antenna in close proximity to different body areas and for different antenna-skin separation. Suitable antenna designs should be capable of screening interference by backscatter from undesired directions.

Frequency-domain characterization of antennas is not appropriate for biomedical imaging applications where near-field conditions occur. Parameters such as gain, radiation efficiency, phase center and radiation pattern refer to far-field conditions and are not suitable in scenarios where the body strongly couples to reactive fields. These parameters rapidly change when the antenna is in close proximity to the body as in typical medical imaging configurations. The dramatic dependence of the antenna behaviour on scanning position and the body area necessitates alternative quality figures. This is validated by numerical tools which simulate the complex electromagnetic problems in the presence of highly detailed voxel models which reproduce the morphology and the electrical dispersion of benign and malignant tissues. But tissue-mimicking phantoms can also enable ad-hoc antenna characterization at the experimental stage.

Although in realistic scenarios tissue morphology changes from patient to patient with consequent variations in antenna response, specific parameters of merit were introduced to characterize antennas in biomedical imaging applications. For this purpose, standard testing configurations were proposed to combine the requirement of approximating a realistic operating situation in a controlled environment (Fig. 32). The following parameters are able to characterize the antenna in terms of properties which are useful for a near-field UWB radar imaging application.

Fidelity Factor: calculated as the maximum magnitude of the cross correlation between the normalized transmitted and the normalized received response across tissue-mimicking multi-layer structures.

Pulse duration: For radar-based breast cancer detection it is important that pulses produced by the antenna are as short as possible with minimal late-time ringing (Gibbins et al. 2010). Therefore pulse duration is a good indication of the ability of the antenna to effectively transmit UWB signals into the medium. The time taken for 99 % of the energy of these pulses to be received can be calculated for different antenna polarizations and across a defined angular beamwidth.

Half Energy Beam (HEB) and *Half Energy Beamwidth (HEBW)*: The HEB and HEBW are based on the Energy Flux Density (EFD) which is calculated by summing time samples of the instantaneous Poynting vector over the duration of the simulation time (Bourqui et al. 2010). In the near-field, the HEB is defined on a plane which is orthogonal to the main radiation beam and situated at a given distance from the antenna aperture as shown in Fig. 33. The HEB describes the region over which the energy is greater than half of the maximum value on the selected plane. Accordingly, the HEBW is defined by the angular width of this region.

Among different screening typologies under investigation, breast cancer imaging has advanced faster and is currently at a clinical trial stages. Hagness et al. pioneered this imaging modality by proposing small UWB microwave radar systems to detect malignant tumors in the breast (Hagness et al. 1998). Encouraging outcomes and the critical necessity for developing early-stage breast cancer screening tools with high sensitivity and specificity prompted significant research efforts which have grown into a well-established microwave application (Hagness et al. 2012). Microwave imaging of breast cancer is either

DOI 10.1007/978-981-4560-75-7_59-1

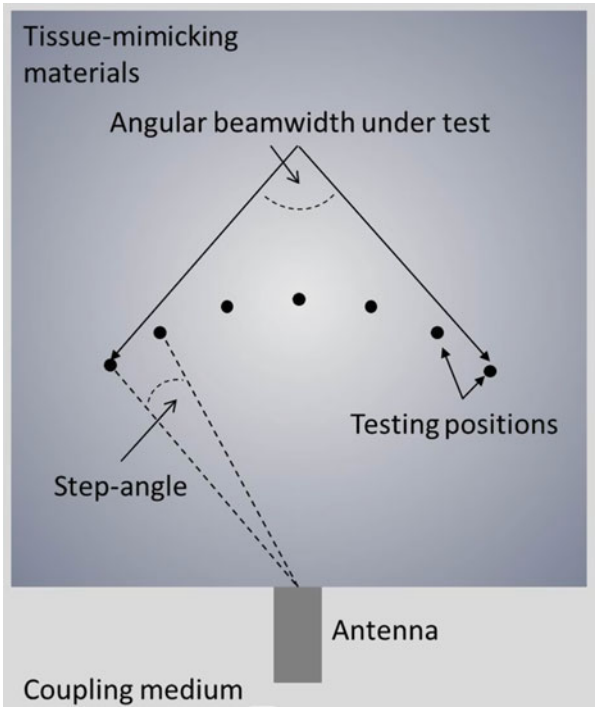


Fig. 32 Testing configuration for imaging-specific parameters

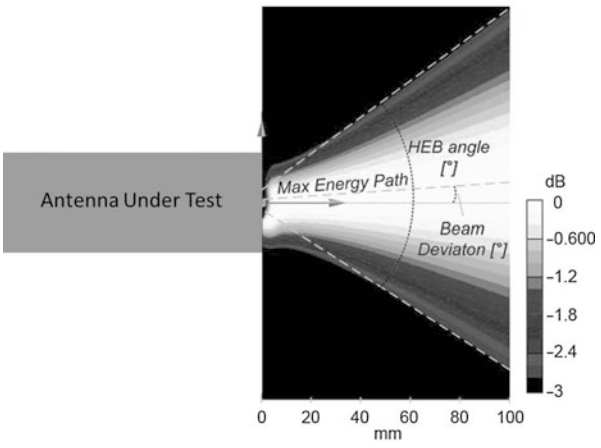


Fig. 33 Half energy area in one plane and HEB representation

based on mapping the permittivity and conductivity distributions as functions of position or the simple detection of malignant tissues. With both approaches measured data are collected by antennas in specific positions and then processed in search of scatterers, i.e., objects with properties which differ from those in the normal or healthy state of the imaged body region. Accordingly, the more electrically similar a tumor is to its surrounding benign tissues, the more challenging it is to detect it. The dielectric contrast between malignant and benign tissues is generally defined as the ratio, often in percentage, of the permittivities (or the conductivities) of the scatterer and the host medium. Recent studies (Sha et al. 2002; Lazebnik et al. 2007a, b; Meaney et al. 2007; Poplack et al. 2007; Halter et al. 2009) indicate that while contrasts

DOI 10.1007/978-981-4560-75-7_59-1

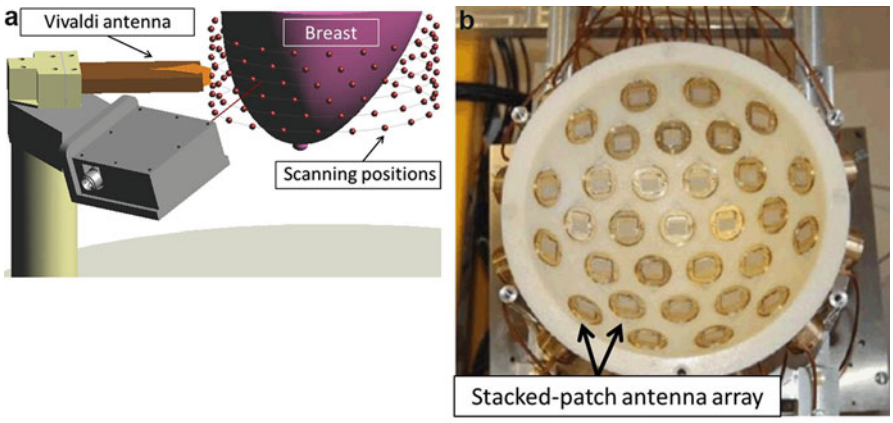


Fig. 34 Breast cancer imaging systems. (**a**) Monostatic scanner at the University of Calgary, Canada; (**b**) Multistatic scanner at the University of Bristol, UK

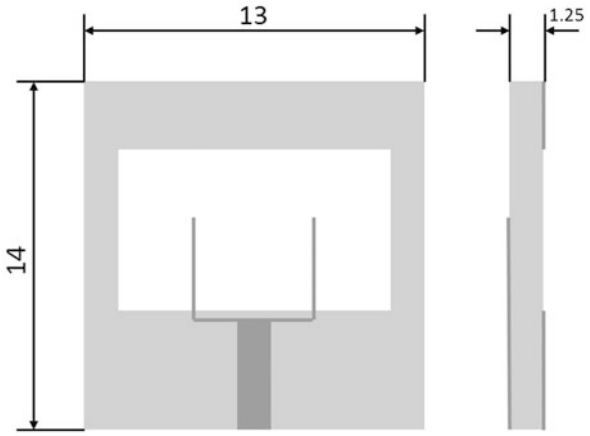


Fig. 35 Geometry of the wide-slot antenna developed at the University of Bristol. All dimensions are in mm

between malignant and adipose breast tissue may be as large as 10, those between malignant and healthy fibroglandular tissues can be as low as 1.1 in both permittivity and conductivity.

Both mono- and multi-static radar architectures have been proposed for microwave breast cancer imaging (Fig. 34). Whether the antenna is a fixed array element displaced around the body area or the antenna moves across different screening positions, its performance must be stable for all configurations.

Numerous antenna designs for radar-based UWB microwave breast imaging have been proposed over the past 15 years. A wide slot antenna with interesting features for UWB cancer detection was developed at the University of Bristol (Gibbins et al. 2010) (Fig. 35). This antenna is surface-mounted on a hemispherical bowl to form a 31-element array and function as a multi-static system with 465 independent measurements (Klemm et al. 2011). The imaging system operates in the full UWB frequency range, between 3 and 10 GHz. Both these antennas mainly consist of printed dielectric substrates with a relative permittivity of 10.2. The geometry was optimized to operate with the antenna face immersed in a matching medium with dielectric properties that mimics that of human breast tissue, in order to reduce reflections by eliminating the air/skin interface (Sill and Fear 2005). The matching medium used is mainly a solution made of paraffin oil and distilled water. This mixture has a relative permittivity of between 9 and 10 in the 2–10 GHz region and an attenuation of 2 dB/cm at 8 GHz. The wide-slot antenna developed at

DOI 10.1007/978-981-4560-75-7_59-1

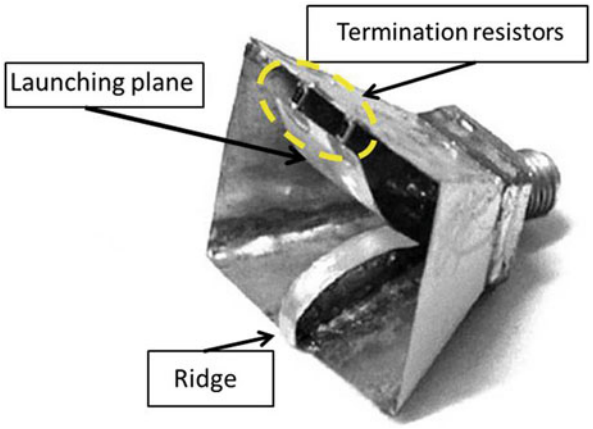


Fig. 36 Modified ridged pyramidal horn antenna with curved launching plane developed at the University of Wisconsin

the University of Bristol consists of an approximately square slot cut out from a ground plane which is printed on one side of a substrate with relative permittivity of 10.2. The slot is fed by a forked microstrip feed that splits just below the slot from a 50 Ω feed into two 100 Ω sections. The fork feed excites multiple resonances with a consequent increase of the impedance matching bandwidth.

A cavity is attached to the antenna on the side where the microstrip feed is located. This small cavity prevents any radiation from the antenna rear side (opposite to skin interface) coupling with other elements in the array. The cavity is absorber lined and also stops interference from spurious radiation on the imaging formation process. The antenna provides suitable bandwidth for use in UWB detection systems. In consideration of the complex array architecture where the antenna operates, inter-element transmission properties were also investigated. The wide-slot antenna presents a stable transfer function over a large angular range across bore-sight direction. This results in a fidelity factor better than 90 % for angles up to 60° from boresight and reduced late-time ringing and distortion of transmitted signals.

For monostatic data collection more voluminous three-dimensional antenna designs may be suitable. A ridged pyramidal horn antenna (Li et al. 2003) was reported by the University of Wisconsin, and resistively loaded dipole, tapered slot, transverse electromagnetic horn, and balanced antipodal Vivaldi antennas have been presented by the University of Calgary (Bourqui et al. 2009). Li et al. proposed a modified ridged pyramidal horn antenna with a curved launching plane (Fig. 36). The bandwidth of horn antennas can be increased significantly by adding metallic ridges to the waveguide and flared sections (Walton and Sundberg 1964). In Li's antenna one of the two ridges is replaced by a curved metallic plane terminated by two 100-Ω resistors that suppress reflections from the end of the launching plane (Fig. 36). The waveguide section was eliminated to reduce size. The pyramidal horn has a depth of 13 mm with a 25 mm × 20 mm aperture and is connected to the outer conductor of the coaxial feed and serves as the ground plane, providing a current return path. The maximum width of the tapered launching plane is 12 mm, whereas the thickness of the ridge is 2 mm. The antenna is made of brass and fed by an SMA connector. Thickness, curvature and shape of the launching plane critically control the input impedance of the antenna. But its tapered profile prevents reflections which can distort the pulse.

In 2010 Bourqui et al. proposed a modified balanced antipodal Vivaldi antenna (BAVA) which includes a profiled high dielectric section placed in the antenna aperture (Bourqui et al. 2010). The BAVA provides a more compact profile combined with near-field focusing of the energy in the endfire direction. This antenna was designed to operate in a monostatic scanner immersed in canola oil. From 1 to 14 GHz, the permittivity and conductivity of this coupling medium varies from 2.55 to 2.35 and 0.01–0.04 S/m, respectively. The antenna consists of three copper layers; the two external layers are the ground planes and the central layer is the conductor (Fig. 37). The copper layers are printed on two dielectric substrates

DOI 10.1007/978-981-4560-75-7_59-1

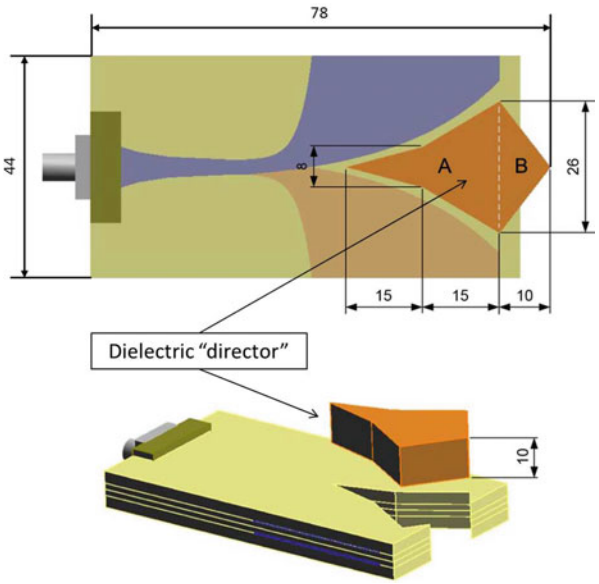


Fig. 37 BAVA including a higher dielectric permittivity director developed at the University of Calgary. All dimensions are in mm

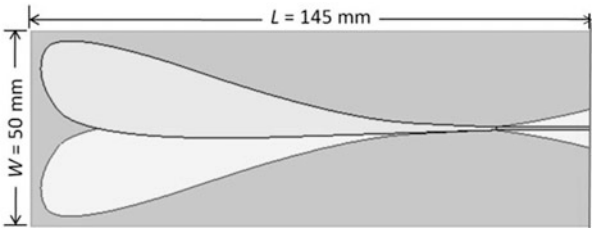


Fig. 38 ParEGO optimized antipodal Vivaldi antenna developed at the Antenna & High Frequency Research Centre, Dublin Institute of Technology, Ireland. All dimensions are in mm

(supportive substrates) and two additional dielectric layers are stacked on each side of the antenna (stacking substrates) to balance the dielectric loading between the conductor and ground planes. As a result of this solution, the usual beam squint observed across frequency in this type of antenna (Langley et al. 1996) is considerably reduced. Moreover, for this application, the energy coupling is improved since contact between the lossy canola oil and the external copper layer is avoided. The director is made of Eccostock HIK material with a relative permittivity of 6. The shape of the director is designed to prevent reflections from its extremities and to enhance near-field focusing. The effect of such narrower beamwidth was verified with the transmitted energy between two antennas being significantly greater in the presence of the director. Furthermore, it was demonstrated that the presence of the director in the aperture of the BAVA increases the backscatter energy from an object placed in front of the antenna, such as a tumor embedded in the breast.

Selective illumination of the breast may assist in tumor localization. In the context of microwave imaging applications, a more focused beam illuminates a smaller volume which can increase the tumor-to-clutter ratio. Ruvio et al. compared three antennas with different coupling mechanisms and focusing capability for a UWB mono-static radar illumination of a tumor embedded in a 2-D multi-layer breast-mimicking phantom (Ruvio et al. 2014). In particular, an antipodal Vivaldi antenna shown in Fig. 38 was optimized by means of a ParEGO algorithm to provide a stable transfer function over the operating

DOI 10.1007/978-981-4560-75-7_59-1

bandwidth ranging from 1 to 3 GHz when immersed in a coupling medium with permittivity of 12. Both numerical and experimental results show the antipodal Vivaldi antenna outperforms a coplanar and a semi-folded monopole as it enables more pronounced clutter mitigation in reconstructions.

Future Directions and Open Problems

The main research activity on UWB antennas is currently driven by emerging applications in biomedicine and security. Since UWB imaging provides low-cost, high resolution imaging, good target recognition and robust immunity to passive jamming, it has the potential to replace or complement the ionizing radiation used in conventional non-invasive imaging applications. Emerging biomedical applications include UWB microwave impedance spectroscopy (non-invasive measurement of glucose and cholesterol levels in blood) and vital signs detection. Security applications include a low-cost alternative detection of concealed metal objects, as well as through-the-wall imaging, body-area networks and asset-tracking. Automotive applications include vehicle entry systems and relay attack avoidance. In the agricultural sector, livestock tracking and feed control methods using UWB are evolving.

Some of the main challenges facing UWB antenna design relate to the environment. For example, close proximity to the body realises dielectric loading which is highly sensitive to proximity distance. There is demand for antennas which are resilient to this loading across all applications. New parameters are required to describe UWB antennas when in close proximity with media which are not known a priori (on-body networks, ground penetrating radar, through-the-wall imaging, microwave breast cancer imaging. . .). These parameters would describe antenna resilience to proximity of these media.

In communications applications, UWB antennas may also be employed in multi-band or cognitive radio applications. Of course, UWB antennas are not confined to 3.1–10.6 GHz with activity below (ground penetrating radar) and above this band. High rate communications in the 57–66 GHz band is the focus of research on UWB antennas on-chip packaging (→On-chip Antennas).

Cross-References

- ▶ Broadband and Multiband Planar Antennas
- ▶ Low-profile Antennas
- ▶ Omni-directional Antennas
- ▶ On-chip Antennas
- ▶ Optimization Methods in Antenna Engineering

References

Agrawall NP, Kumar G, Ray KP (1998) Wide-band planar monopole antenna. IEEE Trans Antennas Propag 46(2):294–295

Allen B, Dohler B, Okon EE, Malik WQ, Brown AK, Edwards JD (2007) Introductions to UWB signals and systems. In: Ultra-wideband antennas and propagation for communication radar and imaging. Wiley, Chichester

Ammann MJ (1999) Square planar monopole antenna. IEE National Conference on Antennas and Propagation:37–40

Ammann MJ (2001) Control of the impedance bandwidth of wideband planar monopole antennas using a bevelling technique. Microw Opt Technol Lett 30(4):229–232

Ammann MJ, Chen ZN (2003a) A wideband shorted planar monopole with bevel. IEEE Trans Antennas Propag 51(4):901–903

Ammann MJ, Chen ZN (2003b) Wideband monopole antennas for multiband wireless systems. IEEE Antennas Propag Mag 45(2):146–150

Angelopoulos ES, Anastopoulos AZ, Kaklamani DI, Alexandridis AA, Lazarakis F, Dangakis K (2006) Circular and elliptical CPW-fed slot and microstrip-fed antennas for ultrawideband applications. IEEE Antennas Wirel Propag Lett 5:294–297

Azevedo S, McEwan TE (1996) Micropower impulse radar. Science & Technology Review: 17–29

Balanis CA (2005) Historical advancement. In: Antenna theory, analysis and design, Thirdth edn. Wiley, Hoboken, p 20

Barret TW (2001) History of ultra wideband communication and radar: part I, UWB. Microw J 44(1):22–54

Bartels RH, Beatty JC, Barsky BA (1996) An introduction to splines for use in computer graphics and geometric modeling. Morgan Kaufmann, Burlington

Bourqui J, Campbell MA, Sill J, Shenouda M, Fear EC (2009) Antenna performance for ultra-wideband microwave imaging. In: IEEE Radio Wireless Symposium, San Diego, US, pp 522–525

Bourqui J, Okoniewski M, Fear EC (2010) Balanced antipodal Vivaldi antenna with dielectric director for near-field microwave imaging. IEEE Trans Antennas Propag 58(7):2318–2326

Cerny P, Mazanek M (2007) Optimization of transient response radiation of printed ultra wideband dipole antennas. Radioengineering 16(2):1–6

Chamaani S, Mirtaheri SA, Abrishamian MS (2011) Improvement of time frequency-domain performance of antipodal Vivaldi antenna using multiobjective particle swarm optimization. IEEE Trans Antennas Propag 59(5):1738–1742

Chen HD, Li JN, Hung YF (2006a) Band-notched ultra-wideband square slot antenna. Microw Opt Technol Lett 48(12):2427–2429

Chen X (2007) Theory of UWB antenna elements. In: Allen B, Dohler B, Okon EE, Malik WQ, Brown AK, Edwards JD (eds) Ultra-wideband antennas and propagation for communications, radar and imaging. Wiley, Chichester, pp 111–145

Chen ZN (2000) Impedance characteristics of planar bow-tie-like monopole antennas. Electron Lett 36(13):1100–1101

Chen ZN, Ammann MJ, Qing XM, Wu XH, See TSP, Cai A (2006b) Planar antennas: promising solutions for microwave UWB applications. Microw Mag 7(6):63–73

Chen ZN, Qing X, Chia MYW (2005) UWB characteristics of disc cone antenna. In: IEEE international workshop on antennas technology, Singapore, pp 97–100

Chen ZN, Wu XH, Li HF, Yang N, Chia MYW (2004) Considerations for source pulses and antennas in UWB radio systems. IEEE Trans Antennas Propag 52(7):1739–1748

Donelli M, Massa A (2005) Computational approach based on a particle swarm optimizer for microwave imaging of two-dimensional dielectric scatterers. IEEE Trans Microwave Theory Tech 53(5):1761–1776

Dubost G, Zisler S (1976) Antennes à large bande. Masson, Paris/New York, pp 128–129

Dumoulin A, John M, Ammann MJ, McEvoy P (2012) Optimized monopole and dipole antennas for UWB asset tag location systems. IEEE Trans Antennas Propag 60(6):2896–2904

Duroc Y, Ghiotto A, Vuong TP, Tedjini S (2007) UWB Antennas: systems with transfer function and impulse response. IEEE Trans Antennas Propag 55(5):1449–1451

DOI 10.1007/978-981-4560-75-7_59-1

Evans JA, Ammann MJ (1999) Planar trapezoidal and pentagonal monopoles with impedance bandwidths in excess of 10:1. In: IEEE international symposium on antennas and propagation, vol 3, Orlando, FL, USA, pp 1558–1561

Foley JD, van Dam A, Feiner SK, Hughes JF, Phillips RL (1993) Introduction to computer graphics. Addison-Wesley, Reading

Ghosh D, De A, Taylor MC, Sarkar TK, Wicks MC, Mokole EL (2006) Transmission and reception by Ultra-Wideband (UWB) antennas. IEEE Antennas Propag Mag 48(5):67–99

Gibbins D, Klemm M, Craddock IJ, Leendertz JA, Preece A, Benjamin R (2010) A comparison of a wide-slot and a stacked patch antenna for the purpose of breast cancer detection. IEEE Trans Antennas Propag 58(3):665–674

Go HC, Jung YW (2004) Multi-band modified fork-shaped microstrip monopole antenna with ground plane including dual-triangle portion. Electron Lett 40(10):575–577

Griffiths LA, Fuse C, Chung YC (2005) BroadBand and multiband antenna design using the genetic algorithm to create amorphous shapes using ellipses. IEEE Trans Antennas Propag 54(10):2776–2782

Guo L, Liang J, Chiau CC, Chen X, Parini CG, Yu J (2007) Performance of UWB disc monopoles in time-domain. Microw Antennas Propag 1(4):955–959

Hagness SC, Fear EC, Massa A (2012) Guest editorial: special cluster on microwave medical imaging. IEEE Antennas Wirel Propag Lett 11:1592–1597

Hagness SC, Taflove A, Bridges JE (1998) Two-dimensional FDTD analysis of a pulsed microwave confocal system for breast cancer detection: fixed-focus and antenna-array s. IEEE Trans Biomed Eng 45:1470–1479

Halter RJ, Zhou T, Meaney PM, Hartov A, Barth RJ, Rosenkranz KM, Wells WA, Kogel CA, Borsic A, Rizzo EJ, Paulsen KD (2009) The correlation of in vivo and ex vivo tissue dielectric properties to validate electromagnetic breast imaging: initial clinical experience. Physiol Meas 30(6):121–136

Hammoud M, Poey P, Colombel F (1993) Matching the input impedance of a broadband disc monopole. Electron Lett 29(4):406–407

Haupt RL (2007) Antenna design with a mixed integer genetic algorithm. IEEE Trans Antennas Propag 55(3):577–582

Honda S, Ito M, Seki H, Jinbo Y (1991) On a broadband disk monopole antenna. Technical Report of Television Society Japan. ROFT 91–55

Honda S, Ito M, Seki H, Jinbo Y (1992) A disk monopole antenna with 1:8 impedance bandwidth and omnidirectional radiation pattern. In: ISAP'92, Sapporo, pp 1145–1148

http://www.decawave.com/products/dwm1000-module

http://www.zebra.com/us/en/products-services/location-solutions/dart-uwb/dart-tag.html

IEEE Computer Society (2011) Part 15.4: low-rate wireless personal area networks (LR-WPANS). IEEE, New York

IEEE Standard 802.15.4a (2007) Part 15.4: wireless medium access control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs)

Jin N, Rahmat-Samii Y (2005) Parallel particle swarm optimization and finite- difference time-domain (PSO/FDTD) algorithm for multiband and wide-band patch antenna designs. IEEE Trans Antennas Propag 53(11):3459–3468

John M, Ammann MJ (2007) Wideband printed monopole design using a genetic algorithm. IEEE Antennas Wirel Propag Lett 6:447–449

John M, Ammann MJ (2009) Antenna optimisation with a computationally efficient multiobjective evolutionary algorithm. IEEE Trans Antennas Propag 57(1):260–263

Johnson JM, Rahmat-Samii Y (1997) Genetic algorithms in engineering electromagnetic. IEEE Antennas Propag Mag 39(4):7–21

Kanda M (1986) Time-domain sensors and radiators. In: Miller EK (ed) Time-domain measurements in electromagnetic. Van Nostrand Reinhold, New York, p 125

Karacolak T, Topsakal E (2006) A double-sided rounded bow-tie antenna (DSRBA) for UWB communications. IEEE Antennas Wirel Propag Lett 5:446–449

Klemm M, Fumeaux C, Baumann D, Craddock IJ (2011) Time-domain simulations of a 31-antenna array for breast cancer imaging. In: IEEE international symposium on antennas and propagation, Spokane, Washington, USA, pp 710–713

Kuo YL, Wong KL (2002) Dual polarised monopole antenna for wireless LAN operation. IEEE Intl. Antennas & Propagat. Symp. Dig. 4:80–83

Kwon HD (2006) Effect of antenna gain and group delay variations on pulse-preserving capabilities of ultrawideband antennas. IEEE Trans Antennas Propag 54(8):2208–2215

Langley J, Hall P, Newham P (1996) Balanced antipodal Vivaldi antenna for wide bandwidth phased arrays. Inst Elect Eng Microwaves Antennas Propag 143(2):97–102

Lazebnik M, McCartney L, Popovic D, Watkins CB, Lindstorm MJ, Harter J, Sewall S, Magliocco A, Brooske JH, Okoniewski M, Hagness SC (2007a) A large-scale study of the ultrawideband microwave dielectric properties of normal breast tissue obtained from reduction surgeries. Phys Med Biol 52(10):2637–2656

Lazebnik M, McCartney L, Popovic D, Watkins CB, Lindstorm MJ, Harter J, Sewall S, Magliocco A, Brooske JH, Okoniewski M, Hagness SC (2007b) A large-scale study of the ultrawideband microwave dielectric properties of normal, benign and malignant breast tissues obtained from cancer surgeries. Phys Med Biol 52(20):6093–6115

Lee JJ, Livingston S (1993) Wide band bunny-ear radiating element. In: IEEE international symposium on antennas and propagation, vol 3, Ann Arbor, US, pp 1604–1607

Li X, Hagness SC, Choi MK, van der Weide D (2003) Numerical and experimental investigation of an ultrawideband ridged pyramidal-horn antenna with curved launching plane for pulse radiation. IEEE Antennas Wirel Propag Lett 2:259–262

Liang J, Chiau CC, Chen X, Parini CG (2004) Printed circular disc monopole antenna for ultra-wideband applications. Electron Lett 40(20):246–1247

Liang J, Chiau CC, Chen X, Parini CG (2005) Study of a printed circular disk monopole antenna for UWB systems. IEEE Trans Antennas Propag 53(11):3500–3504

Lizzi L, Manica L, Massa A (2009) Time-domain analysis for UWB antenna synthesis. In: Proceedingds of the 39th European microwave conference, Rome, pp 93–96

Lizzi L, Oliveri G, Massa A (2012) A time-domain approach to the synthesis of UWB antenna systems. Prog Electromagn Res 122:557–575

Mandke K, Nam H, Yerramneni L, Zuniga C (2003) The evolution of UWB and IEEE 802.15.3a for very high data rate WPAN. [Online]. Available: http://kom.aau.dk/group/03gr993/Papers/802.15/GT3a.pdf. Accessed 21 Aug 2014

Meaney PM, Fanning MW, Raynolds T, Fox CJ, Fang Q, Kogel CA, Poplack SP, Paulsen KD (2007) Initial clinical experience with microwave breast imaging in women with normal mammography. Acad Radiol 14(2):207–218

Meinke H, Gundlach FW (1968) Taschenbuch der Hochfrequenztechnik. Springer, Berlin, New York, pp 531–535

Narbudowicz A, John M, Bao X, Ammann MJ (2012) Vivaldi array for generation of UWB circular polarization. In: IEEE international symposium on antennas and propagation, Chicago, US, pp 1–2

Poplack SP, Tosteson TD, Wells WA, Pogue BW, Meaney PM, Hartov A, Kogel CA, Soho SK, Gibson JJ (2007) Electromagnetic breast imaging: results of a pilot study in women with abnormal mammograms. Radiology 243:350–359

DOI 10.1007/978-981-4560-75-7_59-1

Qing X, Chen ZN, Chung HL (2007) Ultra-wideband circularly polarized wide-slot antenna fed by three-stub hybrid coupler. In: IEEE international conference on ultra-wideband, Singapore, pp 487–490

Qu SW, Li JL, Xue Q, Chan CH, Li S (2009) Wideband and unidirectional cavity-backed folded triangular bowtie antenna. IEEE Trans Antennas Propag 57(4):1259–1263

Rambabu K, Tan AEC, Chan KKM, Chia MYW (2009) Estimation of antenna effect on ultra-wideband pulse shape in transmission and reception. IEEE Trans Electromagn Compat 51(3):604–610

Ruvio G, Solimene R, Cuccaro A, Gaetano D, Browne JE, Ammann MJ (2014) Breast cancer detection using interferometric MUSIC: experimental and numerical assessment. Medical Physics 41(10), 103101/1–11

Schantz HG (2004) Dispersion and UWB antennas. In: Proceedings of the conference on . ultrawideband systems and technologies, conference: Kyoto, Japan

Schantz HG, Fullerton L (2001) The diamond dipole: a Gaussian impulse antenna. In: Proceedings of antennas and propagation society international symposium conference: Boston, MA, USA

Sha L, Ward ER, Story B (2002) A review of dielectric properties of normal and malignant breast tissue. IEEE Southeast conference, Columbia, US, pp 457–462

Shlivinski A (2009) Time-domain circularly polarized antennas. IEEE Trans Antennas Propag 57(6):1606–1611

Sill J, Fear EC (2005) Tissue sensing adaptive radar for breast cancer detection: study of immersion liquids. Electron Lett 41(3):113–115

Siwiak K, McKeown D (2004) Radiation of UWB signals. In: Ultra-wideband radio technology. John Wiley & Sons Ltd, New York, p 111

Telzhensky N, Leviatan Y (2006) Novel method of UWB antenna optimization for specified input signal forms by means of genetic algorithm. IEEE Trans Antennas Propag 54(8):2216–2225

Valderas D, Sancho JI, Puente D, Ling C (2011) Ultrawideband antennas: design and applications. Imperial College Press, London, p 179

Walton KL, Sundberg VC (1964) Broadband ridged horn design. Microw J 4:96–101

Wang S, Wu Q, Su D (2012) A novel reversed T-match antenna with compact size and low profile for ultrawideband applications. IEEE Trans Antennas Propag 60(10):4933–4937

Wu Q, Jin R, Geng J, Su D (2010) On the performance of printed dipole antenna with novel composite corrugated-reflectors for low-profile ultrawideband applications. IEEE Trans Antennas Propag 58(12):3839–3846

Zhang X, Larson LE, Asbeck PM (2003) Design of linear RF outphasing power amplifiers. Artech House Inc, Boston, p 47

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_60-1

Beam-Scanning Leaky-Wave Antennas

Jun H. Choi[a,b]* and Tatsuo Itoh[a]
[a]Department of Electrical Engineering, University of California, Los Angeles, Los Angeles, CA, USA
[b]Department of Electrical Engineering and Computer Science, Center for Science and Technology, Syracuse University, Syracuse, NY, USA

Abstract

This chapter presents the basic principles and recent developments of beam-scanning leaky-wave antennas. The single radiating element is realized by modifying waveguides to leak along the structure producing radiated beams that scan as the operating frequency is varied. This unique characteristic distinguishes leaky-wave antennas from other types of antenna. Leaky-wave antennas are broadly classified into three classes: uniform, periodic, and quasiperiodic. In the following sections, each class will be explained and compared. Samples of novel design techniques and potential applications are also highlighted.

Keywords

Beam scanning; Dispersion diagram; Frequency scanning; Leaky-wave antenna; Metamaterial; Periodic structure

Introduction

Leaky-wave antennas exhibit radiated beams that scan as a function of frequency. The unique frequency-scanning feature is realized by engineering the waveguiding structure to leak energy as the wave propagates down the waveguide. Therefore, leaky-wave antennas are considered traveling wave-type antennas. Since leaky-wave antennas are generally realized by utilizing or modifying conventional waveguiding structures, the key in designing such antennas lies in producing and controlling the leakage. As will be shown in the following sections, leakage may be produced by operating the waveguide in a higher-order propagating mode or spatial harmonics or utilizing unconventional waveguide structures that permit the fundamental mode to radiate. The latter approach includes the use of a relatively new waveguide technology based on the metamaterial concept.

If the field traveling inside the waveguide structure leaks energy, the complex propagation wavenumber of the guided field is expressed by leakage constant α and propagation constant β. The frequency-dependent propagation constant $\beta(\omega)$ mainly dictates the beam-scanning angles and α shapes the amplitude distribution of the antenna. The length of the antenna is usually designed such that most of the energy is leaked out of the antenna by the time the wave reaches to the opposite end, where the leakage constant can be judiciously engineered to meet design-specific requirements. For example, if the antenna dimension is relatively short, large α is preferred to radiate the energy quickly to make the antennas more

*Email: jhcho100@syr.edu
*Email: choijh@ucla.edu

DOI 10.1007/978-981-4560-75-7_60-1

efficient. On the other hand, if a longer antenna dimension is allowed, α can be relaxed to spread the amplitude distribution thereby increasing the radiated directivity of the antenna.

The very first leaky-wave antenna dates back to 1940, when rectangular waveguides were the dominant type of microwave guiding medium (Hansen 1940). Radiation was achieved by opening a narrow slit along the side of the waveguide. As the wave propagated through the medium, partial energy coupled to free space through the narrow slits. The concept was then extended to various structural forms using different waveguide technologies (Hessel 1969; Tamir 1969; Menzel 1979; Oliner and Jackson 2007; Jackson and Oliner 2008; Sutinjo et al. 2008; Caloz et al. 2011; Jackson et al. 2012; Jackson 2013; Menzel 2013). Elaborate details in the theory and analysis of leaky-wave antennas can be found in these book chapters and review articles, which are highly recommended for readers seeking for more in-depth understanding. This chapter is organized to concisely summarize the core fundamentals of leaky-wave antennas. The second half of the chapter highlights samples of recent techniques in designing leaky-wave antennas with specific radiating characteristics. Practical applications will also be presented at the end of the chapter.

Leaky-Wave Antenna Physics and Classifications

A large number of design techniques are available to control and enhance the radiated beam characteristics of leaky-wave antennas. But before considering various improvement options or adding particular features, the antenna has to first properly leak energy. The primary requirement in designing a leaky-wave antenna is therefore to ensure the waveguiding structure is operated in the fast-wave region ($\beta/k_o < 1$). This condition allows the propagation constant of outward pointing wavenumber vector at the interface between the antenna and open boundary to be real, which translates to radiation. To better understand the radiating mechanism, this section discusses the fundamental operating principles of leaky-wave antennas. Then, different realization options will be briefly introduced.

Leaky-Wave Antenna Fundamentals

For a general understanding of the radiating mechanism, consider a lossless two-dimensional waveguide structure with the source located at $z \ll 0$ (Fig. 1). When designing a simple planar antenna radiating toward the upper hemisphere, the lower boundary of the waveguide (at $x = -h$) is assigned with a perfect

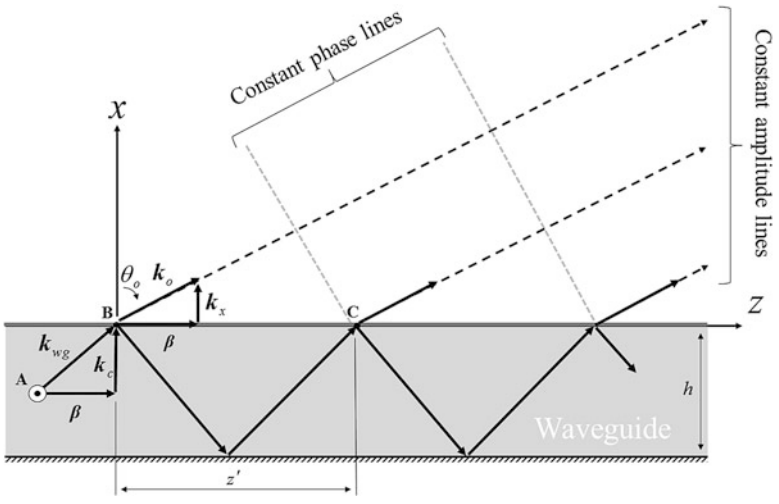


Fig. 1 Side view of an arbitrary two-dimensional leaky-wave antenna

DOI 10.1007/978-981-4560-75-7_60-1

reflecting layer such as a perfect electric conductor (PEC) or perfect magnetic conductor (PMC). To accommodate radiation, the upper boundary of the waveguide (at $x = 0$) should however be a slightly imperfect reflecting boundary (partially leaking boundary) that minimally disturbs the guided field. Although this is a generalized example, the main concept can be applied in understanding the fundamental working mechanisms of leaky-wave antennas. The partially reflective interface can take various forms, such as the interface between different material layers, narrow slots, periodic array of apertures or slots, array of parallel wires, etc. (Oliner and Jackson 2007).

Consider a y-oriented field propagating along the z-direction with the propagation constant $\beta(\omega)$. The magnitude of $\boldsymbol{k}_{wg}$ may vary depending on the material enclosed by the waveguide structure. For example, k_{wg} equals free-space wavenumber (k_o) if the guiding structure is filled with air but becomes larger than k_o if the guiding structure encloses dielectric materials. In this example, lossless condition also results k_{wg} to be real. The cutoff wavenumber k_c is determined by the waveguide dimensions and the propagation mode. If the antenna permits partial energy leakage only at the interface between the antenna and air (in our case at $x = 0$), the reflected field amplitude is reduced in proportion to the leakage amount, α. Starting with the initial field expression at point A,

$$\boldsymbol{\psi}(x,z) = \hat{y}A_o e^{-j(k_c x+\beta z)}, \tag{1}$$

where A_o is the initial amplitude of an arbitrary field. Assuming the antenna interface generates negligible reactive near-field energy storage, partial leakage produces

$$\boldsymbol{\psi}(x,z) = \hat{y}A_o e^{-j(\mathrm{k}_o\cdot\mathrm{r})} = \hat{y}A_o T e^{-j(k_x x+\beta z)} \quad \text{for } x = 0^+, \tag{2a}$$

$$\boldsymbol{\psi}(x,z) = \hat{y}A_o e^{-j\left(\mathrm{k}_{wg}\cdot\mathrm{r}\right)} = \hat{y}A_o \Gamma e^{-j(-k_c x+\beta z)} \text{ for } x = 0^-, \tag{2b}$$

just above and below the interface point B between the antenna and air, respectively. Here, the transmission coefficient (T) and reflection coefficient (Γ) are not strictly due to material discontinuity but are also due to the radiating structure. The coefficients are dictated by impedance mismatch at the interface. Also, the phase-matching condition at the interface requires $\beta(\omega)$ to be continuous for lossless case. Continuing along the propagation path, the fields just above and below the interface at the second interface point C can be treated in a similar manner,

$$\boldsymbol{\psi}(x,z) = \hat{y}A_o T\Gamma e^{-j(k_x x+\beta z)} \text{ for } x = 0^+, \tag{3a}$$

$$\boldsymbol{\psi}(x,z) = \hat{y}A_o \Gamma\Gamma e^{-j(-k_c x+\beta z)} \text{ for } x = 0^-. \tag{3b}$$

The field inside the antenna will continue propagating down the waveguiding structure transversely bouncing back and forth in a zigzag manner until either all the energy is leaked out or the wave is terminated with a matched load. As the wave travels over a unit distance z', the field amplitude is reduced by $\Gamma = \exp(-\alpha z')$. Thus, the field amplitude is exponentially reduced by a unit attenuation constant (α) as the guided field encounters each field-emanating interface. Consequently, the field just above the interface can be expressed as

$$\boldsymbol{\psi}(x,z) = \hat{y}A_o T e^{-\alpha z} e^{-j(k_x x+\beta z)}. \tag{4}$$

The dispersion relation directly above the interface (or aperture) of the antenna is given by

DOI 10.1007/978-981-4560-75-7_60-1

$$k_x = \sqrt{k_o^2 - \beta^2}. \tag{5}$$

To support wave propagation away from the interface, the antenna has to operate in the fast-wave regime where $\beta(\omega) < k_o$. The term "fast wave" describes the faster phase velocity of the wave traveling down the propagating direction (z-direction) relative to that of the speed of light:

$$v_p = \frac{\omega}{\beta(\omega)} > c = \frac{\omega}{k_o}. \tag{6}$$

If $\beta(\omega) > k_o$ (slow wave), k_x becomes imaginary which translates to an evanescent wave along the positive x-direction. Under the latter condition the structure will no longer behave as a leaky-wave antenna. The main beam angle θ_o of the leaky-wave antenna operating under the fast-wave condition is determined by the ratio between $\beta(\omega)$ and k_o as

$$\theta_o(\omega) = \sin^{-1}\left[\frac{\beta(\omega)}{k_o}\right]. \tag{7}$$

The length of the antenna (L) is generally adjusted to radiate about 90 % of its power. An initial length satisfying this condition can be estimated by taking the ratio of the power leaving through the opposite end of the antenna structure over the input power and then equating the ratio to 10 %:

$$L \approx \frac{0.18(\lambda_o)}{(\alpha/k_o)}. \tag{8}$$

Designing leaky-wave antennas with near 100 % efficiency is generally considered impractical because the aperture dimension has to be extremely large.

Leaky-Wave Antenna Classifications

Leaky-wave antennas are classified as uniform, periodic, or quasiperiodic. When choosing the class, it is up to the designer to evaluate the design trade-offs. For example, leaky-wave antennas based on uniform structures are relatively easy to design and fabricate, but usually provide limited frequency-scanning capability. Periodic-type leaky-wave antennas demand a relatively relaxed fabrication requirement in comparison to the quasi-uniform types, but require special design technique to circumvent undesired radiated gain reduction near broadside direction. Quasi-uniform structures can be systematically designed to provide full-frequency-scanning capability, but are comprised of relatively complex structural forms. Although numerous variations are currently available in realizing each class of leaky-wave antennas, the following sections will use single-conductor rectangular waveguides and multiconductor planar waveguides to illustrate the subtle differences between classes.

Leaky-Wave Antenna Designs

In this section, three classes of leaky-wave antennas will be explained using dispersion diagrams (commonly known as k-β diagram). Samples of novel design techniques are also outlined in the second half of the section. A dispersion diagram provides graphical interpretation of the dispersion relations of forward and backward propagating waves in the waveguide. $k_o d$ and βd represents spatial phase-shift quantity per unit distance (d) along the unbound and bound propagating directions, respectively. The dash

DOI 10.1007/978-981-4560-75-7_60-1

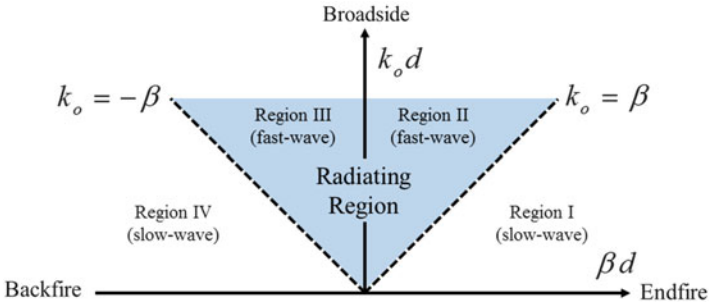


Fig. 2 Dispersion diagram of a 1-D leaky-wave antenna

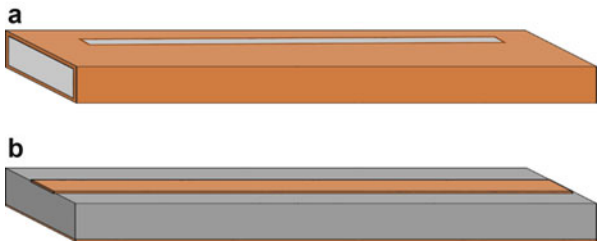


Fig. 3 Sketch of uniform leaky-wave antennas based on (**a**) single-conductor waveguide with a narrow open slot and (**b**) multiconductor microstrip line

line is called the airline or transverse electromagnetic (TEM) line that delineates the radiation (fast-wave) region and guided (slow-wave) region (Fig. 2). The dispersion diagram presents a graphical reference to quickly determine if a given waveguide structure is radiative. For example, even if the structure is not completely shielded and permits radiation, the fields will be confined within the structure when operating in the slow-wave region (regions I and IV). On the other hand, if the same structure is operated inside the fast-wave region (regions II and III), the same waveguiding structure behaves as an antenna and radiates with a main beam direction given by Eq. 7. The plot has been primarily utilized to analyze the dispersion characteristics of periodic structures. But in this chapter it is also used to explain the uniform-type leaky-wave antennas to provide a more consistent comparison between the three leaky-wave antenna classes. When plotting the dispersion diagram for the uniform case, the structure is assumed to be infinitely long with $d \ll \lambda_g$.

Uniform Structure

The fundamental mode of a conventional single-conductor rectangular waveguide shown in Fig. 3a exhibits a dispersion that originates inside the fast-wave region and asymptotically approaches the airline as the frequency (proportional to $k_o d$) increases (Fig. 4a). Therefore, a leaky-wave antenna can be realized by partially opening the closed structure to allow the fundamental mode to couple energy to the surrounding unbound environment. The latter condition assumes the energy coupling aperture or interface minimally perturbs the guided wave. A unidirectionally fed uniform leaky-wave antenna scans the beam from near broadside direction toward the endfire direction as the frequency increases. As shown in Fig. 4a, asymptotic behavior of the dispersion curve for an air-filled waveguide ($\varepsilon_r = 1$) requires relatively wide frequency bandwidth to tilt the radiated beam to near endfire direction. However, if the waveguide is loaded with high dielectric materials ($\varepsilon_r > 1$), the dispersion curve asymptotically approaches the dielectric-loaded line (dashed line). The dispersion curve slope and hence leaky-wave bandwidth can therefore be engineered. Engineering flexibility provides relaxed design criteria for practical systems with

DOI 10.1007/978-981-4560-75-7_60-1

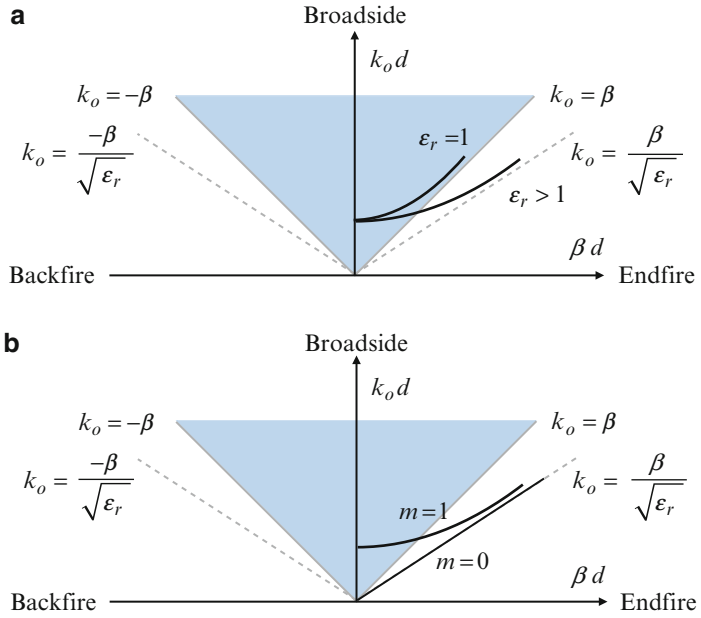


Fig. 4 Dispersion diagrams of uniform leaky-wave antennas based on (**a**) single-conductor waveguide and (**b**) multiconductor microstrip line

limited frequency bandwidth needing to scan the beam from near broadside to endfire. In addition to the fundamental propagating mode, an infinite number of higher-order propagating modes are also generated as the frequency is increased. However, since leaky-wave antennas usually operate in the lower propagation modes, only the dispersion relations for the frequency band corresponding to the lower modes are shown throughout the chapter.

A slightly different approach is required to operate a multiconductor uniform waveguide structure such as a microstrip line shown in Fig. 3b as a radiating structure. A microstrip line is not a completely shielded structure, but the fundamental quasi-TEM mode ($m = 0$) is bounded as shown in Fig. 4b. If however the structure is operating in the first higher-order mode ($m = 1$), radiation occurs in the frequency regions above the first higher-order mode cutoff frequency and within the leaky-wave region. The radiation takes place from open side boundaries of the microstrip line. This antenna was first proposed by Menzel (Fig. 3b) and then extensively analyzed by Oliner (Menzel 1979; Oliner and Lee 1986).

Among the three classes of leaky-wave antennas, uniform leaky-wave antennas have a relatively simpler form factor. However, one of the major drawbacks is the limited beam-scanning range since this class of antennas can only direct the beam toward the endfire direction. Full-frequency beam scanning from backfire to endfire requires additional design techniques such as using more exotic ferrite materials. Detailed explanations of this method will be discussed in section "Ferrite Loaded Leaky-Wave Antenna."

Periodic Structure

Periodic structures can be analyzed in the spatial frequency domain using spatial Fourier expansion or more commonly known as Floquet spatial harmonic expansion (Hessel 1969; Collin 1991; Pozar 1998). For an arbitrary one-dimensional periodic waveguide structure (Fig. 5), the time-harmonic field reacting to the structure can be described by

$$\boldsymbol{\psi}(x, y, z + d) = e^{-jk_{z0}d}\boldsymbol{\psi}(x, y, z). \tag{9}$$

DOI 10.1007/978-981-4560-75-7_60-1

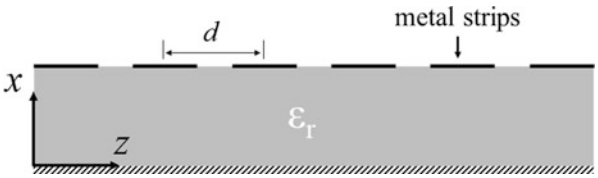


Fig. 5 Example of a periodic waveguide (dielectric waveguide with periodic perturbation along the propagating axis)

The neighboring field at a distance d away only differs by the zero space harmonic complex phase-shift amount of $k_{z0}d = (\beta - j\alpha)d$. The above equation can be expressed in terms of a periodic vector function $\boldsymbol{P}(x, y, z) = \boldsymbol{P}(x, y, z + d)$ as follows:

$$\boldsymbol{\psi}(x, y, z) = e^{-jk_{z0}z}\boldsymbol{P}(x, y, z). \tag{10}$$

Moreover, the periodic function may be expanded in a Fourier series,

$$\mathrm{P}(x, y, z) = \sum_{n=-\infty}^{\infty} \boldsymbol{a}_n(x, y)e^{-j\frac{2\pi}{d}nz}, \tag{11}$$

where $\boldsymbol{a_n}$ is a function describing the spatial harmonic field variation in the x and y directions. Plugging Eq. 11 into Eq. 10 results in

$$\boldsymbol{\psi}(x, y, z) = \sum_{n=-\infty}^{\infty} \boldsymbol{a}_n(x, y)e^{-jk_{zn}z}, \tag{12}$$

where

$$k_{zn} = k_{z0} + \frac{2\pi}{d}n,\ n = 0,\ \pm 1,\ \pm 2,\ \geq \cdots. \tag{13}$$

The corresponding propagation constants of outward pointing wavenumber vectors at the field-emanating interface are

$$k_{xn} = \sqrt{k_o^2 - k_{zn}^2}, \tag{14}$$

which simplifies to

$$k_{xn} = \sqrt{k_o^2 - \beta_n^2} \tag{15}$$

for the lossless case. Radiation conditions are satisfied for all n^{th} spatial harmonics located inside the fast-wave region. This relation can be visualized by plotting the dispersion diagrams of spatial harmonics that repeats in βd with period 2π (Hessel 1969).

Two examples of periodic leaky-wave antennas are shown in Fig. 6. The first antenna is realized by opening periodic slots on the top metallic layer of an air-filled rectangular waveguide. The second example is a planar microstrip line with periodic open stubs etched on a grounded low dielectric substrate. Although a realistic dispersion diagram takes a much more complicated form, a simplified version of an antenna with unidirectional feeding can be drawn assuming the perturbation due to the periodic loading is

DOI 10.1007/978-981-4560-75-7_60-1

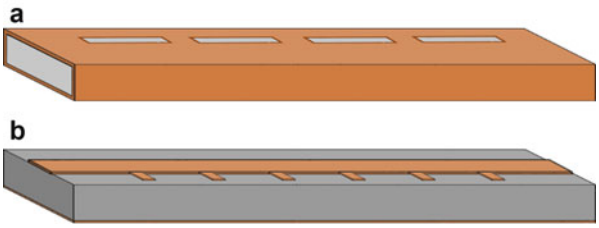


Fig. 6 Sketch of periodic leaky-wave antennas based on (**a**) single-conductor waveguide and (**b**) multiconductor combline structure

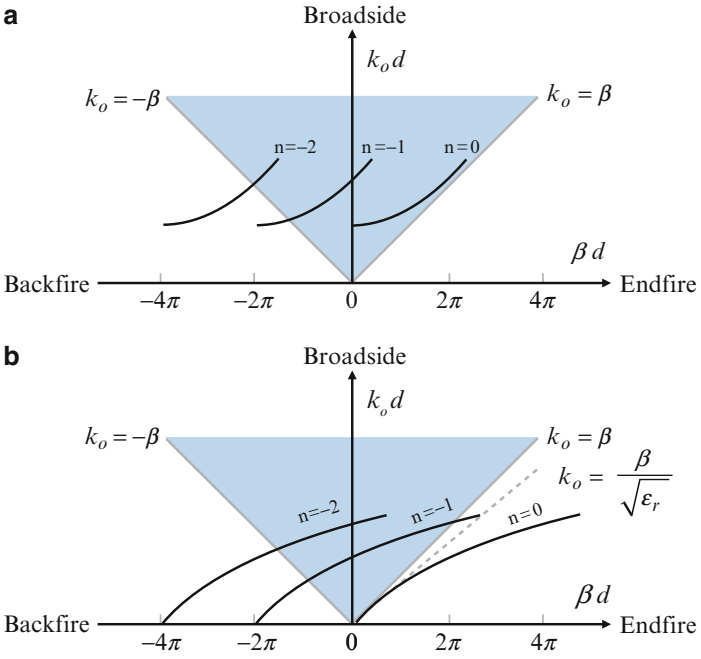


Fig. 7 Dispersion diagrams of periodic leaky-wave antennas based on (**a**) single-conductor waveguide and (**b**) multiconductor combline structure

negligibly small (Fig. 7). Furthermore, plotting only the first few spatial harmonics near $n = 0$ is sufficient to understand the basic properties of the periodic leaky-wave antennas. In the first case, the fundamental mode will appear above the cutoff frequency of the waveguide and asymptotically approach the airline. All other space harmonics have the same dispersion curve but are shifted by $\pm 2\pi$ along the βd axis. Figure 7 shows first three $n \leq 0$ space harmonics that lie within the fast-wave region. For the combed microstrip line structure, the $n = 0$ mode starts at DC ($k_o d = 0$) and then becomes a slow wave as the frequency is increased. Since the fundamental mode never enters the fast-wave region, $n < 0$ modes are used for radiation.

For both structures shown in Fig. 6, additional design schemes are needed if single beam scanning from backfire to endfire direction is desired. For the combline structure, a simple method is to use high dielectric constant material and limit the periodicity to less than $\lambda_o/2$ (Jackson and Oliner 2008). Sketch of the modified dispersion diagram of high dielectric-loaded combline leaky-wave antenna is shown in Fig. 8. Within the highlighted operating bandwidth, single spatial harmonic ($n = -1$) scans the beam from backfire toward endfire directions. For different waveguide technologies, alternate options including the phase-reversal technique (which will be discussed in section "Phase-Reversal Technique") are also available to ensure single-mode radiation.

DOI 10.1007/978-981-4560-75-7_60-1

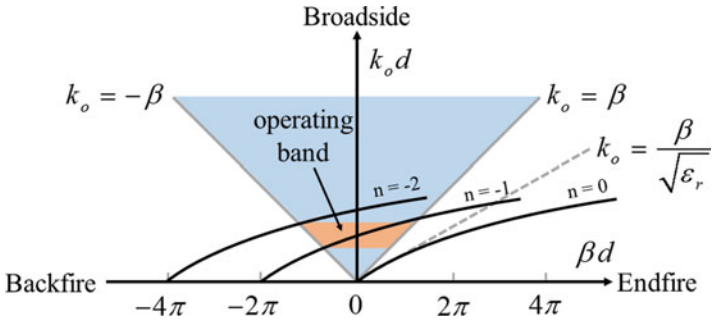


Fig. 8 Dispersion diagram of a periodic leaky-wave antenna based on the combline structure with high dielectric loading

The simplified dispersion curves shown above serve well in illustrating the basic radiating characteristics for periodic-type leaky-wave antennas. However, to be more precise, the partial reflections at each periodic discontinuity cannot be completely disregarded, especially at frequencies where each reflected wave all superimposes constructively creating a stopband region within the desired operational bandwidth. For full-frequency-scanning antennas, the stopband that occurs near $\beta = 0$ point has negative effect on the radiated gain stability. At this point, the traveling waves become a standing wave and excite all the loading elements simultaneously. As an example, consider the periodic combline structure in Fig. 6b operating at the broadside direction. All the stubs are excited in phase with $\beta_{-1} = 0$, and the admittance of individual loading add together, resulting in a short-circuit condition at the location of each stub. The field within the unit cell between adjacent stubs becomes a perfect standing wave instead of a traveling wave. From a practical implementation perspective, the antenna becomes difficult to match since the input impedance would be equivalent to a short circuit, resulting in complete reflection (Jackson and Oliner 2008; Caloz et al. 2011). Therefore, bandgap circumventing techniques such as integrating matching networks between the unit cells (which will be discussed in section "Stopband Countering Technique Near the Transition Frequency") are required to minimize the undesired stopband near $\beta = 0$ point.

Metamaterial (Quasi-uniform) Structure

Recent introduction of metamaterial concepts in the microwave field has revitalized leaky-wave antenna research (Liu et al. 2000; Grbic and Eleftheriades 2002; Lai et al. 2004; Caloz and Itoh 2005; Eleftheriades and Balmain 2006; Engheta and Ziolkowski 2006). Metamaterial based leaky-wave antennas are considered quasi-uniform because the periodicity is substantially smaller than the guided wavelength ($d \ll \lambda_g$), which macroscopically behaves as a uniform homogeneous structure. These antennas can perform frequency scanning from backfire to endfire including broadside direction while operating in the fundamental mode. One of the important achievements in the development of metamaterial-based leaky-wave antenna is the systematic design approach based on equivalent circuit models (Caloz and Itoh 2005). In addition to the dispersion diagrams, circuit analysis provides a very straightforward design procedure in obtaining initial geometrical parameters of the antennas. Samples of metamaterial-based leaky-wave antennas are shown in Fig. 9. Since planar structures have an easier fabrication process and a lower form factor profile, the substrate integrated waveguide (SIW)-based metamaterial leaky-wave antenna is chosen for the first example. SIW is a rectangular waveguide integrated inside the dielectric substrate with closely spaced vias forming the effective side metallic walls (Hirokawa and Ando 1998; Wu et al. 2003; Bozzi et al. 2011). Similar to the uniform and periodic type of rectangular waveguide leaky-wave antenna, SIW metamaterial leaky-wave antennas operate in the fundamental TE_{10} mode (Dong and Itoh 2011). A composite right/left-handed (CRLH) transmission-line (TL)-based leaky-wave antenna shown in Fig. 9b is one of the first microwave metamaterial structures developed in the early 2000s.

DOI 10.1007/978-981-4560-75-7_60-1

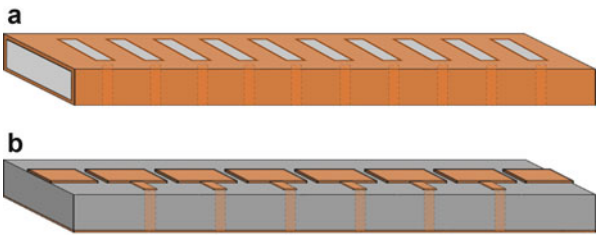


Fig. 9 Sketch of metamaterial leaky-wave antennas based on (**a**) substrate integrated waveguide and (**b**) multiconductor transmission line structure

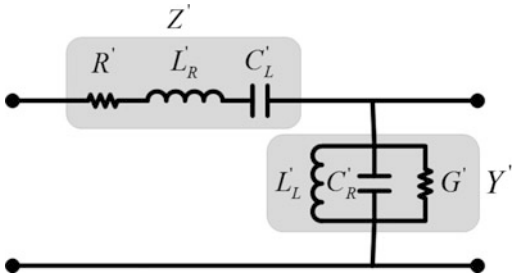


Fig. 10 Metamaterial leaky-wave antenna unit cell

The unit cell of a CRLH metamaterial structure is comprised of both conventional right-handed (series inductance L_R and shunt capacitance C_R) reactive elements and their dual left-handed (series capacitance C_L and shunt L_L) counterparts (Fig. 10). Per-unit-length series impedance and shunt admittance including dissipation and radiation losses are denoted with the constituent unit elements as follows:

$$Z' = R' + j\omega L'_R - j\frac{1}{\omega C'_L}, \tag{16a}$$

$$Y' = G' + j\omega C'_R - j\frac{1}{\omega L'_L}. \tag{16b}$$

The complex propagation constant and characteristic impedance can be obtained from the generalized telegrapher's equations as

$$k'_z = \sqrt{Z'Y'}, \tag{17a}$$

$$Z'_c = \sqrt{\frac{Z'}{Y'}}, \tag{17b}$$

respectively. The dispersion diagram of CRLH structures is bounded by low-frequency and high-frequency stopbands, and typically an interband bandgap is also generated between the series and shunt resonant frequencies. $(k_o d)_{se}$ and $(k_o d)_{sh}$ related to the series and shunt resonant frequencies, respectively, are defined as

DOI 10.1007/978-981-4560-75-7_60-1

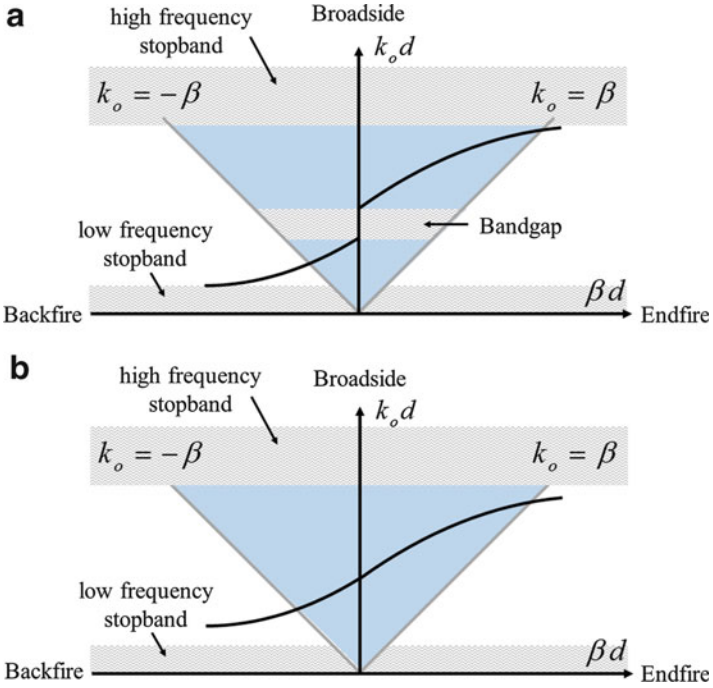


Fig. 11 Dispersion diagrams of metamaterial leaky-wave antenna under the (**a**) unbalanced and (**b**) balanced condition

$$(k_o d)_{se} = \frac{d}{c\sqrt{L'_R C'_L}}, \tag{18a}$$

$$(k_o d)_{sh} = \frac{d}{c\sqrt{L'_L C'_R}}, \tag{18b}$$

where c is the speed of light. Under the balanced condition where the following condition is satisfied:

$$(k_o d)_{se} = (k_o d)_{sh}, \tag{19}$$

a continuous dispersion curve can be obtained, eliminating an undesired bandgap near the transition frequency (Fig. 11). Therefore, a constant gain can be maintained when the scanning beam approaches broadside direction. Also, the characteristic impedance of the periodic structure becomes independent of frequency:

$$Z_c = \sqrt{\frac{L'_R}{C'_R}} = \sqrt{\frac{L'_L}{C'_L}}. \tag{20}$$

This enables a more relaxed antenna feed matching requirement and supports efficient frequency scanning over the entire spatial angle.

Design Techniques

The frequency-scanning property of leaky-wave antennas is a unique and attractive feature. However, additional radiation properties such as gain, side lobe level, polarization, etc., should also be carefully

DOI 10.1007/978-981-4560-75-7_60-1

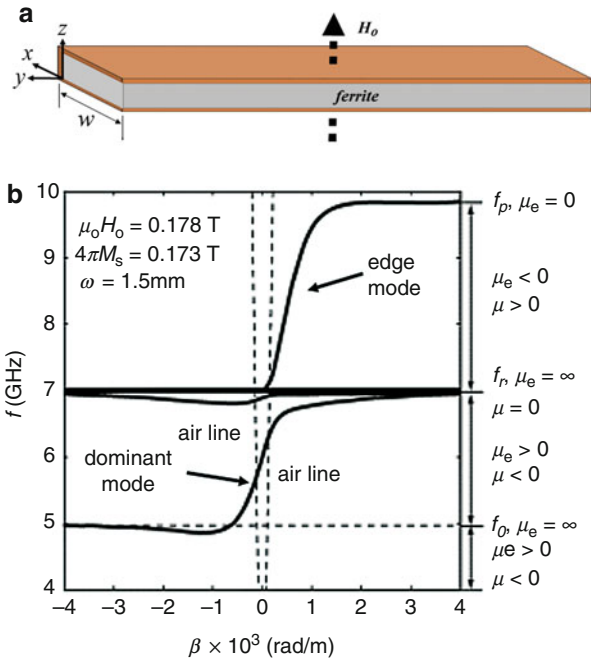


Fig. 12 (**a**) Sketch of a uniform ferrite-loaded leaky-wave antenna and (**b**) its dispersion diagram (Kodera and Caloz 2009, © IEEE)

examined when designing practical antennas. In this section, design examples are collected to highlight novel techniques in controlling and improving the radiated properties of leaky-wave antennas.

Ferrite-Loaded Leaky-Wave Antenna

A drawback of the uniform type of leaky-wave antennas is the limited beam-scanning range. As discussed in section "Uniform Structure," single-conductor-based leaky-wave antennas generally steer the beam toward the endfire direction. However, if the uniform type of leaky-wave antenna is filled with ferrite material and externally biased by a DC magnetic field, full scanning capability from backfire to endfire can be achieved (Baccarelli et al. 1997). A simplified sketch of the ferrite-loaded open waveguide leaky-wave antenna with the magnetic biasing field (H_o) and the corresponding dispersion diagram of the first two modes are shown in Fig. 12. A similar structure was originally developed as a magnetic isolator (Araki et al. 1975). However, it operated as an edge-mode located inside the slow-wave region to prevent radiation (Fig. 12b). If the structure is operated in the lower frequency band, the structure radiates with the dispersion characteristics that resemble that of CRLH leaky-wave antennas (Kodera and Caloz 2009). This fundamental mode provides full-frequency scanning from backward to endfire direction without the bandgap at the transition frequency ($\beta = 0$). The dispersion relation that takes the form of the transcendental equation can be derived using the transverse resonance technique (Baccarelli et al. 1997; Kodera and Caloz 2009).

Phase-Reversal Technique

Periodic leaky-wave antennas normally suffer from multiple spatial harmonic radiations. For example, in Fig. 7a, full-frequency-scanning bandwidth of any radiating mode will always be shared by one or more leaky space harmonics. Phase-reversal technique can be applied to generate a single radiated beam that can scan the entire backfire-to-endfire spatial angle. In this technique, an additional π phase shift per period is provided through structural modifications resulting in an effective phase delay per cell of $\beta_o d + \pi$ instead of $\beta_o d$ (Hessel 1969). The technique was originally utilized for broadside radiating arrays to

DOI 10.1007/978-981-4560-75-7_60-1

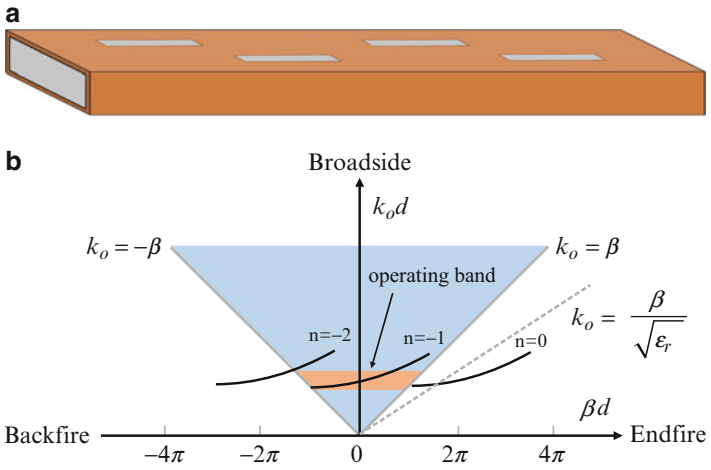


Fig. 13 (**a**) Sketch of a single-conductor periodic leaky-wave antennas with phase-reversal technique and (**b**) corresponding dispersion diagram

suppress the grating lobes without using high dielectric materials or structurally complex corrugated lines (Silver 1949). But the same concept is also adapted in leaky-wave antenna design to ensure single-mode radiation. A waveguide-based phase-reversal leaky-wave antenna is shown in Fig. 13a, where multimode radiation is suppressed by shifting the dispersion curves along the βd-axis, thereby pushing $n = 0$ mode outside the $n = -1$ radiation band (Fig. 13b). Examples of phase-reversal leaky-wave antennas applied in different structural forms including waveguide and planar structures can also be found in Hessel (1969) and Yang et al. (2009).

Stopband Countering Technique Near the Transition Frequency

Another downside of periodic-type leaky-wave antennas is the difficulty in providing efficient radiation toward the broadside direction (Caloz et al. 2011; Burghignoli et al. 2006; Jackson et al. 2012). Figure 14 shows a conventional planar combline leaky-wave antenna. Without any treatment for the transition frequency ($\beta = 0$), the radiated gain drops drastically as the beam approaches broadside direction (Fig. 14b). A simple yet effective approach in countering the stopband at the broadside direction is by incorporating impedance transforming networks composed of a delay line and a quarter-wave transformers or a matching stub between the junctions of unit cells (Paulotto et al. 2009). These internal matching networks minimize the abrupt impedance discontinuities between the unit cells in periodic-type leaky-wave antennas. Figure 15 shows an example of a combline leaky-wave antenna where a section of delay line of length d_1 followed by a quarter-wave transformer is inserted between the unit cells. The impedance looking into the quarter-wave transformer at each unit cell can be matched to the port impedance of Z_o if the delay line length is adjusted so that the impedance looking into the delay line is real-valued and the transformer impedance is designed to

$$Z_{\text{in}_{\lambda/4}} = \sqrt{Z_o \cdot Z_{\text{in}_{\text{delay\_line}}}}. \tag{21}$$

Sinusoidally Modulated Reactance Surface-Based Leaky-Wave Antenna

In section "Periodic Structure," periodic-type leaky-wave antennas are explained using grating-like structures. However, space harmonic generation and corresponding dispersion behavior can also be obtained by periodically modulating the impedance surface. The advantage of this technique is the independent control of the leakage and phase constants along the surface, thereby providing better design

DOI 10.1007/978-981-4560-75-7_60-1

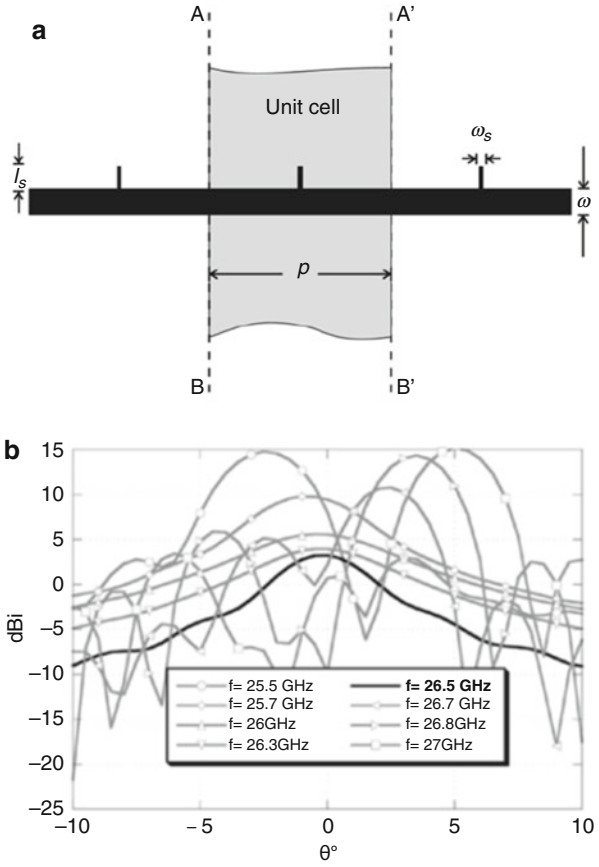


Fig. 14 (**a**) 1-D microstrip combline leaky-wave antenna and (**b**) simulated gain patterns near broadside direction for various operating frequencies (Paulotto et al. 2009, © IEEE)

control of the radiated beamwidth, side lobe level, and scan angle (Patel and Grbic 2011). Theoretical investigation of sinusoidally modulated reactance surface (SMRS) was first reported by A. Oliner in 1959 (Oliner and Hessel 1959). SMRS can be designed to take the following form (Oliner and Hessel 1959; Patel and Grbic 2007, 2011):

$$\eta_{\text{surf}}(z) = j\eta_o X' \left[1 + M\cos\left(\frac{2\pi z}{a}\right)\right], \tag{22}$$

where X' is the average surface reactance normalized by the free-space wave impedance ($X' = X/\eta_o$), M is modulation factor, z is wave propagation direction, and a is periodicity of the sinusoid. Based on this concept, printed leaky-wave antenna is realized by etching metallic strips separated by varying gaps on the top surface of a back-metallized dielectric substrate (Fig. 16). The proposed antenna is designed to radiate via the $n = -1$ space harmonic (Patel and Grbic 2007, 2011). In addition to frequency-scanning capability, this antenna can also be systematically designed to direct a particular frequency beam to a desired elevation angle:

$$\theta_{n=-1} \approx \sin^{-1}\left(\sqrt{1 + X'^2} - \frac{2\pi}{k_o a}\right). \tag{23}$$

DOI 10.1007/978-981-4560-75-7_60-1

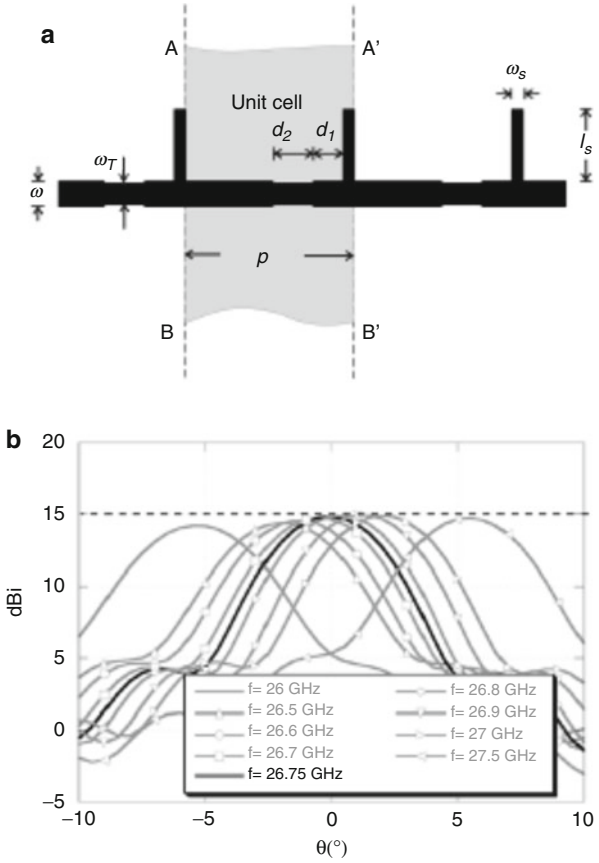


Fig. 15 (**a**) 1-D microstrip combline leaky-wave antenna with matching circuits and (**b**) simulated gain patterns near broadside direction for various operating frequencies (Paulotto et al. 2009, © IEEE)

Fig. 16 Photograph of a 1-D leaky-wave antenna based on sinusoidally modulated reactance surface (Patel and Grbic 2011, © IEEE)

The far-field beamwidth can be independently controlled by changing the modulation factor, which mainly alters the attenuation constant (α) (Patel and Grbic 2007).

Polarization Control

Polarization type and orientation have to be well defined and controlled to maximize the signal transfer. However, for some leaky-wave antennas such as the CRLH type shown in Fig. 9b, the radiation takes place from both interdigital capacitors and stub inductors resulting in an ambiguous polarization orientation (Liu et al. 2000). This section illustrates design approaches that enable better control of the polarization properties. In the first example (Fig. 17), orientation of a linearly polarized field is controlled by placing two CRLH-based leaky-wave antennas side-by-side and driving the antenna with either common-mode or differential-mode excitations (Hashemi and Itoh 2010). If the common-mode excitation is applied, fields radiated by the interdigital capacitors add in phase, whereas the fields radiated by the subinductors add destructively in the far-field. Therefore, electrical polarization is oriented parallel to the antenna scanning plane. The effects are reversed if the coupled CRLH-based leaky-wave antennas are driven with differential feeding, radiating linearly polarized wave oriented along the cross-plane with respect to the antenna scanning plane. In both cases, measured results show the cross-pol level is maintained below 15 dB throughout the observed scan range (around $-30^\circ < \theta_o < +30^\circ$). A simple polarization-orienting mechanism can be realized by adding a rat-race coupler to the input of the antenna and switching the input excitation to either the sum or difference port of the coupler.

Several techniques have been proposed to generate circularly or elliptically polarized fields (Dong and Itoh 2012; Kitatani et al. 2012; Ishii and Fukusako 2012; Lee et al. 2014a, b). Figure 18 shows a recent example based on single radiator SIW-based circularly polarized leaky-wave antenna. This SIW CRLH leaky-wave antenna-inspired structure is a periodic type of leaky-wave antenna where each unit cell is composed of a pair of orthogonally oriented radiating slots with $\pi/2$ phase-shifting separation to generate circular polarization (Lee et al. 2014a). A broadside enhancement technique discussed in section

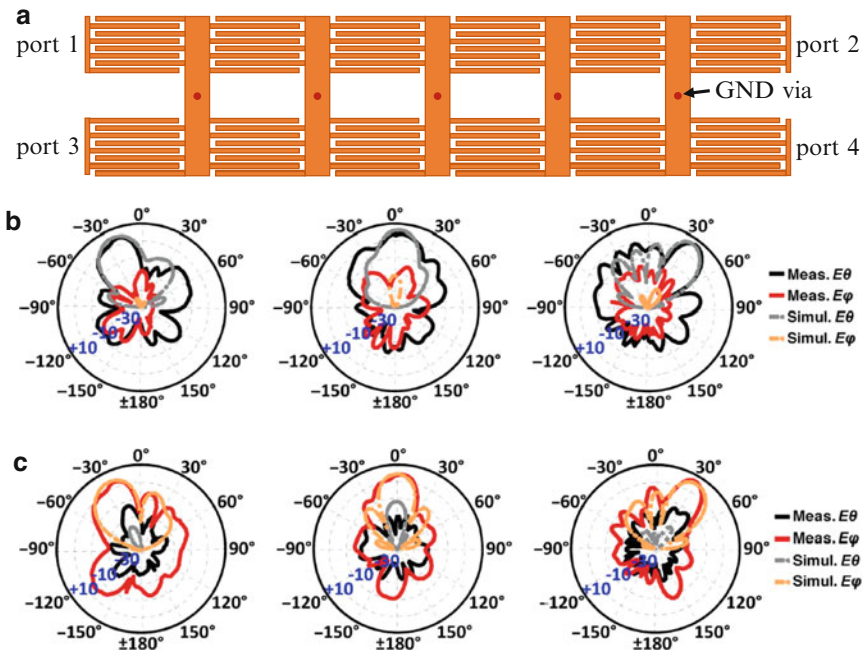


Fig. 17 (**a**) Sketch of a coupled CRLH leaky-wave antenna, (**b**) measured and simulated radiation patterns for common-mode excitations, and (**c**) measured and simulated radiation patterns for differential-mode excitations (Hashemi and Itoh 2010, © IEEE)

DOI 10.1007/978-981-4560-75-7_60-1

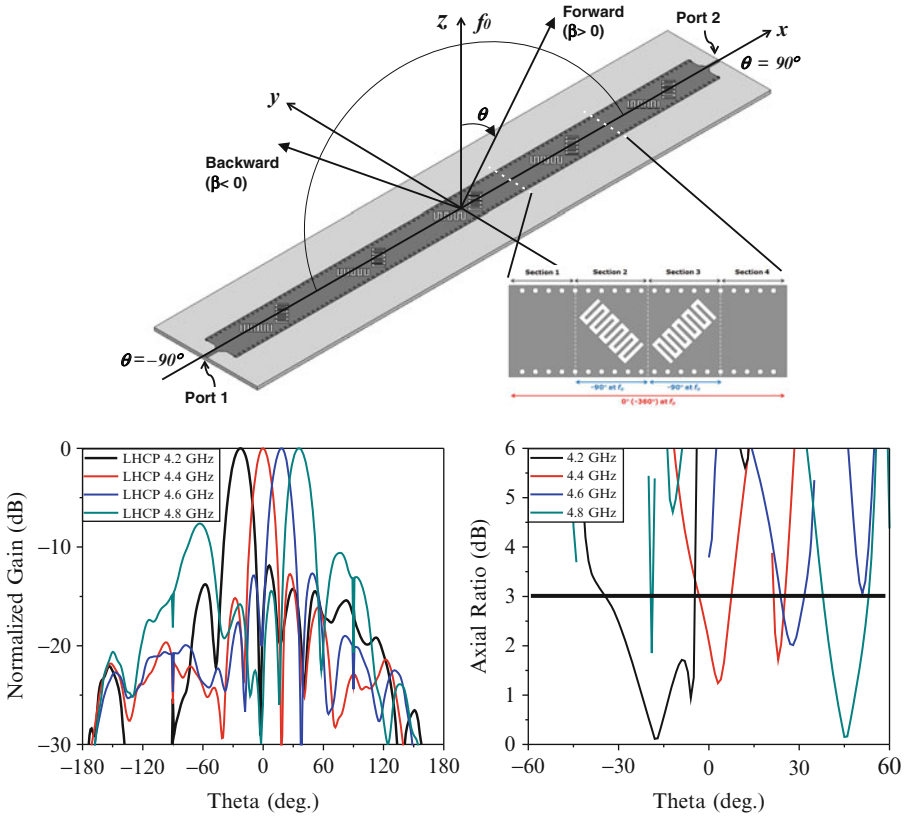


Fig. 18 (**a**) Perspective view of a single-conductor SIW-based circularly polarized leaky-wave antenna, (**b**) measured normalized radiation patterns versus frequency, and (**c**) measured axial ratio versus frequency (Lee et al. 2014a, © IEEE)

"Stopband Countering Technique Near the Transition Frequency" is also incorporated to maintain high gain value throughout the entire scanning angle (Lee et al. 2014b). The measured results for the radiated $n = -1$ space harmonic exhibit wide space angle frequency scanning without gain fluctuation while maintaining an axial ratio under 3 dB throughout the entire operating bandwidth.

Amplitude Control

Far-field radiation patterns are governed by the Fourier relation of the current amplitude distribution on the radiating aperture (Balanis 1997; Stutzman and Thiele 1998). For a given antenna size, the highest directivity is obtained when the radiating aperture is excited with uniform amplitude distribution; however, it suffers from high side lobe levels since the first side lobe is over 4 % of the peak intensity. If the side lobe level is a more critical design requirement, the aperture amplitude distribution can be properly tapered to suppress side lobes. As discussed in section "Leaky-Wave Antenna Fundamentals," leaky-wave antennas normally have exponentially decaying amplitude distribution. However, leakage distribution along the aperture can be varied through structural modifications (Whetten and Balanis 1991; Wu et al. 2009; Siragusa et al. 2012). In Whetten and Balanis (1991) far-field patterns of leaky-wave antennas with different long slot configurations are analyzed. Figure 19 shows three example configurations and their corresponding far-field radiation patterns. All three antennas are 20 λ long and 0.08 λ wide. The first antenna is a straight longitudinal slot with an offset of $x_o = 0.75\ a$, where a is the width of the waveguide. The second antenna is a sine squared curvature antenna with the meander contour of

$$x = \left[0.5 + 0.2\sin^2\left(\frac{\pi}{L}z\right)\right]a \tag{24}$$

DOI 10.1007/978-981-4560-75-7_60-1

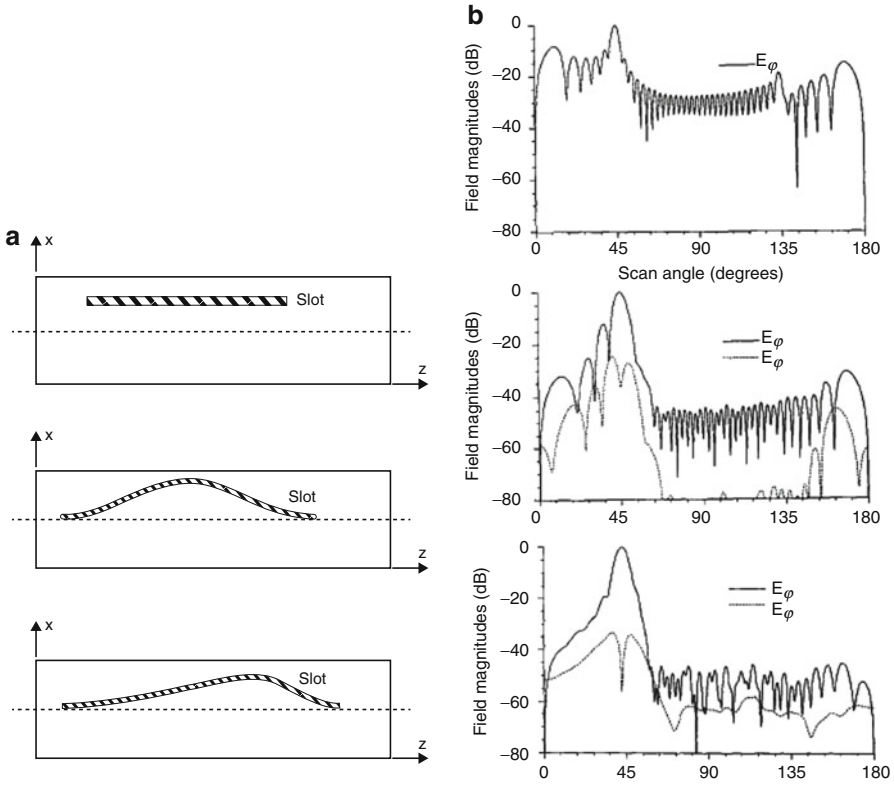


Fig. 19 (**a**) Leaky-wave long slot antennas with different meander patterns and (**b**) corresponding radiation patterns (Whetten and Balanis 1991, © IEEE)

The third antenna is contoured to compensate the exponentially decaying field distribution so that the radiated power has sinusoidal distribution. The antenna with sine squared curvature (second configuration) produces side lobe levels that are lower compared to that of the straight line case. In the last configuration, the side lobes are further reduced where all side lobe levels except for a small shoulder on the main beam are suppressed below 40 dB.

Multiband Leaky-Wave Antennas

Metamaterial-based leaky-wave antennas for multiband applications have been recently proposed and analyzed in Rennings et al. (2006) and Eleftheriades (2007). Figure 20a shows the equivalent circuit model of a unit cell for generalized negative-refractive-index transmission line (also known as extended composite right/left-handed transmission line) that supports dual-band operation. In a lossless symmetric T-model, the propagation constant is

$$\beta d = \cos^{-1}(1 + Z_h Y_v) \tag{25}$$

where Z_h and Y_v are series impedance and shunt admittance, respectively. To close the undesired bandgap at the transition frequencies and obtain continuous beam scanning in both bands, the following condition has to be satisfied (Eleftheriades 2007):

$$L_{hs} C_{hp} = L_{vs} C_{vp}. \tag{26}$$

The dual-band balanced dispersion diagram is shown in Fig. 20b. The Bloch impedance can be approximated to

DOI 10.1007/978-981-4560-75-7_60-1

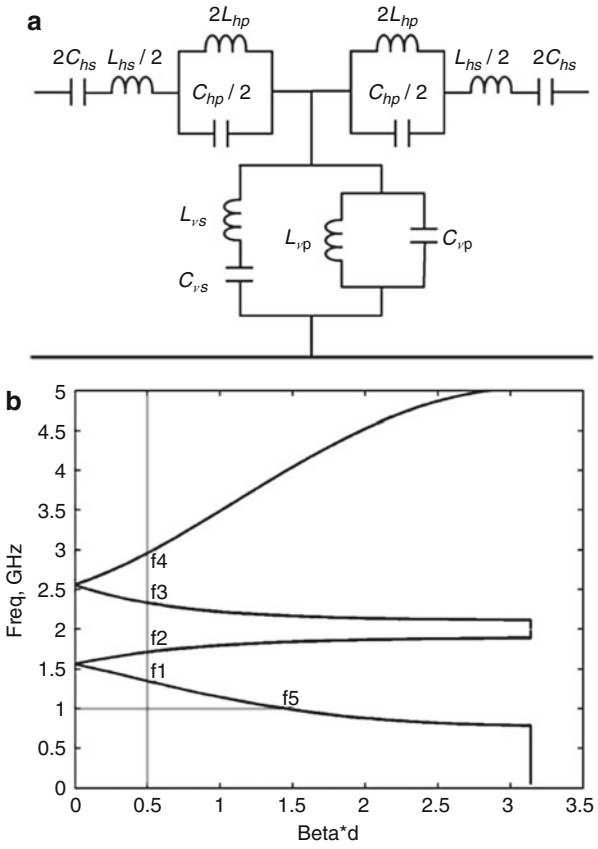


Fig. 20 (**a**) Circuit diagram of an NRI-TL unit cell and (**b**) corresponding dual-band dispersion diagram (Eleftheriades 2007, © IEEE)

$$Z_B \approx \sqrt{\frac{2Z_s}{Y_p}} \tag{27}$$

if the unit-cell dimension is electrically small ($\beta d \ll 1$). Based on the above concept, dual-band leaky-wave antennas have been realized in planar and SIW forms (Ryan and Eleftheriades 2010; Machac and Polivka 2012; Duran-Sindreu et al. 2013). Figure 21 shows an SIW-based dual-band leaky-wave antenna. This antenna provides continuous frequency scanning in the two operating bands, 3.5–4.2 GHz and 5.164–5.6 GHz (Duran-Sindreu et al. 2013).

Power Recycling

Leaky-wave antennas are generally designed to radiate about 90 % of the energy by the time waves reach the opposite end of the antenna structure. To obtain this high level of efficiency, leaky-wave antennas are usually elongated to around 8–10 λ_o. But antennas with such large electrical dimension may be problematic for applications with size constraints. To maintain high efficiency of leaky-wave antennas in a more compact size, unradiated energy of the traveling wave can be recycled and fed back to the antenna through a power combining rat-race coupler as shown in (Fig. 22) (Nguyen et al. 2010). The coupler is designed to constructively add the recycled power (port 3) to the input signal of the system (port 1) through the sum port (port 4). For a lossless system where dielectric and metallic losses are assumed to be zero, the radiated efficiency (η_s) can reach 100 % if the scattering matrix of the power recycling system becomes unitary. Given the scattering matrix of the rat-race coupler

DOI 10.1007/978-981-4560-75-7_60-1

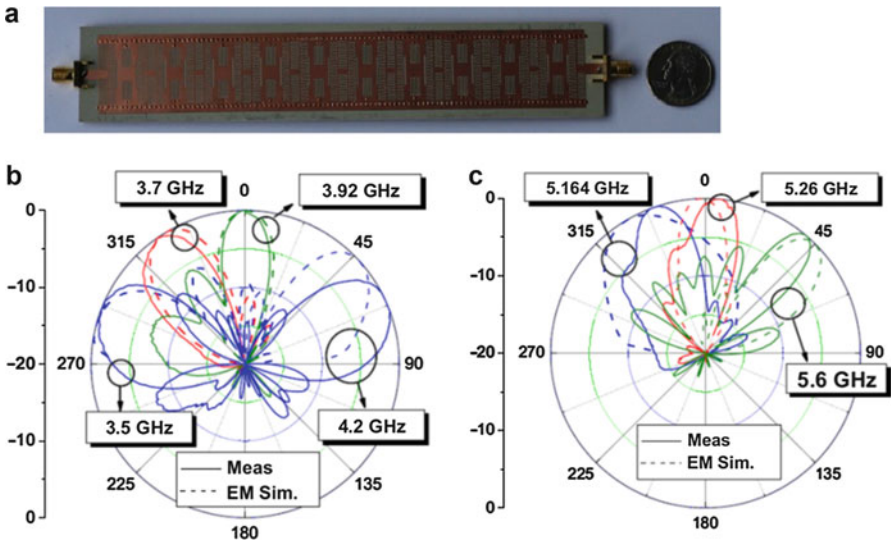


Fig. 21 (**a**) Photograph of a fabricated SIW metamaterial based dual-band leaky-wave antenna, (**b**) normalized simulated and measured radiation patterns for the lower radiation bands, and (**c**) normalized simulated and measured radiation patterns for the upper radiation bands (Duran-Sindreu et al. 2013, © IEEE)

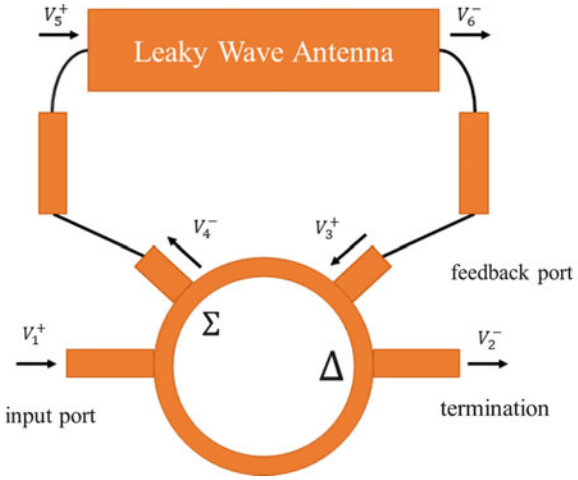


Fig. 22 Sketch of a power recycling leaky-wave antenna using rat-race coupler

$$\begin{bmatrix} V_1^- \\ V_2^- \\ V_3^- \\ V_4^- \end{bmatrix} = \begin{bmatrix} 0 & ja & 0 & -jb \\ ja & 0 & -jb & 0 \\ 0 & -jb & 0 & -ja \\ -jb & 0 & -ja & 0 \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \\ V_3^+ \\ V_4^+ \end{bmatrix}, \tag{28}$$

the required conditions to achieve system efficiency of 100 % for the power recycled leaky-wave antenna are as follows:

$$a = \sqrt{1 - \eta_0} \tag{29a}$$

DOI 10.1007/978-981-4560-75-7_60-1

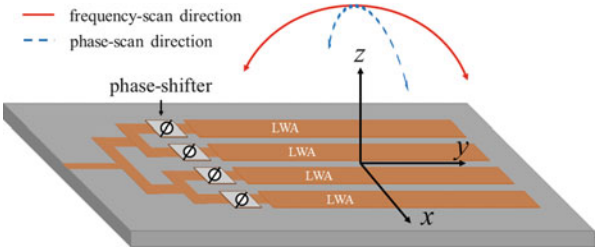


Fig. 23 Sketch of a 2-D leaky-wave antenna in an array arrangement

$$b = \sqrt{\eta_0} \tag{29b}$$

$$\theta = -\frac{\phi}{2} + \frac{3\pi}{4} + m\pi,\ m \in N, \tag{29c}$$

where η_o is the radiated efficiency of isolated leaky-wave antenna without the feedback circuit, θ is the electrical length of the delay lines connecting the rat-race coupler to the leaky-wave antenna, and ϕ is the total phase shift across the antenna. Experimental results for power recycled leaky-wave antenna with $\eta_o = 38$ % accounting all the mismatch and absorptive losses show the system radiated efficiency enhancement by a factor of 1.8 (Nguyen et al. 2010).

Two-Dimensional Leaky-Wave Antennas

In 1-D leaky-wave antennas, radiated directivity along the scanning plane can be simply controlled by varying the aperture length of the antenna. But since the transverse dimension (width) of the antenna is usually narrow and fixed, 1-D leaky-wave antennas produce a fan-beam radiation pattern with narrow beamwidth along the beam-scanning plane but wide beamwidth along its cross-plane. Therefore, 1-D leaky-wave antennas have poor spatial resolution in the plane orthogonal to the scanning plane. For applications requiring high spatial resolution in both dimensions, a pencil beam pattern can be obtained by placing 1-D leaky-wave antennas in an array arrangement (Fig. 23). If phase shifters are added at the input of each leaky-wave antenna, a fixed-frequency beam can be scanned in Z-X plane in addition to the frequency scanning along Z-Y plane (Oliner and Jackson 2007).

When an all passive 2-D periodic structure is operated as a leaky-wave antenna, a unique frequency-scanning conical beam pattern can be generated. An example of 2-D leaky-wave antenna based on the mushroom structure (Sievenpiper et al. 1999) is shown in Fig. 24 (Allen et al. 2004). If the 2-D periodic mushroom structure is fed at the center, radially propagating waves produce a frequency-scanning conical beam pattern with a radiation null at the broadside direction. This null in the pattern is due to destructive interferences of the far-fields generated by radially traveling waves in a 2-D symmetric structure.

Applications

Several interesting applications and devices exploiting the frequency-scanning capability of leaky-wave antennas can be found in Shelley et al. (1991), Tospann et al. (1995), Sato et al. (2007), and Menzel and Moebius (2012). This section presents sample of systems and applications that may benefit from this unique beam-scanning property. Examples of leaky-wave antenna systems that are covered in this section include automotive radar sensor and direction-of-arrival systems. Various fixed-frequency beam-scanning antennas based on the leaky-wave antenna are also presented. In comparison to fixed-frequency beam-scanning phased-array systems that rely on a large number of antenna elements, complex feed network,

DOI 10.1007/978-981-4560-75-7_60-1

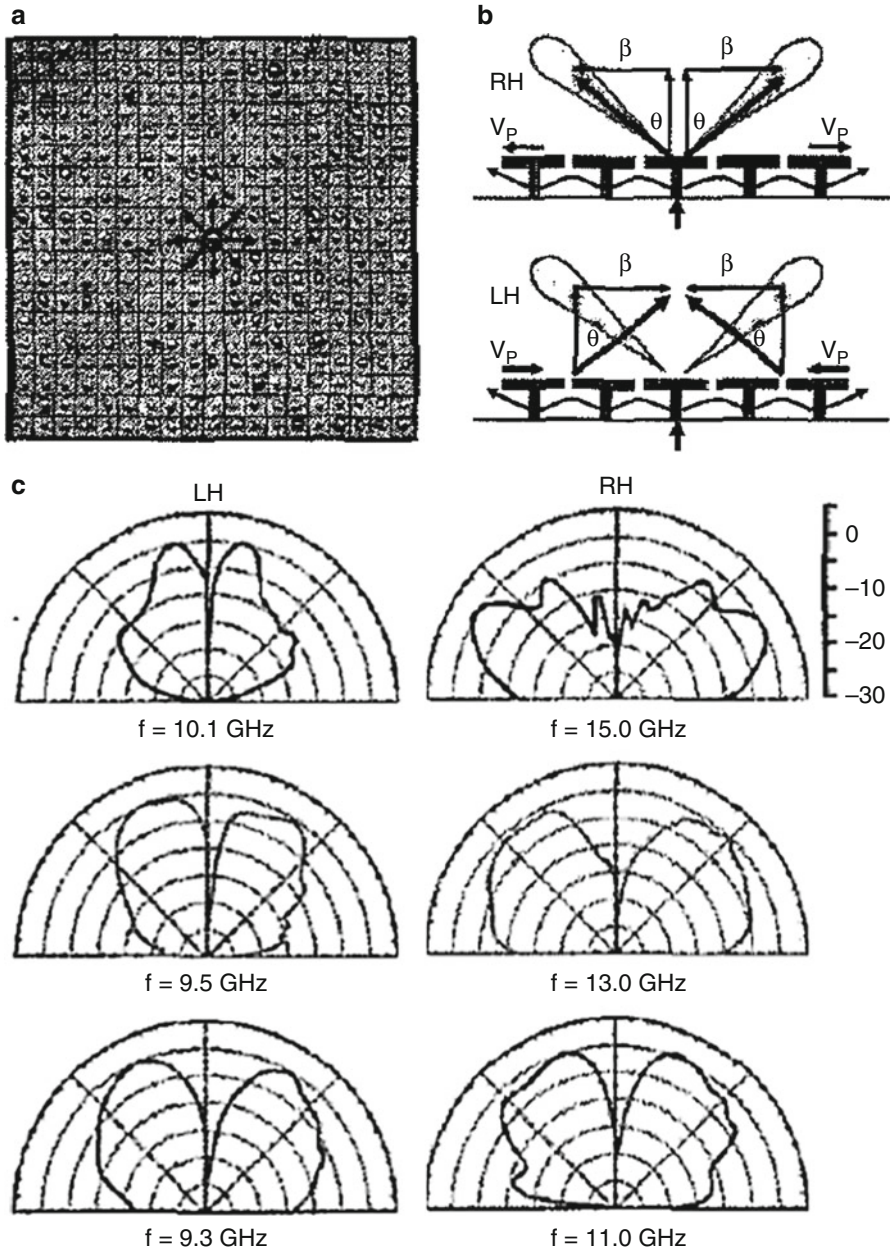


Fig. 24 (**a**) *Top* view of a 2-D leaky-wave antenna based on periodic mushroom structure, (**b**) side view of the antenna illustrating the RH and LH radiation, (**c**) radiation patterns for different operating frequencies (Allen et al. 2004, © IEEE)

and costly/bulky phase shifters, leaky-wave antennas provide an attractive alternative design solution using single radiating element and low-cost tuning schemes without a complex feed network.

Leaky-Wave Antenna-Based Radar Sensor

Beam-scanning leaky-wave antennas spread the radiated beams to different spatial angles with respect to the frequency. However, for single-frequency or narrowband operation, highly directive fixed-frequency beam-scanning antennas can also be realized using leaky-wave antennas. In 2009, Continental ADC commercialized a radar sensor based on a leaky-wave antenna for automobile applications. Beam scanning is achieved by placing a dielectric waveguide close to a rotating drum with periodic metallic ridges (Menzel and Moebius 2012). The wave traveling inside the dielectric waveguide encounters

DOI 10.1007/978-981-4560-75-7_60-1

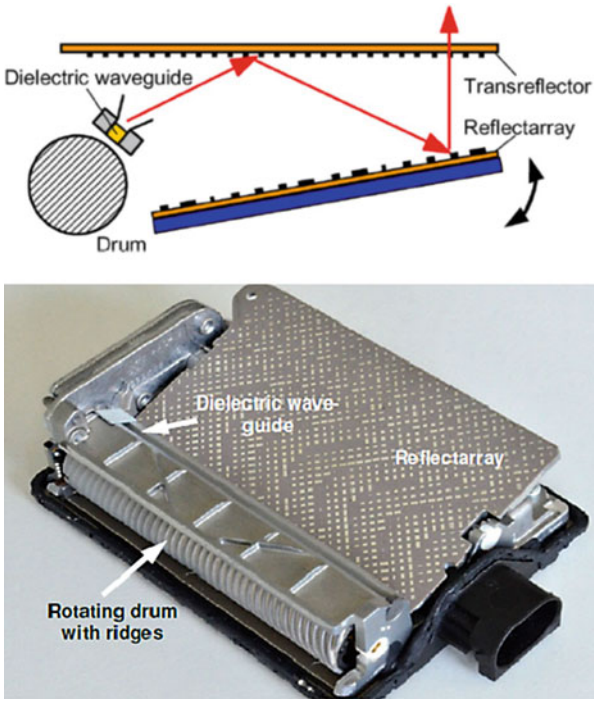


Fig. 25 (**a**) Sketch of ARS300 antenna, (**b**) photograph of ARS300 antenna with the cover removed (Menzel and Moebius 2012, © IEEE)

periodic perturbations from the ridges, leading to leaky radiation. The periodicity of the metallic ridges is varied by rotating the drum, thereby scanning the radiated beam along the azimuth direction (Manasson et al. 1996). Narrow beamwidth along the elevation angles is obtained by adjusting the phase angles of the reflectarray. The elevation angle of the final radiated beam leaving the radar sensor can also be adjusted by mechanically controlling the reflectarray angle. In this design, the transreflector not only serves as a polarizer to direct the radiated waves from the dielectric waveguide to the reflectarray but also works as a transparent cover to the waves leaving and entering the sensor. The sketch of the sensor and a photograph of the actual sensor with the cover removed are also shown in Fig. 25.

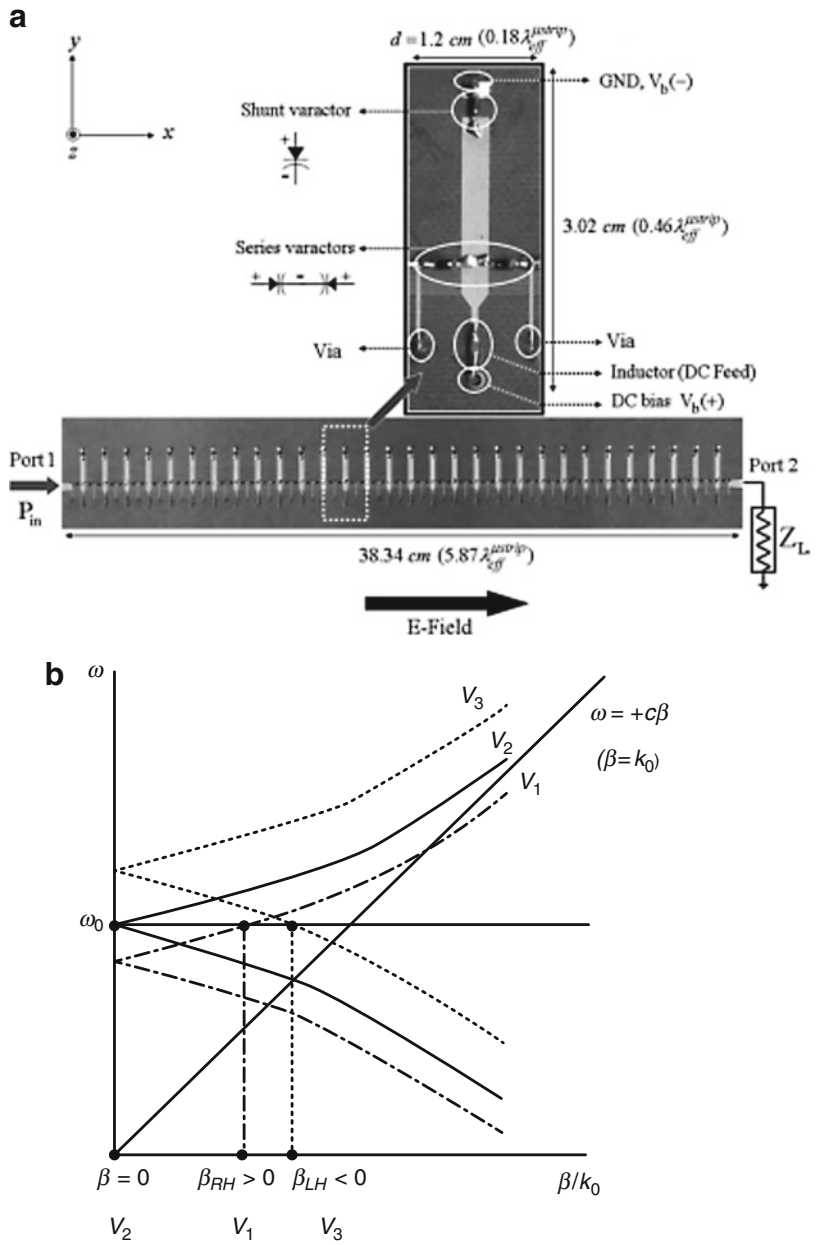
Phase-Shifter Less Electronically or Magnetically Scanning Leaky-Wave Antennas

In the previous section, single-frequency beam scanning is achieved through mechanical actuation. An alternate nonmechanical solution to generate fixed-frequency beam-scanning patterns can be realized using either electrically or magnetically controlled leaky-wave antennas (Baccarelli et al. 1997; Huang et al. 2000; Lim et al. 2004; Sato et al. 2006; Kodera and Caloz 2009; Ueda et al. 2012). In comparison to the conventional phased arrays, these techniques do not require bulky and costly voltage-controlled phase shifters to obtain fixed-frequency beam scanning. In the first example (Fig. 26), varactors are added in each unit cell to electrically tune series-capacitance and shunt-inductance values (Lim et al. 2004). An ideal electrically controlled leaky-wave antenna can shift the dispersion curve along the frequency axis by varying applied reverse bias voltages of the varactors. As a result, the propagation constant at a fixed operating frequency becomes a function of applied DC bias voltage, $\beta(V)|_{\omega o}$. Measured beam patterns of a 30-unit-cell electronically controlled leaky-wave antenna operating at 3.33 GHz with various biasing voltages ($V = 1.5$–18 V) are shown in Fig. 26.

A second example of a fixed-frequency beam-scanning antenna is realized by varying the external magnetic field of a ferrite-loaded asymmetrical CRLH antenna. The antenna is composed of a finite length CRLH TL terminated at both ends with either a short or open circuit. Once the energy is coupled to the

DOI 10.1007/978-981-4560-75-7_60-1

antenna, the wave travels down the waveguide and returns after reflecting back from the boundary. When an external magnetic field is applied perpendicular to the ground plane of a ferrite slab, field displacement occurs where the RF energy tends to be concentrated at one side of the transmission line (Hines 1971; Tsutsumi and Ueda 2004; Ueda et al. 2009). Therefore, nonreciprocal phase shift can be obtained by asymmetrical reactive loading. In (Ueda et al. 2012), CRLH structure is periodically loaded to one side of the transmission line and designed so that the propagation constant value has the same magnitude but opposite sign with respect to the case without the loading. Therefore, the radiated beams caused by the wave traveling in opposite directions will add in phase in the far-field. Without the external magnetic field, the ferrite material will behave as dielectric material and the radiated beam points to the broadside. But the dispersion diagram can be shifted along the βd-axis, depending on the magnitude and direction of the externally applied magnetic field as shown in Fig. 27b. The antenna was designed at around 6 GHz to



DOI 10.1007/978-981-4560-75-7_60-1

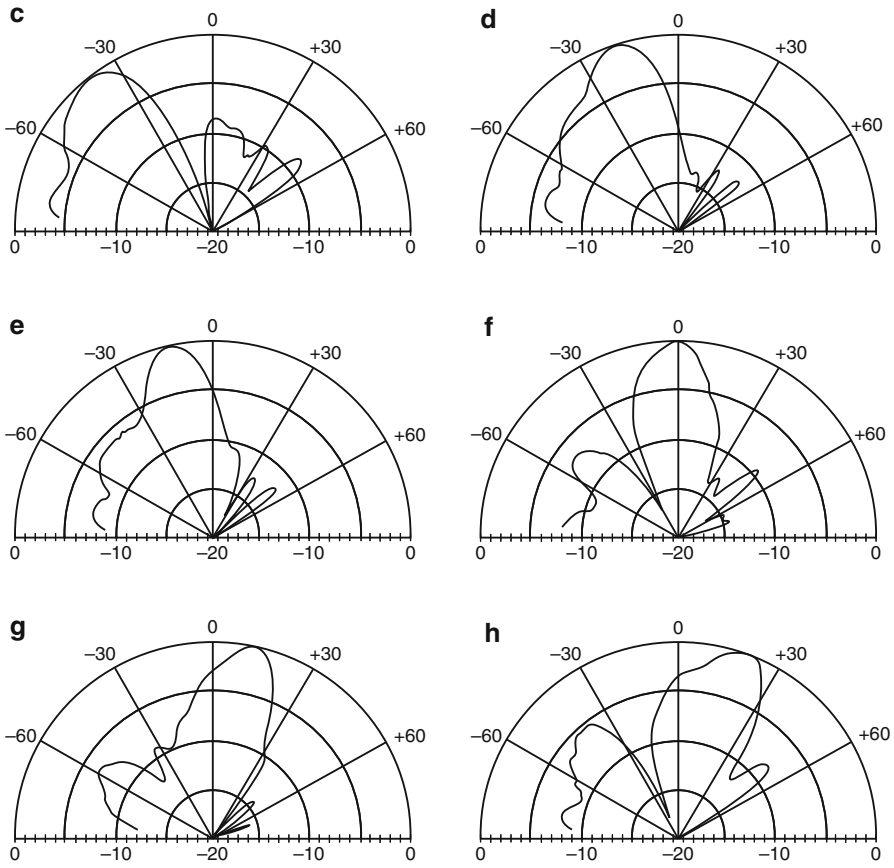


Fig. 26 (**a**) Photograph of an electronically controlled CRLH leaky-wave antenna, (**b**) ideal dispersion diagram illustrating the vertical dispersion shifting concept and measured normalized radiation patterns for various reverse biased voltages: V = (**c**) 18, (**d**) 10, (**e**) 8, (**f**) 3.5, (**g**) 2, and (**h**) 1.5 V (Lim et al. 2004, © IEEE)

avoid large magnetic loss below 4.5 GHz. The realized antenna demonstrated continuous backfire-to-endfire beam steering with more than 40° with almost constant 5 dBi gain.

Direction-of-Arrival Estimation Using Leaky-Wave Antenna

A simple direction-of-arrival (DoA) estimating system can be designed when an electronically scanning leaky-wave antenna is operated in the receiving mode. Although full-range ($-90^\circ \leq \theta_i \leq +90^\circ$) estimation can ideally be achieved using only one power meter attached to either end of the antenna, power detectors are attached at each end of the antenna in Abielmona et al. (2011) to relax the required dynamic range of the varactors. In this scheme, DoA is estimated by measuring the difference of power at the two detectors. Full-range estimation can be obtained by operating the system in either V_F-only mode or V_B-only mode, as illustrated in Fig. 28b. When the angle of the incident wave (θ_i) is aligned to the scan angle (θ_p) of the leaky-wave antenna, the received power at one of the detectors reaches a maximum value. θ_p is varied by changing the applied reverse biasing voltage of the varactors. A leaky-wave antenna-based DoA estimator provides an alternative cost-effective solution in a planar and compact form.

Terahertz Leaky-Wave Antenna

Terahertz (THz) frequencies accommodate larger communication bandwidth and provide high spatial resolution in comparison to microwave frequencies. Since the band is located just below the infrared

DOI 10.1007/978-981-4560-75-7_60-1

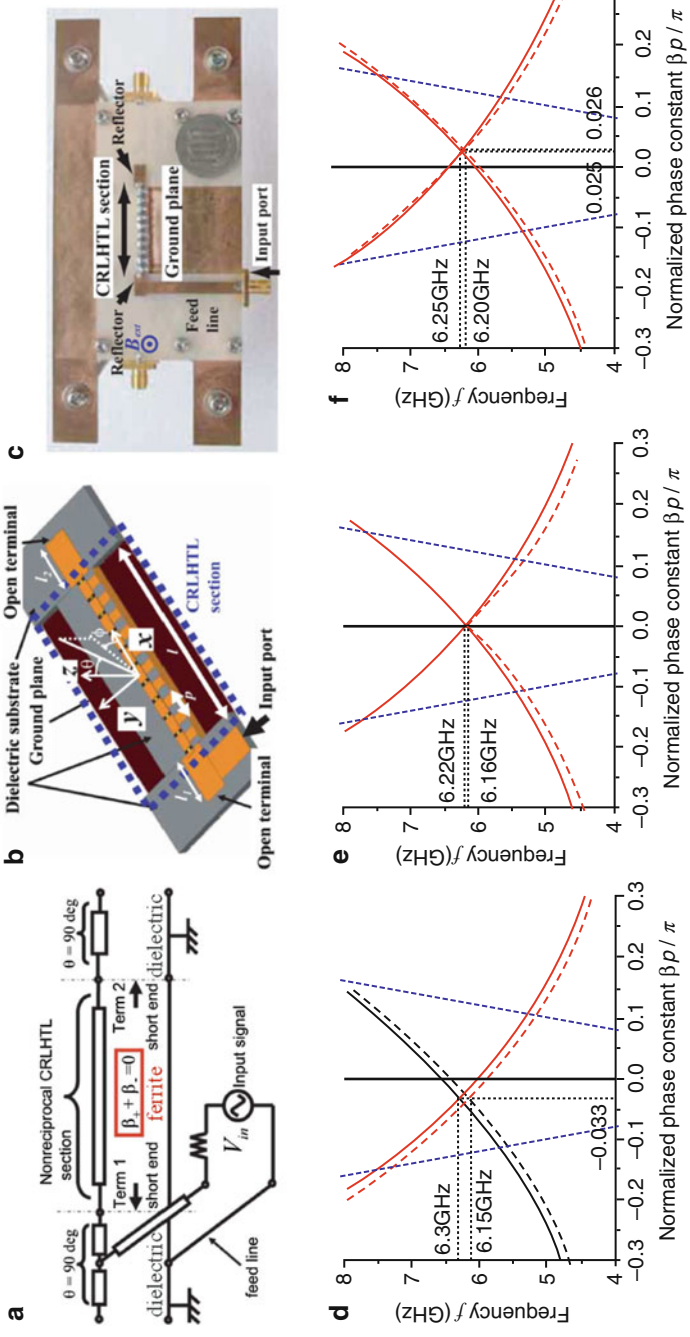


Fig. 27 Nonreciprocal CRLH TL: (**a**) equivalent circuit, (**b**) perspective view, (**c**) photograph of the fabricated circuit and dispersion diagrams of forward and backward propagating waves with various externally applied DC magnetic field: (**d**) $B_{ext} = -55$ mT, (**e**) $Be_{xt} = 0$ mT, and (**f**) $B_{ext} = 34.5$ mT (*solid* and *dash lines* represent measured and simulated results, respectively) (Ueda et al. 2012, © IEEE)

DOI 10.1007/978-981-4560-75-7_60-1

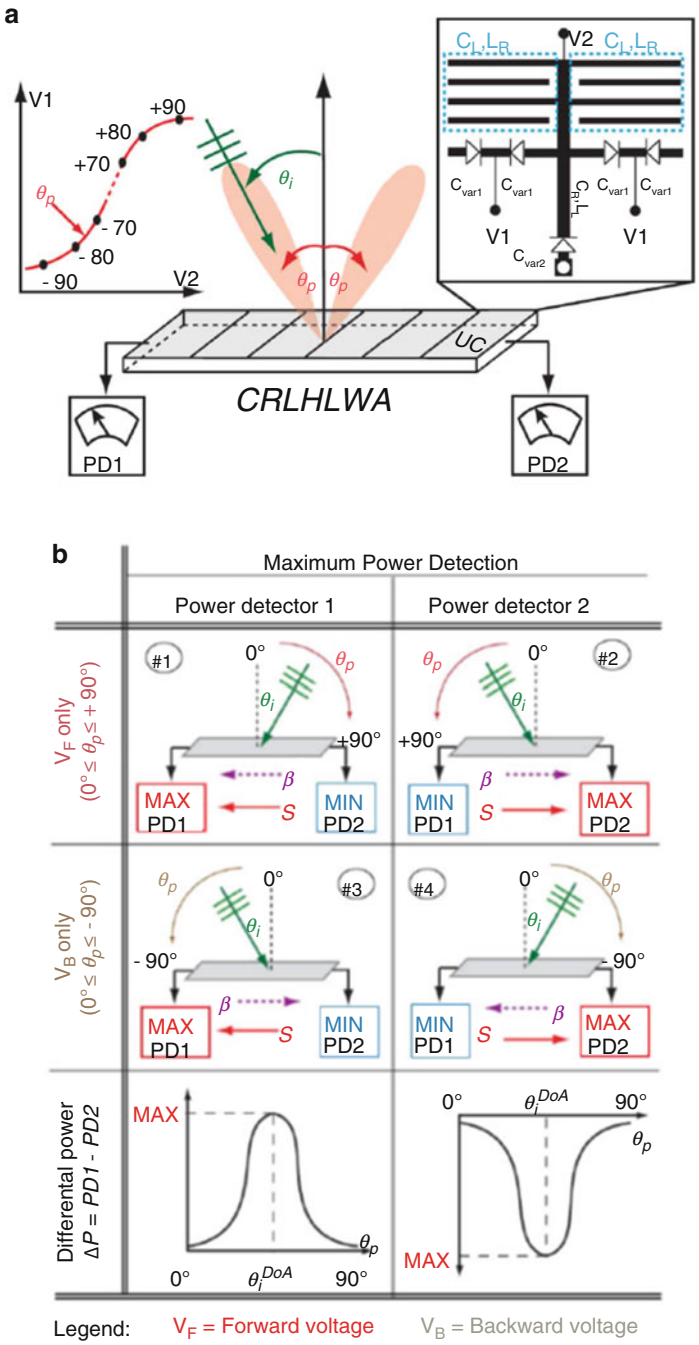


Fig. 28 (**a**) Sketch of a DoA estimation circuit based on electronically scanning leaky-wave antenna, (**b**) DoA estimation scenarios (Abielmona et al. 2011, © IEEE)

band, the nonionizing property is also less harmful to human tissues especially for medical imaging or security scanning applications. Despite these attractive features, realization of practical devices such as THz transceivers and sensors has been partially hampered by the antennas. Recently, leaky-wave antennas have been considered as a candidate for THz applications since directive beam patterns can be obtained in a relatively simple and compact form without a complex feed network. But designing THz antenna systems is nontrivial due to high material loss and difficulty of scaling solid-state electronic devices. In

DOI 10.1007/978-981-4560-75-7_60-1

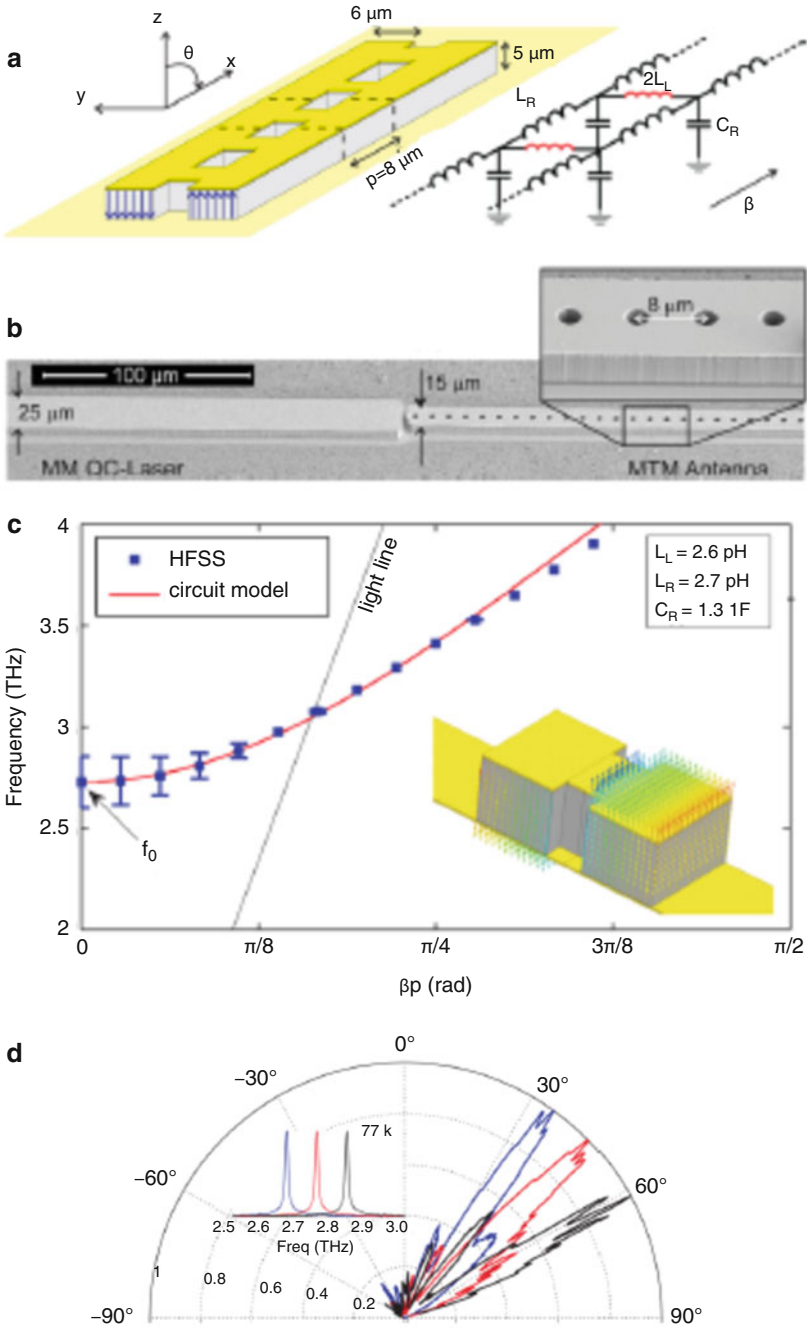


Fig. 29 (**a**) Perspective view and equivalent transmission-line mode of a 1-D THz leaky-wave antenna, (**b**) scanning electron micrograph (SEM) image of a fabricated THz leaky-wave antenna, (**c**) dispersion characteristic with the *inset* showing a unit cell with the odd-mode field profile, and (**d**) measured radiation pattern with the *inset* showing the measured spectra versus varying bias voltages (Reproduction with permission from A. A. Tavallaee, B. S. Williams, P. W. C. Hon, T. Itoh, and Q. S. Chen, "Terahertz quantum-cascade laser with active leaky-wave antenna," *Appl. Phys. Lett.*, vol. 99, no. 14. Copyright (2011), AIP Publishing LLC)

this section, recent achievements of THz leaky-wave antennas that utilize clever techniques that overcome the previously mentioned challenges are highlighted.

The first example shown in Fig. 29 is a leaky-wave antenna with distributed photonic gain excited by a quantum cascade laser (QC-laser) (Tavallaee et al. 2011). The structure resembles that of Menzel's

DOI 10.1007/978-981-4560-75-7_60-1

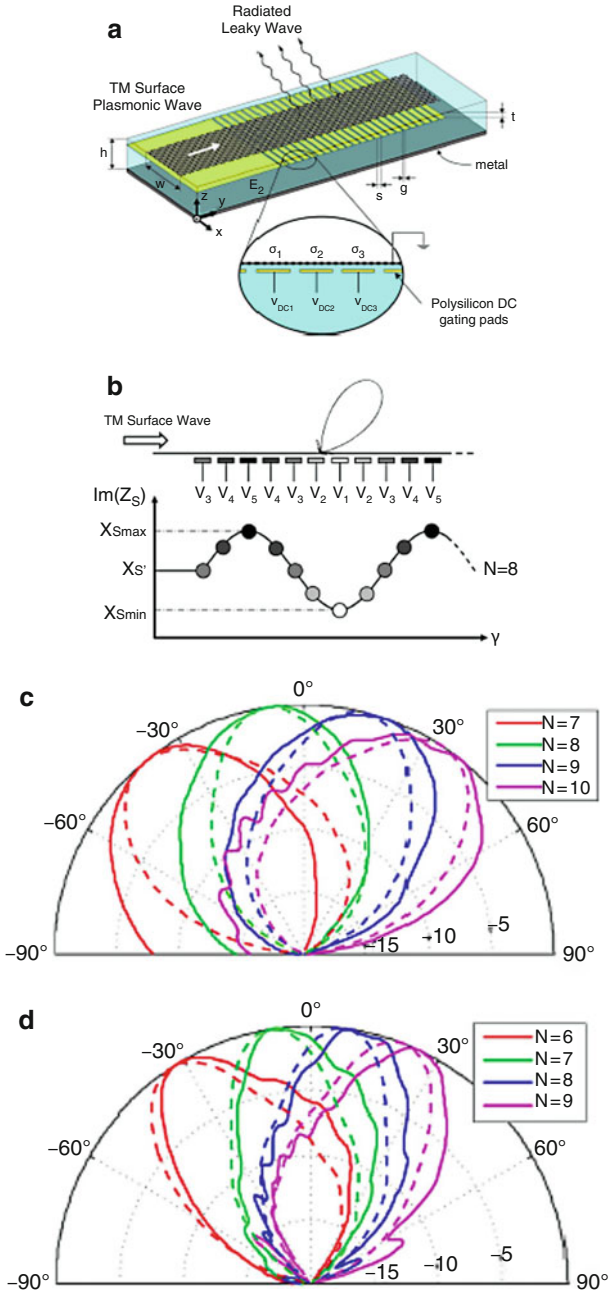


Fig. 30 (**a**) Terahertz leaky-wave antenna based on sinusoidally modulated reactance graphene surface, (**b**) sketch illustrating the relationship between the biasing DC voltage and corresponding graphene reactance profile, (**c**) radiation patterns versus N [parameters used: electron relaxation time (τ) = 1 ps, temperature (T) = 300 K, f_o = 2 THz, V_{DC} = 6.5–45 V, ε_r = 3.8, pad length (s) = 4.8 μm, gap between pads (g) = 0.2 μm], (**d**) normalized radiation patterns versus N using τ = 1 ps, T = 300 K, f_o = 2 THz, V_{DC} = 6.5–45 V, ε_r = 3.8, s = 4.8 μm, and g = 0.2 μm (Esquius-Morote et al. 2014, © IEEE)

antenna (Menzel 1979) and functions in a similar manner, where first higher-order mode is forced to radiate. GaAs/AlGaAs multiple quantum wells comprise the dielectric layer of the antenna and, when biased, provide amplification via stimulated emission of photons due to intersubband THz radiative transition. To provide independent biasing of the antenna, the antenna is proximity coupled to a metal-metal QC-laser cavity. The fabricated antenna provides directive patterns with full width at half maximum of around 15° and a frequency scanning toward endfire direction was also demonstrated around 2.7 THz.

DOI 10.1007/978-981-4560-75-7_60-1

The second example extends on the concept of a sinusoidally modulated surface described in section "Sinusoidally Modulated Reactance Surface-Based Leaky-Wave Antenna" (Oliner and Hessel 1959; Patel and Grbic 2007, 2011) to design a fixed-frequency beam-scanning leaky-wave antenna operating in the THz spectrum. The proposed antenna can be realized by transferring graphene sheet just above a set of polysilicon DC grating pads arranged on a grounded dielectric substrate (Fig. 30). The graphene sheet supports the propagation of transverse-magnetic surface plasmon polaritons, which can be manipulated by applying DC voltage to different pads that alter the complex conductivity σ (or surface impedance $1/\sigma$) of the graphene sheet (Esquius-Morote et al. 2014). In this design, one modulation period p is sampled in N points according to the number of grating pads used. For fixed pad length and distance between the pads, p can be dynamically controlled according to the periodicity imposed by the different V_{DC}. The simulated results validate fixed-frequency beam scanning at 2 THz when varying the sampling points (Fig. 30).

Summary

Beam-scanning leaky-wave antennas have been steadily improving ever since it was first introduced by W. W. Hansen in 1940 (Hansen 1940). With the addition of metamaterial concepts, novel leaky-wave antennas and unique design techniques have recently flourished in both research and industrial communities. As the general supporting technologies improve each year, leaky-wave antennas are expected to mature with improved fabrication technologies, novel realization methods, new exotic materials, etc., delivering yet more unique beam-scanning features with newer functionalities and enhanced radiation characteristics. With the ongoing advancements, leaky-wave antennas will continue to serve as a competitive design solution for applications requiring beam scanning, high directivity, and fabrication simplicity.

Cross-References

- ▶ Metamaterials and Antennas
- ▶ Transmission-Line Based Metamaterials in Antenna Engineering
- ▶ Traveling Wave Antennas (Including Leaky Wave and Surface Wave Antennas)

References

Abielmona S, Nguyen HV, Caloz C (2011) Analog direction of arrival estimation using an electronically-scanned CRLH leaky-wave antenna. IEEE Trans Antennas Propag 59(4):1408–1412

Allen CA, Caloz C, Itoh T (2004) Leaky-wave in a metamaterial-based two-dimensional structure for a conical beam antenna application. In: Proceedings of the IEEE-MTT international symposium, Fort Worth, TX, pp 305–308

Araki K, Koyama T, Naito Y (1975) A new type of isolator using the edge-guided mode. IEEE Trans Microw Theory Tech MTT-23(3):321

Baccarelli P, Nallo C, Freza F, Galli A, Lampariello P (1997) Attractive features of leaky-wave antennas based on ferrite-loaded open waveguides. In: IEEE AP-S international symposium digest, vol 2, Montreal, Canada, pp 1442–1445

Balanis CA (1997) Antenna theory: analysis and design, 2nd edn. Wiley, New York

DOI 10.1007/978-981-4560-75-7_60-1

Bozzi M, Georgiadis A, Wu K (2011) Review of substrate-integrated waveguide circuits and antennas. Microw Antennas Propag IET 5(8):909–920

Burghignoli P, Lovat G, Jackson DR (2006) Analysis and optimization of leaky-wave radiation at broadside from a class of 1D periodic structures. IEEE Trans Antennas Propag 54:2593–2603

Caloz C, Itoh T (2005) Electromagnetic metamaterials: transmission line theory and microwave applications. Wiley-IEEE Press, New York

Caloz C, Jackson DR, Itoh T (2011) Leaky-wave antennas. In: Frontiers in antennas: next generation design & engineering. McGraw-Hill, New York, pp 339–409

Collin RE (1991) Field theory of guided waves, 2nd edn. IEEE Press, New York

Dong Y, Itoh T (2011) Composite right/left-handed substrate integrated waveguide and half mode substrate integrated waveguide leaky-wave structures. IEEE Trans Antenna Propag 59(3):767–775

Dong Y, Itoh T (2012) Substrate integrated composite right/left-handed leaky-wave structure for polarization-flexible antenna application. IEEE Trans Antenna Propag 60(2):760–771

Duran-Sindreu M, Choi J, Bonache J, Martin F, Itoh T (2013) Dual-band leaky wave antenna with filtering capability based on extended-composite right/left-handed transmission lines. In: Proceedings of the IEEE-MTT international symposium, Seattle, WA

Eleftheriades GV, Balmain KG (2006) Negative refraction metamaterials: fundamental principles and applications. Wiley-IEEE Press, New York

Eleftheriades GV (2007) A generalized negative-refractive index transmission-line (NRI-TL) metamaterial for dual-band and quad-band applications. IEEE Microw Wirel Compon Lett 17(6):415–417

Engheta N, Ziolkowski RW (2006) Electromagnetic metamaterials: physics and engineering explorations. Wiley-IEEE Press, New York

Esquius-Morote M, Gomez-Diaz JS, Perruisseau-Carrier J (2014) Sinusoidally modulated graphene leaky-wave antenna for electronic beamscanning at THz. IEEE Trans THz Sci Technol 4(1):116–122

Grbic A, Eleftheriades GV (2002) A backward-wave antenna based on negative refractive index L-C networks. In: Proceedings of the IEEE antennas propagation society international symposium digest, vol 4, San Antonio, TX, pp 340–343

Hansen WW (1940) Radiating electromagnetic waveguide. US Patent 2.402.622

Hashemi MRM, Itoh T (2010) Coupled composite right/left-handed leaky-wave transmission-line based con common/differential-mode analysis. IEEE Trans Microw Theory Tech 58(12):3645–3656

Hessel A (1969) General characteristics of traveling-wave antennas. In: Collin RE, Zucker FJ (eds) Antenna theory, part II. McGraw-Hill, New York

Hines ME (1971) Reciprocal and nonreciprocal modes of propagation in ferrite stripline and microstripline devices. IEEE Trans Microwave Theory Tech MTT-19(5):442–451

Hirokawa J, Ando M (1998) Single-layer feed waveguide consisting of posts for plane TEM wave excitation in parallel plate. IEEE Trans Microw Theory Tech 46(5):625–630

Huang L, Chiao J, Lisio P (2000) An electronically switchable leaky wave antenna. IEEE Trans Antennas Propag 48(11):1769–1772

Ishii M, Fukusako T (2012) Circularly polarized leaky wave antenna using composite right/left-handed transmission line. In: PIERS proceedings, Moscow, Russia, pp 972–975

Jackson DR, Oliner AA (2008) Leaky-wave antennas. In: Balanis C (ed) Modern antenna handbook. Wiley, New York

Jackson DR, Caloz C, Itoh T (2012) Leaky-wave antennas. Proc IEEE 100(7):2194–2206

Jackson DR (2013) Recent advances in leaky-wave antennas. In: Proceedings of the URSI international symposium electromagnetic theory (EMTS), Hiroshima, Japan, pp 9–12

Kitatani K, Mitsuhata Y, Okamura Y (2012) Circularly polarized leaky wave antenna without phase shifter using CRLH transmission line. In: European conference on antennas and propagation (EUCAP), Prague, Czech, pp 255–259

Kodera T, Caloz C (2009) Uniform ferrite-loaded open waveguide structure with CRLH response and its application to a novel backfire-to-endfire leaky-wave antenna. IEEE Trans Microw Theory Tech 57(4):784–795

Lai A, Caloz C, Itoh T (2004) Composite right/left-handed transmission line metamaterials. IEEE Microw Mag 5(3):34–50

Lee H, Choi JH, Kasahara Y, Itoh T (2014a) A circularly polarized single radiator leaky-wave antenna based on CRLH-inspired substrate integrated waveguide. In: Proceedings of the IEEE-MTT international symposium, Tampa, FL, to be published

Lee H, Choi JH, Wu CM, Itoh T (2014b) A compact single radiator circularly polarized leaky-wave antenna based on CRLH-inspired substrate integrated waveguide. IEEE Trans Antenna Propag IEEE Trans Antenna Propag, submitted

Lim S, Caloz C, Itoh T (2004) Metamaterial-based electronically controlled transmission line structure as a novel leaky-wave antenna with tunable radiation angle and beamwidth. IEEE Trans Microw Theory Tech 52(12):2678–2690

Liu L, Caloz C, Itoh T (2000) Dominant mode (DM) leaky-wave antenna with backfire-to-endfire scanning capability. Electron Lett 38(23):1414–1416

Machac J, Polivka M (2012) A dual band SIW leaky wave antenna. In: Proceedings of the IEEE-MTT international symposium, Montreal, Canada

Manasson V, Sadovnik L, Mino R (1996) MMW scanning antenna. IEEE AES system magazine, pp 29–33

Menzel W (1979) A new travelling wave antenna in microstrip. Arch Elektron Uebertrag Tech 33(4):137–140

Menzel W, Moebius A (2012) Antenna concepts for millimeter-wave automotive radar sensors. Proc IEEE 100(7):2372–2379

Menzel W (2013) Planar leaky-wave antennas – early concept and actual result. In: European microwave conference, Nuremberg, Germany, pp 483–486

Nguyen HV, Parsa A, Caloz C (2010) Power-recycling feedback system for maximization of leaky-wave antennas radiation efficiency. IEEE Trans Microw Theory Tech 58(7):1641–1650

Oliner A, Hessel A (1959) Guided waves on sinusoidally-modulated reactance surfaces. IRE Trans Antennas Propag 7:201–208

Oliner AA, Lee KS (1986) The nature of the leakage from higher modes on microstrip line. In: IEEE MTT-S international microwave symposium digest, Baltimore, MD, pp 57–60

Oliner AA, Jackson DR (2007) Leaky-wave antennas. In: Volakis J (ed) Antenna engineering handbook, 4th edn. McGraw-Hill, New York

Patel AM, Grbic A (2007) A printed leaky-wave antenna with a sinusoidally modulated surface reactance. In: Proceedings of the IEEE AP-S international symposium, Charleston, SC, pp 3820–3023

Patel AM, Grbic A (2011) A printed leaky-wave antenna based on a sinusoidally-modulated reactance surface. IEEE Trans Antennas Propag 59(6):2087–2096

Paulotto S, Baccarelli P, Frezza F, Jackson DR (2009) Novel technique for open-stopband suppression in 1-D periodic printed leaky-wave antenna. IEEE Trans Antennas Propag 57(7):1894–1906

Pozar DM (1998) Microwave engineering, 2nd edn. Wiley, New York

Rennings A, Otto S, Mosig J, Caloz C, Wolff I (2006) Extended composite right/left-handed metamaterial and its application as quadband quarter-wavelength transmission line. In: Proceedings of the Asia-Pacific microwave conference (APMC), Yokohama, pp 1405–1408

Ryan CGM, Eleftheriades GV (2010) A dual-band leaky-wave antenna based on generalized negative-refractive-index transmission-lines. In: IEEE antennas and propagation society international symposium (APS-URSI), Toronto, Canada, pp 1–4

Sato K, Matsuzawa S, Inoue Y, Nomura T (2006) Electronically scanned left-handed leaky wave antenna for millimeter-wave automotive applications. In: Proceedings of the IEEE international workshop antenna technology, White Plains, NY, pp 420–423

Sato K, Yonak SH, Nomura T, Matsuzawa S, Iizuka H (2007) Metamaterial for automotive applications. In: Proceedings of the IEEE antenna propagation society international symposium digest, Honolulu, HI, pp 1144–1147

Shelley MW, Mason AG, Markus K (1991) Frequency scanning antennas for aeronautical communications. In: Proceedings of the IEE colloquium on satellite antenna technology in the 21st century, London, UK, pp 12/1–12/4

Sievenpiper D, Zhang D, Broas RFJ, Alexopolus NG, Yablonovitch E (1999) High-impedance electromagnetic surfaces with a forbidden frequency band. IEEE Trans Microw Theory Tech 47:2059–2074

Silver S (1949) Microwave antenna theory and design. McGraw-Hill, New York

Siragusa R, Perret E, Lemaitre-Auger P, Van Nguyen H, Tedjini S, Caloz C (2012) A tapered CRLH interdigital/sub leaky-wave antenna with minimized side-lobe levels. IEEE Antennas Wirel Propag Lett 11:1214–1217

Stutzman WL, Thiele GA (1998) Antenna theory and design, 2nd edn. Wiley, New York

Sutinjo A, Okoniewski M, Johnson RH (2008) Radiation from fast and slow traveling waves. IEEE Antennas Propag Mag 50(4):175–181

Tamir T (1969) Leakey-wave antennas. In: Collin RE, Zucker FJ (eds) Antenna theory, part II. McGraw-Hill, New York

Tavallaee AA, Williams BS, Hon PWC, Itoh T, Chen QS (2011) Terahertz quantum-cascade laser with active leaky-wave antenna. Appl Phys Lett 99:1411–15

Tospann FJ, Pirkl M, Gruner W (1995) Multifunction 35 GHz FMCW radar with frequency scanning antenna for synthetic vision application. In: Proceedings of the SPIE conference synthetic vision for vehicle guidance and control, vol 243, Orlando, FL, pp 28–37

Tsutsumi M, Ueda T (2004) Nonreciprocal left-handed microstrip lines using ferrite substrate. In: IEEE MTT-S international microwave symposium digest, vol 1, Fort Worth, TX, pp 249–252

Ueda T, Horikawa K, Akiyama M, Tsutsumi M (2009) Nonreciprocal phase-shift composite right/left handed transmission lines and their application to antennas. IEEE Trans Antennas Propag 57(7):1995–2005

Ueda T, Yamamoto S, Kado Y, Itoh T (2012) Pseudo-traveling wave resonator with magnetically tunable phase gradient of fields and its applications to beam steering antennas. IEEE Trans Microw Theory Tech 60(10):3043–3054

Whetten F, Balanis CA (1991) Meandering long slot leaky-wave antennas. IEEE Trans Antennas Propag 39(11):1553–1559

Wu K, Deslandes D, Cassivi Y (2003) The substrate integrated circuits – a new concept for high-frequency electronics and optoelectronics. In: Proceedings of the 6th international conference telecommunications in modern satellite, cable and broadcasting service, vol 1, Nis, Yugoslavia

Wu JW, Wang CJ, Jou CF (2009) Method of suppressing the side lobe of a tapered short leaky wave antenna. IEEE Antennas Wirel Propag Lett 8:1146–1149

Yang N, Caloz C, Wu K (2009) Fixed-beam frequency-tunable phase reversal coplanar stripline antenna array. IEEE Trans Antennas Propag 57(3):671–681

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_61-2

Reconfigurable Antennas

Joseph Costantine[a]*, Youssef Tawk[b] and Christos G. Christodoulou[c]
[a]Electrical and Computer Engineering Department, American University of Beirut, Beirut, Lebanon
[b]Electrical and Computer Engineering Department, Notre Dame University-Louaize, Zouk Mosbeh, Lebanon
[c]Electrical and Computer Engineering Department, The University of New Mexico, Albuquerque, NM, USA

Abstract

In this chapter, a summary of the various components, categorization, and design process of reconfigurable antennas is presented. The various applications where reconfigurable antennas have been implemented and the association of various antenna properties with new communication applications are discussed. The chapter takes into consideration new and evolving applications and associate novel antenna designs with these applications. The control, modeling, and optimization of reconfigurable antennas are also detailed and multiple mechanisms are presented. Finally this chapter emphasizes on the development of reconfigurable antennas to service futuristic and evolving practical wireless communication applications and propose optimal solutions for more efficient and holistic designs.

Keywords

Reconfigurable antennas; Wireless communications; MIMO systems; Mobile terminals; Cognitive radio; Wireless LAN; Wireless sensing; Terahertz communication; Millimeter waves; Space communication; Small satellites; Self-biasing; Graph models; Arduino; Neural networks; FPGA; PIN diodes; Varactors; RF MEMS; Photoconductive switches; Thermal switches; Actuators; Motors; Plasmonics; Graphene; Liquid crystals

Introduction

New emerging applications in wireless communications require advanced antenna systems that are able to satisfy the needs of such applications. Reconfigurable antennas appear to offer a solution that allows the integration of multiple radios into a single platform. The integration of multiple radios with reconfigurable antennas maximizes connectivity without resorting to multiple antenna components. The ability of reconfigurable antennas to change their functionality on demand allows them to dynamically cater for multiple wireless services without increasing the real estate required to accommodate multiple antennas.

Designing any antenna requires the understanding of its fundamental characteristics such as gain, impedance matching, and radiation properties. For a reconfigurable antenna, these characteristics become tight constraints, especially when an antenna designer is trying to come up with a reliable design that is predictable for every state of its reconfiguration. Reconfiguring an antenna is achieved by rearranging the antenna currents or reconfiguring its radiating edges (Bernhard 2007; Balanis 2011; Christodoulou et al. 2012; Costantine et al. 2013c). The change in the antenna functionality allows it to be used for multiple wireless communications applications.

*Email: jcostantine@ieee.org

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_61-2

An antenna design can be made reconfigurable by utilizing various reconfiguration techniques such as switches or mechanical actuators. The most widely reconfigured antenna type is the microstrip antenna, where the integration of reconfiguring components into the antenna structure becomes part of the antenna design process. Typically, a reconfigurable patch antenna consists of a number of separate metalized regions that lie on a plane and are connected together using tuning components (Costantine et al. 2013c). By dynamically controlling the state of the tuning components, the different metalized sections can be activated in a very selective way, thus altering the radiation performance of the total antenna.

In general, every reconfigurable antenna design starts with a predetermined objective that aims to satisfy the existing technical constraints. However, every antenna engineer targets a certain antenna property to reconfigure. An example of such property can be frequency of operation, radiation pattern, and/or polarization. Once this property is determined, the designer works to achieve this objective by rearranging the different antenna components correspondingly.

In this chapter, the different methods of designing reconfigurable antennas that do not resort to redundant components while maintaining an optimal performance are discussed. The applications of reconfigurable antennas in practical wireless communication applications are presented. This chapter also discusses the modeling, control, and optimization of reconfigurable antenna structures.

How to Design Reconfigurable Antennas?

The different methods adopted in the design of reconfigurable antennas are based on the reconfiguration techniques as well as the properties of each reconfiguration. Reconfigurable antennas can be assembled into various groups based on the properties of each reconfiguration (Costantine et al. 2013c). These groups are shown in Fig. 1 (Costantine et al. 2013c) and are arranged as follows:

- *Group 1: Frequency reconfigurable antennas*
  Antennas under this group can change their frequency of operation based on the user's demand. They reconfigure their operation to function in multiple frequency bands. Such antennas are widely useful in wireless communication applications that require a change in operating frequencies and to switch from one channel into another. Cognitive radio is an example application for this antenna group.
- *Group 2: Radiation pattern reconfigurable antennas*
  Antennas under this group can change their radiation pattern while maintaining a fixed frequency of operation. These antennas reshape their radiation patterns to block a signal or to allow radiation in a certain predetermined direction. Mobile antenna systems can be proposed as examples of such antenna group.
- *Group 3: Antennas with reconfigurable polarization*
  Antennas under this group can change their polarization type while maintaining their fixed frequency and radiation pattern. These antennas reshape their radiation characteristics to exhibit multiple

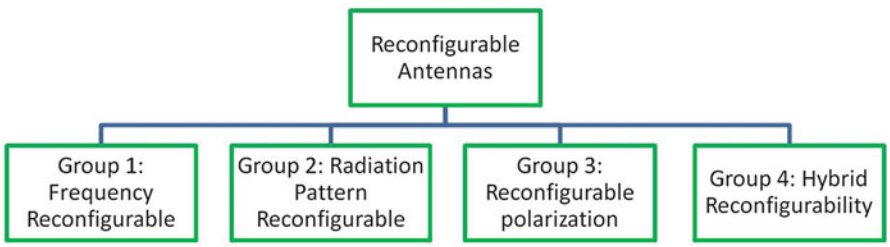


Fig. 1 Categorization of reconfigurable antennas (Costantine et al. 2013c)

DOI 10.1007/978-981-4560-75-7_61-2

polarization schemes. MIMO (multiple input multiple output) channels are application examples of such antenna group.

– *Group 4: Antennas with hybrid reconfiguration techniques*

  Antennas under this group can simultaneously change multiple characteristics in their operation. These antennas can, for example, change their operating frequency as well as their polarization scheme for each frequency of interest. They can also reshape their radiation pattern while changing their operating frequencies or polarizations.

Achieving the required reconfiguration functionality requires an appropriate selection of the reconfiguration mechanism that will achieve the desired functionality or property. A reconfiguration mechanism is selected to satisfy imposed constraints while completing the antenna design task efficiently. There are numerous reconfiguration techniques that can be incorporated into the antenna design in order to redistribute its currents. The reconfiguration techniques are also divided into multiple categories as shown in Fig. 2 (Costantine et al. 2013c; Christodoulou et al. 2012).

Electrical reconfiguration techniques are based on the use of switches to connect and disconnect antenna parts as well as to redistribute the antenna currents. Radio-frequency microelectromechanical systems (RF MEMS) have been proposed for integration into reconfigurable antennas since 1998 (Brown 1998). Many designs have resorted to RF MEMS to reconfigure their performance (Huff and Bernhard 2006; Erdil et al. 2007; Nikolaou et al. 2009; Cetiner et al. 2010; Grau et al. 2010; Yang et al. 2011; Zohur et al. 2013). RF MEMS-based reconfigurable antennas rely on the mechanical movement of these switches to achieve reconfiguration. The isolation of RF MEMS is very high, and they require minimal power consumption. The switching speed of RF MEMS is in the range of 1–200 μsec which may be considered low for some applications (Christodoulou et al. 2012).

Faster electrical reconfiguration techniques lead to PIN diodes or varactors. These types of switches are faster and more compact alternatives to RF MEMS. The switching speed of a PIN diode is in the range of 1–100 nsec (Christodoulou et al. 2012). Reconfigurable antennas resorting to PIN diodes (Chen et al. 2007; Chen and Row 2008; Kim et al. 2008; Wu and Ma 2008; Sarrazin et al. 2009; Shelley

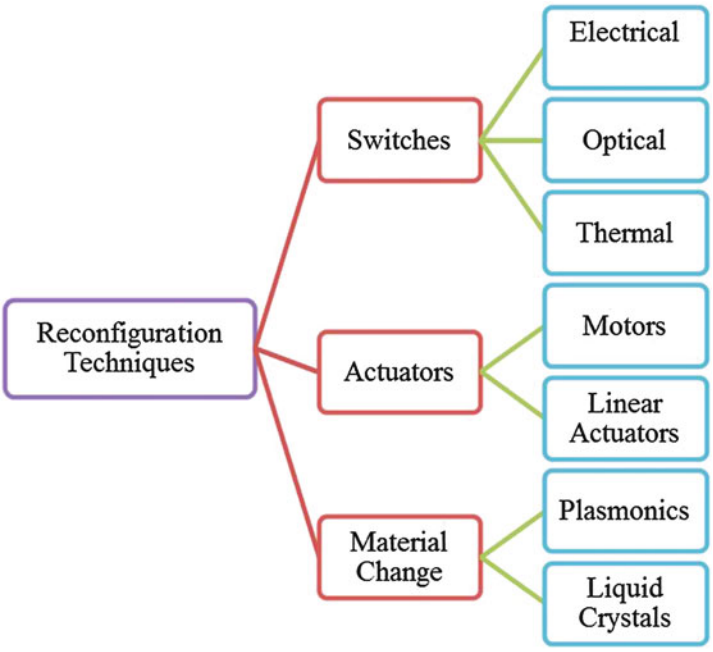


Fig. 2 Different reconfiguration techniques

DOI 10.1007/978-981-4560-75-7_61-2

et al. 2010; Perruisseau-Carrier et al. 2010; Piazza et al. 2010; Qin et al. 2010b; Quin et al 2012; Hinsz and Braaten 2014) have a more dynamic reconfiguration capability. Other reconfigurable antennas resort to the use of varactors (Daviu et al. 2007; Jeong et al. 2008; Yang et al. 2008; Jiang et al. 2009; Oh et al. 2010; Tawk et al. 2012a; Bai et al. 2013; Onodera et al. 2013; Ramadan et al. 2014) where varying the biasing voltage can result in varying the capacitance of the corresponding varactor. Such antennas enjoy a wide tuning ability that is based on integrating a variable capacitance into the antenna structure. It is important to indicate that while electrical switching components may present an efficient reconfiguration mechanism, they require an appropriate design of their biasing networks.

The integration of photoconductive switches into antenna structures allow the use of optical switching mechanisms. Photoconductive switches incorporated into an antenna structure become conductive once they are subjected to a laser beam (Tawk et al. 2010, 2012b; Jin et al. 2011; Sathi et al. 2012; Patron et al. 2014; Pendharker et al. 2014). The laser beam originates from integrated laser diodes (Tawk et al. 2012b). The integration of all these components into compact antenna structures, while taking into consideration power consumption issues, constitutes challenges of such reconfiguration techniques. Thermal switches (Jiang and Yang 2013) have also been proposed as candidates for antenna reconfiguration. Other reconfiguration techniques are based on graphene plasmonics (Perruisseau-Carrier et al. 2013), liquid crystals (Liu and Langley 2008), and mechanical reconfiguration techniques (Rodrigo et al. 2012; Liyakath et al. 2013; Mehdipour et al. 2013; Costantine et al. 2014).

As a summary, an antenna designer applies a reconfiguration technique to any antenna structure and thus transforms it into a reconfigurable antenna. The application of an appropriate reconfiguration technique allows the implementation of the corresponding antenna in a practical wireless communication application that requires an adaptive RF front-end. The cost of adding reconfigurability to the antenna behavior can be determined as a function of multiple factors (Christodoulou et al. 2012) as discussed below:

(a) Factor 1: A complexity is added to the antenna structure due to the presence of biasing networks that are required to control the state of the reconfiguration components.
(b) Factor 2: The incorporation of active components results in increased power consumption.
(c) Factor 3: Undesired effects such as generation of harmonics due to the addition of nonlinear active components.
(d) Factor 4: The speed of the reconfiguration leading to the required functionality.

Why Design Reconfigurable Antennas?

The new era of antenna design must generate antennas that are cognitive and adjust to the environment and ever-changing surrounding conditions. Also, there is a need for antennas that can overcome failures and swiftly respond to new developments. Cognitive radio, multiple input multiple output (MIMO) channels, on-body networks, satellites, and space communication platforms are all possible applications for the integration of highly versatile, reliable, and efficient reconfigurable antennas. In this section, the design of reconfigurable antennas that are proposed to service different practical wireless communication applications is discussed.

Reconfigurable Antennas for MIMO Communication Systems

A MIMO communication system is based on the need for multiple antennas that are installed at both sides of the communication channel. The purpose of such installation is to enable the communication channel to transmit and receive multiple signals at the same time. This ability increases the spectral efficiency and

DOI 10.1007/978-981-4560-75-7_61-2

capacity, thus overcoming issues such as multipath and fading. On the other hand, a large improvement over the channel capacity can be achieved by incorporating reconfigurable antennas that exhibit a reconfiguration in their radiation patterns and polarizations (Cetiner et al. 2004; Piazza et al. 2008; Li et al. 2009; Grau et al. 2010; Qin et al. 2010a; Christodoulou et al. 2012).

As an example, an antenna shown in Fig. 3 (Quin et al 2012) exhibits a reconfigurable radiation pattern. The antenna is a microstrip U-slot patch antenna that uses eight PIN diodes to achieve pattern reconfiguration. Eight shorting posts are implemented around the patch to change the operating mode of the antenna from monopolar patch mode to normal patch mode (Quin et al 2012). The two modes of operation are designed to operate in similar frequency ranges.

The proposed antenna can electrically reconfigure the radiation pattern between conical and bore-sight patterns with an overlapping impedance bandwidth. Four antenna elements with the same design concept are then employed in a 2 × 2 MIMO-OFDM prototype. The capacity results are compared with omnidirectional antennas, and channel capacity enhancements are witnessed in both line-of-sight and non-line-of-sight indoor environments (Quin et al 2012).

Another example of an antenna designed for MIMO channels is an electronically steerable passive array radiator (ESPAR). This antenna shown in Fig. 4 (Zhou et al. 2014) consists of one driven monopole in the center surrounded by a ring of six uniformly spaced (quarter wavelength spacing) parasitic monopoles. Each parasitic element is terminated with a load that consists of a 0.3 nH inductor and a PIN diode. This load is essential to provide the reactance required to either make the monopole a director when it is inductive or a reflector when it is capacitive (Zhou et al. 2014). If three adjacent passive elements were configured as directors and the remaining three as reflectors, the antenna system would be able to generate a directive beam at a chosen direction. Shifting the configuration circularly by one element each time allows ESPAR to generate six patterns with the beam angles 60° apart in azimuth. Two prototypes are fabricated and are optimized to maximize pattern flexibility.

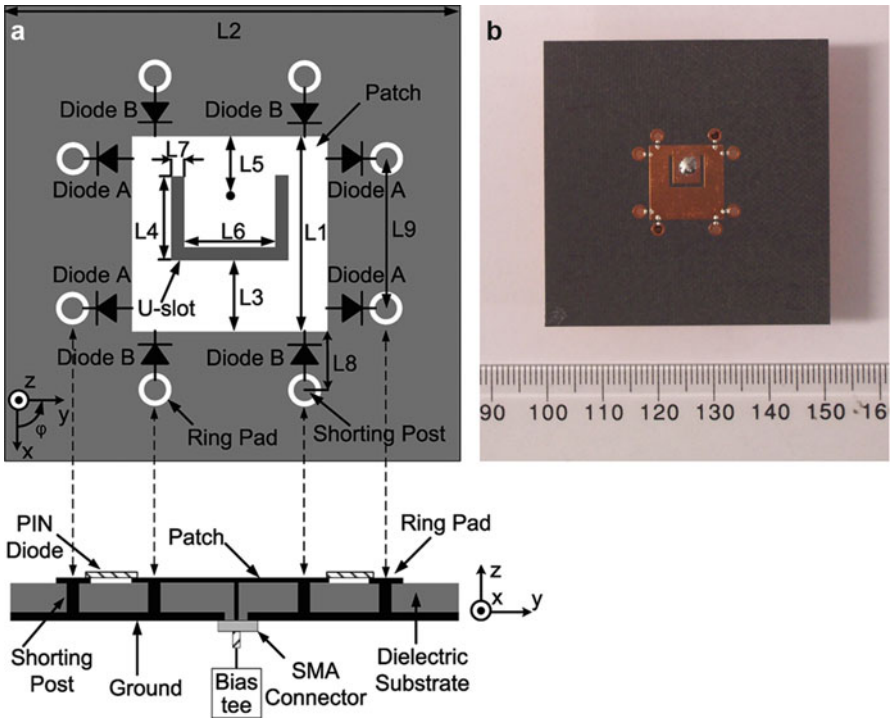


Fig. 3 Antenna with reconfigurable radiation pattern (Quin et al. 2012)

DOI 10.1007/978-981-4560-75-7_61-2

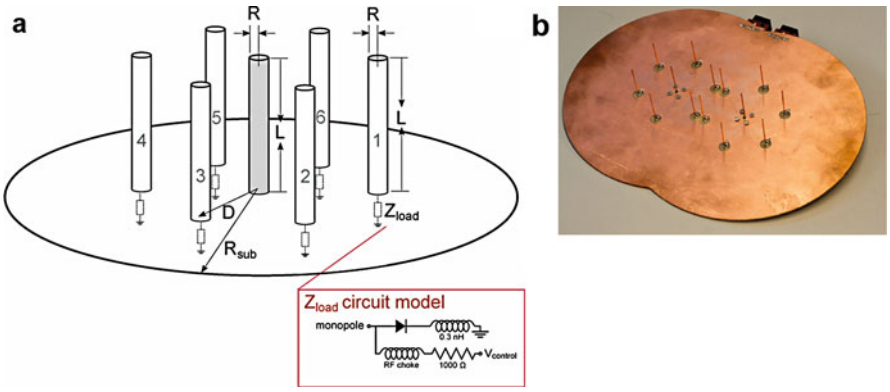


Fig. 4 Proposed and fabricated ESPAR for MIMO communications (Zhou et al. 2014)

Multiple techniques and algorithms are also discussed in literature to facilitate the reconfigurable antennas state selections allowing a smoother reconfiguration transition. For example, spatial correlation and average signal-to-noise ratio (SNR) information are utilized in Piazza et al. (2009) to select the antenna radiation pattern at the receiver. It is shown that the approach proposed (Piazza et al. 2009) achieves capacity gain in a multielement reconfigurable antenna system without changing the data frame of a conventional MIMO system (Piazza et al. 2009).

Reconfigurable Antennas for Mobile Terminals

Mobile terminals such as smartphones, GPS receivers, laptops, or other compact devices make use of multiband antennas to operate over multiple wireless protocols. Newly adopted smartphone platforms allow users to jump from one operational protocol to another with ease and fast selection ability. This multi-standards operation requires the integration of antennas that are able to reconfigure their operations swiftly and dynamically. Thus, reconfigurable antennas are proposed as suitable candidates for this application.

Multiple experiments have been carried out to prove the advantages of using a reconfigurable antenna setting over multiple antennas in mobile terminals. For example, in Yang et al. (2009) a GPS receiver is exposed to a jamming signal when the receiver is using either a multiband antenna or a reconfigurable antenna. The multiband antenna used in this experiment operates over both the GPS and wireless LAN frequencies, while the reconfigurable antenna can selectively change its operation from GPS to WLAN (Yang et al. 2009). It is found that, in the case where the reconfigurable antenna is used, no degradation in the signal quality is noticed. However, the multiband antenna showed a poor out-of-band rejection. Additional filtering would be required for the multiband antenna to attain better noise rejection (Yang et al. 2009). In general, frequency reconfigurable antennas possess inherent band-pass characteristics and generally have excellent out-of-band rejection without filters (Yang et al. 2009). Their higher-order resonances are typically far away from the operating band and can be removed with much less selective filters and at a lower cost (Yang et al. 2009).

The antenna shown in Fig. 5 (Li et al. 2012) is an example of a frequency reconfigurable antenna designed for mobile phones. The antenna design is composed of a folded loop that is reconfigured using one PIN diode and a tuning pad. A matching bridge is also incorporated for increased bandwidth. The PIN diode switches allow the antenna to be transformed from a folded loop into an inverted "F" antenna (IFA) (Li et al. 2012). As a result, hepta-band coverage is achieved in a compact volume. This reconfigurable antenna structure allows operation over GSM 850, GSM 900, GPS, DCS, PCS, UMTS, and WLAN (Li et al. 2012).

DOI 10.1007/978-981-4560-75-7_61-2

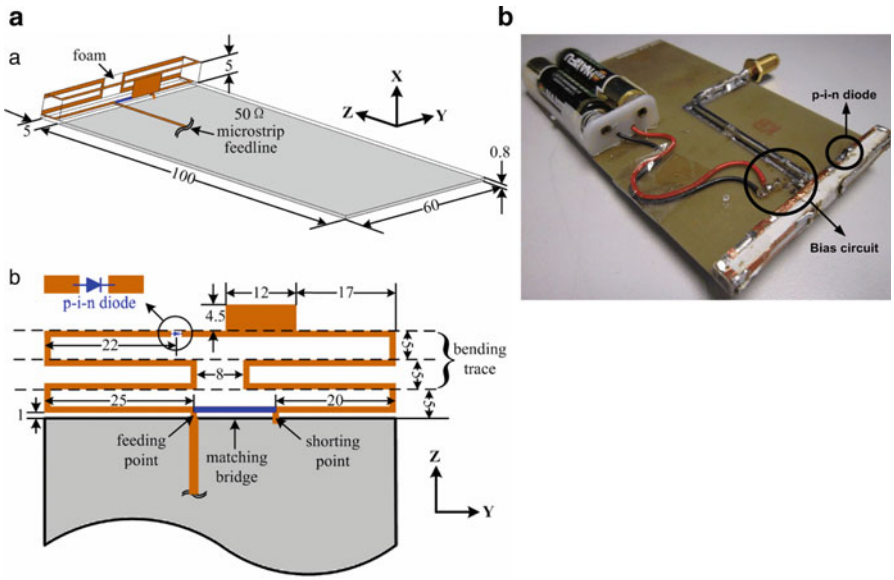


Fig. 5 Simulated and fabricated antenna system as proposed in Li et al. (2012)

Another antenna proposed as a suitable candidate for smartphones is a folded coupled-fed narrow-frame antenna for hepta-band WLAN/LTE (Ban et al. 2014). The antenna shown in Fig. 6 (Ban et al. 2014) is reconfigured using one PIN diode that changes its resonant mode by modifying the electrical length of the lower band in two working states (Ban et al. 2014). The antenna is proven to operate over GSM 850/900/1800/1900/UMTS/LTE2300/2500 (Ban et al. 2014).

Reconfigurable Antennas for Cognitive Radio Applications

A cognitive radio requires antenna systems that have the ability to identify changes in a communication setting and be able to react accordingly. Cognitive radio is based on the concept of monitoring a channel and searching for unused spaces in the spectrum. Once unused spaces are identified, a reconfigurable antenna is activated to tune its operation to broadcast, thus increasing the communication efficiency (Tawk et al. 2014a).

Channel sensing can be achieved through a wideband antenna or a reconfigurable narrowband antenna. The communication process on the other hand requires a reconfigurable antenna to be able to change its operation in order to manage the dynamically discovered white spaces in the spectrum. A cognitive radio device's operation is based on a cycle that is shown in Fig. 7 (Tawk et al. 2014a). This cycle starts by observing the channel activity, which is achieved by a sensing (wideband) antenna. In the second step of the cycle, a cognitive processor decides which part of the spectrum is suitable for communication. The processor next activates the communicating antenna to achieve the required mode of communication. In the last step, the processor achieves cognition by learning from previous channel activities. This cycle allows the cognitive radio device to self-decide and optimally self-reconfigure its hardware in order to realize the selected mode of communication (Jayaweera and Christodoulou 2011; Tawk et al. 2014a).

Multiple reconfigurable antenna designs are presented in literature for implementation on a cognitive radio platform. An example of such antennas is the one shown in Fig. 8 (Rajagopalan et al. 2014). The antenna presented is a frequency reconfigurable E-shaped patch antenna that provides wideband frequency switching using MEMS. The antenna is proposed for applications on cognitive radio platforms with large terminals such as laptops or as an array element (Rajagopalan et al. 2014). The presented E-shaped patch offers a simple single-layer, single-feed structure with a wide instantaneous bandwidth (Rajagopalan et al. 2014).

DOI 10.1007/978-981-4560-75-7_61-2

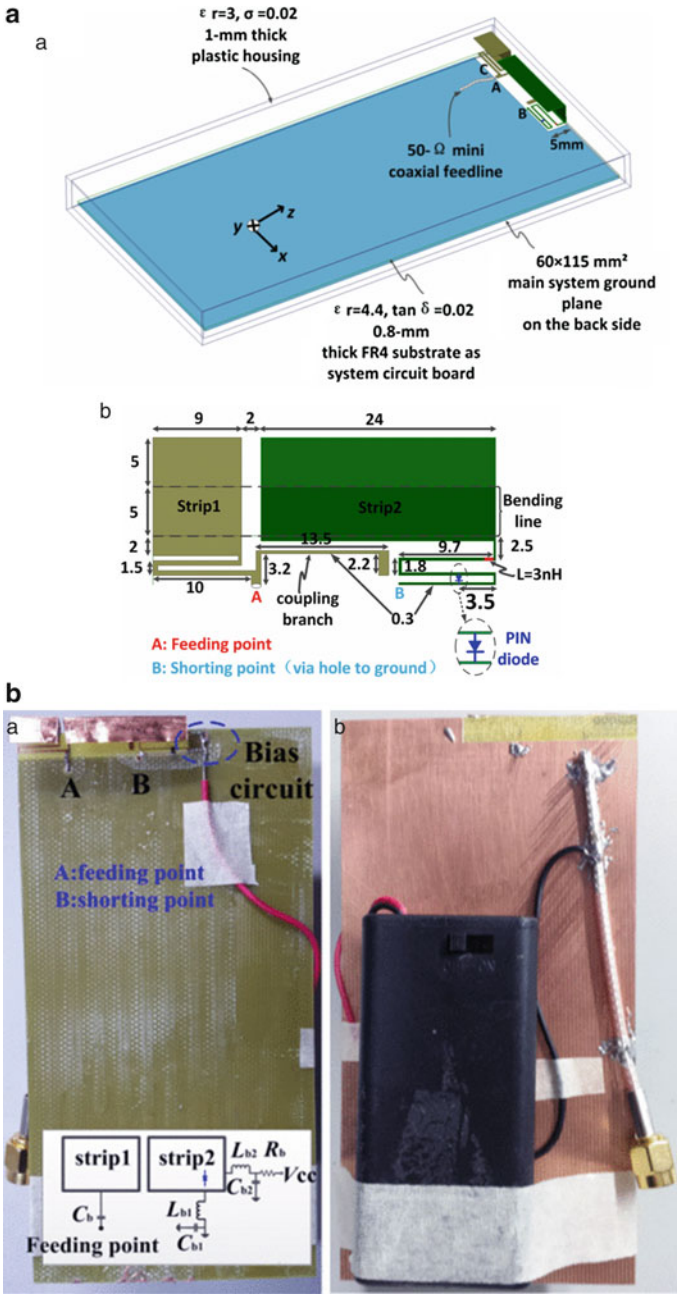


Fig. 6 Narrow-frame reconfigurable antenna proposed in Ban et al. (2014)

A reconfigurable antenna design that is based on switching between wide- and narrowband frequency operation is shown in Fig. 9 (Augustin et al. 2014). The antenna is a single-port antenna design that achieves switching between wideband and narrowband operation by resorting to a slot-based resonator filter. The filter consists of two quarter wavelength resonators that generate two stop bands far spread out to form a passband. Two PIN diodes are embedded across the slot entrance from the CPW. In addition to that, on either side of the CPW, two 0.2 mm slots incorporated with capacitors are introduced. The antenna design also resorts to an elliptically curved slot to reduce back-lobe radiation as well as to enhance the cross-polar level of the antenna (Augustin et al. 2014). An antenna based on the same design concept of using a slot resonator filter is also proposed in Chacko et al. (2015) with an added feature of polarization diversity.

DOI 10.1007/978-981-4560-75-7_61-2

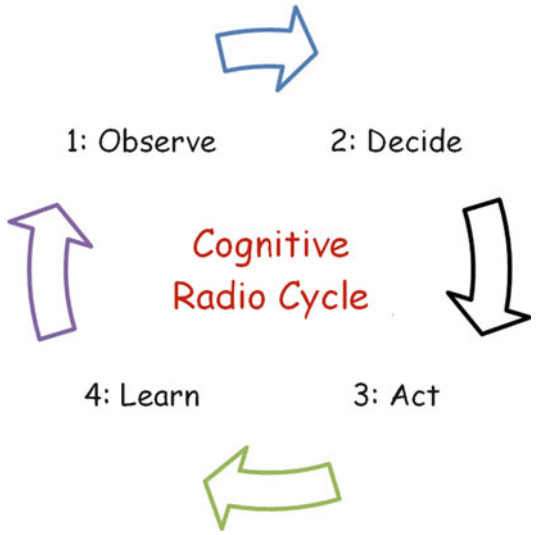


Fig. 7 The cognitive radio cycle (Tawk et al. 2014a)

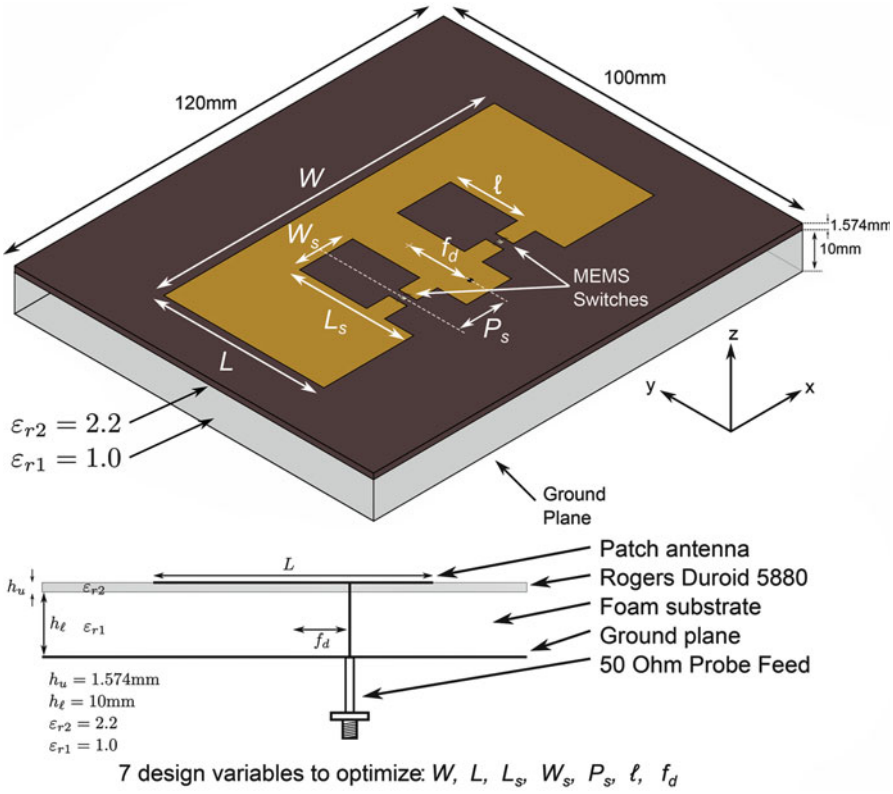


Fig. 8 Reconfigurable E-shaped patch (Rajagopalan et al. 2014)

Another antenna system proposed for cognitive radio applications is based on the integration of band-pass and band-stop filters in either the ground plane of the antenna or the feeding line. The antenna proposed is shown in Fig. 10 (Mansoul et al. 2014). It is composed of a radiating element in the form of an inverted U, fed by a microstrip line on its upper side. In order to achieve the frequency reconfiguration, four horizontal slots with integrated switches are disposed in three rows and incorporated onto the antenna ground plane. The upper and bottom slots are used as stop-band filters in order to suppress the whole

DOI 10.1007/978-981-4560-75-7_61-2

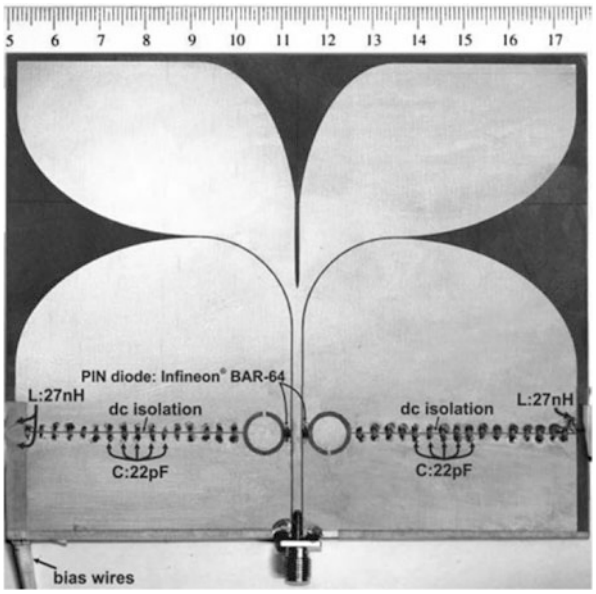


Fig. 9 Cognitive radio reconfigurable antenna proposed in Augustin et al. (2014)

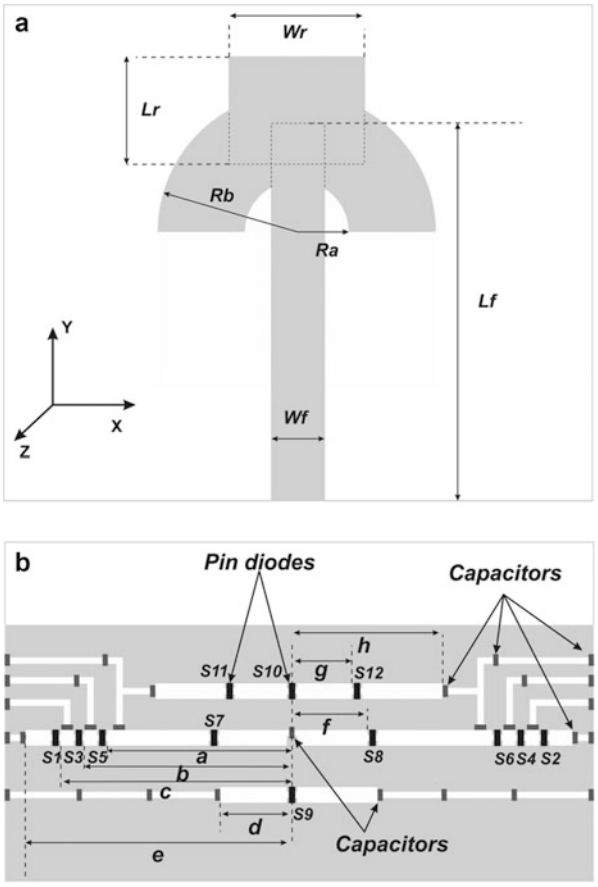


Fig. 10 (**a**) *Front* view, (**b**) *back* view of the proposed antenna (Mansoul et al. 2014)

considered band. The upper slot attenuates the lower frequencies in the band, while the bottom one suppresses the high frequencies (Mansoul et al. 2014). Once the considered band is suppressed, the

DOI 10.1007/978-981-4560-75-7_61-2

introduction of the middle slots in the antenna's ground plane creates a high operating band whose frequency depends on its length. Thus, by controlling the slot length by means of switches, the antenna's operating band can be tuned accordingly (Mansoul et al. 2014).

Reconfigurable Antennas for MIMO-Based Cognitive Radio Applications

Various antenna systems can be proposed for cognitive radio applications (Tawk et al. 2014b). However, it is critical to understand that adding cognition to antenna systems and allowing them to communicate over unused frequency gaps do not solve all other issues in the spectrum. Hence, it is important to design antennas that address spectrum concerns such as fading or multipath in a cognitive radio setup. A MIMO-based antenna can be proposed for cognitive radio applications as shown in Fig. 11 (Tawk et al. 2014b). The reason behind adopting a MIMO-based antenna for cognitive radio is to combat fading, ensure reliable communication between the end users, and improve spectrum usage efficiency.

The MIMO-based cognitive radio antenna system shown in Fig. 11 consists of two sensing antennas as well as two reconfigurable "filtennas" packed all together on the same substrate (Tawk et al. 2014b). The reconfigurable antennas resort to PIN diodes to tune their frequency band-pass operation. A reconfigurable antenna, in a cognitive radio setting, is able to react to the observations of the wideband sensing antenna, based on the recommendations of the cognitive processor as described in Fig. 7. The evaluation of the performance of the proposed antenna systems in mitigating fading in a rich multipath environment is done by observing the values of the envelope correlation coefficient ρ_e and the mean effective gain (MEG) (Vaughan and Andersen 1987; Tawk et al. 2014b).

Another MIMO-based reconfigurable cognitive radio antenna is shown in Fig. 12 (Hussain and Sharawi 2015). The antenna system design is based on a compact meandered line planar reconfigurable two-element MIMO system. The antenna system is made tunable in different bands using PIN and varactor diodes. The antenna operates at lower frequency bands starting from 580 to 680 MHz and from 834 to 1,120 MHz. The important feature of the proposed design is that the reference ground plane of the reconfigurable antenna is optimized to work as a sensing antenna to scan the frequency spectrum while operating as a ground plane for the reconfigurable antenna during the communication stage (Hussain and Sharawi 2015).

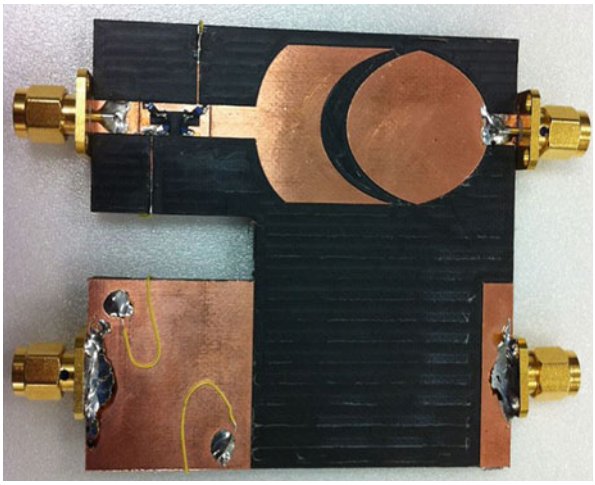


Fig. 11 MIMO-based reconfigurable filtenna (Tawk et al. 2014b)

DOI 10.1007/978-981-4560-75-7_61-2

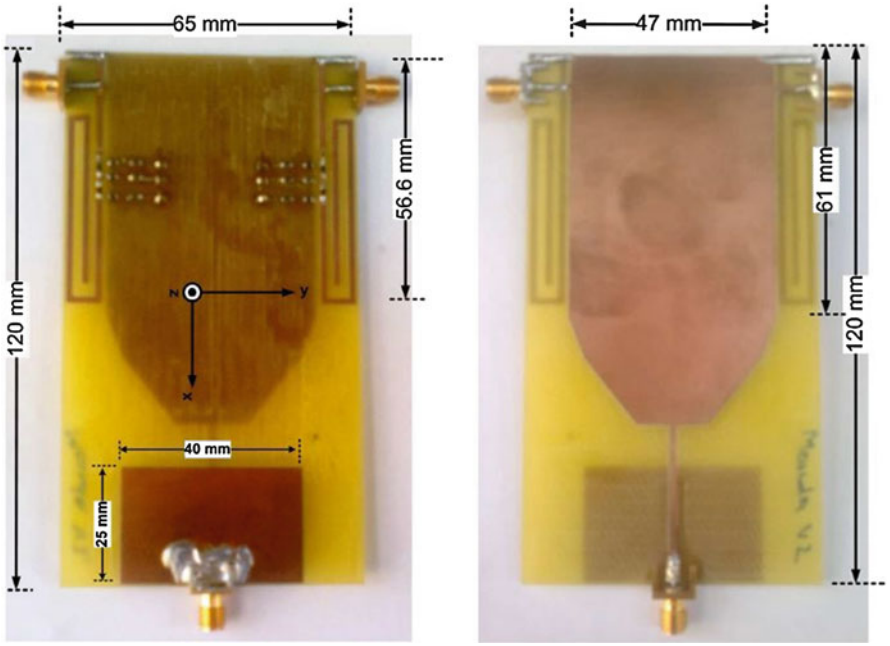


Fig. 12 *Front* and *back* view of the proposed MIMO-based reconfigurable cognitive radio antenna system (Hussain and Sharawi 2015)

Reconfigurable Antennas with Wireless Local Area Network (WLAN) Band Rejection

A very common problem is the interference that occurs between existing ultra-wideband (UWB) antennas and signals from WLAN communication networks. Since UWB antennas possess a large bandwidth, their in-band interference is usually a result of increased traffic over 802.11 n usage or from WLANs that allocate different channels in their upper operational band (5.15–5.825 GHz) (Federal Communications Commission 2002; Anagnostou et al. 2014). The problem of interference between WLAN and UWB nearby signals can be solved by proposing UWB antennas with permanent WLAN band rejection. The problem arises when rejection is not permanently needed and, in this case, dynamic or reconfigurable rejection becomes a necessity (Anagnostou et al. 2014).

An example of a reconfigurable antenna that achieves WLAN rejection is shown in Fig. 13 (Anagnostou et al. 2014). This UWB antenna uses microelectromechanical switches (MEMS) to reconfigure its band rejection on demand in the 5.15–5.825 GHz 802.11 a/n WLAN bands. The antenna design is a planar disc-shaped UWB monopole. The antenna design lacks a backside ground plane which makes it ideal for monolithically integrated MEMS. A quartz substrate is used for the antenna and the high-resistive bias lines. This led to the monolithic on-wafer integration of the whole system (Anagnostou et al. 2014).

Reconfigurable Antennas for Wireless Sensing

Wireless sensing can be applied in applications such as industrial process control, battlefield surveillance, habitat monitoring, and implantable medical telemetries (Jiang and Yang 2013), to mention a few. An example of a reconfigurable antenna that has been proposed for wireless sensing is shown in Fig. 14 (Jiang and Yang 2013). The antenna is a conventional printed dipole with a T-matching network. The antenna's impedance is conjugate matched with an RFID integrated chip (IC) as shown in Fig. 14. The antenna system is printed on an FR4 substrate. The realized gain of the tag antenna and the sensitivity of the IC determine the read range of the tag antenna. The RFID microchip is a nonlinear load whose complex

DOI 10.1007/978-981-4560-75-7_61-2

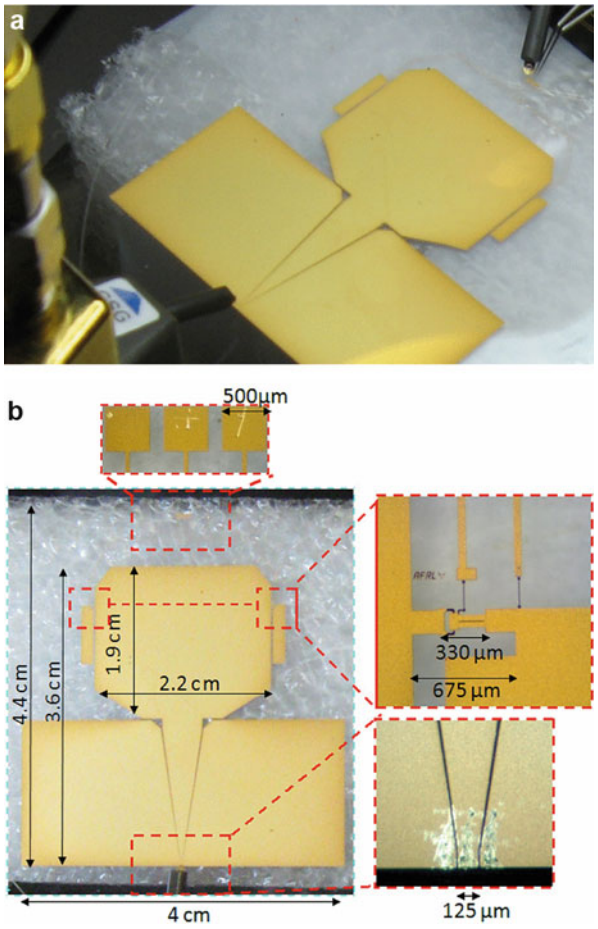


Fig. 13 Antenna with WLAN band rejection discussed in Anagnostou et al. (2014)

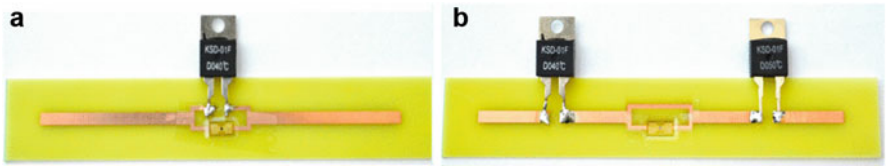


Fig. 14 The proposed reconfigurable sensing antenna with a thermal switch at the center of the dipole (Jiang and Yang 2013)

impedance varies with the frequency and the input power (Jiang and Yang 2013). The reconfigurable antenna can monitor the temperature of the surrounding environment by integrating thermal switches in series or parallel. When the temperature exceeds the specific threshold, the switch will change its status, which leads to a significant change in the antenna performance (Jiang and Yang 2013).

Reconfigurable Antennas for Terahertz (THz) Communication Applications

Communication at THz frequencies has attracted the attention of many antenna researchers especially after the rise of graphene as an antenna design material. The ability of graphene to support slow surface plasmonic modes at THz frequencies, allows the miniaturization of various communication devices for communication and sensing purposes (Tamagnone et al. 2013). The behavior of the surface conductivity of graphene in the THz region enables the creation of plasmonic modes in radiating structures

DOI 10.1007/978-981-4560-75-7_61-2

(Tamagnone et al. 2013). The design of antennas using graphene provides a good total efficiency, good matching, high miniaturization, and inherent reconfiguration capabilities. The antenna structures take advantage of resonant plasmonic modes propagating on graphene sheets to implement actual radiators, able to couple electromagnetic energy from small THz sources to free space (Tamagnone et al. 2013). An example of the use of graphene for reconfigurable antenna design over THz frequencies is shown in Fig. 15 where two self-biasing stacked graphene layers (patches) are designed to achieve dynamical tuning (Tamagnone et al. 2013). The two graphene layers considered are separated by a thin film of Al_2O_3. A DC voltage is applied between the two graphene sheets into the thin insulating layer. With this geometry, the magnetic permeability can be tuned dynamically by varying the voltage between the graphene layers.

A graphene-based leaky-wave antenna has also been proposed based on the graphene modes of operation for THz applications (Perruisseau-Carrier et al. 2013). The antenna can be composed of a graphene sheet deposited on a dielectric substrate with several independent polysilicon DC grating pads beneath it (Perruisseau-Carrier et al. 2013). The desired graphene surface reactance profile is achieved by applying different bias voltages to the different pads to create the periodic reactance modulation. By modifying the modulation period, the pointing angle of the radiated beam can be controlled (Perruisseau-Carrier et al. 2013). The aforementioned work is also extended to the design of a reconfigurable beam reflect-array antenna (Carrasco et al. 2013). The complex conductivity of graphene can be varied when an electric field is applied allowing the control of the phase of the reflected wave at each element of the reflect-array. The proposed patches support plasmonic slow-wave propagation at THz frequencies, drastically reducing the physical dimensions of the cell (Carrasco et al. 2013).

Reconfigurable Antennas for Millimeter-Wave Communication Applications

Communication between 30 and 300 GHz is defined as communication over millimeter (mm)-wave frequencies. The mmWave spectrum has become of high interest for many system designers for its many benefits that are mainly focused on its wide bandwidths (Federal Communications Commission 1997). Such wide bandwidths are useful and practical when it comes to high speed data transmission and video distribution. The propagation characteristics of mmWaves are very particular in the sense that mmWave signals can only travel a few miles or less. These waves do not penetrate solid materials very well and are heavily affected by environmental factors such as raindrops and fog (Federal Communications Commission 1997).

Many reconfigurable antennas have been developed for operation over mmWave frequencies. An example of such antennas is the one shown in Fig. 16 (Lafond et al. 2013). This antenna is composed of

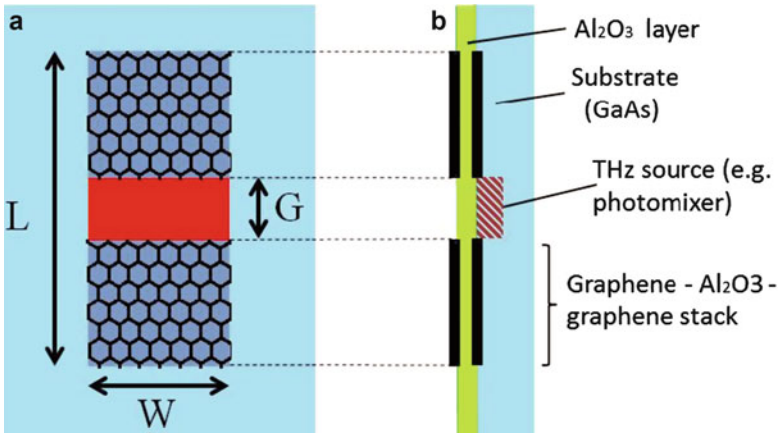


Fig. 15 The proposed double-layered graphene antenna. (**a**) *Top* view. (**b**) Cross section (Tamagnone et al. 2013)

DOI 10.1007/978-981-4560-75-7_61-2

Fig. 16 Antenna prototype proposed in Lafond et al. (2013)

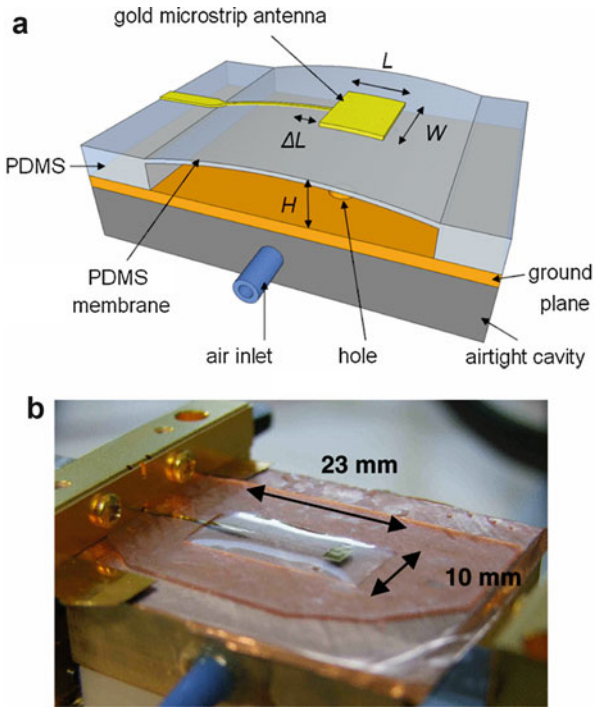


Fig. 17 PDMS antenna proposed in Hage-Ali et al. (2010)

16 different patches that are fed by a corporate feeding network. The different patches are placed in a semicircle manner with 10° apart from each other. The patches are fed by ridged source waveguides and are matched by using monolithic microwave integrated chip (MMIC) power amplifiers. The whole system is then connected to a lens that is illuminated by the various patch sources. The lens is a plate Luneburg lens that is defined as an inhomogeneous lens with infinity focus points. Once a patch illuminates the lens, the other patches are loaded with 50 Ω (Lafond et al. 2013).

Reconfigurable antennas for mmWaves' applications can also be designed using liquid crystals (Hu et al. 2008) or mechanical reconfiguration techniques similar to the one shown in Fig. 17 (Hage-Ali et al. 2010). The antenna presented (Hage-Ali et al. 2010) is a millimeter-wave frequency-agile

DOI 10.1007/978-981-4560-75-7_61-2

microstrip antenna that is printed on an ultrasoft elastomeric polydimethylsiloxane (PDMS) substrate. The microstrip patch antenna is supported by a PDMS membrane suspended over an air cavity. The distance between the patch and the ground plane is tuned using pneumatic actuation, taking advantage of the extreme softness of the PDMS membrane. The tuning of the antenna height results in tuning of the antenna's operating frequency. A continuous frequency shift varying from 55.35 to 51 GHz has been obtained for a tuning range of height between 200 and 575 μm. In all configurations, the antenna remains matched and its radiation characteristics are stable (Hage-Ali et al. 2010).

Reconfigurable Antennas for Space Communication Applications

Space communication is based on deployable antennas that are packed during launch and then deployed once in space. Many types of deployable antennas have been used on orbit. Reflector types constitute the widest category (Tadashi et al. 2004). Other deployable structures made with folded hoops or ribs are also used for space communications (Zheng et al. 2008). Some researchers have also resorted to tape springs and neutrally stable material to design their structures (Murphey and Pellegrino 2004). Helical antennas (Balanis 2012) are also a popular choice of satellite antennas due to their natural circular polarization and wide bandwidth (King and Wong 1980).

Traditional deployable antennas present valid options for space communication. However, the rise of small satellites' development and the launch of clusters of CubeSats have increased the need for more agile antennas. Currently dipole antenna arrays in various configurations constitute the most common antennas deployed on CubeSats (Muri et al. 2010). The merging of reconfigurable and deployable antennas appears to be a solution to the increased demand for communication agility in space.

Mechanical antenna reconfiguration techniques merged with deployment techniques seem to present an answer to reconfigurability in space. Other electrical reconfiguration techniques can also be explored especially with the advancement in satellite power handling technology. Mechanical reconfiguration mechanisms for deployable antennas are based on the nature of the antennas, their composite material, and their deployment mechanism. For example, the reconfiguration of deployable antennas that are designed with bistable composite tape springs can be achieved using one-way linear or rotary actuator (Murphey et al. 2010; Jeon and Murphey 2011; Costantine et al. 2012b). Antennas made with neutrally elastic mechanism (NEM) tape springs require two-way actuators to be reconfigured (Murphey et al. 2010; Jeon and Murphey 2011; Costantine et al. 2012b).

Another reconfiguration technique can be based on deploying in progression. For example, a helical antenna shown in Fig. 18 (Costantine et al. 2013a) deploys in stages by using actuators. The sequential deployment tunes the antenna's operating frequency. The concept is based on the fact that only the part of the antenna that is above the deployed ground plane radiates (Costantine et al. 2013a). In general,

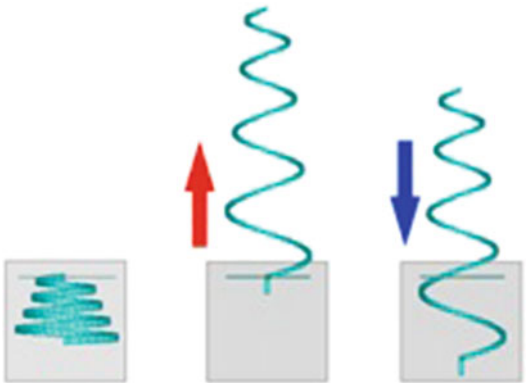


Fig. 18 Helical antenna deployed in progression (Costantine et al. 2013a)

DOI 10.1007/978-981-4560-75-7_61-2

reconfiguring antennas for space communications is a difficult task with many constraints, such as maintaining antenna performance without draining existing power and reducing reliability. However, space communications agility is a great addition that constitutes a major advancement in available communication technology.

Reconfigurable Antennas for Multiple Wireless Communication Applications

Several reconfigurable antenna concepts have been proposed in literature for multiple wireless communication applications. An example of such antennas is the one shown in Fig. 19 (Hinsz and Braaten 2014). The reconfigurable antenna shown is an autonomously controlled frequency reconfigurable antenna. The presented antenna does not require external and separate control signals to achieve reconfiguration. The antenna design concept consists of a power divider, band-pass filter, voltage doubling circuitry, and reconfigurable antenna controlled with voltage. The system is based on the concept that once RF power is provided to the summing junction of the power divider, part of it is sent to the reconfigurable antenna and the band-pass filter. If the signal is in the passband of the filter, RF power is provided to the voltage doubling circuit and converted to a positive DC voltage proportional to the input RF power (Hinsz and Braaten 2014). As a result, the antenna can then be reconfigured based on this output voltage. If the RF signal driving the antenna is not in the passband of the filter, then it gets rejected and the antenna will not be reconfigured.

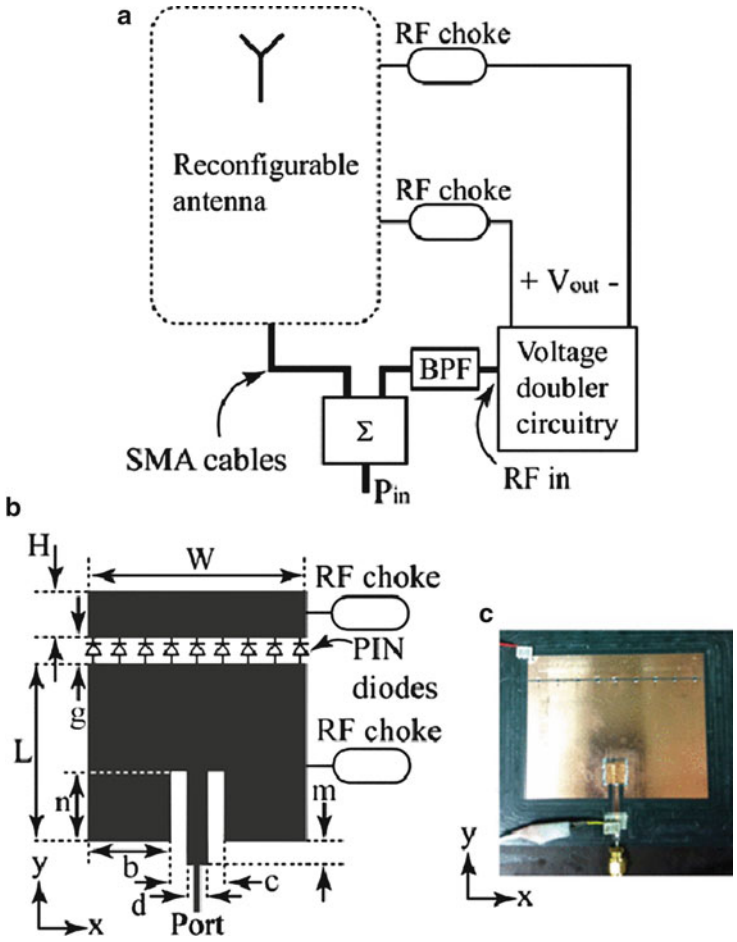


Fig. 19 Layout of the autonomously reconfigured antenna (Mazlouman et al. 2011)

DOI 10.1007/978-981-4560-75-7_61-2

Mechanically reconfigurable antennas are also proposed for various applications that can tolerate these systems. Antennas with electroactive polymers (Mazlouman et al. 2011), shape-memory alloys (Mazlouman et al. 2011), or actuators (Costantine et al. 2014) are mainly proposed as techniques that are alternative to switching reconfiguration methods for multi-application reconfigurable antennas.

Modeling, Control, and Optimization of Reconfigurable Antennas

Reconfigurable antennas are required to transition from one state into another to cater for varying communication channels requirements. To that aspect, antenna designers have resorted to various optimization algorithms to smooth the state transition in a reconfigurable antenna's operation. Genetic algorithms, simulated annealing, ant-colony optimization, self-organized maps, particle swarm optimization, the cross-entropy method, and the self-adaptive-induced mutation algorithm are proposed to smooth transitions between various states of a reconfigurable antenna (Coleman et al. 2004; Min et al. 2004; Bossard et al. 2005; Akdagli et al. 2007; Jin and Rahmat-Samii 2007; Connor 2008; Skinner et al. 2009; Song and Murch 2014; Cao et al. 2014). These algorithms also minimize the negative effect of any reconfiguration technique on the antenna performance. A comparison between these optimization techniques shows that a particular optimization algorithm cannot be separated from the rest as the best fit, before selecting a specific reconfigurable scheme (Langoni et al. 2009).

Additionally, the complexity of some large reconfigurable antenna designs is another aspect that requires consideration. An increased complexity results in unwanted costs and losses (Costantine et al. 2013c). Several approaches have been proposed to reduce the complexity without affecting the reliability of any reconfigurable antenna system (Mehmood and Wallace 2010; Obeidat et al. 2010; Costantine et al. 2011a, 2012a, c; Mehmood et al. 2014). Most of these approaches are based on applying on antenna structures various techniques or models that are already known in other fields of study (Costantine et al. 2013c). Learning state selection approaches can be used to learn the antenna behavior for various reconfiguration states and then regenerate them on demand (Gulati and Dandekar 2013). Neural networks constitute a great example of such learning approaches. Neural networks can be used as an antenna behavioral predictor that allows the generation of various antenna states based on previous learning experience (Patnaik et al. 2005). The merging of neural networks and graph models has resulted in computational reduction and optimal antenna regeneration (Costantine et al. 2011b).

Graph Modeling Reconfigurable Antennas

Graphs are used as models that help automate and optimize the antenna performance. The user controlling the reconfiguration process of an antenna resorts to software to control the tuning of the antenna (Costantine et al. 2013c). In fact, the rise of reconfigurable antennas is based on the fact that these antennas are software controlled and can be automated easily. Graph models (Cormen et al. 2001) have been used to improve the software control process of reconfigurable antennas and to allow the employment of various algorithms into the antennas' automation and optimization process (Costantine et al. 2013c).

Graphs are mathematical illustrations of various systems. They are symbolic representations of relationships between different points in a system. A graph can also be a description of a communication protocol. In particular, a suitably designed graph can precisely describe and direct the changing network topology of a self-organizing system (Klavins et al. 2004; Klavins 2006; Napp et al. 2006; Klavins 2007) such as a reconfigurable antenna.

A graph is a collection of vertices that are connected by lines called edges (Costantine et al. 2011a). Each vertex can represent a part of an antenna or an endpoint of a switch (Costantine et al. 2013c). The

DOI 10.1007/978-981-4560-75-7_61-2

connection between the different antenna parts or the activation of the switch is represented by the presence of an edge connecting the corresponding vertices (Costantine et al. 2013c). Several rules are introduced to model reconfigurable antennas (Costantine et al. 2013c). These rules allow the designer to relate each possible topology to a corresponding electromagnetic performance. Once a graph is drawn, an adjacency matrix can be formed based on this graph. Thus, the graph has transformed an antenna into a matrix where various algorithms can be applied.

Graph models can also be used to investigate the presence of redundant components in an antenna structure. An element is defined as redundant if its presence gives the antenna more functions than required and its removal does not affect the antenna's performance (Costantine et al. 2011a). A redundancy reduction approach is proposed in Costantine et al. (2011a) to eliminate redundant components from reconfigurable antenna structures for more efficient designs. Removing redundant components from an antenna structure reduces its complexity and allows for a more efficient antenna reconfiguration. For example, reducing the number of switches used to reconfigure an antenna reduces the nonlinearity effects and the interference from the corresponding switching components' biasing lines. The planar antenna shown in Fig. 20 (Costantine et al. 2011a) is built out of a hexagonal main patch and six trapezoidal parts placed around it. Each trapezoidal part is connected to the main patch by a switch. Thus, this antenna originally utilizes six switches to tune its frequency operation. The graph model of this antenna shown in Fig. 20a is composed of six vertices (P_1, P_2, P_3, P_4 P_5, P_6) representing the trapezoidal

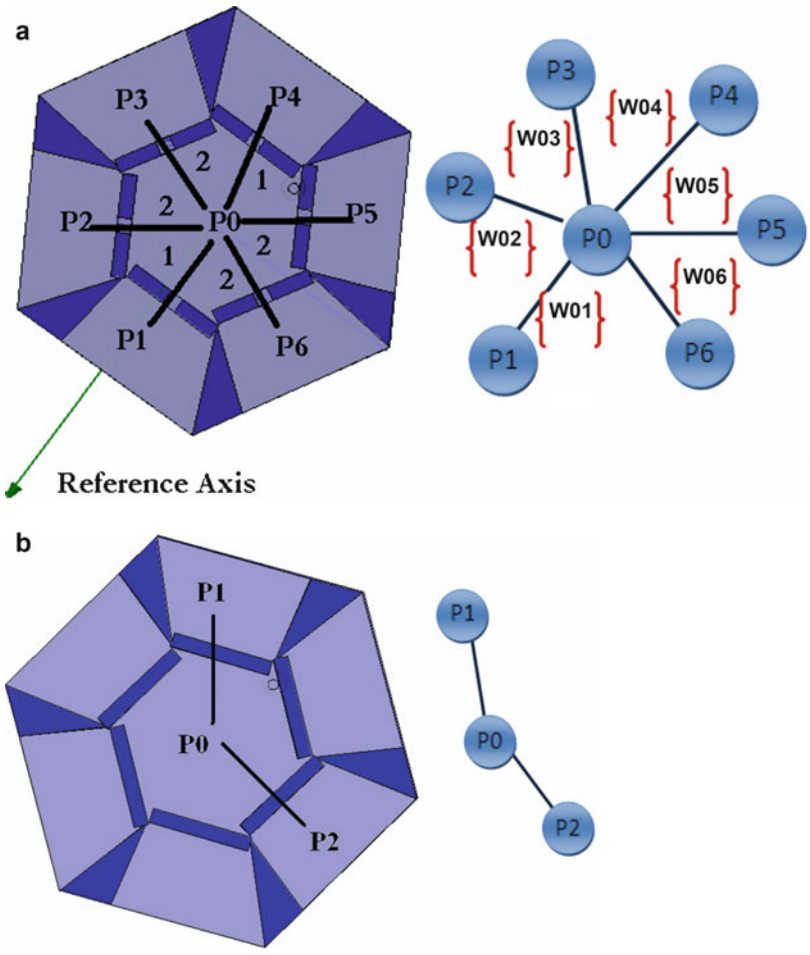


Fig. 20 (**a**) Reconfigurable antenna before the redundancy reduction approach, (**b**) reconfigurable antenna after the redundancy reduction approach (Costantine et al. 2011a)

DOI 10.1007/978-981-4560-75-7_61-2

parts connected by six edges to the seventh vertex (P_0) that represents the main patch (Costantine et al. 2011a). After the application of the graph-based redundancy reduction approach, the number of used switches is transformed from six to two as shown in Fig. 20b. It is important to note that the functioning of the antenna is preserved and its optimized response is compared with the original redundant antenna response (Costantine et al. 2011a). The redundancy reduction approach is also tested on antennas with a larger number of switches, and as a result, more efficient designs are created for an easier system control and integration (Costantine et al. 2011a).

Reconfigurable antennas can achieve the operation at one frequency with multiple switch configurations. These configurations are defined as "equivalent configurations" (Costantine et al. 2012c). Reducing redundant components and eliminating unwanted switches reduce the antenna's general complexity and also reduce the number of equivalent configurations. On the other hand, it is important to address the effect of switch failure on the antenna's reliability, after eliminating the redundant switching components from the antenna structure. It is shown that while the physical antenna redundancy is reduced, the equivalent configurations present are sufficient to maintain acceptable antenna reliability (Costantine et al. 2012c).

The reliability of a reconfigurable antenna mainly depends on the number of antenna configurations at a certain frequency and the probability to achieve them. However, it is inversely proportional to the number of edges needed. The solution is to design reconfigurable antennas with several equivalent configurations but only with a small number of connections. Thus, the reliability of a reconfigurable antenna can be calculated as in Eq.1 (Costantine et al. 2012c). On the other hand, an antenna's general complexity is represented by the total number of edges in a graph as shown in Eq. 2 (Costantine et al. 2012c). Antenna's frequency-dependent complexity is the measure of the maximum number of edges in a particular configuration as defined in Eq. 3 (Costantine et al. 2012c). It can be deduced from these equations that the reliability of a reconfigurable antenna is inversely proportional to its frequency-dependent complexity (Costantine et al. 2012c). This conclusion affects the design process of an antenna researcher that looks to maximize the reliability of its design while reducing its complexity.

$$R(f) = \frac{\sum_{i=1}^{Nc(f)} \sum_{j=1}^{NE_i(f)} P(E_{ij})}{\sum_{i=1}^{Nc(f)} NE_i(f)} \times 100 \tag{1}$$

where:

$R(f)$ = the reconfigurable antenna reliability at a particular frequency f.
$NC(f)$ = the number of configurations achieving the frequency f.
$NE(f)$ = the number of edges for different configurations at the frequency f.
$P(E)$ = probability of achieving the edge $E = 1 - P$ (a switch failing).

$$C = NE \tag{2}$$

where NE represents the total number of edges in a graph for all possible connections.

$$C(f) = \underset{i=1, NC(f)}{\text{Max}} (NE_i(f)) \tag{3}$$

DOI 10.1007/978-981-4560-75-7_61-2

where:

$C(f)$ represents the complexity of the antenna system at a frequency f.
$NC(f)$ represents the number of equivalent configurations at a frequency f.
$NE_i(f)$ represents the number of edges at the configuration i for a frequency f.

The graph-based analysis can also be extended to reconfigurable antenna arrays where their complexity, reliability, and switch failure correction techniques are correlated and analyzed (Costantine et al. 2012a).

The importance of the graph-based analysis is that it gives the antenna designer a strong tool to use at the initial design process of reconfigurable antennas. It is executed so that optimal reconfigurable antenna designs are created with less redundancy. Reducing redundant components while maintaining reliability creates systems that have less loss and fewer costs and are efficient and easier to control and automate.

Software Control of Reconfigurable Antennas

Many tools can be used to automate the software control of reconfigurable antennas. The use of field-programmable gate arrays (FPGA), a simple microcontroller, or an Arduino board (Arduino 2013) constitutes valid approaches that have been used in the past. For example, the antenna discussed in Shelley et al. (2010) uses an FPGA to activate and deactivate PIN diodes that connect different parts of its structure as shown in Fig. 21 (Costantine et al. 2013c). The use of an FPGA in such a system allows the antenna user to control the operation of this antenna through simple commands. In another example, LabVIEW and an NPN Darlington array control the rotation mechanism of a stepper motor that is used to reconfigure an antenna structure (Tawk et al. 2011). The simplicity of such an approach makes the software-controlled reconfiguration process a task that can be executed by any antenna designer without very complicated programming skills.

Arduino boards also allow antenna designers to program their antenna systems, control and automate them in a compact setting as discussed in Costantine et al. (2014). The user in such systems controls the operation of the antenna and decides when and how to change its function. In other cases, detection of surrounding activities is needed to change an antenna's operation. An example of such system is one where a change in temperature triggers thermal switches installed onto the antenna structure (Jiang and

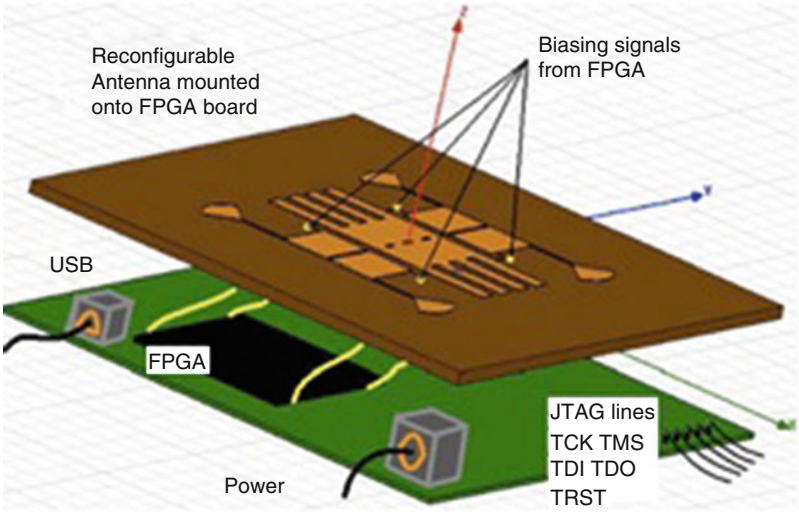


Fig. 21 FPGA controlling a PIN diode frequency reconfigurable antenna (Costantine et al. 2013c)

DOI 10.1007/978-981-4560-75-7_61-2

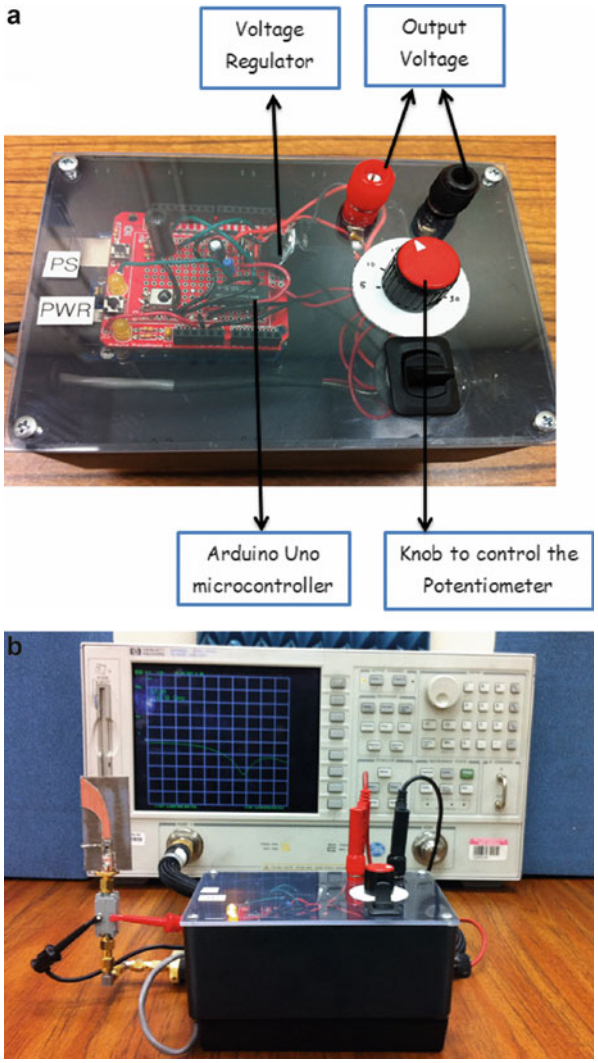


Fig. 22 The motion detection circuit connected to the varactor-based reconfigurable "filtenna" (Costantine et al. 2013b)

Yang 2013). Motion detection is also a trigger mechanism that can be used to bias and tune a reconfiguring component. An infrared (IR) motion detection sensor is connected to the biasing network of a varactor, incorporated onto a reconfigurable antenna as shown in Fig. 22 (Costantine et al. 2013b). Once a movement is detected in a predetermined area, the biasing circuit changes voltage levels. This action results in antenna frequency tuning (Costantine et al. 2013b).

Graph models also help in the software control process by minimizing redundant components and allowing the application of graph algorithms to optimize performance. For example, a shortest path algorithm (Costantine et al. 2013c) applied to a mechanically reconfigurable antenna expedites its performance by always searching for the shortest path. This results in the fastest reconfiguration process possible with the reconfiguration technique (Costantine et al. 2013c).

Neural Networks Applied to Reconfigurable Antennas

The use of software to control reconfigurable antennas has encouraged antenna researchers to employ neural networks (NNs) to control their antennas through FPGAs. NNs are used on reconfigurable antennas as antenna synthesizers. A NN is trained to associate all the different configurations of a reconfigurable antenna with its different operating frequencies (Costantine et al. 2011b). NN can be

DOI 10.1007/978-981-4560-75-7_61-2

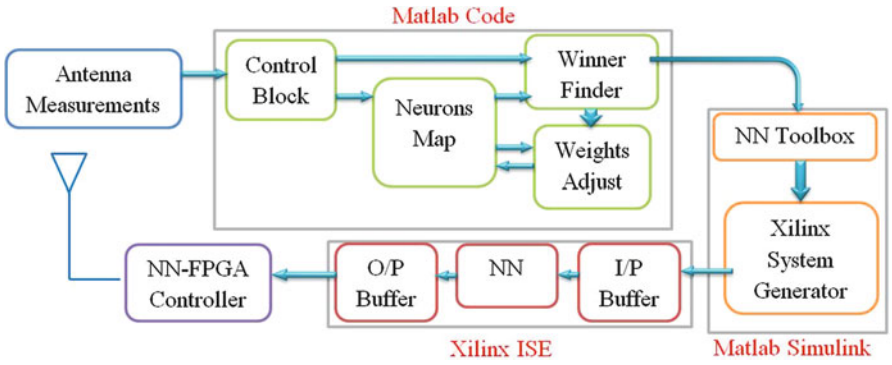


Fig. 23 Block diagram of the NN operation (Costantine et al. 2011b)

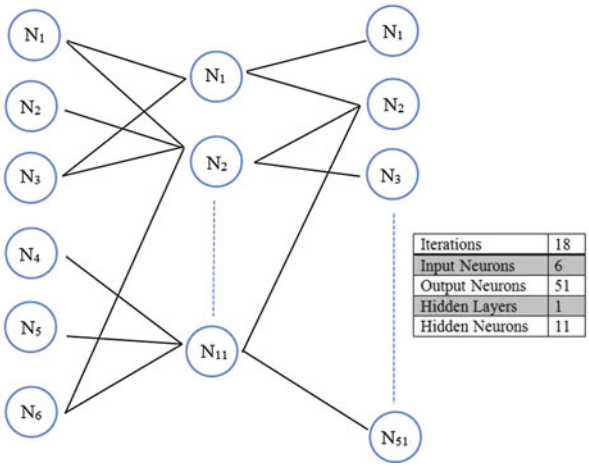


Fig. 24 The NN structure of the antenna in Fig. 21a (Costantine et al. 2011b)

designed using a standard back-propagation technique, which provides meaningful cues that steer the design to convergence (Lyke 2002). Applying NNs to reconfigurable antennas results in the association of different antenna configurations with different frequency responses called clusters. The association of such frequency clusters with corresponding antenna configurations trains the NN to be able to configure the antenna and regenerate frequency responses on demand.

The NN architecture has three layers: input layer, hidden layer, and output layer. The input layer represents, in the case of a switch-reconfigured antenna, the number of switches existing in that antenna. The number of hidden neurons in the hidden layer represents the accuracy needed to regenerate the frequency response of the antenna in question. The number of output layer neurons is determined depending on the number of points required to recreate, up to an acceptable accuracy the return loss of the antenna (Zuraiqi 2012).

The application of neural networks requires a lot of data training and software handling. The general methodology is summarized in the block diagram of Fig. 23 (Costantine et al. 2011b). A neural network is built and trained in MATLAB Simulink. Training the NN requires adjusting the connection weights and specifying the number of hidden neurons. Error back-propagation (BP) is used to train the NN in this case. BP compares neural networks' output to the measured antenna response and calculates an error adjustment for each of the nodes in the network. The neural network adjusts the connection weights according to the error values assigned to each node (Zuraiqi 2012). After the weight adjustment process, a Xilinx system generator creates the NN VHDL code to be used to control an FPGA (Zuraiqi 2012).

DOI 10.1007/978-981-4560-75-7_61-2

As an example, applying neural networks to a reconfigurable antenna with six switches such as the one shown in Fig. 20a results in six input neurons, 11 hidden neurons, and 51 output neurons. The NN structure for this antenna is shown in Fig. 24. After removing the redundant components from this antenna as shown in Fig. 20b, the input layer of the neural network shrinks from six to two and the number of hidden neurons will have to shrink from 11 to 8 (Costantine et al. 2011b). As a result the training process requires less time and the regeneration of antenna states based on NN is faster and more efficient.

Summary

In this chapter, the questions that any antenna designer asks in regards to reconfigurable antenna design are addressed. The chapter addresses the need to design such antenna structures and presents the reader with a plethora of applications in which reconfigurable antennas are proposed as potential solutions. Reconfigurable antennas constitute a major part of future communication systems. They are the basis for many novel applications that are based on cognition and reaction to evolving/changing conditions in any RF or communication environment. A static communication system that relies on one or more fixed antennas no longer caters for the evolving needs of today's communication technology.

A reconfigurable antenna that is used for any wireless communication application is not complete without the possibility of its software control and automation. The learning of communication activities and the application of previously learned experiences into future action allow the reconfigurable antenna to leap from a simple RF front-end in a communication device into a real cognitive structure.

Future terrestrial and space communications will be based on antennas that adapt and learn. The adoption of reconfigurable antennas into commercial devices will increase with the progression of available technology. Therefore, it is crucial to understand these structures and optimize their behavior for a more efficient futuristic wireless communication protocols.

Cross-References

- Antenna Design for Diversity and MIMO Application
- Antennas in Hand-Held Devices
- Commercial Antenna Design Tools
- Frequency Selective Surfaces
- Microstrip Patch Antennas
- Millimeter-Wave Antennas and Arrays
- MIMO Systems and Antennas for Terminals (Including Portable Devices Such as Handsets, iPad Laptops)
- Multi-Beam Antenna Arrays
- Reconfigurable Antenna Arrays for Wireless Communications
- Space Antennas including Terahertz Antennas
- Ultra Wideband Antennas

DOI 10.1007/978-981-4560-75-7_61-2

References

Akdagli A, Guney K, Babayigit B (2007) Clonal selection algorithm for design of reconfigurable antenna array with discrete phase shifters. J Electromagn Waves Appl 21(2):215–227

Anagnostou DE, Chryssomallis MT, Braaten B, Ebel JL, Sepulveda N (2014) Reconfigurable UWB antenna with RF-MEMS for on-demand WLAN rejection. IEEE Trans Antennas Propag 62(2):602–608

Arduino (2013) Arduino Uno. Internet: http://www.arduino.cc/en/Main/arduinoBoardUno. Accessed 20 May 2013

Augustin G, Chacko BP, Denidni TA (2014) Electronically reconfigurable uni-planar antenna for cognitive radio applications. IET Microw Antennas Propag 8(5):367–376

Bai Y, Xiao S, Liu C, Shuai X, Wang B (2013) Design of pattern reconfigurable antennas based on a two-element dipole array model. IEEE Trans Antennas Propag 61(9):4867–4871

Balanis CA (2011) Modern antenna handbook. Wiley-Interscience, Hoboken, NJ, USA

Balanis CA (2012) Antenna theory: analysis and design. Wiley-Interscience, Hoboken, NJ, USA

Ban YL, Chen ZX, Chen Z, Kang K, Li JLW (2014) Reconfigurable narrow-frame antenna for heptaband WWAN/LTE smartphone applications. IEEE Antennas Wirel Propag Lett 13:1365–1368

Bernhard JT (2007) Reconfigurable antennas. Morgan and Claypool publishers, San Rafael, CA, USA

Bossard JA, Werner DH, Mayer TS, Drupp RP (2005) A novel design methodology for reconfigurable frequency selective surfaces using genetic algorithms. IEEE Trans Antennas Propag 53(4):1390–1400

Brown ER (1998) RF –MEMS switches for reconfigurable integrated circuits. IEEE Trans Microwave Theory Tech 46(11–2):1868–1880

Cao K, Jiang H, Chen G, Cui P, Xiong T (2014) Self-adaptive induced mutation algorithm for reconfigurable antenna systems. IEEE Antennas Wirel Propag Lett 13:237–240

Carrasco E, Tamagnone M, Perruisseau-Carrier J (2013) Tunable graphene-based reflectarray element for reconfigurable beams. In: 7th European conference on antennas and propagation, Gothenburg, Sweden, pp 1779–1782

Cetiner BA, Jafarkhani H, Qian JY, Yoo HJ, Grau A, De Flaviis F (2004) Multifunctional reconfigurable MEMS integrated antennas for adaptive MIMO systems. IEEE Commun Mag 42(12):62–70

Cetiner BA, Crusats GR, Jofre L, Biyikli N (2010) RF MEMS integrated frequency reconfigurable annular slot antenna. IEEE Trans Antennas Propag 58(3):626–632

Chacko BP, Augustin G, Denidni TA (2015) Electronically reconfigurable uni-planar antenna with polarization diversity for cognitive radio applications. IEEE Antennas Wirel Propag Lett 14:213–216

Chen RH, Row JS (2008) Single-fed microstrip patch antenna with switchable polarization. IEEE Trans Antennas Propag 56(4):922–926

Chen SH, Row JS, Wong KL (2007) Reconfigurable square-ring patch antenna with pattern diversity. IEEE Trans Antennas Propag 55(2):472–475

Christodoulou CG, Tawk Y, Lane SA, Erwin SR (2012) Reconfigurable antennas for wireless and space applications. Proc IEEE 100(7):2250–2261

Coleman CM, Rothwell JE, Ross JE (2004) Investigation of simulated annealing, ant-colony and genetic algorithms for self-structuring antennas. IEEE Trans Antennas Propag 52(4):1007–1014

Connor JD (2008) Antenna array synthesis using the cross entropy method. PhD dissertation, Florida State University, Tallahassee

Cormen TH, Leiserson CE, Rivest RL, Stein C (2001) Introductions to algorithms. MIT Press, Cambridge, MA, USA

Costantine J, al-Saffar S, Christodoulou CG, Abdallah CT (2011a) Reducing redundancies in reconfigurable antenna structures using graph models. IEEE Trans Antennas Propag 59(3):793–801

Costantine J, Tawk Y, Al-Zuraiqi E, Barbin SE, Christodoulou CG (2011b) Applying graph models and neural networks on reconfigurable antennas for cognitive radio applications. In: IEEE APS topical conference on antennas and propagation in wireless communications (APWC), pp 909–912, Torino, Italy

Costantine J, Tawk Y, Christodoulou CG (2012a) Complexity versus reliability in arrays of reconfigurable antennas. IEEE Trans Antennas Propag 60(11):5436–5441

Costantine J, Tawk Y, Christodoulou CG, Banik J, Lane S (2012b) CubeSat deployable antenna using bistable composite tape-springs. IEEE Antennas Wirel Propag Lett 11:285–288

Costantine J, Tawk Y, Christodoulou CG, Lyke JC, De Flaviis F, Grau Besoli A, Barbin SE (2012c) Analyzing the complexity and reliability of switch-frequency-reconfigurable antennas using graph models. IEEE Trans Antennas Propag 60(2–2):811–820

Costantine J, Kabalan KY, ElHajj A, Tawk Y, Christodoulou CG (2013a) A reconfigurable/deployable helical antenna for small satellites. In: IEEE international symposium on antennas and propagation, Orlando, FL, USA, pp 390–391

Costantine J, Tawk Y, Christodoulou CG (2013b) Motion-activated reconfigurable and cognitive radio antenna systems. IEEE Antennas Wirel Propag Lett 12:1114–1117

Costantine J, Tawk Y, Christodoulou CG (2013c) Design of reconfigurable antennas using graph models. Morgan and Claypool, San Rafael, CA, USA

Costantine J, Tawk Y, Woodland J, Floam N, Christodoulou CG (2014) Reconfigurable antenna system with a movable ground plane for cognitive radio. IET Microw Antennas Propag 8(11):858–863

Daviu EA, Fabres MC, Bataller MF, Jimenez AV (2007) Active UWB antenna with tunable band-notched behavior. IEEE Electron Lett 43(18):959–960

Erdil E, Topalli K, Unlu M, Civi OA, Akin T (2007) Frequency tunable patch antenna using RF MEMS technology. IEEE Trans Antennas Propag 55(4):1193–1196

Federal Communications Commission (1997) Millimeter wave propagation: spectrum management implications. FCC Bulletin 70

Federal Communications Commission (2002) First report and order revision of part 15 of the commission's rule regarding ultra wideband transmission systems. FCC 02–48

Grau A, Romeu J, Lee M, Blanch S, Jofre L, De Flaviis F (2010) A dual linearly polarized MEMS-reconfigurable antenna for narrowband MIMO communication systems. IEEE Trans Antennas Propag 58(1):4–16

Gulati N, Dandekar KR (2013) Learning state selection for reconfigurable antennas: a multi-armed bandit approach. IEEE Trans Antennas Propag 62(3):1027–1038

Hage-Ali S, Tiercelin N, Coquet P, Sauleau R, Preobrazhensky V, Premod P (2010) A millimeter-wave inflatable frequency-agile elastomeric antenna. IEEE Antennas Wirel Propag Lett 9:1131–1134

Hinsz L, Braaten BD (2014) A frequency reconfigurable transmitter antenna with autonomous switching capabilities. IEEE Trans Antennas Propag 62(7):3809–3813

Hu W, Cahill R, Encinar JA, Dickie R, Gamble H, Fusco V, Grant N (2008) Design and measurement of reconfigurable millimeter wave reflectarray cells with nematic liquid crystal. IEEE Trans Antennas Propag 56(10):3112–3117

Huff GH, Bernhard JT (2006) Integration of packaged RF-MEMS switches with radiation pattern reconfigurable square spiral microstrip antennas. IEEE Trans Antennas Propag 54(2):464–469

Hussain R, Sharawi MS (2015) A cognitive radio reconfigurable MIMO and sensing antenna system. IEEE Antennas Wirel Propag Lett 14:257–260

Jayaweera SK, Christodoulou CG (2011) Radiobots: architecture, algorithms and real-time reconfigurable antenna designs for autonomous, self-learning future cognitive radio. University of New Mexico. Technical report, EECE-TR-11-0001

Jeon S, Murphey TW (2011) Design and analysis of a meter-class CubeSat boom with a motor-less deployment by bi-stable tape springs. In: 52nd AIAA structures, structural dynamics, and materials conference, Denver, CO, USA

Jeong WS, Lee SY, Lim WG, Lim H, Yu JW (2008) Tunable band-notched ultra wideband (UWB) planar monopole antennas using varactor. In: 38th European microwave conference, Amsterdam, The Netherlands, pp 266–268

Jiang Z, Yang F (2013) Reconfigurable sensing antennas integrated with thermal switches for wireless temperature monitoring. IEEE Antennas Wirel Propag Lett 12:914–917

Jiang H, Patterson M, Zhang C, Subramanyan G (2009) Frequency tunable microstrip patch antenna using ferroelectric thin film varactor. In: IEEE national aerospace and electronics conference, Fairborn, OH, USA, pp 248–250

Jin N, Rahmat-Samii Y (2007) Advances in particle swarm optimization for antenna designs: real-number, binary, single-objective and multiobjective implementations. IEEE Trans Antennas Propag 55(3):556–567

Jin GP, Zhang DL, Li RL (2011) Optically controlled reconfigurable antenna for cognitive radio applications. IET Electron Lett 47(17):948–950

Kim B, Pan B, Nikolaou S, Kim YS, Papapolymerou J, Tentzeris MM (2008) A novel single-feed circular microstrip antenna with reconfigurable polarization capability. IEEE Trans Antennas Propag 56(3):630–638

King H, Wong J (1980) Characteristics of 1 to 8 wavelength uniform helical antennas. IEEE Trans Antennas Propag 28:291–296

Klavins E (2006) Self-assembly from the point of view of its pieces. In: American control conference, Minneapolis, MN, USA, p 7

Klavins E (2007) Programmable self assembly. IEEE Control Syst Mag 27(4):43–56

Klavins E, Ghrist R, Lipsky D (2004) Graph grammars for self assembling robotic systems. IEEE Int Conf Robot Autom 5:5293–5300

Lafond O, Himdi M, Merlet H, Lebars P (2013) An active reconfigurable antenna at 60 GHz based on plate inhomogeneous lens and feeders. IEEE Trans Antennas Propag 61(4):1672–1678

Langoni D, Weatherspoon MH, Ogunti E, Foo SY (2009) An overview of reconfigurable antennas: design, simulation, and optimization. In: IEEE 10th annual wireless and microwave technology conference (WAMICON)

Li Z, Du Z, Gong K (2009) Compact reconfigurable antenna array for adaptive MIMO systems. IEEE Antennas Wirel Propag Lett 8:1317–1321

Li Y, Zhang Z, Zheng J, Feng Z, Iskander MF (2012) A compact hepta-band loop inverted F reconfigurable antenna for mobile phone. IEEE Trans Antennas Propag 60(1):389–392

Liu L, Langley RJ (2008) Liquid crystal tunable microstrip patch antenna. IET Electron Lett 44(20):1179–1180

Liyakath RA, Takshi A, Mumcu G (2013) Multilayer stretchable conductors on polymer substrates for conformal and reconfigurable antennas. IEEE Antennas Wirel Propag Lett 12:603–606

Lyke JC (2002) A cellular automata FPGA architecture that can be trained with neural networks. Aerosp Conf Proc 5:2347–2354

Mansoul A, Ghanem F, Hamid MR, Trabelsi M (2014) A selective frequency-reconfigurable antenna for cognitive radio applications. IEEE Antennas Wirel Propag Lett 13:515–518

Mazlouman SJ, Mahanfar A, Menon C, Vaghan RG (2011) A review of mechanically reconfigurable antennas using smart material actuators. In: 5th European conference on antennas and propagation, Rome, Italy, pp 1076–1079

Mehdipour A, Denidni TA, Sebak A, Trueman CW, Rosca LD, Hoa SV (2013) Mechanically reconfigurable antennas using an anisotropic carbon-fibre composite ground. IET Microw Antennas Propag 7(13):1055–1063

Mehmood R, Wallace JW (2010) Diminishing returns with increasing complexity in reconfigurable aperture antennas. IEEE Antennas Wirel Propag Lett 9:299–302

Mehmood R, Wallace JW, Jensen MA (2014) Key establishment employing reconfigurable antennas: impact of antenna complexity. IEEE Trans Antennas Propag 13(11):6300–6310

Min Z, Xiao-Wu L, Guang-Hui W (2004) Preliminary research of the reconfigurable antenna based on genetic algorithms. In: 2004 third international conference on computational electromagnetics and its applications, Beijing, China, pp 137–140

Muri P, Obulpathi C, McNair J (2010) Enhancing small satellite communication through effective antenna system design. In: 2010 military communications conference – unclassified program, San Jose, CA, USA

Murphey TW, Pellegrino S (2004) A novel actuated composite tape-spring for deployable structures. Technical report, AIAA, TR-1528

Murphey TW, Jeon S, Biskner A, Sanford G (2010) Deployable booms and antennas using bi-stable tape-springs. In: 24th AIAA/USU conference on small satellites

Napp N, Burden S, Klavins E (2006) The statistical dynamics of programmed assembly. In: IEEE international conference on robotics and automation, Orlando, FL, USA, pp 1469–1476

Nikolaou S, Kingsley ND, Ponchak GE, Papapolymerou J, Tentzeris MM (2009) UWB elliptical monopoles with a reconfigurable band notch using MEMS switches actuated without bias lines. IEEE Trans Antennas Propag 57(8):2242–2251

Obeidat KA, Raines BD, Rojas RG, Strojny BT (2010) Design of frequency reconfigurable antennas using the theory of network characteristic modes. IEEE Trans Antennas Propag 58(10):3106–3113

Oh SS, Jung YB, Ju YR, Park HD (2010) Frequency-tunable open ring microstrip antenna using varactor. In: International conference on electromagnetics in advanced applications, Sydney, Australia, pp 624–626

Onodera S, Ishikawa R, Saitou A, Honjo K (2013) Multi-band reconfigurable antennas embedded with lumped-element passive components and varactors. In: Proceedings of the 2013 Asia-Pacific microwave conference, pp 137–139

Patnaik A, Anagnostou DE, Christodoulou CG, Lyke JC (2005) Neurocomputational analysis of a multiband reconfigurable planar antenna. IEEE Trans Antennas Propag 53(11):3453–3458

Patron D, Daryoush AS, Dandekar KR (2014) Optical control of reconfigurable antennas and application to a novel pattern-reconfigurable planar design. IEEE J Lightw Technol 32(20):3394–3402

Pendharker S, Shevgaonkar RK, Chandorkar AN (2014) Optically controlled frequency-reconfigurable microstrip antenna with low photoconductivity. IEEE Antennas Wirel Propag Lett 13:99–102

Perruisseau-Carrier J, Pardo-Carrera P, Miskovsky P (2010) Modeling, design and characterization of a very wideband slot antenna with reconfigurable band rejection. IEEE Trans Antennas Propag 58(7):2218–2226

Perruisseau-Carrier J, Tamagnone M, Gomez-Diaz JS, Esquius-Morote M, Mosig JR (2013) Resonant and leaky-wave reconfigurable antennas based on graphene plasmonics. In: Proceedings of the 2013 international symposium on antennas and propagation, Orlando, FL, USA, pp 136–137

Piazza D, Kirsch NJ, Forenza A, Heath RW, Dandekar KR (2008) Design and evaluation of a reconfigurable antenna array for MIMO systems. IEEE Trans Antennas Propag 56(3):869–881

Piazza D, Kountouriotis J, D'amico M, Dandekar KR (2009) A technique for antenna configuration selection for reconfigurable circular patch arrays. IEEE Trans Wirel Commun 8(3):1456–1467

DOI 10.1007/978-981-4560-75-7_61-2

Piazza D, Mookiah P, D'amico M, Dandekar K (2010) Experimental analysis of pattern and polarization reconfigurable circular patch antennas for MIMO antennas. IEEE Trans Antennas Propag 59(5):2352–2362

Qin PY, Jay Guo Y, Liang CH (2010a) Effect of antenna polarization diversity on MIMO system capacity. IEEE Antennas Wirel Propag Lett 9:1092–1095

Qin PY, Weily AR, Guo YJ, Bird TS, Liang CH (2010b) Frequency reconfigurable quasi-yagi folded dipole antenna. IEEE Trans Antennas Propag 58(8):2742–2747

Quin P, Guo YJ, Weily AR, Liang C (2012) A pattern reconfigurable U-slot antenna and its applications in MIMO systems. IEEE Trans Antennas Propag 60(2–1):516–528

Rajagopalan H, Kovitz JM, Rahmat-Samii Y (2014) MEMS reconfigurable optimized E-shaped patch antenna design for cognitive radio. IEEE Trans Antennas Propag 62(3):1056–1064

Ramadan AH, Costantine J, Al-Husseini M, Kabalan KY, Tawk Y, Christodoulou CG (2014) Tunable filter-antennas for cognitive radio applications. Progr Electromagn Res 57:253–265

Rodrigo D, Jofre L, Cetiner B (2012) Circular beam-steering reconfigurable antenna with liquid metal parasitics. IEEE Trans Antennas Propag 60(4):1796–1802

Sarrazin J, Mahe Y, Avrillon S, Toutain S (2009) Pattern reconfigurable cubic antenna. IEEE Trans Antennas Propag 57(2):310–317

Sathi V, Ehtheshami N, Nourinia J (2012) Optically tuned frequency reconfigurable microstrip antenna. IEEE Antennas Wirel Propag Lett 11:1018–1020

Shelley S, Costantine J, Christodoulou CG, Anagnostou DE, Lyke JC (2010) FPGA-controlled switch-reconfigured antenna. IEEE Antennas Wirel Propag Lett 9:355–358

Skinner DE, Connor JD, Foo SY, Weatherspoon MH, Powell N (2009) Optimization of multi-band reconfigurable microstrip line-fed rectangular patch antenna using self-organizing maps. In: IEEE 10th annual wireless and microwave technology conference (WAMICON), Clear Water, FL, USA

Song S, Murch RD (2014) An efficient approach for optimizing frequency reconfigurable pixel antennas using genetic algorithms. IEEE Trans Antennas Propag 62(2):609–620

Tadashi T, Miura K, Natori M, Hanayama E, Inoue T, Noguchi T, Miyahara N, Nakaguro H (2004) Deployable antenna with 10-m maximum diameter for space use. IEEE Trans Antennas Propag 52(1):2–11

Tamagnone M, Gomez Diaz JS, Perruisseau-Carrier J, Mosig JR (2013) High-impedance frequency-agile THz dipole antennas using graphene. In: 7th European conference on antennas and propagation, Gothenburg, Sweden, pp 533–536

Tawk Y, Albrecht AR, Hemmady S, Balakrishnan G, Christodoulou CG (2010) Optically pumped frequency reconfigurable antenna design. IEEE Antennas Wirel Propag Lett 9:280–283

Tawk Y, Costantine J, Avery K, Christodoulou CG (2011) Implementation of a cognitive radio front-end using rotatable controlled reconfigurable antennas. IEEE Trans Antennas Propag 59(5):1773–1778

Tawk Y, Costantine J, Christodoulou CG (2012a) A varactor based reconfigurable filtenna. IEEE Antennas Wirel Propag 11:716–719

Tawk Y, Costantine J, Hemmady S, Balakrishnan G, Avery K, Christodoulou CG (2012b) Demonstration of a cognitive radio front end using an optically pumped reconfigurable antenna system (OPRAS). IEEE Trans Antennas Propag 60(2–2):1075–1083

Tawk Y, Costantine J, Christodoulou CG (2014a) Cognitive radio antenna functionalities: a tutorial. IEEE Antennas Propag Mag 56(1):231–243

Tawk Y, Costantine J, Christodoulou CG (2014b) Reconfigurable filtennas and MIMO in cognitive radio applications. IEEE Trans Antennas Propag 62(3):1074–1083

Vaughan RG, Andersen JB (1987) Antenna diversity in mobile communications. IEEE Trans Veh Technol 36:149–172

DOI 10.1007/978-981-4560-75-7_61-2

Wu SJ, Ma TG (2008) A wideband slotted bow-tie antenna with reconfigurable CPW-to slotline transition for pattern diversity. IEEE Trans Antennas Propag 56(2):327–334

Yang SLS, Kishk AA, Lee K-F (2008) Frequency reconfigurable U-slot microstrip patch antenna. IEEE Antennas Wirel Propag Lett 7:127–129

Yang S, Zhang C, Pan HK, Fathy AE, Nair VK (2009) Frequency-reconfigurable antennas for multiradio wireless platforms. IEEE Microw Mag 10(1):66–83

Yang XS, Wang BZ, Yeung SH, Xue Q, Man KF (2011) Circularly polarized reconfigurable crossed-yagi patch antenna. IEEE Antennas Propag Mag 53(5):65–80

Zheng F, Chen M, Li W, Yang P (2008) Conceptual design of a new huge deployable antenna structure for space application. IEEE Aerosp Conf 1–7

Zhou Y, Adve RS, Hum SV (2014) Design and evaluation of pattern reconfigurable antennas for MIMO applications. IEEE Trans Antennas Propag 62(3):1084–1092

Zohur A, Mopidevim H, Rodrigo D, Unlu M, Jofre L, Cetiner BA (2013) RF MEMS reconfigurable two-band antenna. IEEE Antennas Wirel Propag Lett 12:72–75

Zuraiqi EA (2012) Neural network field programmable gate array (FPGA) controllers for reconfigurable antennas. PhD dissertation, University of New Mexico

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_62-1

Radial Line Slot Antennas

Hiroyuki Arai*
Yokohama National University, Yokohama, Kanagawa, Japan

Abstract

This chapter presents the development history of radial line slot antenna. The antennas are used for circular polarized high-gain antenna in microwave and millimeter wave bands. Linearly polarized arrays are also provided by changing slot array arrangement. The feeding structures are given by a coaxial probe and cavity resonators, which is presented in detail. The applications of radial line slot antennas are satellite communication antenna, plasma etching, and high power use. This chapter also describes these applications.

Keywords

Radial line; Slot antenna; Circular polarization; Linear polarization; Direct broadcast satellite; Plasma etching

History of Radial Line Slot Antenna

Radial line is an infinite parallel plate waveguide excited by a feeding probe at the center. The propagating outward cylindrical wave is multimode in azimuth direction, and antenna elements on the top plate are arrayed to control the radiation pattern and polarization. The first application of radial line antenna was designed as the planar circularly polarized slot array which generates sum or difference far-field patterns for monopulse tracking. The antenna is a concentric slot array designed to radiate right-hand circular polarization in the frequency range 10.2–10.8 GHz (Kelly and Goebels 1964), and difference patterns are obtained by a phasing circuit of circular waveguide (Subbarao and Fusco 2003).

A radial line slot antenna (RLSA) is a circularly polarized high efficiency planar array. Two types of RLSAs, a single-layered and a double-layered one, were released commercially for 12 GHz band direct broadcast satellite (DBS) reception. The former adopts nonuniform slots excited by a radially outward traveling wave, while the latter has uniform ones excited by an inward traveling wave. The RLSA was initially developed as a double-layered folded waveguide for the uniform aperture distribution not disturbed by slot array near the feeding probe. A radially inward traveling transverse electromagnetic (TEM) mode is propagating inside the upper waveguide, excited by the outer rim of a twofold radial waveguide fed by a coaxial probe at the center of lower waveguide. Slots are arrayed in spirally on the top plate to radiate a circularly polarized broadside beam in X band (Ando et al. 1985). The measured performance of an antenna with a diameter of 0.6 m is gain of 35 dBi, axial ratio of 1 dB, and the aperture efficiency of 57 % at 12.2 GHz (Ando et al. 1986). For the compact antenna design, a single-layered radial line slot antenna (RLSA) utilizes a radially outward traveling wave, and the slot length is varied to synthesize uniform aperture illumination. The DBS reception antenna consists of back plate, foam, slotted plate, and radome as shown in Fig. 1. The radial line is terminated by a matching spiral, in place of a

*Email: arai@ynu.ac.jp

DOI 10.1007/978-981-4560-75-7_62-1

Fig. 1 Radial line slot antenna (Courtesy of Tokyo Tech)

dummy load. It suppresses reflection and radiates all the residual power in circular polarization. For a model antenna of 600-mm diameter, the efficiency of 65 % was obtained (Takahashi et al. 1991, 1992). Merging the single-layered RLSA in the double-layered one, dual circularly polarized RLSA was proposed for DBS antenna at 12 GHz. A double coaxial waveguide provides an outward traveling wave in upper layer and an inward wave transferred from lower waveguide. Polarization isolation is more than 20 dB in DBS frequency band (Takahashi et al. 1995). A lightweight portable planar slot array antenna at X band for fixed satellite communications is composed of two RLSAs: one for reception band with left-handed circular polarization (LHCP) and the other one for transmission band with right-handed circular polarization (RHCP). A lightweight two-layered dielectric structure is used to achieve and optimize the RLSA weight. A radiation efficiency of more than 70 % is achieved due to the low dielectric constant substrate (Gonzalez et al. 2011).

Linear polarization is given by another slot array for RLSA; however, its concentric slot arrangement increases reflection from radiating elements at the feed point due to in-phase condition. The poor input S11 response is typical for waveguide broadside arrays and is suppressed by tilting the beam from the boresight (Ando et al. 1988; Takada et al. 1992). This effect is also suppressed by reflection-canceling slots; four slots form one slot set as a unit radiator of linear polarization. The antenna input reflection of S11 is improved from −2 to −10 dB, and antenna efficiencies of 48 and 54 % are measured for 400- and 600-mm-diameter antennas, respectively (Davis et al. 1997, 1999). A circularly polarized conical beam is also provided by another slot arrangement of RLSA (Takada et al. 1994). A dual linearly polarized single-layered RLSA is given by the slots arrayed in concentric ellipses and feeding the structure with two ports at the ellipses foci. Antenna ports are accessed through a 180° hybrid ring to obtain the dual behavior (Pla et al. 2006).

For a concentric array radial line slot antenna (CA-RLSA), a rotating mode should be excited. A pair of slots should have the same excitation amplitude by setting two slots with the same inclined angle to radial direction. This constraint is not applied to arbitrary polarization angle, i.e., the arbitrary phase, for slot pairs on the radial waveguide. The spiral arrangement of slot pairs is used to control its phase for the RLSA. By exciting a rotating mode inside the radial waveguide, the concentric slot arrangement can be used. Feeding circuits for the rotating mode excitation are a cum-shaped phase shifter and a cavity resonator at the feeding point (Sumiyoshi et al. 1993; Arai and Goto 1995). Another feeding circuit consists of a coplanar waveguide (CPW), ring slot, and cavity, which is planar and suitable for integrating and minimizing even in a millimeter-wave band. The experimental results agree well with the simulation (Sudo et al. 2007; Ronghong et al. 2001). Matching slots are also used for small aperture CA-RLSA. In contrast to large aperture RLSAs based upon continuous source model, the parameters of discrete slots in

DOI 10.1007/978-981-4560-75-7_62-1

the finite aperture are directly determined for uniform excitation. A very high efficiency of 85 %, and 82 %, is measured for 24-cm and 16-cm-diameter antennas, respectively (Akiyama et al. 1988).

The design of slot arrangement was investigated by a measurement-based method (Sasazawa et al. 1988), and then the slot coupling to the radial waveguide was approximated as a parallel plate waveguide with a periodic boundary condition to simulate a sectoral region of the circular aperture in an RLSA. A set of integral equations for the fields in the slot is derived, based on the field equivalence theorem, and is solved numerically by Galerkin's method. Theoretical results effectively express the measured effects of the various antenna design parameters on the slot coupling (Hirokawa et al. 1990). Using the wall impedance approach, the analysis of a single-layered RLSA is divided into two separate problems of the cavity and its external environment. A field matching method is described to analyze a recessed circular cavity radiating into a radial waveguide. This analysis was developed for linearly polarized RLSA (Bialkowski et al. 2002). The full-wave analysis and design of RLSAs was given by MoM-based software (Takada et al. 1997; Albani et al. 2006, 2009). The linearly polarized RLSAs are also applied to full-wave analysis for beam-tilting pattern (Herranz et al. 2010). Although slot antennas are usually modeled as perfectly electric conductors, for accurate antenna design and optimization, ohmic loss effects cannot be neglected. This is especially true in millimeter- and submillimeter-wave applications; Green's functions accounting for the finite conductivity of metals are derived analytically and used in the MoM admittance matrix expressions. Several RLSA realizations are analyzed to investigate the effect of ohmic losses in a practical antenna design (Mazzinghi et al. 2010; Albani et al. 2011).

The RLSA has been developed for DBS applications; however, the other is the antennas in millimeter wave bands. The conventional structure is composed of three separate parts, two conducting disks, and a foaming dielectric material, while another is a PTFE substrate for mass production. The antenna characteristics are measured in the 60-GHz band, and an excellent efficiency of 50 % at 33 dBi is obtained (Akiyama et al. 2000). A single-layered RLSA etched on a substrate and fed by a rectangular waveguide is also presented in the 60-GHz band. The antenna has rectangular waveguide feed structure using a rectangular waveguide-to-radial line transition. The prototype antenna of 10-cm diameter was tested and the gain of 30 dBi was measured at 60 GHz (Kim et al. 2002). At a 270-GHz band, the RLSA is integrated in low-temperature co-fired ceramic (LTCC). The impedance matching is enhanced by loading each radiating slot individually with a strip pair and employing an integrated feeding transition to an external standard waveguide in measurement. The antenna prototype exhibits boresight gain of 27.6 dBi with a 3-dB gain bandwidth of 9.8 GHz (Junfeng et al. 2013). Lightweight RLSAs with honeycomb-type parallel plate are also designed and demonstrated for a satellite on board antenna with a diameter of 900 mm, weight of 1.16 kg, and high gain of 44.6 dBi at 32 GHz (Nguyen et al. 2013).

Array Design

The radiation pattern of RLSA is designed by slot arrangement on the radial waveguide. Amplitude and phase of each slot are adjusted to obtain a required pattern and polarization. A popular application of high-gain circular polarization is given by a pair of nonresonant slots cut on the radial axis with orthogonal angles and quarter guided wavelength ($\lambda_g/4$) spacing as shown in Fig. 2. The λ_g is larger than the wavelength in free space not to excite grating lobes, and a slow-wave structure is inserted between the top and bottom plate. The nonresonant slot is used to control its excitation coefficient in wide range and the orthogonal slot and $\lambda_g/4$ spacing is determined to excite circular polarization. The $\lambda_g/4$ spacing also satisfies the reflection cancelation by the reflected waves from both slots with out of phase. A slot pair on the axis ρ is cut with the angle of $\pm 45^\circ$, and the directions of radiating electric fields are denoted as E_1 and E_2. The phase of E_2 delayed in $\pi/4$ gives the clockwise polarization location, while a slot pair on the axis ρ'

DOI 10.1007/978-981-4560-75-7_62-1

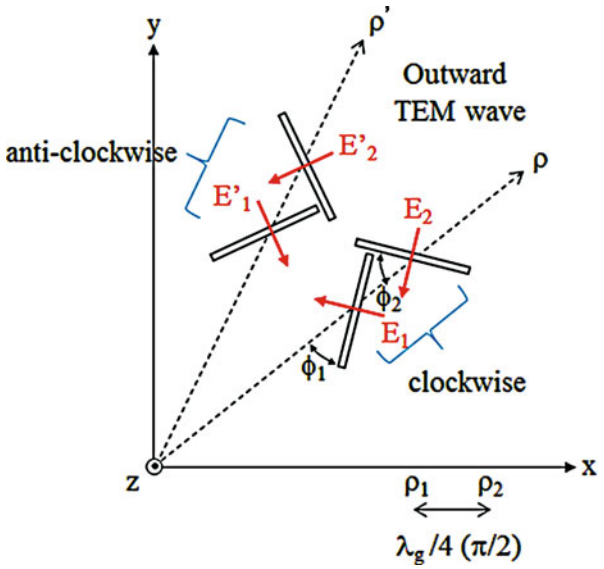


Fig. 2 Geometry of slot arrangement

is for the anticlockwise one. As shown in Fig. 2, the angle of slot cut, ϕ_1 and ϕ_2, determines the right- or left-handed circular polarization.

The coupling factor of the slot depends on the magnetic field strength parallel to the slot propagating inside the radial waveguide. A large number of slot pairs for a pencil boresight beam give very small difference in their amplitude excitation coefficient for a quarter wavelength distance along the radial direction, and then a pair of slots is designed as the same length. The array parameter S_φ and S_ρ are less than λ_g not to excite grating lobes. The slot pairs are placed on a spiral to provide the phase difference between adjacent ones to keep rotational phase angle for the circular polarization. The direction of TEM wave propagating inside radial waveguide changes the right- or left-handed circular polarization. The slot arrangement in Fig. 3 gives right-handed circular polarization for outward propagation wave. The spiral arrangement parameter is given by the position of slot pair along the ρ axis to satisfy the condition of $(\rho'_1 - \rho_1)/\lambda_g = (\Phi_2 - \Phi_1)/2\pi$ for the circular polarization as shown in Fig. 4.

The aperture distribution is designed based on the traveling-wave excitation inside radial waveguide. The amplitude of nonresonant slot pair is controlled by the length, which causes phase shift by the radiation from slots and requires the phase control for uniform phase distribution. The phase between adjacent slot pairs is controlled by its spacing. A theory of the slot design for uniform aperture distribution proposed in (Takahashi et al. 1991) is given in the following procedure. As the first step, a unit slot pair is characterized by a parallel plate waveguide with periodic side wall conditions. Assuming of no reflections from slot pairs, an incident unit power is split into the slot radiation $1-\exp(-2\alpha S_\rho)$ and the transmission $\exp(-2\alpha S_\rho)$, where α and S_ρ are a coupling factor and a spacing between slot pairs along the ρ axis, respectively. The phases of the radiated and transmitted field are denoted as Φ and $-k_o S_\rho/\zeta$ by expressing the slow-wave factor as ζ. The coupling factor α for desired aperture distribution is derived as a function of slot location ρ as shown in Fig. 5a, where the antenna radius of 0.3 m and the power dissipation of 15 % at the termination are assumed at 12 GHz. This curve is determined not to disturb the rotational symmetry of the field excessively. The slot length and position are determined by nonuniform array designs, assuming the variation of these parameters is very small (Hirokawa et al. 1992). Once the slot position ρ_s is set, the coupling factor $\alpha(\rho_s)$ (m) is given by the solid curve denoted as (i) in Fig. 5a; the slot length L_s/λ_o is chosen by the dashed curve as (ii). The slow wave factor ζ and the phase Φ of the radiated field are obtained in Fig. 5b. This procedure is repeated by the end slot. As the last adjustment, all the slot pairs are

DOI 10.1007/978-981-4560-75-7_62-1

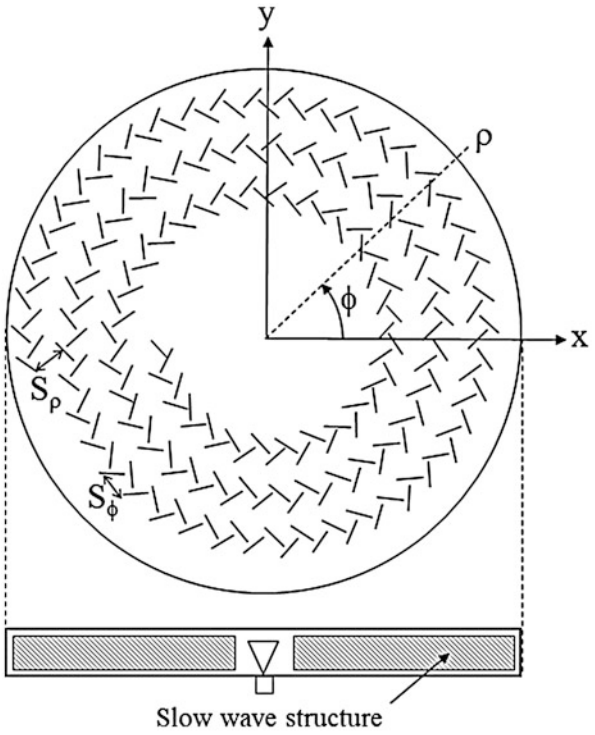


Fig. 3 Spiral slot array for circularly polarization

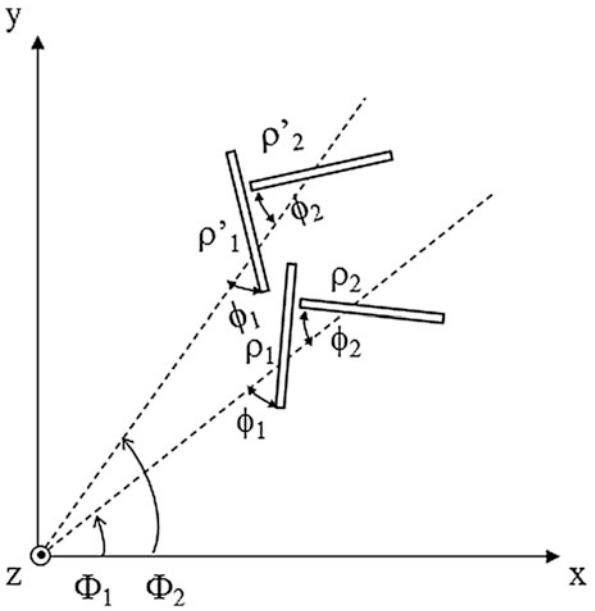


Fig. 4 Slot parameters for circularly polarization

shifted to compensate the phase Φ to obtain uniform phase distribution. This procedure is also applied to a Taylor distribution to suppress the sidelobes (Takahashi et al. 1997; Hirokawa and Ando 2000).

Small-sized aperture antennas have a problem to obtain uniform amplitude distribution by strong excitation of slots near the waveguide center. Another slot arrangement is given by concentric array as

DOI 10.1007/978-981-4560-75-7_62-1

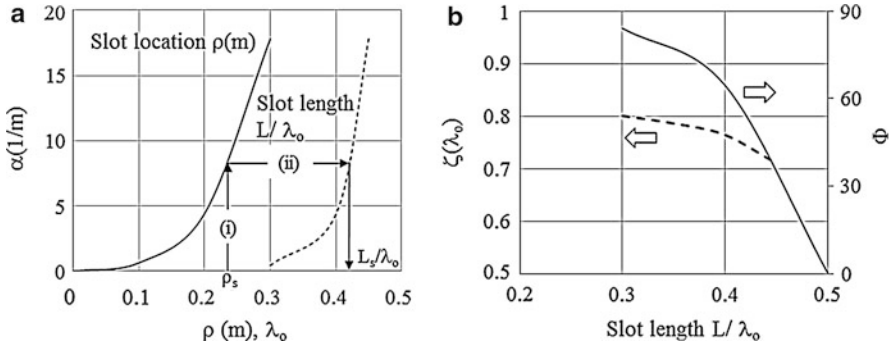


Fig. 5 Slot design parameters, (**a**) coupling factor, (**b**) slow-wave factor

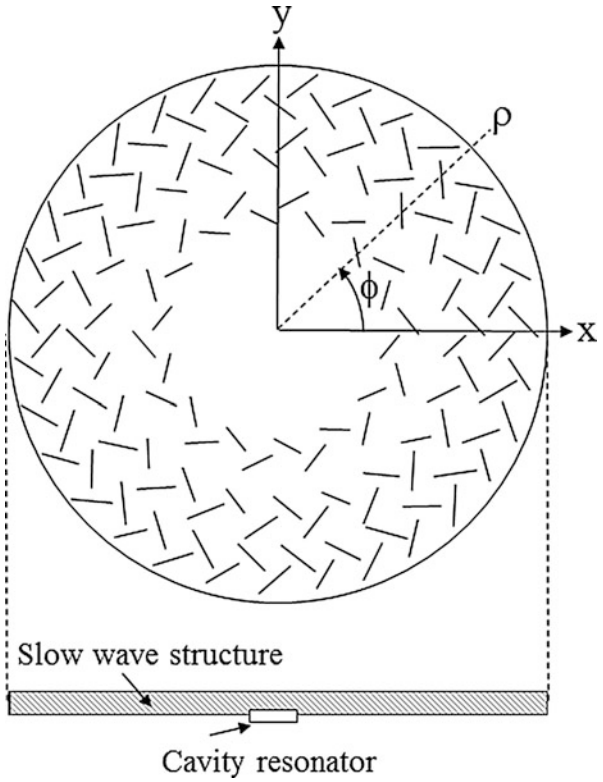


Fig. 6 Concentric slot array

shown in Fig. 6. The slot pairs are arrayed on concentric rings to keep the rotational symmetry and uniform aperture distribution along radial direction. This array needs feed circuits to excite rotating mode in azimuth direction. The feeding structures for this mode excitation are a cavity resonator (Arai and Goto 1995) and a ring or crossed slot cut on the bottom plate (Kigure et al. 1998; Sudo et al. 2002). The concentric array provides a conical beam by the TEM wave excitation inside waveguide. A probe feed at the waveguide center same as the spiral arrangement array is also used.

A linear polarized RLSA is given by the slot arrangement as shown in Fig. 7 (Ando et al. 1988). A slot pair is placed with $\lambda_g/2$ spacing to make the x-directed linear polarization given as the sum of E_1 and E_2. Slots on the x axis parallel to x direction are not coupled to propagating wave, and only y-directed ones radiate in the boresight. The $\lambda_g/2$ spacing of slot pairs degrades input impedance matching at the feed due to reflection waves from the slot in phase, and then reflection-canceling slots, a slot pair with $\lambda_g/4$ spacing,

DOI 10.1007/978-981-4560-75-7_62-1

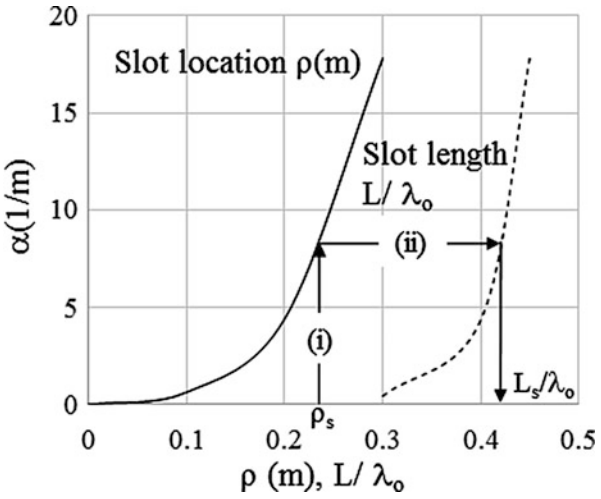


Fig. 7 Linear polarization array

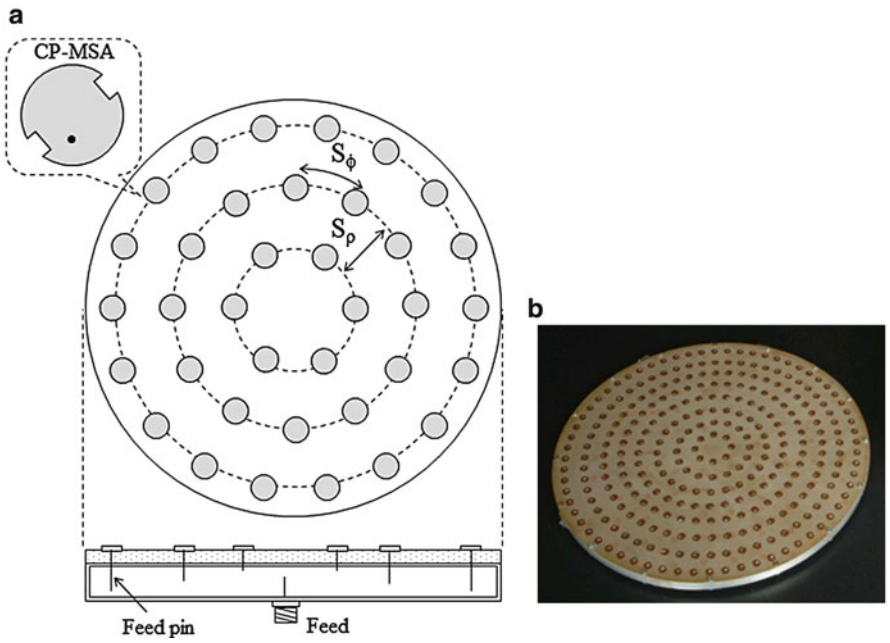


Fig. 8 Radial line waveguide with microstrip antenna, (**a**) antenna geometry, (**b**) photo of antenna (Courtesy of Saitama University)

are introduced to improve input reflection performance (Takada et al. 1992). A tilted beam is also effective to suppress the reflections (Takada et al. 1988).

Other radiating elements for radial waveguide are a helical and a microstrip antenna for circular polarization (Nakano et al. 1992; Shibata et al. 1993). Each antenna element is excited by different length probe for amplitude control, and the phase is changed by its rotation angle. Figure 8 shows RLSA with microstrip antenna radiating elements. The fine-tuning of aperture distributions is easy by adjusting two excitation parameters.

DOI 10.1007/978-981-4560-75-7_62-1

RLSA Applications

RLSAs were introduced for high-density plasma sources of the electron cyclotron resonance (ECR) plasma system, providing low-profile antenna structure, high-power handling capability, and uniform aperture field distribution (Takahashi et al. 1993). The RLSA system eliminates a uniform and strong magnetic field generated by a coil outside the vacuum chamber which has been used for horn antenna systems to improve the uniformity of the plasma on a wafer. The RLSA in a cylindrical vacuum chamber was used in the plasma process (Ohmi 1997). Microwave energy is radiated into the vacuum chamber through a quartz glass. The measured electron density is raised more than 10^{12} cm^{-3}. The plasma is excited by the surface of the glass wafer plate and diffuses towards the electrode. The electron density below the quartz glass surface is far beyond the cutoff density. Microwave incident on the plasma is completely dissipated by high-density plasma near the quartz plate.

The radial line slot antenna provides high-density plasma system at a temperature of 400 °C as an advanced gate dielectric film, resulting in low density of interface trap and bulk charge, lower leakage current than jet vapor deposition silicon nitride, and thermally grown silicon oxide with same equivalent oxide thickness. It raised breakdown field intensity with no stress-induced leakage current, very little trap generation even in high-field stress, and excellent resistance to boron penetration and oxidation (Sekine et al. 2000).

For high-power microwave applications, RLSA is fed from a double-layered radial line waveguide to realize the directional radiation with high efficiency. A reflection-canceling slot is added to improve the antenna impedance matching. An antenna operating at 9.42 GHz was designed with a length of 15 cm and an aperture radius of 27 cm; the high-power experiment proves that this antenna has a power-handling capacity of more than 1 GW (Cheng et al. 2013).

For an application of near-field focusing, the RLSA is designed to control the side lobe level and beamwidth of the normal component of the electric field with respect to the radiating aperture. The requirements on the near-field pattern are provided over a focusing plane at a given distance from the radiating aperture. A set theoretic approach is then used to derive the aperture field distribution fitting the requirements over the near field. Then, the aperture field distribution is synthesized by accurately placing and sizing the slots of the antenna. The spillover efficiency is maximized during the design process. The antenna is centrally fed by a simple coaxial probe. The antenna design is validated by a prototype and measurements at 12.5 GHz (Ettorre et al. 2014).

Summary

This chapter described the development history, the theory, and application of RLSAs. The unique features of RLSA are a simple feeding circuit and hundreds of radiating elements. The array parameters are designed by several approximate methods and also by full-wave analysis to control precisely. The low transmission loss waveguide is the advantage of this antenna, which provides high-gain antennas in high-frequency range and expands the application in wide areas.

Cross-References

► Antennas in High-Power High-Voltage Systems and Electromagnetic Pulse Systems
► Circularly Polarized Antennas
► Low-Profile Antennas

DOI 10.1007/978-981-4560-75-7_62-1

- ▶ Microstrip Patch Antennas
- ▶ Millimeter-Wave Antennas and Arrays
- ▶ Spiral, Helical and Rod Antennas
- ▶ Waveguide Slot Antennas and Arrays
- ▶ Waveguide Slot Array Antennas

References

Akiyama A et al (1988) Numerical optimisation of slot parameters for a concentric array radial line slot antenna. Microwaves Antennas Propag IEE Proc 145(2):141–146

Akiyama A et al (2000) High gain radial line slot antennas for millimetre wave applications. Microwaves Antennas Propag IEE Proc 147(2):134–138

Albani M et al (2006) An efficient full-wave method of moments analysis for RLSA antennas. IEEE Trans Antennas Propag 54(8):2326–2336

Albani M et al (2009) Asymptotic approximation of mutual admittance involved in MoM analysis of RLSA antennas. IEEE Trans Antennas Propag 57(4):1057–1063

Albani M et al (2011) Rigorous MoM analysis of finite conductivity effects in RLSA antennas. IEEE Trans Antennas Propag 59(11):4023–4032

Ando M et al (1985) A radial line slot antenna for 12 GHz satellite TV reception. IEEE Trans Antennas Propag 33(12):1347–1353

Ando M et al (1986) Characteristics of a radial line slot antenna for 12 GHz band satellite TV reception. IEEE Trans Antennas Propag 34(10):1269–1272

Ando M et al (1988) A linearly polarized radial line slot antenna. IEEE Trans Antennas Propag 36(12):1675–1680

Arai H, Goto N (1995) Novel feed circuit for radial-line waveguide. In: Proceedings of the ninth international conference on antennas and propagation, Eindhoven, vol 1, pp 231–234

Bialkowski ME et al (2002) Analysis of a circular patch antenna radiating in a parallel-plate radial guide. IEEE Trans Antennas Propag 50(2):180–187

Cheng WY et al (2013) Designs and experiments of a novel radial line slot antenna for high-power microwave application. IEEE Trans Antennas Propag 61(10):4940–4946

Davis PW et al (1997) Experimental investigations into a linearly polarized radial slot antenna for DBS TV in Australia. IEEE Trans Antennas Propag 45(7):1123–1129

Davis PW et al (1999) Linearly polarized radial-line slot-array antennas with improved return-loss performance. IEEE Antennas Propag Mag 41(1):52–61

Ettorre M et al (2014) On the near-field shaping and focusing capability of a radial line slot array. IEEE Trans Antennas Propag 62(4):1991–1999

Gonzalez JMF et al (2011) Lightweight portable planar slot array antenna for satellite communications in X-band. IEEE Antennas Wirel Propag Lett 10:1409–1412

Herranz JI et al (2010) Optimization of beam-tilted linearly polarized radial-line slot-array antennas. IEEE Antennas Wirel Propag Lett 9:1165–1168

Hirokawa J, Ando M (2000) Sidelobe suppression in 76 GHz post-wall waveguide-fed parallel plate slot arrays. IEEE Trans Antennas Propag 48(11):1727–1732

Hirokawa J et al (1990) Analysis of slot coupling in a radial line slot antenna for DBS reception. Microwaves Antennas Propag IEE Proc H 137(5):249–254

Hirokawa J et al (1992) Waveguide-fed parallel plate slot array antenna. IEEE Trans Antennas Propag 40(2):218–223

Junfeng X et al (2013) Xianming Qing 270-GHz LTCC-integrated strip-loaded linearly polarized radial line slot array antenna. IEEE Trans Antennas Propag 61(4, Part 1):1794–1801
Kelly KC, Goebels F (1964) Annular slot monopulse antenna arrays. IEEE Trans Antennas Propag 12(4):391–403
Kigure Y et al (1998) A radial line slot antenna fed by a ring slot coupled planar circuit. In: Korea Japan joint conference '98, Sep, Pusan
Kim Y et al (2002) 60 GHz band radial line slot array antenna fed by rectangular waveguide. Electron Lett 38(2):59–60
Mazzinghi A et al (2010) Influence of the finite slot thickness on RLSA antenna design. IEEE Trans Antennas Propag 58(1):215–218
Nakano H et al (1992) Low-profile helical array antenna fed from a radial waveguide. IEEE Trans Antennas Propag 40:279–284
Nguyen T et al (2013) An equivalent double layer model for a fast design and analysis of high gain multilayer radial line slot antennas. IEICE Trans Commun 96(11):2891–2900
Ohmi T (1997) Preface in New era of semiconductor manufacturing I. Ultra Clean Tech 9(Suppl 1):1–5
Pla V et al (2006) Dual linearly polarized single-layer radial-line slot-array antenna. IEEE Antennas Wirel Propag Lett 5(1):120–122
Ronghong J et al (2001) A feeding circuit with CPW for CA-RLSA. IEEE Trans Antennas Propag 49(12):1862–1867
Sasazawa H et al (1988) Slot coupling in a radial line slot antenna for 12-GHz band satellite TV reception. IEEE Trans Antennas Propag 36(9):1221–1226
Sekine K et al (2000) Highly robust ultrathin silicon nitride films grown at low-temperature by microwave-excitation high-density plasma for giga scale integration. IEEE Trans Electron Device 47(7):1370–1374
Shibata O et al (1993) Radiation properties of microstrip array antennas fed by radial line. Electron Commun Jpn 76(12):93–102, Wiley InterSci, pt. 1
Subbarao B, Fusco VF (2003) Probe-fed circularly polarised monopulse radial line slot antenna. Electron Lett 39(21):1495–1496
Sudo K et al (2002) An analysis and a design for excitation of a rotating mode in a radial waveguide by a cross slot-coupled rectangular waveguide. In: IEICE technical report, AP2001-206, Feb (in Japanese), Tokyo
Sudo K et al (2007) Reduction of azimuthal amplitude ripple in the rotating-mode feed to a radial waveguide by using a crossed dog-bone slot. IEEE Trans Antennas Propag 55(9):2618–2622
Sumiyoshi H et al (1993) The feed circuit for exciting a rotating electric field in a radial waveguide. In: IEICE spring national conference, B-51, Mar (in Japanese), Nagoya
Takada J et al (1988) A beam-tilted linearly polarized radial line slot antenna. In: IEICE spring national conference, B-107, Mar (in Japanese), Tokyo
Takada J et al (1992) A reflection cancelling slot set in a linearly polarized radial line slot antenna. IEEE Trans Antennas Propag 40(4):433–438
Takada J et al (1994) Circularly polarised conical beam radial line slot antenna. Electron Lett 30(21):1729–1730
Takada J et al (1997) Method of moments analysis of a small aperture radial line slot antenna using the rectangular cavity Green's function. Microwaves Antennas Propag IEE Proc 144(6):498–500
Takahashi M et al (1991) A slot design for uniform aperture field distribution in single-layered radial line slot antennas. IEEE Trans Antennas Propag 39(7):954–959
Takahashi M et al (1992) Characteristics of small-aperture, single-layered, radial-line slot antennas. Microwaves Antennas Propag IEE Proc 139(1):79–83

DOI 10.1007/978-981-4560-75-7_62-1

Takahashi M et al (1993) A planar antenna for ECR plasma system. In: IEICE spring national conference, B-65, Mar (in Japanese), Nagoya

Takahashi M et al (1995) Dual circularly polarized radial line slot antennas. IEEE Trans Antennas Propag 43(8):874–876

Takahashi M et al (1997) Low sidelobes for radial line slot antennas. Electron Commun Jpn 80(7):70–76, Wiley InterSci, pt. 1

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_64-1

Millimeter-Wave Antennas and Arrays

Wonbin Hong*
Samsung Electronics, Suwon, Republic of Korea

Abstract

The growing interest associated with exploiting the millimeter-wave spectrum for future wireless communication devices and networks has rendered the ensuing need for in-depth research and development of advanced millimeter-wave antennas. This chapter provides an overview of the background and motivation of ultrafast, low-latency millimeter-wave wireless applications as well as key characteristics of the spectrum. Major antenna design considerations and techniques are introduced and discussed in detail. The discussion is followed with state-of-the-art millimeter-wave antenna design techniques applicable to next-generation wireless devices. The chapter is concluded with discussions of future directions.

Keywords

Millimeter waves; Planar antennas; Beam steering; Phased array; 60 GHz; 5G communications

Introduction

Millimeter waves (mmWaves) are collectively defined as frequency spectrums in which the free-space wavelength can be described using millimeter units. Formally denoted as an extremely high frequency (EHF) by the International Telecommunications Unit (ITU), the spectrum ranges from 30 to 300 GHz. The very concept of mmWave radio technology commands over a century-long history with early applications in radio astronomy and military applications half a century ago. However, to the majority of the demographic, mmWave remained a distant, obscure terminology until recent years.

It was not until the relatively recent remarkable technological advancements in silicon and compound semiconductors, coupled with the relentless demand for sophisticated wireless data infrastructure, that mmWave technologies gained traction among diverse technology circles. As of present, potentially disruptive technologies, such as extreme high-data-rate low-latency wireless communication (Chen and Niknejad 2011; Emami et al. 2011; Ghosh et al. 2014; Roh et al. 2014; Talwar et al. 2014), advanced resolution radar imaging (Valdes-Garcia et al. 2013), and inter- and intra-chip communication networks (Baek and Hong 2013), are primary examples of how mmWave radio and radars are reshaping the paradigm of wireless technologies. For each of these applications, increasing the operating frequency by several orders of magnitude is a viable and an effective solution to the current worldwide frequency spectrum overcrowding issues below 6 GHz. Moreover, the channel capacity denoted in bits per seconds established in the classical work by Shannon and Hartley (Shannon 1948) is directly proportional to the channel bandwidth. For instance, the short-range, high-data-rate IEEE 802.11ad WiGig standard can achieve PHY (physical layer) rates that are ten times or more higher than conventional Wi-Fi by utilizing more than 2 GHz of channel bandwidth centered around the unlicensed 60 GHz IMS (industrial, medical, scientific) band. The latest 4G Long-Term Evolution (LTE) cellular technology continues to become

*Email: wonbin.hong@samsung.com

DOI 10.1007/978-981-4560-75-7_64-1

increasingly sophisticated and efficient using innovative technologies such as carrier aggregation (CA), single-user and multiuser multiple-input multiple-output (SU-MIMO/MU-MIMO), and heterogeneous networks (HetNets) using macro and small cells. Encouraged by notable advancements in cost-efficient mmWave RFIC technologies (Natarajan et al. 2009; Cohen et al. 2013; Tsukizawa et al. 2013; Kong et al. 2013; Bereka et al. 2014; Okada et al. 2014), ultra-wideband mmWave wireless communication technologies that can augment LTE networks are being actively researched and experimented by numerous research institutions and private sectors (Sun et al. 2014; Taori and Sridharan 2014; Chin et al. 2014; Hong et al. 2014a). By exploiting the large available bandwidth at mmWave, low real-time latency access and backhaul networks with more than 10 Gbps peak data rates are projected to be the backbone of future fifth-generation wireless communications.

The benefits of mmWave technologies do not come without challenges. The inherent difference at mmWave compared to frequencies below 10 GHz demands unique design considerations in hardware circuitry. The antenna is one of the most critical elements that affect the overall performance of mmWave radios. In this chapter, section "Theory" will first discuss and understand the unique nature of mmWave spectrum and major factors that must be taken into consideration during the antenna design. Afterward, key mmWave design techniques and schematics will be presented in detail in section "Key mmWave Antenna Design Techniques". Section "Notable mmWave Antenna Applications" will exemplify notable mmWave antenna topologies currently employed in high-volume applications such as consumer electronics and will discuss the top-down design process. Section "Conclusion" presents a number of important future trends in mmWave antenna technology, and the chapter is concluded.

Theory

Key Properties of mmWaves

At mmWave spectrum, there are several notable loss factors that will significantly affect the behavior of radio transmission and reception. A number of key propagation loss factors that affect radio signal transmission and reception in the following will be first identified and discussed.

Free-Space Propagation Loss

Free-space propagation loss is defined as the propagation loss that is dependent on the distance and operating frequency of two isotropic transmit and receive antennas. Equation 1 describes the free-space propagation loss in closed-form expression,

$$\text{Free} - \text{Space}\ \ \text{Propagation}\ \ \text{Loss} = \left(\frac{4\pi R}{\lambda}\right)^2 \tag{1}$$

where R is the distance between transmit and receive antennas and lambda is the free-space wavelength at the operating frequency. Equation 1 can be rewritten in dB scale and frequency for better understanding as shown in Eq. 2:

$$\text{Free} - \text{Space Propagation Loss}\ (\text{dB}) = \ 20\log F + 20\log R + 92.4 \tag{2}$$

where frequency F is expressed in GHz and distance R is expressed in kilometers (km). It can be intuitively interpreted that the free-space path loss will increase in logarithmic scale of the operating frequency. The effect of free-space propagation loss on the radio link budget can be visualized using the following Eq. 3, where P_{RX} is defined to be the receiving power (dB) and P_{TX} is defined to be the transmitting power (dB):

DOI 10.1007/978-981-4560-75-7_64-1

$$P_{RX} = P_{TX}\left(\frac{\lambda}{4\pi R}\right)^2 \tag{3}$$

Precipitation Loss

At mmWave frequencies, the sizes of the raindrops are typically within several multitudes of the free-space wavelength. This causes noticeable scattering during radio signal transmission in the presence of rain. Figure 1 presents the radio signal attenuation per kilometer (*km*) as a function of rain rate and rain intensity. As it can be observed, radio signals below 10 GHz generally experience limited levels of attenuation as a function of rain rate. In contrast for mmWave radio transmission, the specific region and the yearly rain rate must be taken into account from the beginning of the signal link budgeting and the antenna design as this is one of the major factors that account for signal fading.

Atmospheric Absorption and Gaseous Loss

Atmospheric constituents tend to absorb radio signals at much higher rates at mmWave frequencies compared to lower microwave frequencies. This tendency is loosely proportional to signal frequency, and a number of notable attenuation peaks can be observed as presented in Fig. 2. This can be explained by the natural mechanical resonance behaviors of oxygen (O_2) molecules and water (H_2O) molecules at certain mmWave frequencies. For example, the oxygen absorption rates at 30 and 60 GHz are 0.02 and 15 dB/km, respectively. The O_2 molecules reach maximum absorptions around 60.3 GHz. The inherent nature of severe gaseous losses at certain mmWave frequencies are the primary factors behind majority of the devised close-range communication scenarios. Otherwise, spectral regions within the mmWave frequency bands with limited gaseous losses commonly denoted as "atmospheric windows" are utilized for transmissions requiring midrange and above scenarios.

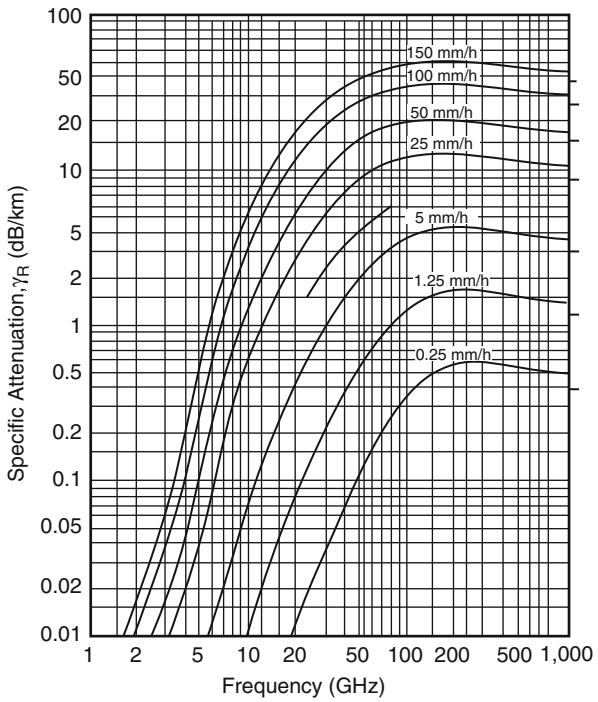


Fig. 1 Precipitation loss as a function of rain rate (Marcus and Pattan 2005)

DOI 10.1007/978-981-4560-75-7_64-1

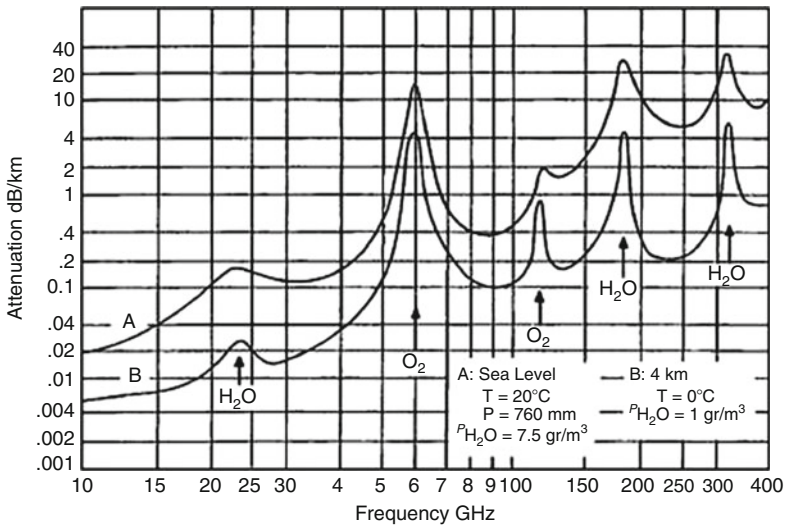


Fig. 2 Atmospheric absorption (Marcus and Pattan 2005)

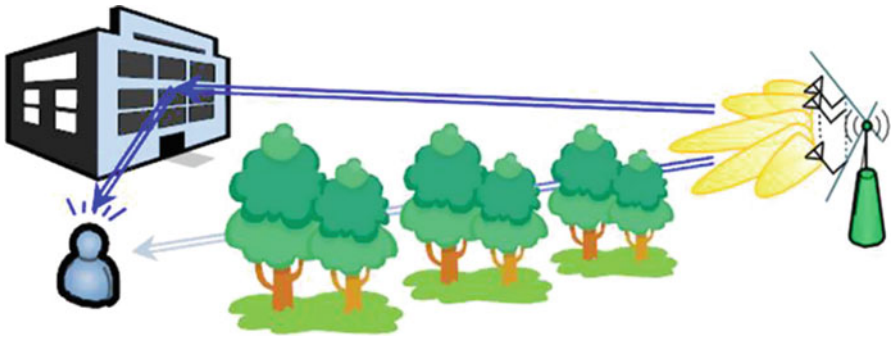


Fig. 3 Illustration of foliage loss

Foliage Loss

Losses incurred by foliage can significantly affect long-term fading of the mmWave channel. Foliage loss illustrated in Fig. 3 is dependent on the density of the foliage, the tree species, humidity, etc., and non-negligible (Marcus and Pattan 2005). The empirical foliage loss can be expressed as Eq. 4 as follows where f is defined as the operating frequency in MHz and D is denoted as the depth of foliage transverse in meters when $D < 400$ m:

$$L_{foliage} = 0.2f^{0.3}D^{0.6} \text{ dB} \tag{4}$$

Material Loss

It is imperative to carefully consider signal attenuation loss incurred by various materials during the design stage. With few exceptions, radio signals beyond 5 GHz are subject to an exponential increase in attenuation as the signal waves propagate through the stated material. This tendency is presented in Fig. 4. Consequently, proper material selection becomes one of the most critical design aspects for mmWave antennas as the material composition of the antenna substrate, the radome, and the connectors is closely linked to the radiation efficiency. Furthermore, the importance of low-loss interconnects and antenna feeding and packaging technologies is emphasized to minimize attenuations of hard-fought mmWave

DOI 10.1007/978-981-4560-75-7_64-1

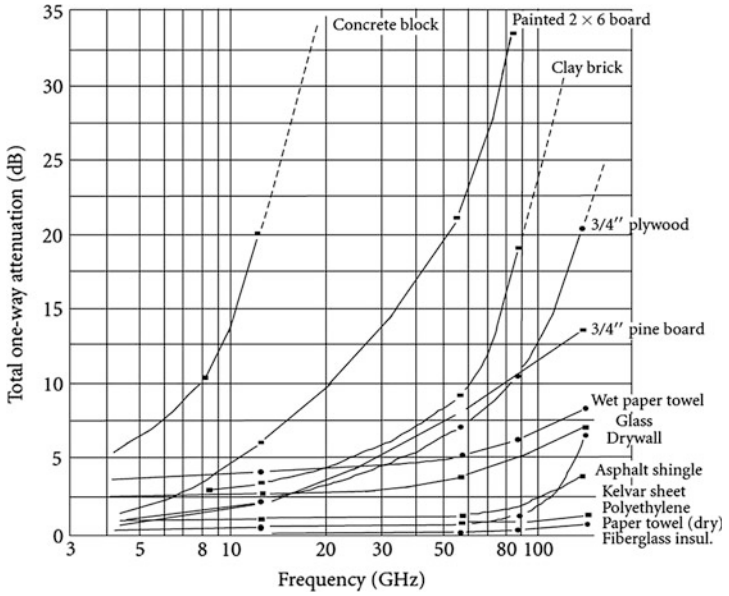


Fig. 4 One-way signal attenuation as a function of frequency

signals that were received and generated by the antenna and the RFIC, respectively. It is imperative that antenna and mmWave radio link designers must be accurately aware of the loss tangent properties of materials that the radio waves are likely to encounter during propagation scenarios. As it is not common for majority of the material vendors and manufacturers to possess material properties beyond 10 GHz, permittivity and loss tangent extraction and calculation techniques (El Sabbagh et al. 2004; Fratticcioli et al. 2004; Fang et al. 2004) must be considered and selectively applied for each design scenario.

Antenna Gain

As it can be observed from the previous subsection, notable levels of signal attenuation from numerous factors are one of the most compelling properties at the mmWave frequency spectrum. In order to establish successful radio transmission, the collective signal attenuation must be less than the radio link budget. The radio link budget can be expressed as described in Eq. 5 below:

$$\text{Radio Link Budget (dB)} = \text{PTX} + \text{GTX} + \text{GRX} - R_{\text{sensitivity}}\ \text{(dB)} \tag{5}$$

$R_{\text{sensitivity}}$ is defined as the minimum received power that is detectable for the specified bit error ratio (BER) or packet error ratio (PER). $R_{\text{sensitivity}}$ incorporates the various signal attenuation loss factors described earlier. The term should not be confused with the receiver noise figure or the signal-to-noise ratio (SNR) as these terms are also inclusive of the $R_{\text{sensitivity}}$ calculation. The designing of the mmWave RX antenna is closely related to $R_{\text{sensitivity}}$ as the signal attenuation loss in the antenna feedlines degrades the noise figure of the receiver. PTX is defined as the transmit power and can be understood as either the total power transmitted at the TX antenna port or the total power generated by the power amplifier. In general, the amplitudes of the two tend to be different as internal signal losses such as routing losses between the power amplifier and the TX antenna or switch and duplexer losses tend to be non-negligible at mmWave frequencies. Conclusively speaking, maximizing the radio link budget (dB) is the principal priority in mmWave radio system design.

DOI 10.1007/978-981-4560-75-7_64-1

Based on Eq. 5, increasing P_{TX} is an intuitive method in achieving the desired radio link budget and mmWave radio signal coverage. However, it remains an extremely difficult technical challenge to achieve an ideal balance between permissible levels of power consumption and power-added efficiency (PAE) using present-day semiconductor technologies. The scarcity of available power consumption capacity is evidently rampant particularly in the consumer electronics arena. For example, the recent design trend featuring continuously improved hardware capacity for slimmer smartphones results in significant strain on realizing improved battery life for mobile and wearable devices. Utilizing III–V compound semiconductor technologies in conjunction with CMOS processing for better RF output power and efficiency at mmWave can potentially alleviate this constraint. However, identifying cost-effective fabrication and packaging schematics remains a pertinent issue at present.

The significant power constraints related to PTX exemplifies the importance of antenna gain at mmWave frequencies. For better understanding of the effect of the antenna gain, Eq. 3 can be reorganized to be function of the antenna aperture as presented in Eq. 6:

$$P_{RX} = P_{TX}\left(\frac{1}{4\pi R^2}\right)\left(\frac{\lambda^2}{4\pi}\right) \tag{6}$$

For resonant antennas, the apertures of mmWave antennas are within the sub-mm and mm range, which are typically a fraction of low-microwave-frequency antenna apertures. This decrease in antenna aperture at mmWave naturally results in smaller captured energy at the receiver side which explains the inverse relationship between P_{RX} and antenna operating frequency. Employing omnidirectional antennas similar to those that are typically used for Wi-Fi and cellular communications, the relatively small antenna gain combined with signal attenuation loss at mmWave can potentially reduce received signal power levels in the orders of 20log$F$ where F is operating frequency in GHz. For example, for identical P_{TX}, the P_{RX} at 30 GHz will be 20 dB less than that at 3 GHz. Consequently, these design parameters become significant limiting factors in transmitting and receiving radio signals beyond very close range, line-of-sight (LoS) distance. This phenomenon justifies many of the mmWave applications designated for reactive and Fresnel zones for which the antenna design aspects will be discussed later in this chapter. Evidently, the mmWave antenna aperture becomes an ever-important piece in the puzzle. As antenna aperture is directly linked to its directivity and gain, Eq. 6 can be rewritten in the form of Friis transmission formula as shown in Eq. 7:

$$P_{RX} = P_{TX}G_{TX}G_{RX}\left(\frac{\lambda}{4\pi R}\right)^2 \tag{7}$$

G_{TX} and G_{RX} are denoted as the respective antenna gain for transmit and receive antennas. Furthermore, Eq. 7 can be reorganized to illustrate how transmit and receive antenna apertures affect P_{RX} as shown in Eq. 8 where $A_{e,\,Tx}$ and $A_{e,\,Rx}$ are defined as the effective antenna aperture for transmit and receive antennas, respectively:

$$\begin{aligned} P_{RX} &= P_{TX}\left(\frac{4\pi A_{e,Tx}}{\lambda^2}\right)\left(\frac{4\pi A_{e,Rx}}{\lambda^2}\right)\left(\frac{1}{4\pi R^2}\right)\left(\frac{\lambda^2}{4\pi}\right) \\ &= P_{TX}A_{e,Tx}A_{e,Rx}\left(\frac{1}{\lambda^2 R^2}\right) \end{aligned} \tag{8}$$

DOI 10.1007/978-981-4560-75-7_64-1

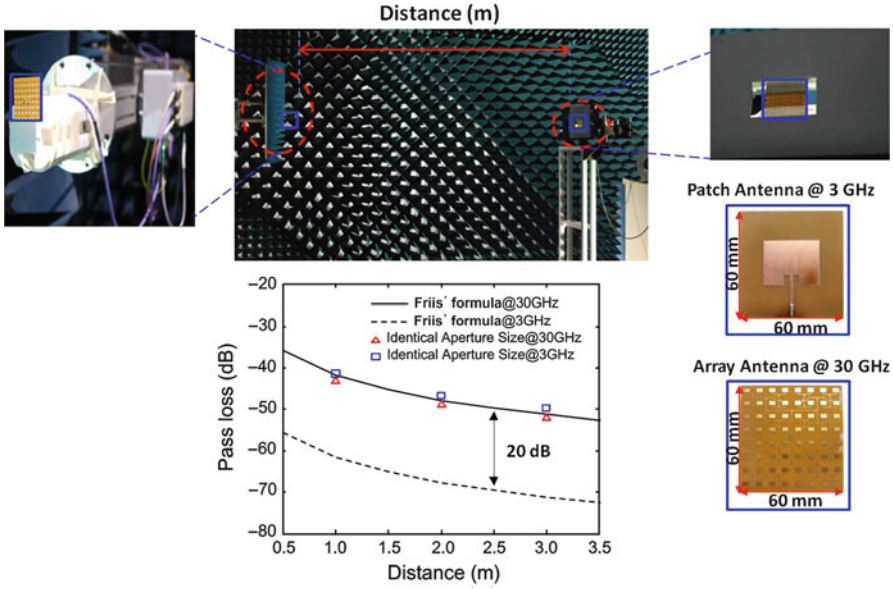


Fig. 5 Demonstrating the effectiveness of increasing the antenna aperture

From Eq. 8, P_{RX} can be interpreted as the product of the effective apertures of the two antennas and is inversely proportional to the square of the operating frequency and communication distance. Figure 5 illustrates a simple demonstration of the effectiveness of increasing the antenna aperture at mmWave. As an example, a rudimentary patch antenna at 3 GHz is designed. When applying an identical patch antenna topology at 30 GHz, the antenna aperture would decrease in accordance to the square of the antenna scaling factor, eventually resulting in a 20 dB difference in P_{RX}. However, by designing an antenna topology at 30 GHz which features an antenna aperture size that is identical to that at 3 GHz, the consequent antenna gain enhancement compensates for the inherent path loss difference associated within $R_{\text{sensitivity}}$ as described earlier. This demonstration underlines the critical importance of exploiting numerous antenna gain improvement design methodologies for mmWave communications. It also explains why majority of mmWave antenna designs employ various aperture and phased-array antenna design schematics. At mmWave frequencies, losses within the antenna substrate material interconnect insertion losses and polarization mismatch can be mostly attributed to the reduction in antenna gain.

Antenna Radiation Characteristics

The antenna gain is inherently dependent on the directivity of the antenna. The antenna directivity is defined to be the ratio of the radiation intensity in a given direction (θ,φ) in the far-field region to the radiation intensity value which is averaged over a surface area of sphere. Resonant-type or electrically small antennas tend to radiate equally in all directions and effectively have very low directivity which is close to 1 in amplitude or 0 dBi. For applications where the antennas mostly operate within the near-field or very close range, the directivity of mmWave antennas is less of a design priority. In contrast for scenarios where the mmWave transmitter and receiver are separated by more than a few meters in distance, the directivity of the antenna plays a substantial role in establishing a radio link. It is possible to categorize how the antenna radiation characteristics affect the overall quality of mmWave radios based on the deployments of the transmitter and receiver.

DOI 10.1007/978-981-4560-75-7_64-1

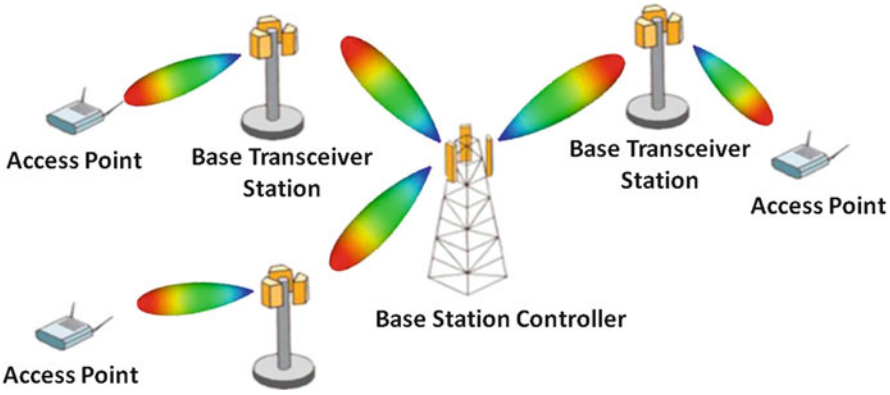


Fig. 6 Example of a mmWave fixed-beam application

Fixed-Beam Applications (Point to Point)

The RX and TX antennas are respectively configured based on careful radio link estimations using SNR, effective isotropic radiated power (EIRP), and received signal strength indication (RSSI). The deployments of the mmWave transmitter and receiver are carefully devised often using iterative ray tracing and channel propagation modeling and eventually set at stationary positions. Such deployment schematics are often used for close- and midrange network architectures such as mmWave backhaul/fronthaul as well as very close-range applications such as chip-to-chip (C2C) or board-to-board communication scenarios. An exemplary user scenario is illustrated in Fig. 6.

The beamwidth and directivity of the stationary mmWave antennas must be carefully designed to suppress unexpected misalignment between the transmitter and receiver. The sensitivity of the antenna misalignment for LOS communications becomes a greater issue as the communication range increases and the antennas become more directive. This becomes an important design consideration in particular due to the fact that majority of point-to-point mmWave communications are designed for outdoor environments where anything as subtle as wind, snow, birds, and any other expected environmental distractions can offset the antenna alignments by a few millimeters. The beamwidths of the mmWave antennas must be designed to withstand such degradations to preserve the quality of service (QoS). Some of the recent mmWave antenna designs for point-to-point communication circumnavigate this issue by employing precise mechanical beam steering. Electrical beam steering methods such as phased array are also being researched in recent years. However, interconnect losses between the antenna elements and the radio-frequency front end become increasingly problematic as the number of antenna elements increases.

Antenna Beamforming Applications Using Beam Steering

For situations where LOS environment propagation environments or point-to-point communication conditions cannot be guaranteed, it is essential to able to redirect the mmWave signals in order to search and reestablish a connection link within a very short time. Widely referred as to antenna beamforming technology, mmWave signals with narrow beamwidths generated by directive antennas are steered across the elevation and azimuth angle to transmit and receive radio signals using the most efficient propagation path. Antenna beamforming is highly effective for indoor environments where there tends to be limited wall penetrations, high absorptions, and various multipaths due to signal reflecting off furniture, ceiling, humans, etc., at mmWave spectrum. Compared to frequencies commonly used for present-day cellular and Wi-Fi applications (below 6 GHz), radio waves and mmWave frequencies rely on reflection when establishing links for NLOS propagation conditions as illustrated in Fig. 7. Antenna beamforming is also a

DOI 10.1007/978-981-4560-75-7_64-1

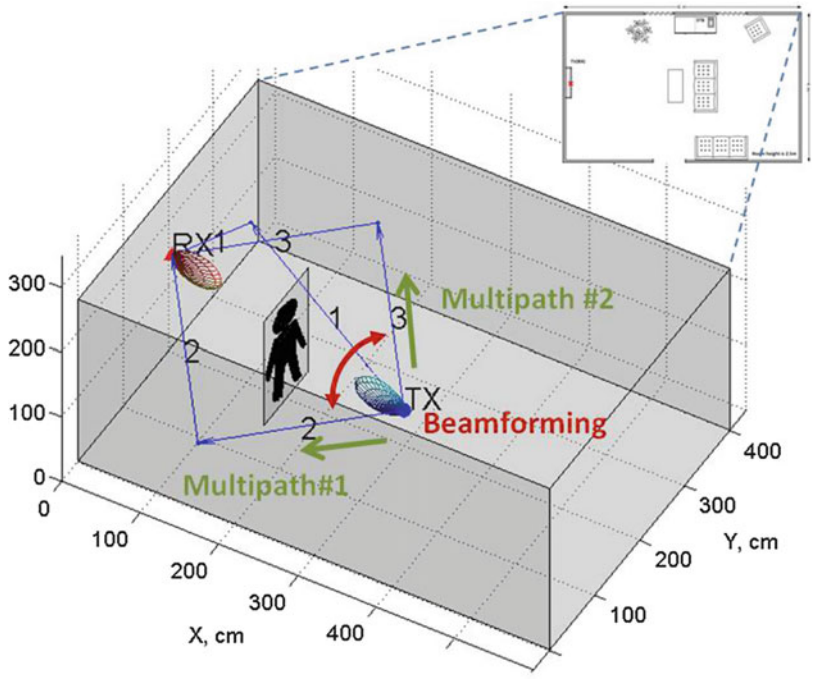


Fig. 7 Indoor mmWave beam steering scenario

critical and useful methodology for outdoor and indoor-to-outdoor applications. Obstacles such as pedestrians, vehicles, lightning posts, and foliage can instigate double knife-edge diffractions (Vogler 1982) causing "point of failures" in fronthaul, backhaul, and access networks. By spatially reusing the atmosphere, mmWave outdoor communication systems can not only circumvent signal blockage but also significantly enhance the network capacity and data rates.

Encouraged by the advancements of silicon technology and antenna packaging techniques, electrical beam steering has been increasingly favored over conventional mechanical beam steering at mmWave spectrum due to advantage in form factor, economic cost, and accuracy. Electrical beam steering can be broadly categorized into two techniques as follows:

Switched Beam Steering Switched beam steering techniques combine the radio wave signals in the array using a fixed set of weight vectors and time or phase delays. This is widely used for access points and cellular base stations where the direction of arrival (DoA) is well understood. Each antenna element must be designed to the radiation within a predefined "sector of coverage" and oftentimes utilize active RF components, such as switches, MEMS, combiners, etc., for real-time beam steering. Due to the small antenna aperture at mmWave frequencies, the overall antenna dimension employing switched beam steering can be advantageous for maximized far-field radiation coverage as illustrated in an example in Fig. 8. mmWave sub-array schematics can be employed to enhance the directivity targeted for each sector. Similar to point-to-point communication applications, the directivity of the mmWave antenna must be carefully designed to avoid unexpected nulls in the combined far-field radiation patterns. This is especially the case for military and space applications where design tolerances are usually very tight. However, for commercial mmWave applications, such as mmWave access points and sensors, the antenna pattern requirements are typically lower. Low cost, design complexity, ease of manufacturing, and form factor must be carefully considered. An antenna array with simplified corporate feeding networks (Pozar 1992), sub-array overlapping, or interleaved array arrangements (Abbaspour-Tamijani

DOI 10.1007/978-981-4560-75-7_64-1

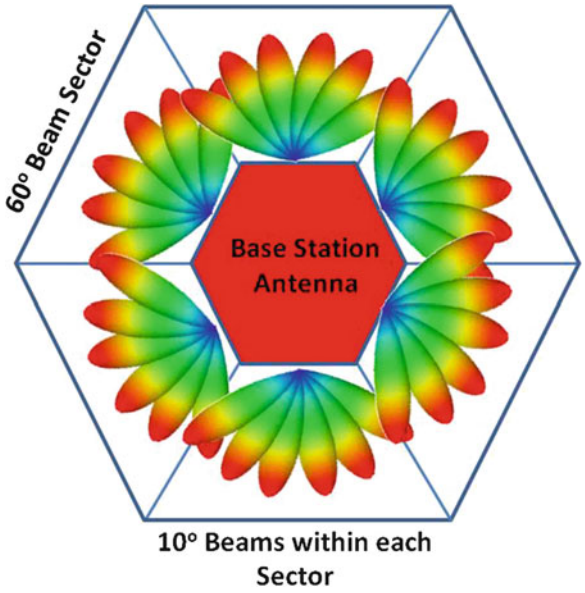


Fig. 8 Sectorized switched beam steering

and Sarabandi 2003) can be very instrumental in reducing the number of active RF components. For interleaved mmWave array, the mutual loading between adjacent antenna elements is dependent on the state of the antenna beam and methods to mitigate the mutual coupling effects must be taken into consideration. Furthermore, fixed beam steering techniques can be applied in conjunction with adaptive beam steering for mmWave applications requiring beyond long-distance coverage and wide beam steering range.

Adaptive Beam Steering Initially developed in the 1960s for military applications such as sonar and radar, adaptive beam steering was initially devised to mitigate the effect of signal jamming. The concept has continued to evolve over the decades and is now commonly defined as a technique to automatically adopt in real time based on predefined criterion to constructively combine the signals in a direction exhibiting the least amount of propagation attenuation.

A line array consisting of N number of identical point sources that are uniformly spaced as a function of d is represented in Fig. 9. Assuming that the array is receiving an incoming signal $S_n(t)$ at the strongest direction θ, the slanted angle θ introduces a delay in the time of arrival of the signal between the reference signal point source $P_0(t)$ and $P_m(t)$ denoted as τ where $\tau_n = n\tau$. τ is defined as

$$\tau = \frac{d \sin \theta}{c} \tag{9}$$

The received signals can be superimposed in a constructive manner in the direction of θ by subsequently adding true-time delays so that the signals arrive in a coherent manner. Then $S_n(t)$ can be expressed as follows in Eq. 10:

$$S_n(t) = S\left(t - \tau_n - \tau_{\text{shift},n}\right) \cos\left(\omega_{RF}\left(t - \tau_{\text{shift},n}\right)\right) \tag{10}$$

Majority of mmWave communication applications fit the model of narrowband conditions in which the frequency bandwidth is a fraction of the center frequency ω_{RF}. In such cases, the true-time delay can be approximated with a constant phase shift instead. Denoted as phased array (Natarajan 2007), Eq. 10 can

DOI 10.1007/978-981-4560-75-7_64-1

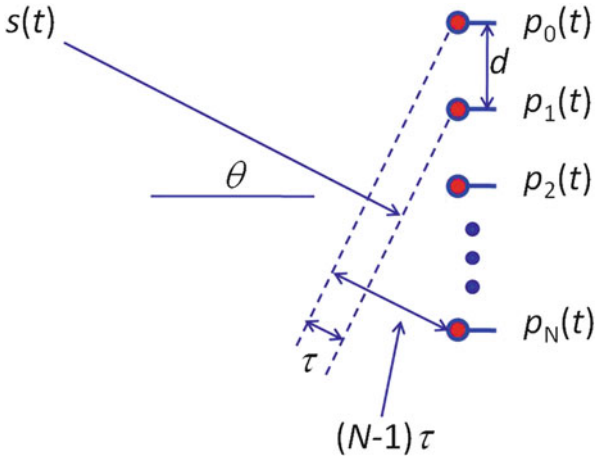


Fig. 9 An illustration of a uniformly spaced lined array

then be reorganized in the frequency domain and can replace the true-time delay with phase shifts φ_n at each point source. Then $S_n(t)$ can be expressed as follows in Eq. 11:

$$S_n(t) = S(t - \tau_n)\cos\left(\omega_{RF}(t - \tau_n) - \phi_n\right) \tag{11}$$

where φ is

$\phi_n = (N - n)\phi$

The total received signal of the phased array in the direction of θ can be expressed as the summation of $S_n(t)$ as follows in Eq. 12:

$$\begin{aligned} S_{\text{array}}(t) &= \sum_{n=0}^{N-1} S(t - \tau_n)\cos\left(\omega_{RF}(t - \tau_n) - \phi_n\right) \\ &= \sum_{n=0}^{N-1} S(t - n\tau)\cos\left(\omega_{RF}t - n\omega_{RF}\tau - (N - n)\phi\right) \end{aligned} \tag{12}$$

Assuming narrowband conditions,

$$S(t) \approx S(t - n\tau) \approx S(t - \tau) = S_0(t)$$

It is possible to rewrite S_{array} *(t)* as the following:

$$\left|S_{\text{array}}(t)\right| = \left|\frac{\sin\dfrac{N(\omega_{RF}\tau - \phi)}{2}}{\sin\dfrac{\omega_{RF}\tau - \phi}{2}} S_0(t)\right| \tag{13}$$

Equation 13 infers that for planar antenna arrays where $d > \lambda/2$, the far-field radiation pattern of the phased array will be subject to grating lobes. Phased-array antennas where $d << \lambda/2$ can potentially result in excessive mutual coupling and introduce nonlinear terms in Eq. 13. Hence, it is generally the rule of thumb to approximately configure $d = \lambda/2$. The 3 dB beamwidth is dependent on N and θ and be approximated as in Eq. 14. Based on the rules of antenna reciprocity, the beam steering behavior remains identical during transmission mode:

DOI 10.1007/978-981-4560-75-7_64-1

$$BW_{3\text{dB}} = \frac{0.87\lambda}{Nd\cos(\theta)} \tag{14}$$

The derivation of the phased-array antenna pattern demonstrates that each phase shifter located at respective point sources must be variable from 0 to 2π to scan the main beam of the antenna array across the hemisphere (i.e., $-\pi/2 \leq \theta \leq \pi/2$). For practical use cases, the scanning angle of the main beam is reduced due to the rise in side lobes and can be detrimental in distinguishing the most coherent propagation path for beamforming algorithms using the maximum power criterion. The effect of antenna beam steering range and the beamwidth on mmWave beamforming protocols can be understood in a qualitative manner.

Figure 10 exemplifies a set of beamforming protocols for the 60 GHz-based IEEE 802.11ad standard which is devised for uncompressed high-definition (HD) video streaming and high-speed data transfer. Understanding the specifics of the protocol is beyond the scope of this handbook, and the readers are referred to Jo et al. (2014) for details. The basic principle of concept can be described as below:

- To define a set of antenna steering directions θ for **a spherical sector coverage**
  - To cover the given spherical sector with minimum antenna gain
  - To minimize the number of antenna positions used
- Then, the generated set of the antenna positions is applied at both TX and RX sides, and *an iteration over all mutual TX and RX antenna positions* from this set is used to define a combination giving the maximum transmission power.

The radiation patterns during the quasi-omnidirectional antenna configuration and directional TX/RX antenna configuration are highly correlated to the efficiency of the adaptive beam steering algorithm illustrated in Fig. 10. Oftentimes, it is necessary to introduce variations in the antenna beamwidth between the sector level sweep (SLS) and the beam refinement protocol (BRP) frame. A main beam of an $N \times M$ phased-array antenna can be characterized by its respective beamwidth BW_{ev} and BW_{az} in the elevation and azimuth plan as shown in Fig. 11. During the adaptive beamforming optimization process, the SLS frame initiates link detection by selectively training the TX and RX antenna. Antenna array configurations with predefined BW_{ev} and BW_{az} are used in search of the best sector. The BRP frame uses an iterative procedure to "train" the RX and TX antenna in identifying the optimal propagation path in that certain sector and oftentimes employ identical BW_{ev} and BW_{az} used during the SLS . In situations where the received power levels fall beyond the dynamic range of the wireless link budget, the protocol reverts back to the SLS frame using a refined main beam exhibiting different BW_{ev} and BW_{az}. The dynamic scalability of BW_{ev} and BW_{az} is correlated to the efficiencies of the SLS and BRP frames.

The steering angle θ must be designed to cover the maximum physically allowable range while minimizing the beam steering iterations for a timely transmit and receive training field (TRN-T/TRN-R). The resolution of the antenna array beam steering is a function of antenna spacing d and the angular resolution (or the number of maximum allowable bits) of the corresponding phase shifters. Pencil-beam and fan-beam antenna patterns are commonly preferred for mmWave-based adaptive beam steering for this matter. The final decision will be made based on the specific application.

Key mmWave Antenna Design Techniques

This section discusses a number of antenna design parameters and ensuing techniques that are particularly critical at mmWave frequencies. For wireless communication applications, low-profile antenna topologies

DOI 10.1007/978-981-4560-75-7_64-1

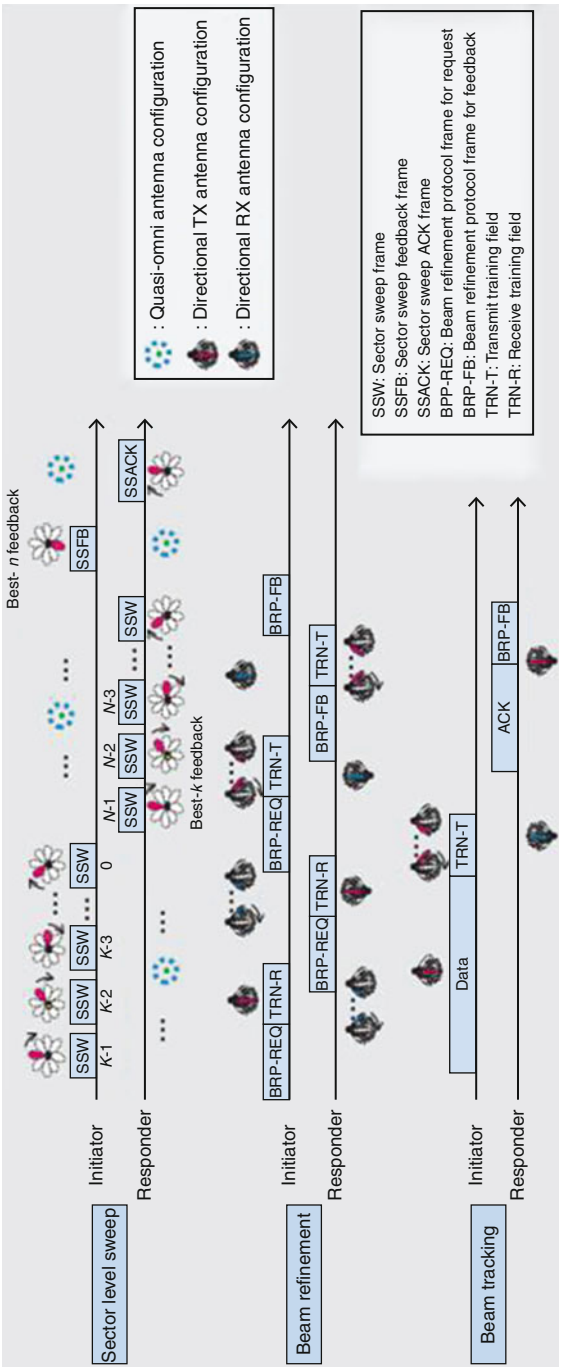


Fig. 10 A summarized illustration of the adaptive beamforming protocols used for WiGig/IEEE 802.11ad (Jo et al. 2014)

DOI 10.1007/978-981-4560-75-7_64-1

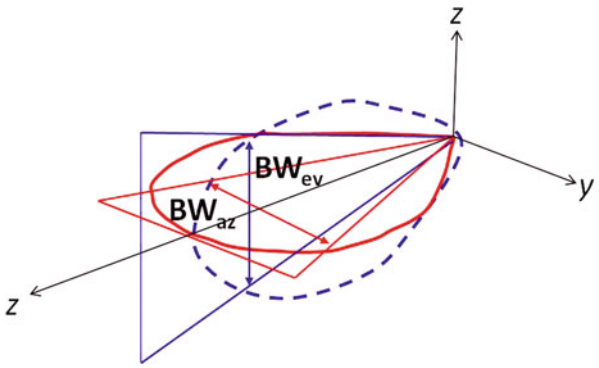


Fig. 11 Definition of the antenna beamwidth

Table 1 Summarized comparison of mmWave antenna types

Antenna type	Ease of fabrication	Real estate	Gain	Bandwidth
Planar and printed antenna	Feasible	Small	Moderate	Narrow
Dielectric resonator antenna	Moderate	Moderate	Moderate	Low
Substrate-integrated waveguide antenna	Moderate	Moderate	Moderate	Large
Reflector antenna	Difficult	Large	High	Moderate
Slot antenna	Feasible	Moderate	Low	Moderate
Lens antenna	Difficult	Large	High	Narrow

are preferred due to its ease of fabrication, assembly, and integration with the rest of the wireless system. Oftentimes, the reduction of the antenna topology is realized at the cost of important antenna characteristics such as directivity, gain, and bandwidth. Therefore it becomes imperative to clarify the advantages and disadvantages of each antenna designs when faced with specific allowable antenna real estate. Fabrication and assembly cost is another important design consideration in consumer electronics and automobile industry and this is closely correlated to the selection and lamination of the antenna material and type of the antenna. To assist in designing the appropriate type of antenna for a specific application, several mmWave antenna types that are frequently used are illustrated and summarized in Table 1.

The aforementioned antenna types can be used independently or in a combination in the form of multipurpose antenna arrays. The design procedure can be collectively characterized as listed below:

1. Identify the target frequency of operation and required antenna bandwidth.
2. Determine the antenna gain utilizing the wireless link budget for a target communication distance and coverage.
3. Estimate the optimal directivity, radiation pattern, and polarization based on the environment encompassing the antenna.
4. Assess the maximum allowable antenna real estate and determine the best placement within the wireless device.
5. Assess the antenna feeding mechanism and calculate the feedline loss.
6. Select the type of antenna to be used for the specific application.
7. Choose the antenna substrate material, methods of packaging, fabrication, and integration with the mmWave RFIC.
8. Devise methods of antenna performance evaluation and identify design parameters for occasional fine-tuning.

DOI 10.1007/978-981-4560-75-7_64-1

In-depth discussions on the design procedure will be presented in the following subsections.

mmWave Antenna Packaging and Integration

Antenna Packaging

Methods of integrating the antenna with the rest of the wireless module are one of the most distinguishable differences between mmWave RF front ends and those designed at microwave frequencies (typically below 10 GHz). Antennas for preexisting wireless communications such as 3G/4G or legacy Wi-Fi are regarded as an independent design block within the entire wireless architecture. The antenna, RFIC, digital IC, and the subsequent interconnects are separately designed based on predefined and oftentimes standardized hardware interfaces. The antenna and RFIC are typically connected via industry standard interconnects such as coaxial RF connectors, surface-mount (SMT) contact clips, printed circuit board (PCB) lines, and RF switches that are frequently used in conjunction with external microwave matching networks. The major contributors to the insertion loss between the antenna and RFIC are input reflection coefficient properties and the dielectric loss of interconnect and the matching networks. However, this composition becomes problematic when the target operating frequency is located at mmWave spectrum as the insertion loss is increased dramatically compared to the microwave spectrum. Emphasis on containing the insertion loss of the interconnects between the antenna port and the RFIC pad becomes one of the primary concerns for mmWave antenna design. Hence, the basic rule of thumb for mmWave antennas is to minimize the physical distance between the antenna port and the RFIC pad.

In general, there are two major philosophies for mmWave antenna packaging – the system on chip (SoC) and the system in package (SiP). SoC antenna packaging is realized in the form of on-chip antennas which are directly incorporated within multilayer metals using silicon technologies such as CMOS or III–V semiconductors. An on-chip antenna configuration using 130 μm CMOS technology is exemplified in Fig. 12. CMOS implementations are advantageous over III–V technologies for its scalability, higher levels of integration with the rest of the microwave integrated circuits (MMIC), and lower cost. On the other hand, III–V technologies are superior in terms of higher electron mobility, power handling capacity, and lower substrate loss. The low resistivity of the substrate used for on-chip antennas is the primary attribution to its poor antenna gain and high insertion loss across the antenna feeding networks. The resistivity is in the orders of 10 Ω-cm for high-volume CMOS technologies and approximately in the orders of 10^7–10^9 Ω-cm for III–V technologies. mmWave antennas designed using low-resistivity

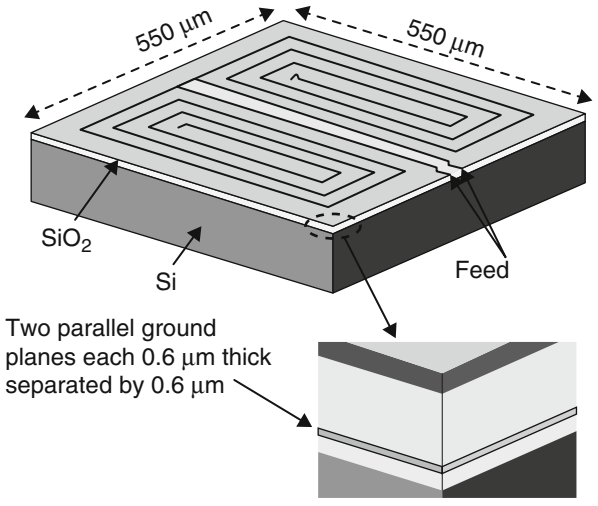


Fig. 12 Design example of an on-chip slot antenna (Behdad et al. 2007)

DOI 10.1007/978-981-4560-75-7_64-1

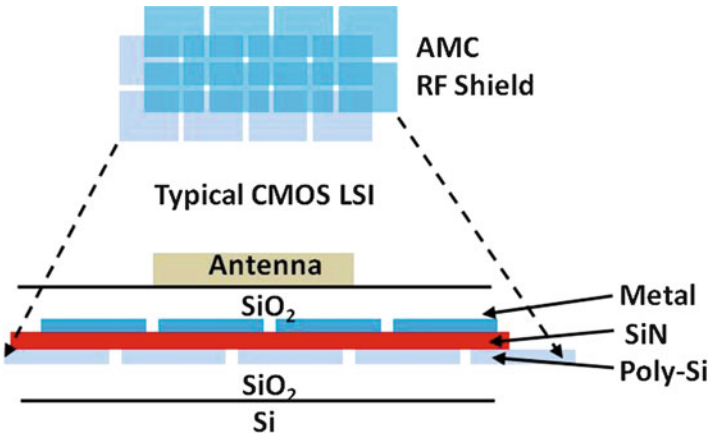


Fig. 13 SoC antenna implemented on an artificial magnetic surface (AMC)

substrates are subject to severe signal attenuation and this is exacerbated by the thin metallic layers incorporated using standard foundry processes. The thickness of the metallic layers is fractions of the skin depth at mmWave frequencies and the antenna topology is further penalized with substantial substrate losses. On-chip antennas employing high-resistivity silicon substrates are effective methods in reducing the signal attenuation constant. However, the high fabrication cost remains an obstacle for mass market applications such as consumer electronics. In recent years, periodic electromagnetic bandgaps in the form of metamaterial surface or artificial magnetic conductor (AMC) surfaces have been postulated for on-chip antennas as presented in Fig. 13 (Lin and Ooi 2009; Chu et al. 2009).

The main motivation behind SoC antenna packaging is to realize the entire RF front-end architecture in a completely monolithic fashion. As previously explained, the reduced physical distance between the mmWave antenna element and the rest of the MMIC does not always warrant improved antenna feedline losses. For applications where multiple number of antenna elements are required, SoC antenna may not be justifiable in the sense of engineering and economics. For example, in 60 GHz phased-array applications, the increased area incurred by the inclusion of multiple on-chip antenna elements within a silicon RFIC die will dramatically drive the unit price of the RFIC die upward. Attributed to these trade-off relations, SoC antenna packaging schematics are better suited for mmWave spectrum beyond 90 GHz or for very close-range mmWave wireless communication scenarios.

Another efficient method is to directly integrate the antenna with the rest of the wireless module. Defined as system in package (SiP), the packaging methodology is illustrated in Fig. 14 using multilayer lamination topologies.

The lateral integration schematic presented in Fig. 14a is suitable for applications where an end-fire radiation mode is advantageous. The mmWave antenna element is designed and placed in very close distance to the RFIC using microstrip, stripline, or coplanar waveguide (CPW)-based antenna feeding networks. This topology is especially compatible with dielectric resonators antennas (DRA), Yagi-Uda antennas, and horizontal dipole antenna types. Oftentimes, an additional transition such as a balun or a matching subnetwork is required to fine-tune the characteristic input impedance of the antenna port. High-permittivity substrates such as ceramic, high-end organic materials, and polymers are selected for the chosen antenna designs such as DRA and subsequently integrated with the surface of the mmWave carrier board using surface-mount technology. For planar antenna types such as Yagi-Uda antennas and horizontal dipole antennas, utilizing the top surface of the PCB carrier board is more advantageous in terms of design simplicity, cost, and overall aperture of the mmWave antenna module. Figure 15 shows a unique antenna packaging method – for mmWave frequencies where the profile of the resonant antenna is small

DOI 10.1007/978-981-4560-75-7_64-1

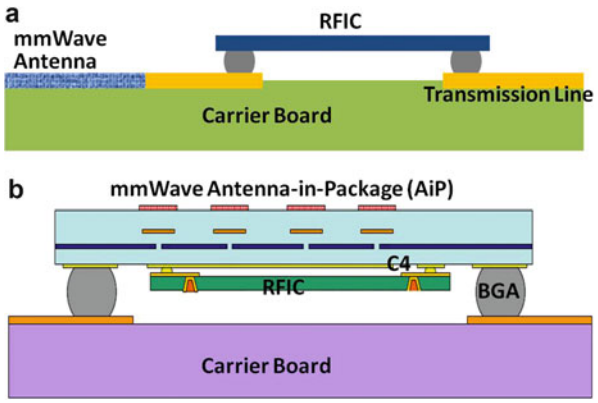


Fig. 14 SiP mmWave antenna integration schematic. (**a**) Lateral integration. (**b**) Vertical integration

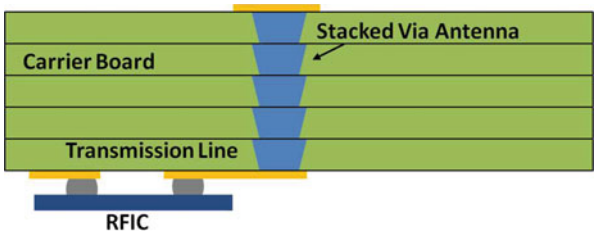


Fig. 15 The end-fire monopole antenna realized using stacked vias

enough to be fully embedded *within* the antenna carrier board. Using multilayer lamination technology, the antenna element is constructed using vertical interconnects such as stacked vias, through vias, and buried vias. The antenna element topology is equivalent to a vertically oriented monopole antenna and will radiate in the end-fire direction.

The effectiveness of the lateral integration schematic tends to degrade as the number of antenna elements increases. This is primarily attributed to the increased complexity and physical length of the antenna feeding networks as the antenna elements must be mutually separated at a predefined minimum pitch. Consequently, the lateral integration schematics are commonly used for mmWave wireless communication scenarios where the total number of RF paths and the antenna element is relatively small.

Figure 14b illustrates a vertical integration schematic. This type of antenna packaging configuration offers a wider range of antenna type selections compared to the lateral integration schematic. The antenna elements and the RFIC are typically located on opposite faces of the antenna module substrate for two distinctive reasons: (1) minimized interconnection insertion loss due to the employment of vias and (2) minimized antenna array footprint and antenna module dimension. Assembling the RFIC with the antenna module in an inverted manner is denoted as flip-chip bonding, also known as controlled collapse chip connection, or C4. The solder bumps are deposited on the top side of the RFIC during the final wafer processing process constructing RF ball pads. The RFIC is eventually inverted or "flipped" over so that the solder bumps come in direct contact with the signals pads of the antenna module. Subsequently solder is reflowed to construct the interconnect. At mmWave spectrum, flip-chip bonding tends to be preferred over conventional wire bonding attributed to superior rigidity and parasitic effects. The parasitic inductance of bond wires at mmWave frequencies is approximately 1 nH per 1 mm bond wire length. This amount of parasitic is sufficient to cause detrimental detuning of the antenna impedance matching. The limited controllability on the exact wire bonding procedure is another challenge when the RFIC die requires large number of bonding connections with the carrier board.

DOI 10.1007/978-981-4560-75-7_64-1

Planar antennas with broadside pattern properties such as patch antennas and slot antennas are particularly advantageous with the vertical integration schematic. The die size of mmWave RFICs is within the range of a few mm^2, while the mmWave antenna array topology tends to be much larger. Therefore, in addition to the vertical interconnects, lateral antenna feedlines such as microstrip, striplines, or CPW transmission lines are used to route the RF signals and align the ball pads of the RFIC to the corresponding antenna element ports. Likewise, the insertion loss attributed by the antenna feeding network is proportional to the number of RF paths and antenna elements.

Antenna Material

The substrate material for the mmWave antenna modules is a critical design parameter in terms of cost-effectiveness, design flexibility, and antenna performance. Advanced antenna materials such as low/high-temperature co-fired ceramics (LTCC/HTCC) and organic printed circuit boards represented by liquid crystal polymers (LCP) feature scalable permittivity and low loss tangent values at high RF frequencies. For such reasons, they are preferred over high-volume PCBs such as FR-4 for high-performance mmWave antenna designs. In addition, the fabrication and lamination processes using advanced antenna materials oftentimes exhibit more precise and sophisticated design rules such as smaller producible line widths and finer via pad diameters compared to FR-4. These advantages collectively grant mmWave antenna engineers enhanced degrees of freedom during the design stage. Ease of integration with various lumped components or ICs (die) is another important criterion in selecting among available antenna materials. The designed antenna aperture must feature minimum warpage to avoid damaging surface-mount lump components. Warpage can be quantitatively characterized using parameters such as the Young's modulus (Wang et al. 2003). Thermal stability and rigidity is another important design consideration. The excessive heat created by the MMIC circuit is typically harnessed or funneled using thermal vias or external heat sinks. In cases where the mmWave antenna modules are extremely small in dimension, the antenna material is exposed to higher heat temperatures and can result in unexpected mechanical deformation. Such thermal expansions of the antenna material can lead to damages or cracks in the solder balls between the antenna pads and the IC. Antenna materials with coefficient of thermal expansion (CTE) values that are closer to that of the IC lessen the probability of such incidents from occurring.

In terms of scales of economy, the material and fabrication cost of advanced antenna materials can frequently be several times higher than that of FR-4. The lamination can be simplified by reducing the number of layers or the total antenna module dimension for further cost savings.

mmWave Wireless Module Integration

mmWave wireless modules can be in large classified into three major components – the antenna module, RFIC, and the digital IC. The three components are ultimately integrated together on a carrier board which contains external peripheral components such as peripheral component interconnect express (PCIe) or universal serial bus (USB) interfaces. Extensive interconnections transporting numerous signals and reference in the form of RF, IF (intermediate frequency), reference clocks, IC power supply voltage, and reference signal ground coexist within wireless communication modules. The complexity of the interconnections and RF architecture must be taken into account in devising an appropriate method of integration. Three distinctive methods of mmWave wireless module integration are illustrated in Fig. 16.

The key differentiating features between the wireless module integration technique presented in Fig. 16a compared with Fig. 16b and c are the composition of the wireless module and the ensuing assembly procedure. This type of wireless module topology is preferred when the different material substrates are used for the antenna module and the carrier board, respectively. The antenna module entails a vertical integration schematic oftentimes using advanced multilayer laminations such as LTCC, HTCC,

DOI 10.1007/978-981-4560-75-7_64-1

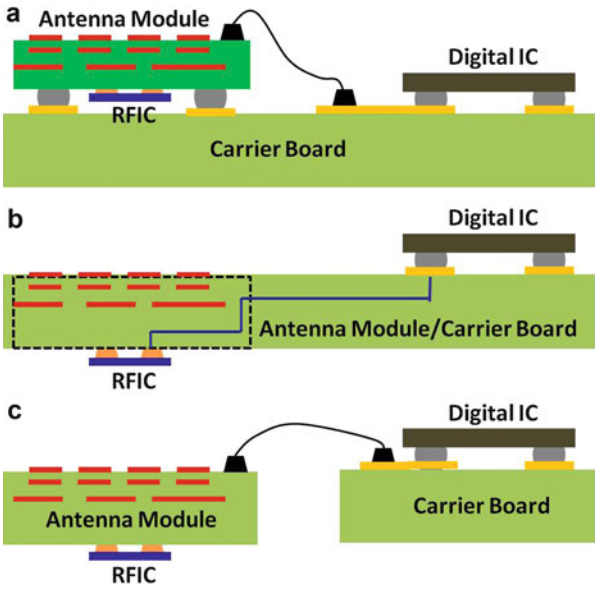


Fig. 16 mmWave wireless module integration techniques

or organic PCBs. The three major components identified as the antenna module, RFIC, and the carrier boards are subsequently assembled to realize an integrated wireless module using the following three-step assembly procedure:

1. Flip-chip attach the RFIC to the antenna module.
2. Surface mount the antenna module containing the RFIC with the carrier board using grid-array assembly technologies such as ball grid array (BGA) or land grid array (LGA).
3. Surface mount external peripherals such as connectors and lumped components on to the carrier board.

The finalized wireless module is connected with the digital IC either via a microwave transmission line or coaxial-type transmission line depending on the physical separation distance between the two. Within the wireless module, the IF signals, reference clock, power supply voltage, and the signal ground reference are interconnected from the digital IC via the antenna module and the carrier board. The flip-chip solder balls must be meticulously designed to minimize RF reflections and impedance mismatches. The electrical properties of the vertical interconnect represented by the flip-chip solder ball can be modeled using numerical electromagnetic computations based on computer-aided design (CAD) renderings. Figure 17 illustrates a flip-chip solder ball transition between the antenna module and the RFIC pad. The reactance of the transition tends to be inductive (several nH) and this can be compensated using passive matching networks such as shunt stubs which are attached to the antenna feedlines. The solder ball layout constructed on the bottom of the antenna module must be identically mirrored on the top surface of the RFIC die. For mmWave antenna modules, optimizing the solder ball layout is one of the most critical design stages out of the entire process. The following key design parameters are affected by the solder ball layout:

– The RFIC die size and front-end architecture
– Antenna material, manufacturer, and fabrication method
– Characteristic impedance of the vertical interconnect at the target operating frequency

DOI 10.1007/978-981-4560-75-7_64-1

Fig. 17 Layer view of the flip-chip solder ball transition

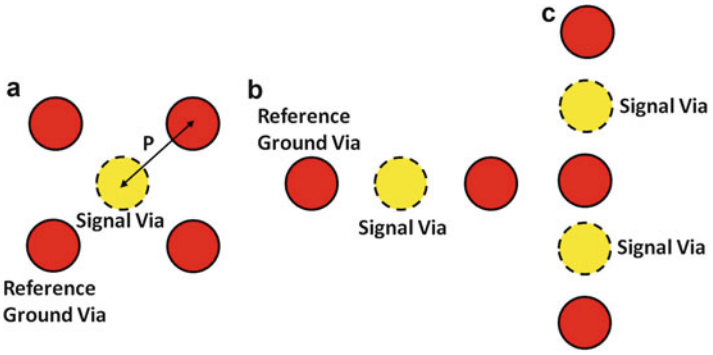


Fig. 18 Topologies of the solder ball layout. (**a**) Quasi-TEM mode. (**b**) CPW mode. (**c**) Multiple signal bump layout

Accurate characterization of the solder ball transition is extremely critical as the cascaded LNA, PA, and switches which will most likely not feature purely resistive 50 Ω input impedance. It is imperative to implement reactive matching networks on the solder ball layout. The topology of the solder ball layout must be devised so that each signal pad remains balanced and operate in the quasi-TEM mode. Each signal pad must be encompassed by a minimum of four reference ground pads as illustrated in Fig. 18a. Closely resembling a coaxial topology, the pitch *P* between the signal pad and the reference ground pad determines the characteristic impedance of the vertical interconnect as well as the permittivity of the solder balls, encapsulation, underfill, and antenna material. The minimum allowable pitch *P* is frequently limited by the fabrication accuracy and resolution of the PCB lamination technology involved. In situations where permissible real estate of the RFIC or the antenna module cannot afford four reference ground pads, the solder ball layout can be further simplified to a CPW configuration as illustrated in Fig. 18b. This rule of thumb can be further expanded for solder ball layouts containing multiple signal pads as shown in Fig. 18c. Failure to properly shield the signal pads with sufficient number of reference ground pads will pick up stray electric fields and induce unwanted noise in the antenna feeding network.

As discussed previously, the mechanical rigidity of the flip-chip solder balls are affected by the specific substrate material of the antenna module and heat generated by the RFIC. The flip-chip solder balls can be significantly enhanced mechanically, owing to recent advances in flip-chip underfill materials and processes.

mmWave antenna modules using the three-step configuration generally require several hundreds of grid-array assemblies to construct an electrical, mechanical, and thermal passage between the RFIC and the carrier board via the antenna module. Needless to say, any minor damaging of the grid arrays can potentially be related to a direct failure of the entire mmWave module. The importance and the sensitivity of the grid array result frequently require extensive levels of quality assurance inspections involving time-consuming and costly testing such as X-ray scanning. Compounded with the usage of advance antenna materials, the fabrication cost incurred by the ball grid arrays may exceed target price ranges for selected high-volume applications.

DOI 10.1007/978-981-4560-75-7_64-1

Figure 16b presents a wireless module integration technique which can be instrumental in resolving cost-related issues discussed in the three-step configuration. As opposed to a separate antenna module and carrier board configuration, the stated topology fully embeds and integrates the antenna module inside the carrier board. This approach eliminates the need for a discrete antenna module and the entire lamination can be realized monolithically. The overall assembly process of the wireless module is simplified compared to the three-step configuration as described in the following:

1. Flip-chip attach the RFIC to the monolithic antenna module/carrier board.
2. Surface mount external peripherals such as connectors and lumped components.

It should be noted that the specific sequence between the two assembly procedures can be interchanged depending the feasibility of applying the stencils for the flip-chip assembly process with or without surface-mount components. As it can be observed, the two-step wireless module configuration entirely eliminates the need for any grid-array assemblies. Consequently, the overall time and cost of the wireless module integration is noticeably reduced, translating to a more competitive unit cost for high-volume applications. Bypassing the grid-array assembly enables integration flexibility with the rest of the wireless communication hardware. This integration technique enables a multilayer mmWave antenna design to be realized within a wireless device hardware without any additional discrete material cost and assembly process. The whole motherboard and the wireless module can be integrated as one, simultaneously designed and fabricated monolithically. Figure 16b is extremely advantageous for scenarios where the characteristics of the selected mmWave antenna design, the antenna material, its specific user scenario, and the antenna placement are well established. One notable limitation comes with the need to implement the digital IC in close proximity to the antenna for super heterodyne wireless architectures. In addition, in situations where a faulty mmWave RFIC die or an error within the antenna module is identified, the entire carrier board must be replaced as opposed to just replacing the antenna module in Fig. 16a. Figure 16c mitigates this issue with a stand-alone antenna module.

The two-step wireless module configuration is valid under the precondition that the antenna module, carrier board, and occasionally the motherboard use identical or very similar substrate materials, laminations, and design rules. This can be rephrased to state that the antenna material is dependent on the material composition of the carrier board and the motherboard. Few exceptions withstanding, conventional high-volume PCB materials such as FR-4 are likely to be used in consumer electronics, telecommunications, and automobile industries. As the material loss tangent for conventional PCBs can be in the orders of 0.005 and higher, it is imperative to devise methods in reducing the insertion loss incurred by the antenna feeding networks within the antenna module/carrier board. Utilizing the CPW topologies as opposed to microstrip topologies and efficient antenna feedline routing and layout are effective methods.

Hybrid mmWave Antenna Packaging Technologies

It becomes possible to apply hybrid packaging and integration approach for mmWave frequencies that are high enough for resonant-type antenna elements to be deduced to extremely small form factors.

Figure 19 illustrates a mmWave wire bonding antenna topology. The inductive reactance and the capacitive reactance respectively derived from the wire bond and the signal pad on the PCB carrier board are utilized at a target resonant frequency. As the antenna geometry is quite fragile, additional encapsulation such as a polymer-based casing is required to protect the mmWave wire bond antenna from unexpected deformation. The mmWave wire bond antenna is instrumental especially when dealing with a limited number of end-fire radiation mode antenna elements typically at frequencies beyond 80 GHz. The design flexibility is relatively limited as the wire bonding process is subject to few variations.

DOI 10.1007/978-981-4560-75-7_64-1

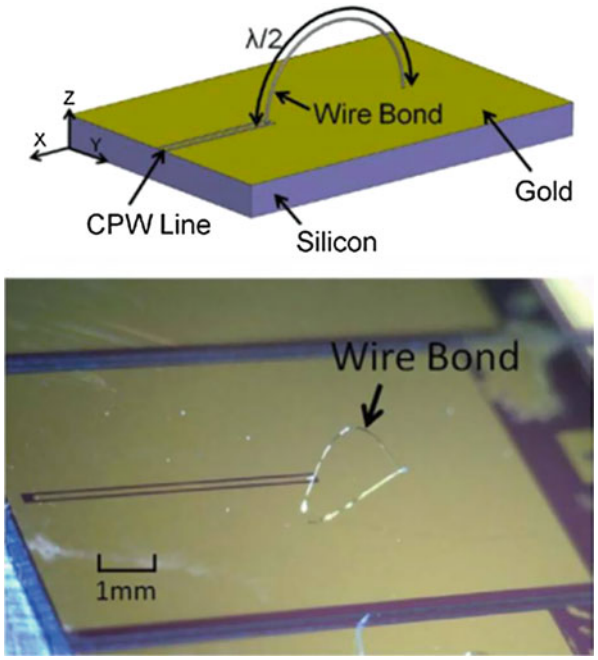


Fig. 19 Diagram of the mmWave wire bond antenna and its fabricated prototype (Willmot et al. 2009)

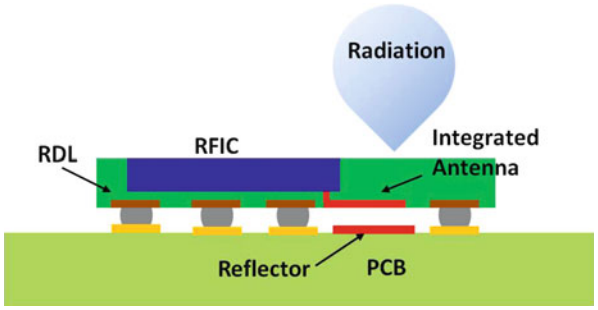


Fig. 20 Illustration of the mmWave eWLB antenna

Another notable hybrid packaging and integration technique is described as the embedded wafer-level ball grid array (eWLB). This approach is derived by introducing fan-out regions inside a conventional wafer-level package (WLP) as illustrated in Fig. 20. This is especially beneficial for multi-chip packaging applications such as high-power communications in the automobile and telecommunication arena. The concept of the eWLB antenna is to devise a planar antenna topology inside the fan-out region of the package by utilizing the redistribution layer (RDL) (Al Henawy and Schneider 2011). This type of packaging technique is best fit for broadside antenna modes. The degree of freedom becomes somewhat limited when designing antenna array geometries as the topology must be devised with the antenna elements encompassing the IC situated at the center. Nonetheless, the eWLB antennas are one of the most effective methodologies for close- and midrange mmWave communications beyond V-band frequencies.

Wideband mmWave Antenna Design

The acute disparity between the available spectrum below 10 GHz and the required frequency bandwidth for multi-Gbps wireless communication underlines the need of facilitating mmWave frequencies. In accordance to the classical Shannon-Hartley information theory, one of the most notable methods of

DOI 10.1007/978-981-4560-75-7_64-1

significantly enhancing the wireless capacity at mmWave frequencies lies in the expansion of the frequency bandwidth. Bandwidth enhancement techniques that are relatively well understood and frequently applied at microwave frequencies can also be incorporated in mmWave frequencies. Additional design considerations which are unique to mmWave applications such as the aforementioned packaging and integration design parameters must be further taken into account. The laminations and multilayer package topologies can be capitalized in devising a robust, efficient, and broadband antenna design topology. Single-layer mmWave antennas such as dipole antennas, Yagi-Uda antennas, and traveling wave antenna types can also be designed to be compatible with the rest of the antenna module package at the cost of increased lateral dimension.

Generally speaking, there are two ways of achieving broadband multilayer antennas at mmWave frequencies. The first approach involves the employment of low-permittivity antenna substrate materials. Antenna designs based on low-permittivity dielectrics feature lower Q which translates to a more usable antenna bandwidth. The second approach involves increasing the thickness of the dielectric (Katehi and Alexopoulos 1983). However, this tends to induce surface-wave excitations which are especially detrimental for planar antenna geometries. For TE and TM mode surface waves, the cutoff frequencies can be expressed as follows in Eq. 15 (Harrington 1961):

$$f_c = \frac{nc}{4h\left(\sqrt{\varepsilon_r} - 1\right)} \tag{15}$$

where c is the speed of light and n is an integer number for the corresponding TE and TM surface modes. Parameter h is defined as the height of the dielectric substrate of the antenna. More surface modes can exist as n increases. In addition, the increase in height of the dielectric substrate enables lower-order surface modes to couple easier. In majority of situations, the dimension of the mmWave antenna module can be regarded as a finite-size dielectric substrate. In such cases, the excitation of surface-wave modes causes undesired diffractions which are difficult to characterize. The edge diffractions eventually result in radiation nulls in the main lobe, rise of side lobes, and deteriorated gain and polarization. The effect of surface modes on the far-field behavior for mmWave antennas is analyzed in Liu et al. (2009). The trade-off related to increasing the thickness of the dielectric substrate for bandwidth enhancement can be circumvented to a certain extent by slowly varying the excitation across the designed mmWave antenna element. As a rule of thumb, antenna substrate profiles with $h < 0.01\lambda$ are recommended for planar antenna type topologies.

Selection of the optimum permittivity of the antenna material substrate and its height is succeeded by the designing of the mmWave antenna geometry. For rudimentary planar antennas such as patch antennas, achieving more than 10 % bandwidth may be technically challenging. Generating multiple numbers of closely spaced antenna mode resonances by carefully controlling the parasitic capacitance is an effective technique for multilayer laminations. Figure 21 illustrates variations of mmWave planar antenna schematics based on parasitic capacitance coupling methods. Figure 21a is described as the vertically stacked antenna geometry and is among the most effective methods for broadband applications requiring more than 15 % bandwidth. The ensuing design consists of multiple parallel RLC circuits, respectively represented by patch antennas. The combined RLC circuit is excited by series capacitance coupling which is predominantly dependent on the separation gap between the patch antennas. In reality, this separation gap is highly correlated to the specific antenna substrate material, the available substrate height, and the lamination and fabrication method. Therefore the fine-tuning range of the separation gap is typically discrete. Figure 21b illustrates a horizontally coupled antenna configuration. As the name indicates, the electric field generated by the parasitic capacitive coupling is contained in the horizontal plane. As a result, this type of antenna configuration is especially suitable for linearly polarized antennas

DOI 10.1007/978-981-4560-75-7_64-1

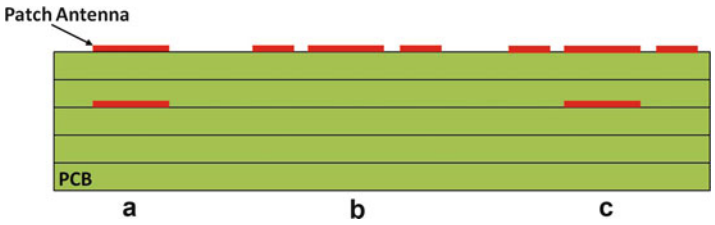


Fig. 21 Bandwidth enhancement techniques for mmWave multilayer antennas

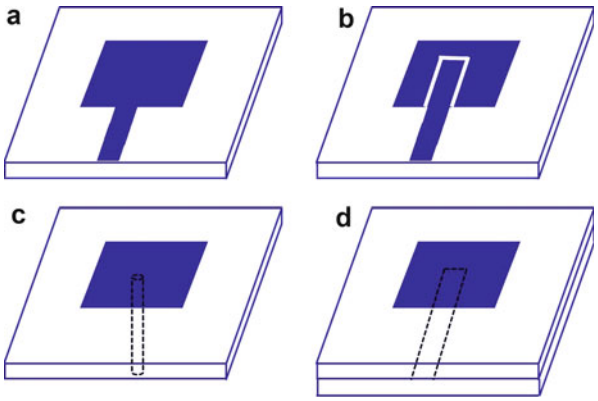


Fig. 22 Illustrations of planar antenna feeding techniques. (**a**) Microstrip edge feed. (**b**) CPW feed. (**c**) Probe feed. (**d**) Capacitive feed

having the E-plane located in the direction that is transverse to the horizontal separation gap. This type of bandwidth enhancement is useful for mmWave antennas requiring thin antenna substrates or limited layers of lamination. The drawback compared to the vertically stacked antenna geometry is the increased dimension in the horizontal plane. Additional antenna miniaturization techniques may be required for antenna array applications. The vertically stacked antenna and the horizontally coupled antenna technique can be used in conjunction as described in Fig. 21c. This configuration further enhances the antenna aperture efficiency and can be very effective for planar antenna designs at mmWave spectrum that require more than 20 % bandwidth. The trade-off is the increased complexity in the design which can lead to possible degradation of the fabrication yield. When applying conventional, high-volume PCB technology, the fabrication tolerance is oftentimes in the range of tens of microns, and this can result in detuning the parasitic capacitive coupling by more than 30 %. One way to stabilize the fabrication process and avoid costly fine-tuning the antenna design is to first design and fully characterize the vertically stacked antenna and the horizontally coupled antenna geometry respectively and separately. Consequently the two methods can be used in conjunction once the dominant design parameters that affect the antenna bandwidth are well analyzed.

The antenna feeding is another critical consideration for broadband mmWave antennas. It can be largely classified into uniplanar and multilayer antenna feeding techniques as categorized in Fig. 22. Figure 22a represents a microstrip edge feed method which is compatible for patch antennas and planar monopole antennas. The microstrip feed is characterized as a series inductive feed and this method is favored for its simplicity as both the antenna and feedline geometry can be fabricated in simultaneously. The drawback is the design inflexibility incurred by the limited number of available design parameters and the increasing far-field radiation contribution of the microstrip feed as the footprint of the mmWave antenna becomes smaller. Therefore it is important to limit the length of the microstrip feed at mmWave spectrum. Figure 22b complements the restrictions and the limitations of the microstrip feed by utilizing a

DOI 10.1007/978-981-4560-75-7_64-1

CPW feed technique. Similar to the Fig. 22a, the antenna feedline is situated on the same plane which contains the antenna. Compared to the microstrip feeding method, the CPW method offers more design flexibility and reduced erratic far-field radiation by the antenna feedlines. The uniplanar antenna techniques illustrated collectively suffer from drawbacks related to increase in the lateral real estate. For mmWave phased-array applications, the complexity of the antenna feeding networks is alleviated by devising sub-array topologies using corporate and serial feedings. In addition, the inductive nature of the antenna feedline typically limits the antenna bandwidth to less than 10 %. Figure 22c represents a probe feed method. For mmWave multilayer laminations, the probe is realized using signal through-hole vias. Quasi-TEM mode can be realized by encircling the signal via with multiple numbers of ground vias to closely emulate a coaxial vertical transition. This method benefits from negligible antenna feed radiation and extremely small profile. The input impedance matching of the antenna element to the probe is controlled by adjusting the contact location of the probe in the lateral direction. However, the bandwidth remains relatively narrowband. A capacitive antenna feeding method is illustrated in Fig. 22d. The antenna element is excited via electromagnetic coupling and feature enhanced design flexibility. Shielded microstrip or slot lines can be designed for the antenna feed transmission line. Using this method, it is possible to excite dual antenna resonance by adjusting the antenna feed geometry and is one useful way of enhancing the bandwidth. Figure 23 is denoted as the aperture-coupled feed (Pozar 1985). As one of the most favored and widely used mmWave antenna feeding techniques, the feeding structure consists of a transmission line and an aperture. The antenna element located at the top of the multilayer lamination is excited due to the electromagnetic coupling between the apertures. The dimension of the aperture can be carefully designed so that it contributes constructively in the main direction of the mmWave antenna element. Naturally this method is preferred for broadside antenna modes and is best fit for patch antenna types. The input impedance of the antenna can be matched by adjusting the location where the

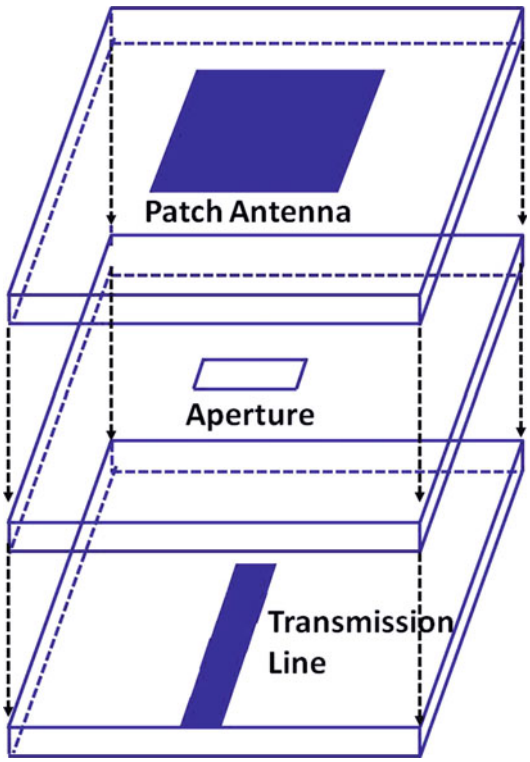


Fig. 23 Diagram of the aperture-coupled feed

DOI 10.1007/978-981-4560-75-7_64-1

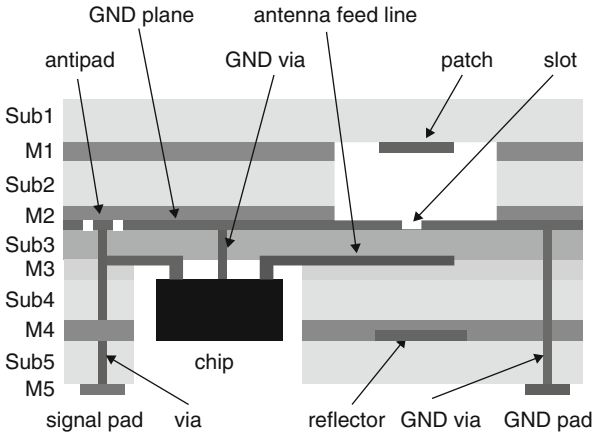


Fig. 24 Cross-sectional view of the integrated mmWave antenna using aperture-coupled patch antenna (Liu et al 2011)

transmission line and the aperture intersect on the horizontal plane. As multiple numbers of antenna resonances can be realized through the aperture and the antenna element, the aperture-coupled feed can be a very powerful technique in realizing broadband characteristics.

In order to better understand how the antenna design topologies and the feeding methods are used in practical scenarios, one of the most notable state-of-the-art broadband antennas designed for mmWave communications (Liu et al 2011) will be investigated. Figure 24 illustrates a cross-sectional view of an integrated mmWave antenna based on a five-layer PCB lamination. The antenna substrate material for the third layer denoted as Sub3 is chosen to feature dielectric permittivity which is approximately 12 % lower than the rest of the antenna substrate materials. This enables the antenna feedline designed on the 3rd layer (M3) of featuring an adequate width-to-height ratio for improved impedance matching. To further improve the antenna feedline efficiency, low-loss adhesives are applied on Sub3, whereas relatively lossy adhesives are applied elsewhere to reduce the overall cost and enhance manufacturability. As it can be observed, portions of Sub4 and Sub5 are removed to form an open cavity. The form factor of the open cavity is designed so that the RFIC can be completely inserted inside the mmWave antenna module. This configuration eliminates the need for vertical signal transitions using via holes and also eliminates requirement of associated matching networks. Signal pads and reference ground pads are implemented on the bottom layer M5 for ball-grid-array assembly to the carrier board.

The presented mmWave antenna renders an aperture-coupled feed technique to achieve broadband performance. The mmWave patch antenna element is designed and implemented on the M1 layer. A sealed air cavity is created by removing portions of Sub2. This essentially lowers the effective permittivity of the patch antenna resulting in lower Q. The ground plane located in M2 functions as the reference ground for the patch antenna and also is an effective way of shielding erratic radiation and signal cross talk within the antenna module. The aperture located on the M2 layer is designed based on a set of resonant, half-wavelength slots as shown in Fig. 25. By merging the two antenna resonance modes, more than 15 % 2:1 VSWR bandwidth can be achieved. The slots are designed to have slightly different resonant frequencies in comparison to the patch antenna. The exact position of the two slots are determined so that the surface waves generated by the slots and the patch antenna are destructively added. As a result, the antenna gain can be further improved. The slots are fed using a set of transmission lines situated in M3. The metallic pattern located in M4 functions as the reflector for the slots since a slot antenna is inherently bidirectional.

DOI 10.1007/978-981-4560-75-7_64-1

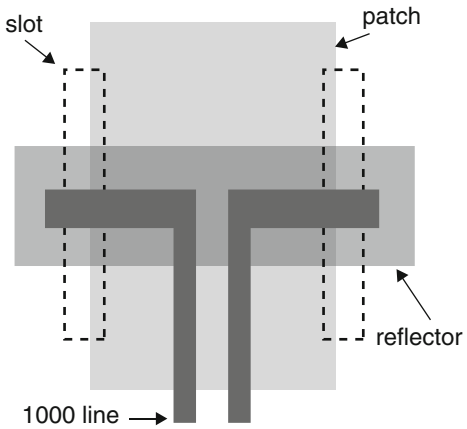


Fig. 25 Feeding structure used for the mmWave aperture-coupled patch antenna in (Liu et al 2011)

mmWave Antennas Design Technologies for Directional and Reconfigurable Radiation Patterns

The applicability of classical omnidirectional antennas is greatly limited to very close-range mmWave communication scenarios due to numerous propagation loss factors discussed in detail in section "Theory." In addition mmWave omnidirectional antennas are prone to receiving undesired incoming stray signals across a diverse range of angle of arrivals (AoA), which can significantly affect the noise figure of the receiver circuitry. This phenomenon, denoted as desense, can degrade the sensitivity of the wireless architecture by up to 10 dB or more under real-life propagation environments and is one of the primary concerns during the design phase. Spatial diversity techniques in the form of antenna beam steering are instrumental in circumventing the influence of interferers especially for multipath propagation environments. Given ideal antenna design conditions, a beam steering antenna such as in the form of a phased array featuring a complete spherical radiation pattern coverage and fine angular resolution would be the most effective solution. Unfortunately in most realistic antenna design conditions, such topologies cannot be realized due to constraints associated with dimension, economical, and technical costs. One practical approach involves enabling mmWave antenna elements featuring sectorized radiation patterns. This technique is particularly useful in propagation environments where the angles of arrival and departure can be estimated based on statistical spatial channel models. Directional mmWave antenna elements can be largely classified into broadside and end-fire topologies and will be discussed in detail as follows.

mmWave Broadside Antenna

For planar mmWave antennas exhibiting broadside antenna mode, the main radiation lobe is situated along the axis that is approximately perpendicular to the plane which contains the antenna geometry. Broadside antenna mode is instrumental when the transmitting and the receiving planar antenna topologies are configured to be facing one another. In another scenario, which is confined to indoor propagation environments, the two broadside antennas can be implemented adjacent to one another and establish a NLOS wireless link through utilizing the ceiling as an effective reflector. When multiple broadside antenna elements are employed to devise antenna arrays, the accuracy of the mmWave NLO antenna beam steering is affected by the relative surface roughness of the ceiling at the operating carrier frequency.

Conventional planar patch antennas are one of most widely used broadside antennas at mmWave. The antenna gain is typically in the range of 3–5 dBi for each patch antenna element under the precondition

DOI 10.1007/978-981-4560-75-7_64-1

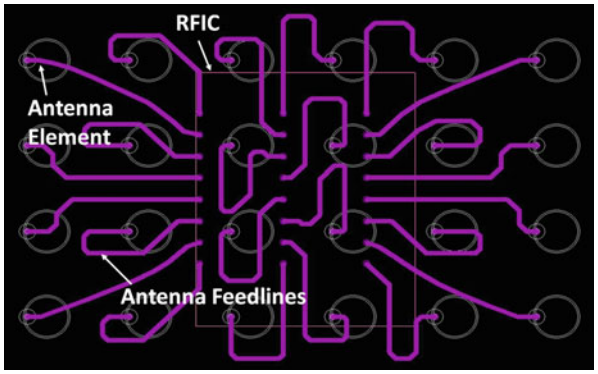


Fig. 26 Layout of a mmWave broadside antenna array and the ensuing antenna feedlines

that relatively low-loss antenna substrate materials are used. mmWave patch antennas can be further expanded into a two-dimensional array in order to attain greater antenna directivity and gain in the broadside direction. In accordance to the well-understood phased-array theory, the array factor is proportional to the logarithmic value of the total number of antenna elements. This relation infers that it is possible to continue enhancing the antenna gain and directivity in the broadside direction through the addition of antenna elements. However, in reality, the effectiveness of this method is eroded beyond a certain number of antenna elements. This is due to the growing disparity between the footprint of the mmWave RFIC die and the broadside antenna array. As illustrated in Fig. 26, the physical lengths of the antenna feedlines which route the mmWave signals between the RFIC signal pads and the respective antenna element ports are proportional to the total number of antenna elements comprising the broadside antenna array. The cumulative insertion loss incurred in the antenna feedline network continues to rise until a certain threshold where it equals the array factor as presented in Fig. 27. Beyond this threshold, the gain of the broadside antenna array starts to decrease as the insertion loss of the antenna array becomes greater than the antenna array factor. Increasing the mmWave RFIC die introduces complicated problems associated with extremely expensive unit costs and unexpected warpage as is deemed unrealistic. Due to this limitation, it becomes difficult to attain an antenna gain beyond 30 dBi using mmWave patch antenna arrays.

In situations where the antenna gain and the directivity need to be enhanced further, an additional dielectric lens can be implemented with the original antenna geometry. This design consists of a mmWave broadside patch antenna situated on the upper layers of the PCB lamination. Illustrated in Fig. 28, a dielectric lens featuring an elliptical or quasi-elliptical shape having a radius R is integrated with a cylindrical-shaped dielectric with an extension height denoted as h The curvature of the dielectric functions as a plano-convex lens, introducing true-time delays (TTD) to form a highly directional antenna radiation pattern. The variations of extension height h modify the elliptical shape of the lens, leading to different broadside directivity. Optimization of the integrated lens antenna is performed by modifying extension height h to control the antenna directivity and antenna array beam steering range. The loss tangent of the dielectric lens and the internal signal reflections in the dielectric lens body are the two primary factors that degrade the antenna gain. The broadside antenna element which is feeding the integrated lens is typically placed at the integrated lens focus to mitigate internal reflections. A fully fabricated mmWave integrated lens prototype and variations of dielectric lens are respectively presented in Fig. 29.

For mmWave access points, backhauls, and compact mobile terminals, the volumetric convex integrated lens solution can be potentially problematic. Designs based on spatial true-time-delay units employing bandpass filter technologies can be effective flat lens design techniques. Also denoted as

DOI 10.1007/978-981-4560-75-7_64-1

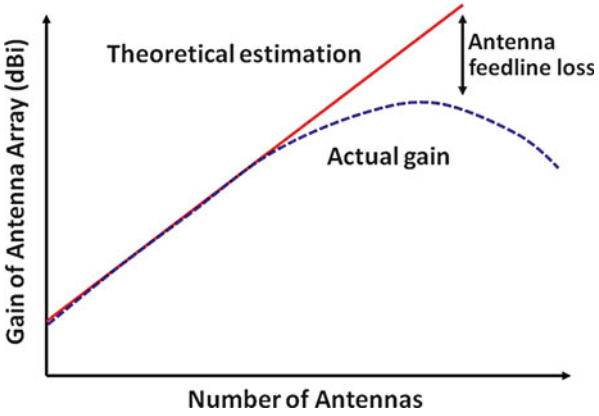


Fig. 27 Gain limitation of large-scaled mmWave broadside antenna arrays due to antenna feedline loss

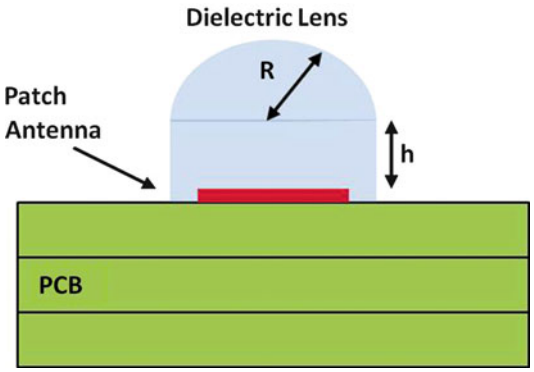


Fig. 28 Integrated lens antenna concept

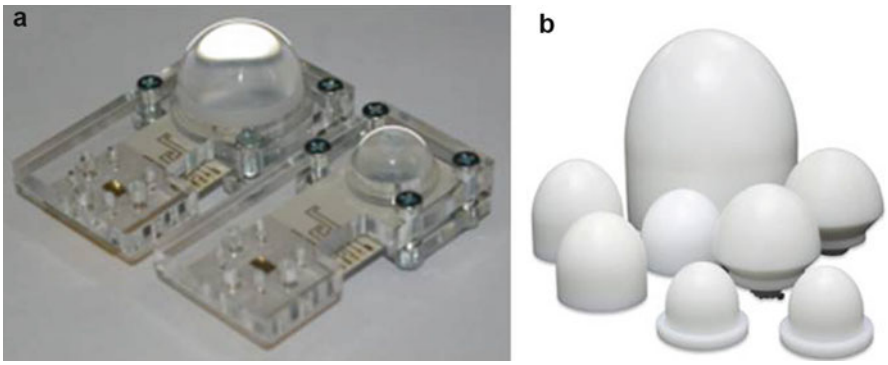


Fig. 29 (**a**) Fully integrated lens antenna (Artemenko et al. 2011). (**b**) Dielectric lens variation (Designed by Radio Gigabit Inc ©)

mmWave transmit array, the topology comprises of single or multilayer quasiperiodic resonant elements arranged to locally compensate the phase delay associated with different path lengths. An illustrative concept depicting the transmit array is presented in Fig. 30. The resonant elements constituting the transmit array layers rephrase the incoming spherical waves and retransmit the mmWave signals in the form of plane waves. The required phase adjustments are dependent on the distance the signal has traveled from the broadside antenna to each of the resonant elements. R_i denoted in Fig. 30 is defined as the vector of *i*th element from the phase center of the broadside antenna. r_i is denoted as the position vector to the *i*th

DOI 10.1007/978-981-4560-75-7_64-1

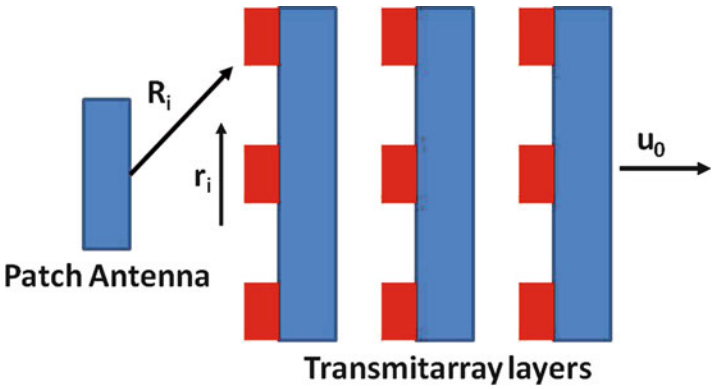


Fig. 30 Transmit array antenna (Ryan et al. 2010)

resonant element from the center of the transmit array lens and u_o is the direction of the transmitted main beam of the antenna. k is denoted as the propagation constant. Defined in Black and Wiltse (1987), Hristov and Herben (1995), and Ryan et al. (2010), the required phase compensation ϕ_i at each resonant element can be expressed as Eq. 16 below:

$$\phi_i + k(R_i + r_i \cdot u_o) = 2\pi n \quad \text{where n} = 0, 1, 2, 3 \ldots \tag{16}$$

Transmit array lenses offer the benefits of reduced profile and wide beam steering scanning ranges. However, due to the RLC circuitry of the resonant elements, they are commonly narrowband and may require extensive multilayer fabrication procedures. The gain enhancement effect is highly sensitive to the physical alignment of the transmit array lens and the broadside antenna, so additional efforts focused on eliminating mechanical jitters and vibrations may be required.

mmWave End-Fire Antenna

End-fire antenna modes are effective solutions for mmWave device-to-device (D2) communications such as file transfers or ultrahigh definition (UHD) seamless video streaming scenarios. The axis of the main lobe is parallel to the plane containing the planar mmWave antenna geometry. The mmWave Vivaldi antennas, (Yngvesson et al. 1989) which are classified as traveling wave antennas, are advantageous when the number of available layers within a mmWave antenna module is limited. This type of antenna geometry is most compatible with the lateral integration schematic described in the previous subsection and widely used when the mmWave RFIC in implemented on the same plane of the carrier board. The antenna gain is proportional to its length which is usually a multiple of the wavelength. For resonant-type end-fire antenna designs, the mmWave Yagi-Uda and the planar dipole antennas respectively illustrated in Fig. 31a and b are by far the most preferred topologies for low-profile applications (Alhalabi and Rebeiz 2009). The mmWave Yagi-Uda antenna largely consists of three discrete sub-elements. The driven element consists of a planar dipole antenna topology which employs a half-wavelength balun on one of the two branches. The balun enables a 180° phase shift and drives the dipole into the desired antenna mode. The driver element electromagnetically couples to the driven element and enhances the directivity in the boresight angle of the Yagi-Uda. The reflector suppresses the back radiation of the driven element which is induced by the inherent nature of bidirectional planar dipoles and redirects the radiating field in the boresight direction. This design is particularly favored for its relatively wide bandwidth and monolithic fabrication. The planar dipole antenna topology illustrated in Fig. 31b is typically applied in applications where an end-fire antenna mode featuring low height profile and wide beamwidth in the elevation plan ("fan-beam patterns" are preferred. Conventional planar dipole antennas feature antenna

DOI 10.1007/978-981-4560-75-7_64-1

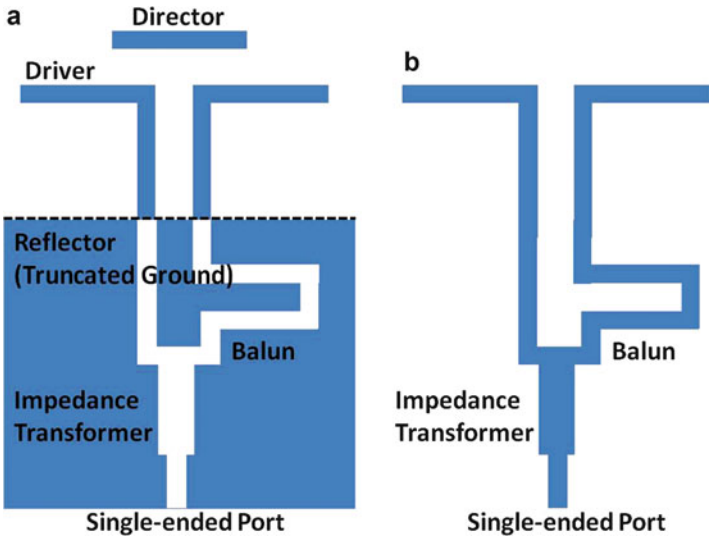


Fig. 31 (**a**) Yagi-Uda antenna. (**b**) Planar dipole antenna

gain values of less than 2 dBi which in turn results in relatively low EIRP for mmWave communications. The EIRP can be doubled when the mmWave planar dipole antenna is used in conjunction with a multi-RF chain phased-array transceiver IC. The half-wavelength balun is completely eliminated and each branch of the planar dipole antenna is respectively connected to each RFIC signal port. Assuming sufficient angular resolution is available by the RF phase shifter, a 180° phase difference between the two RF signals can be fed to each branch. The phase difference drives the planar dipole antenna topology into an antenna mode with a 3 dB enhancement in the radiated power.

Despite the numerous benefits, Yagi-Uda and planar dipole antennas are often subject to performance degradations in real-life mmWave applications. One of the most notable issues can be attributed various metallic layers which may be implemented underneath the antenna module. For compact mobile terminals such as smartphones and wearable devices, the footprint of the wireless module is extremely limited and the ensuing circuitries are densely spaced. The end-fire mmWave antenna will have to be placed in very close proximity to other signal tracelines and metallic layers for reference ground and heat sink. Yagi-Uda and planar dipole antennas inherently require the removal of metallic layers beneath the antenna topology in order to operate in the end-fire radiation mode. In the presence of a metallic ground underneath the mmWave Yagi-Uda and planar dipole antenna, the far-field radiation patterns may experience severe distortions as illustrated in Fig. 32. Methods of mitigating this issue will be discussed in detail in section "Notable mmWave Antenna Applications."

Pattern-Reconfigurable mmWave Antennas

With the advent of durable, efficient active components such as phase shifters, switches, PIN diodes, and microelectromechanical systems (MEMS), reconfigurable antennas are increasingly becoming a practical solution for mmWave communications. In Yang et al. (2005), the patch antenna is composed of two orthogonal quasi-Yagi-Uda antennas printed on the antenna substrate material. Electromagnetic bandgap (EBG) structures are controlled and modified depending on the state of the MEMS switches.

Another state-of-the-art pattern-reconfigurable technique proposed in Marnat et al. (2013) controls the antenna pattern by reorienting the antenna geometry as illustrated in Fig. 33. A broadside antenna mode is enabled when the mmWave bow-tie antenna is placed in parallel to the antenna substrate material. By controlling the buckled cantilever plate (BCP), the designed antenna can be mechanically rotated by 90° to transform into an end-fire antenna mode. For low-latency Gbps communications, the relatively slow

DOI 10.1007/978-981-4560-75-7_64-1

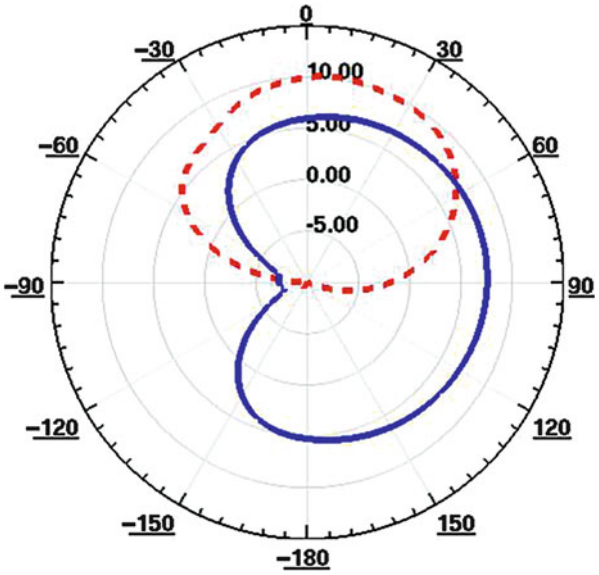


Fig. 32 End-fire pattern distortion when implementing Yagi-Uda and planar dipole antennas above a metallic surface. *Solid line*: elevation pattern in free space. *Dotted lines*: elevation pattern with the inclusion of the metallic surface

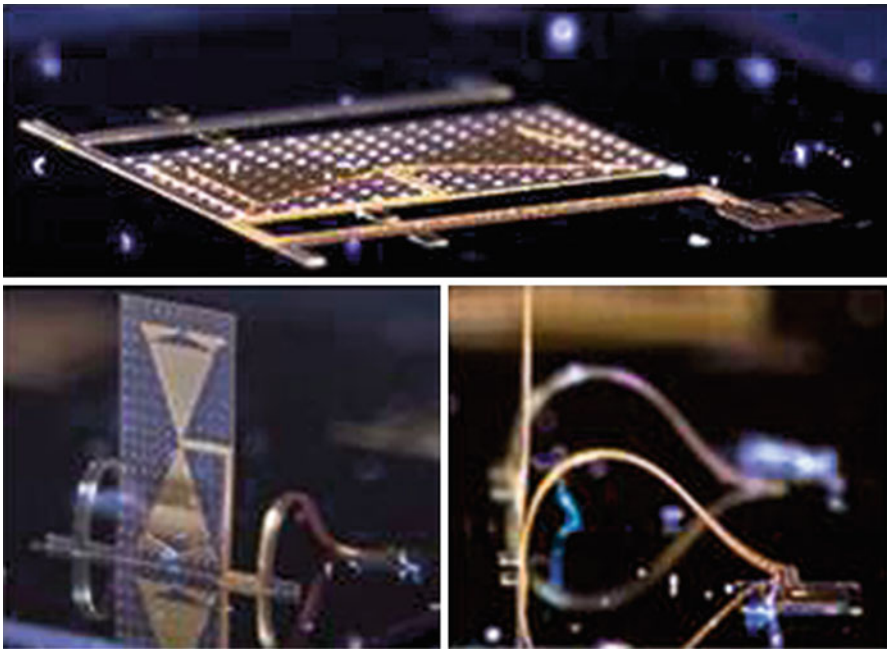


Fig. 33 Photographs of the pattern-reconfigurable mmWave bow-tie antenna (Marnat et al. 2013)

switching speed needs to be further researched as well as addressing the fragility of the entire topology. For high-volume applications, it is imperative for pattern-reconfigurable antennas to feature a small real estate and efficient packaging schematics. Consequently, future active components required for mmWave pattern-reconfigurable antennas will most likely need to be integrated inside the RFIC as opposed to majority of the present-day solutions which are based on discrete, off-chip active components. It is imperative that the mmWave antenna topology must be conceived concurrently with the entire RF front-end architecture.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_64-1

Notable mmWave Antenna Applications

This section will present two mmWave antenna designs in detail which are respectively applied to ongoing progress of Gbps mmWave communication applications in the global wireless and consumer electronics industry. Many of the antenna designs introduced in this section have either been incorporated or are projected to be applied to present and upcoming wireless mobile terminals. The readers are recommended to refer to Hong et al. (2012, 2013, 2014b) for further detail.

Phased-Array Antenna for 60GHz WiGig/IEEE 802.11ad

Introduction and Background

The explosive proliferation of wireless local area networks (WLAN) and high-quality multimedia data demands Gbps wireless data communication with extremely latencies less than 1 ums. Among numerous IEEE wireless standards, the WiGig/IEEE 802.11ad specializes in ultrafast data transfer and seamless streaming of audio and video files at close-range communication scenarios using the unlicensed 60 GHz frequency band.

The concept and notion of utilizing 60 GHz has been perceived more than two decades ago among research institutions and universities and have been followed with rigorous proof-of-concept (PoC) research and investigations. Over the years, 60 GHz antennas have continued to advance at an unprecedented pace. Consequently, as antenna element technology begins to mature, the main focus of research and development efforts has shifted to the realization of low-profile, high-performance 60 GHz antenna modules. Articulating the balance between performance and cost competitiveness is increasingly crucial.

Antenna Module Configuration

To mitigate the high cost associated with advanced antenna materials and ball-grid-array (BGA) assemblies, the state-of-the-art 60 GHz antenna design employs a two-step wireless module integration schematic introduced in section "Key mmWave Antenna Design Techniques." Figure 34 illustrates the lamination configuration. The 60-GHz array antenna package is wholly designed using HDI FR-4 PCB substrates. Consequently, the antenna module is homogeneously embedded within the 12 layer FR-4 carrier board, enabling a simplified, monolithic fabrication and assembly. The 60-GHz RFIC is flip-chip attached to the bottom layer M12 of the antenna module while lumped elements are implemented through SMT. The sequence between the two assembly processes is interchangeable. The lamination of the

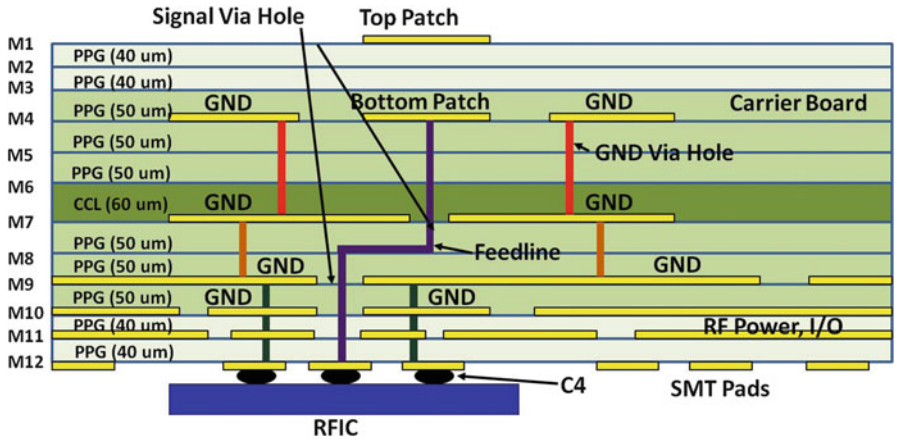


Fig. 34 Lamination configuration of the 60 GHz AiP (Hong et al. 2013)

DOI 10.1007/978-981-4560-75-7_64-1

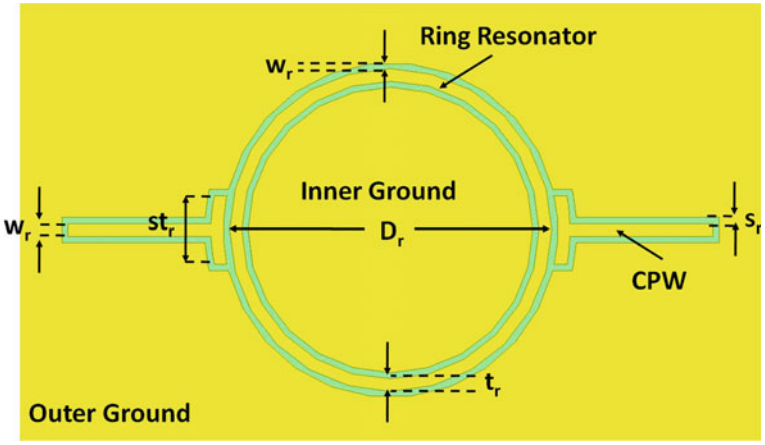


Fig. 35 Topology of the microstrip ring resonator (Hong et al. 2013)

antenna module consists of one substrate core and the prepregs. The substrate core is situated at the center of the lamination. Five prepregs are adhesively bonded together using 0.254 mm thick epoxy resins and stacked above the substrate core. Similarly, another five prepregs are adhesively bonded underneath the substrate core. Metal layers from M2 to M11 are implemented using 20 m thick copper. Metal layers M1 and M12 are each composed of 25 m thick Au plating and finished with 20 m thick solder resist. Laser and mechanical micromachining processes are utilized for the construction of through, buried, and stacked via holes for each substrate prior to lamination.

It is rarity for PCB vendors to characterize the electrical properties of the antenna substrate material beyond 10 GHz. The inherent frequency-dependent nature of pertinent material properties such as permittivity and dielectric loss tangent must be well defined prior to the 60 GHz antenna design process. The deviations in the electrical characteristics between the core substrate and prepregs are reported by the FR-4 supplier to be well less than 1 % up to 10 GHz. Therefore for brevity, the FR-4 PCB substrate is modeled as a homogeneous, isotropic material up to 60 GHz in this subsection. A number of weakly coupled, coplanar waveguide (CPW)-fed microstrip ring resonators are designed to extract the permittivity of the FR-4 substrate. The top view of the ring resonator layout situated on M1 is visualized in Fig. 35. Design parameters are optimized to be $w_r = 0.2$ mm, $s_r = 0.1$, $t_r = 0.2$ mm, $s_{tr} = 1.1$ mm, and $D_r = 5.18$ mm. The bottom ground is implemented on M4. One such measured and simulated S21 amplitudes are presented in Fig. 36. Measurement results indicate the calculated average permittivity $\varepsilon_r = 3.92 \pm 0.13$ among three identical ring resonator samples at 58.3 GHz. Calculation of the loss tangent at 60 GHz is difficult due to limitations in accurate extraction of conductor and radiation loss. As a result, the simulated loss tangent is iteratively matched to the measured results of various transmission lines. The loss tangent is confirmed to be $tan\delta = 0.027 \pm 0.07$ at 60 GHz.

The antenna module is largely classified into three distinct portions based on functionality. The antennas are allocated in the upper portion of the module to avoid interference with the RFIC, occupying layers from M1 to M7. The antenna feedlines which route the 60-GHz signal paths from each of the RFIC ports to the corresponding antenna elements are situated in the mid-portion of the antenna module, ranging from M7 to M9. Lastly, power distributed networks (PDN), control lines, low-speed and high-speed interfaces for the RFIC, and digital blocks are located on the bottom portion of the antenna module occupying layers at M10 and M11. The antenna ports which are each electrically connected to corresponding RFIC ports through flip-chip assembly are designed on M12. M7 and M9 are designated as RF grounds to maximally isolate the antennas and antenna feedlines from unexpected spurious modes and cross talks that may occur in the bottom portion of the antenna module after integration with the RFIC.

DOI 10.1007/978-981-4560-75-7_64-1

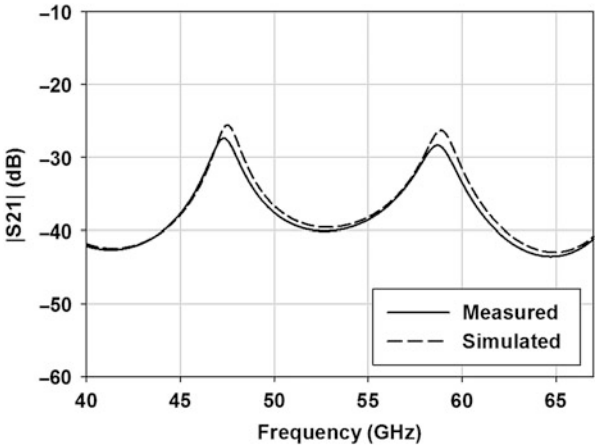


Fig. 36 Measured and simulated S-parameter amplitudes of the microstrip ring resonator (Hong et al. 2013)

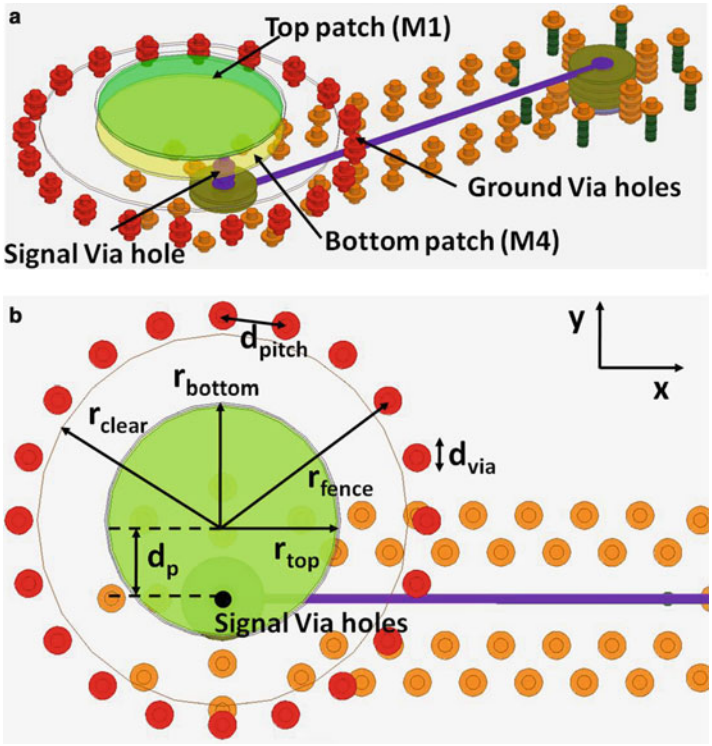


Fig. 37 Illustration of the antenna element. (**a**) 3D view. (**b**) Top view (Hong et al. 2013)

60 GHz Antenna Element Design

A circular stacked patch antenna model discussed in section "Key mmWave Antenna Design Techniques" is devised to achieve the required −10 dB S11 bandwidth from 57 to 66 GHz and broadside radiation characteristic while keeping the lateral footprint to be minimal. Figure 37 illustrates the detailed topology. Two separate circular patches are designed and implemented on M1 and M4 and denoted as the top and bottom patch, respectively. The top and bottom patches are vertically aligned in a way that their center coordinates are identical in the xy plane. A probe-fed mechanism is realized for the bottom patch by connecting a signal via hole at a certain distance d_p from the center of the bottom patch. As a result, a first

resonance is triggered by the bottom patch. The top patch is implemented above the bottom patch with the distance that equals the combined thickness prepreg substrates between M1 and M4. The two patches are not physically in contact with one another. The configuration produces a secondary parasitic resonance that is primarily controlled by the top patch antenna dimension. In comparison to advanced antenna materials such as LTCC or LCP, FR-4 PCB is relatively more prone to substrate thickness variations during the lamination process due to its inferior mechanical and thermal properties such as Young's modulus and CTE (coefficient of thermal expansion). Any unexpected variations and especially reduction in the combined thickness of the prepreg substrates during the fabrication process will have a direct effect on the coupling capacitance between the top and bottom patches. Therefore a parametric study regarding the location of the signal via hole (d_p) and the radii of the top and bottom patches r_{top} and r_{bottom} is performed using numerical simulators to find the optimal design parameter combination for situations where the combined prepreg substrates change within the range of ± 30 um. Minimum bandwidth buffers are incorporated into the target design specification in preparation of any frequency deviations caused by various possible scenarios such as fabrication error, measurement discrepancies, and numerical errors. The finalized design parameters of the circular stacked patch antenna are $d_p = 0.42$ mm, $r_{\text{top}} = 0.61$ mm, and $r_{\text{bottom}} = 0.64$ mm.

An approach involving the utilization of fence of ground via holes surrounding the antenna element introduced is adapted in this design due to its relatively small profile and ease of implementation. First, a clearance with a radius $r_{\text{clear}} = 1.03$ mm which encompasses the bottom patch is designed. The region outside the clearance in M4 is metallized to function as an RF ground. The dimension of r_{clear} is found to have limited effect on the reflection coefficient of the antenna element under conditions that remains greater than 0.95 mm. Afterward, fences of ground via holes penetrating from M4 to M7 are designed to surround the antenna element with a circular radius r_{fence} of 1.15 mm and via hole pitch $d_{\text{pitch}} = 0.344$ mm. The diameter of the ground via holes in this subsection is set to $d_{\text{via}} = 75$ um throughout this subsection. Figure 38 presents the simulated electric fields and the corresponding three-dimensional radiation patterns of the antenna element as a function of the presence or absence of the designed fence of ground via holes. In the presence of implementing fence of ground via holes, it can be observed that the fringing fields on the edges of the FR-4 substrate are relatively reduced in magnitude. The effect of the fence of ground via holes is further ascertained by the reduction levels of the ripples in the simulated three-dimensional radiation patterns throughout the frequency of interest.

The relatively high material loss tangent of the FR-4 substrate at mmWave frequencies has been the primary restricting factor for widespread usage at 60 GHz frequency band. Excessive levels of signal loss incurred by the antenna feedline can potentially have an adverse effect on the overall system link budget, resulting in a more stringent design requirement for the RFIC. Therefore it is imperative to first closely examine the impedance matching and insertion loss properties of the antenna feedline structure prior to integration with the antenna element.

A stand-alone topology of the antenna feedline is first designed for verification purpose as illustrated in Fig. 39. The antenna feedline consists of two vertical coaxial transitions that are each connected to the corresponding ends of a stripline structure located in M8. The impedance of the stripline is designed to 50 Ω by adjusting the width $W_{st} = 50$ um. Fences of ground via holes are implemented along the edges of the stripline with via hole-to-stripline distance $d_{\text{strip}} = 150$ um and minimum via hole-to-via hole pitch $d_{\text{pitch1}} \leq 310$ um. Careful modeling of the fence of ground via holes surrounding the stripline is determined to be an extremely critical factor in preventing two commonly occurring degradations: (1) anomalous resonances ("suck out") and (2) mutual coupling and signal cross talk. The two vertical coaxial transitions are identically designed using one 75 um diameter signal via hole and a series of ground via holes situated around the signal via hole with $d_{\text{pitch2}} = 452$ um, functioning as a coaxial outer conductor. This configuration translates to a calculated input impedance of less than 26.5 Ω. The

DOI 10.1007/978-981-4560-75-7_64-1

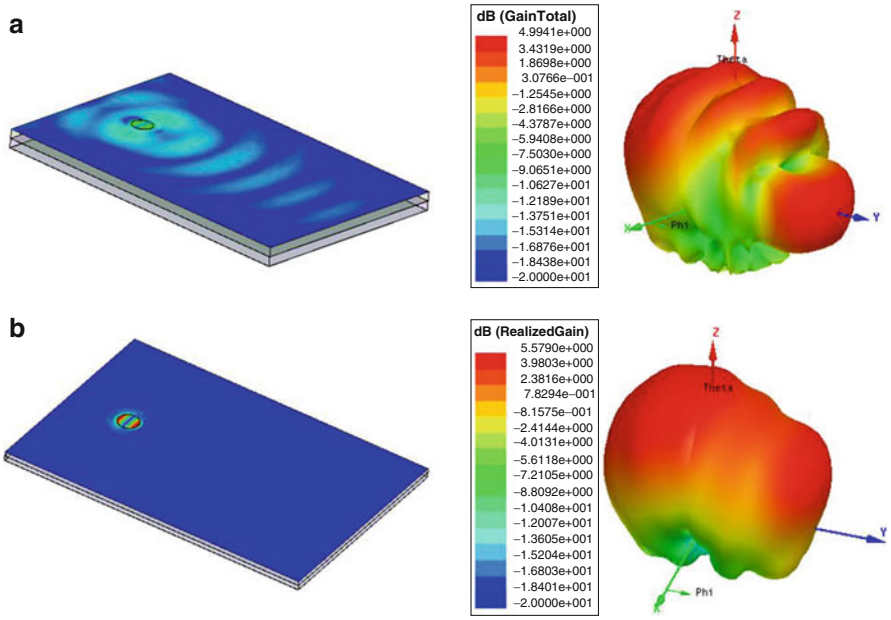


Fig. 38 Simulated electric fields and 3D radiation patterns. (**a**) Without fence of ground via holes. (**b**) With fence of ground via holes (Hong et al. 2013)

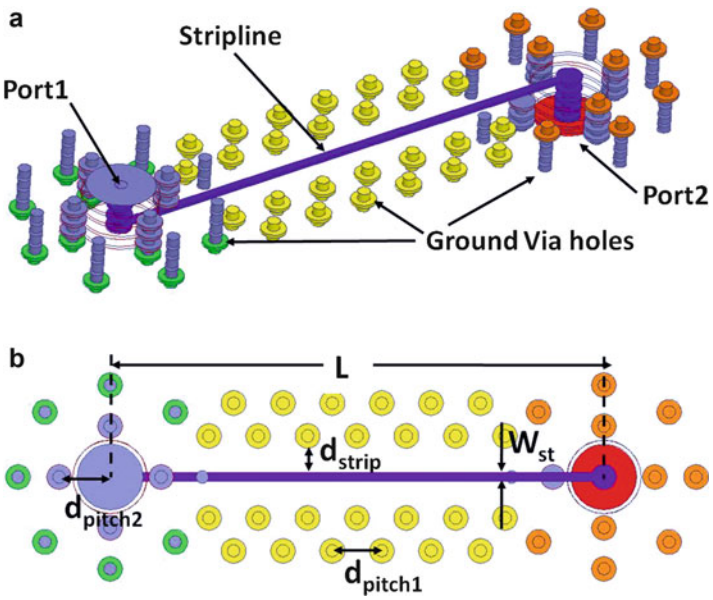


Fig. 39 Measured S-parameter amplitudes of the antenna feedlines (Hong et al. 2013)

simulated insertion loss is approximately 0.3 dB due to the small electrical height of the vertical transitions. Two GSG (ground-signal-ground) ports are connected to each of the vertical transitions at M1 and M12, respectively. Simulation indicates negligible difference when the two GSG ports are located on the same layer such as M12. Four antenna feedline test samples with length $L = 5, 10, 15, 20$ mm are collectively fabricated. The opposite locations of the two GSG ports hinder the usage of conventional RF probe stations. A multi-port in-house RF probe station which is capable of performing simultaneous S-parameter measurements in both the front and backside of the DUT (device under test) is custom

DOI 10.1007/978-981-4560-75-7_64-1

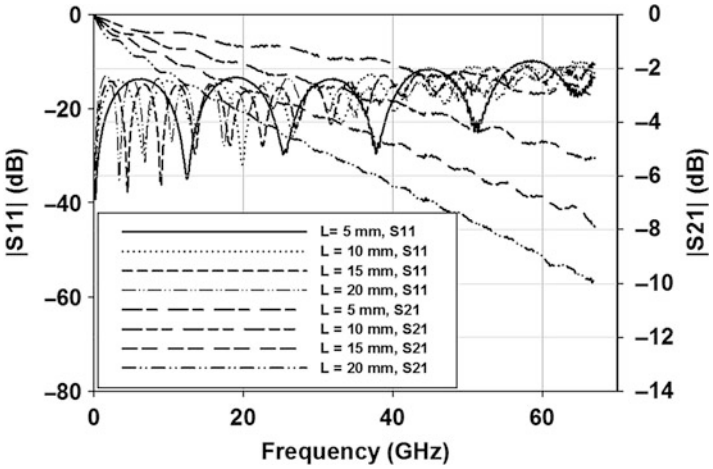


Fig. 40 Measured S-parameters of the fabricated antenna feedline test samples

designed and built. Short-open-load-thru (SOLT) on-wafer calibration is performed prior to S-parameter illustration of the antenna feedline: (a) 3D view and (b) top view.

The measured amplitude of the S-parameters of the fabricated antenna feedline test samples are plotted in Fig. 40. It can be observed that the devised antenna feedline structure features less than −10 dB input reflection coefficients (S11, S22) within the frequency of interest. From the insertion loss (S21, S12) of the four test samples, the average unit loss is calculated to be 0.55 dB/mm. The relatively high unit loss compared to LTCC- or LCP-based antenna feedlines is primarily attributed by the loss tangent of the FR-4 substrate.

Figure 41 shows the measured and simulated far-field radiation patterns of the antenna element. Due to mechanical restrictions of the antenna measurement setup, the radiation patterns are measured within the range of ±90° in the elevation plane. The measured antenna element features an omnidirectional co-polarized radiation patterns in both the E- and H-planes as predicted by HFSS simulations. Cross-polarized measured and simulated radiation patterns indicate a much weaker correlation due to the following attributes: (1) the omission of the antenna measurement environment such as RF cables, adaptors, and chuck in the simulated numerical simulation model due to computation limitations; (2) measurement errors due to various scatterers such as RF cables, chuck, RF probes, etc.; and (3) limited dynamic range of the antenna measurement setup. Additional measures are being investigated to improve the accuracy between the two results. A pair of identical standard gain V-band horn antennas is used to measure the gain of the antenna element at boresight. First a CW signal is transmitted from the TX horn antenna and the RX horn antenna is used to record the received power. Afterward, the RF horn antenna is replaced with the AUT, and the received power is recorded in a similar manner. Losses in the RF probe, cables, and connectors are calibrated out during the gain calculation process. The loss incurred in the antenna feedline is not subtracted. The measured average gain is determined to be 4.8 dBi. Using the directivity derived from the simulated far-field radiation patterns, the average radiation efficiency of the devised antenna element is calculated to be approximately 71 % between 57 and 66 GHz.

Phased-Array Antenna Module

The designed antenna element and antenna feedline structure is employed to devise a 4×2 antenna array, composed of eight antenna elements within the 60 GHz antenna module for phased-array applications. Each of the antenna elements are independently connected with corresponding GSG ports located in M12. Therefore the antenna array also consists of eight independent antenna feedlines. A 20×20 mm^2 region at the center of the FR-4 carrier board is allocated for direct integration of the antenna array. Details of the

DOI 10.1007/978-981-4560-75-7_64-1

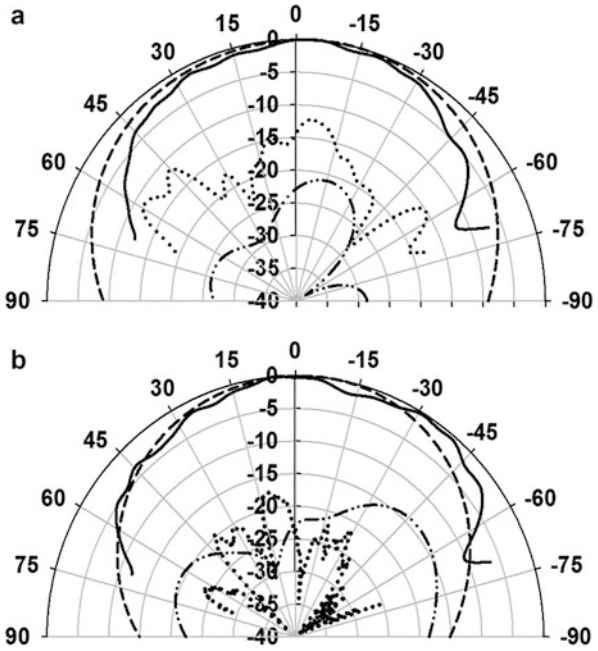


Fig. 41 Measured and simulated far-field radiation patterns of the antenna element. (**a**) E-plane. (**b**) H-plane. *Solid*: measured Co-pol. *Dash*: simulated Co-pol. *Dotted*: measured X-pol. *Dash-dot-dot*: simulated X-pol (Hong et al. 2013)

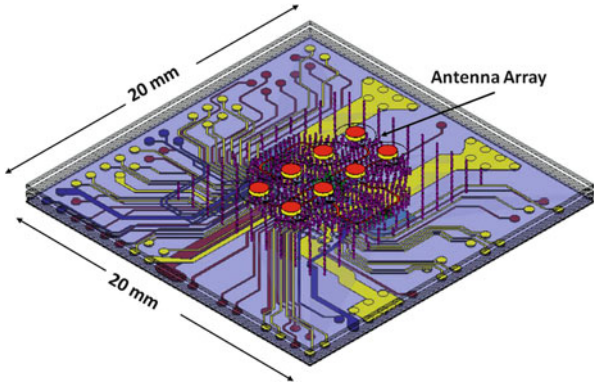


Fig. 42 Simulation setup of the eight-element antenna array. The region of the FR-4 carrier board is not shown for brevity (Hong et al. 2013)

PDN and low-speed and high-speed interfaces are included in the modeling and simulation process of the 60 GHz array antenna as illustrated in Fig. 42. The antenna element spacing are configured to $d_{\text{ant}} = 2.5$ mm in both the x and y direction to achieve minimum mutual coupling while avoiding undesired grating lobes during beam steering scenarios. The physical dimensions of the antenna elements which compose the array antenna module remain identical.

The antenna array feedlines are designed to exhibit identical phase properties and minimum insertion loss. Figure 43 presents the sliced top view of the modeled antenna array feedline layout. The antenna array feedlines are modeled using a commercial CAD program in a quadrant symmetric manner. The maximum deviation among the antenna array feedlines is 14 um. Fences of ground via holes are densely implemented with maximum via hole-to-via hole pitch of 310 um. The ground via holes are determined to be instrumental in shielding the antenna array feedlines from undesired spurious emissions that occur

DOI 10.1007/978-981-4560-75-7_64-1

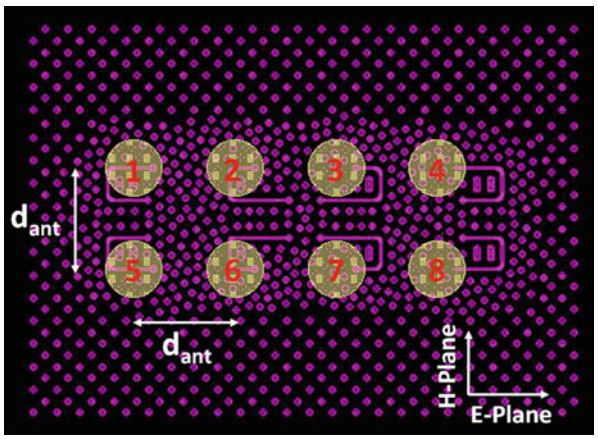


Fig. 43 Sliced top view of the antenna array. M1 and M8 are overlapped (Hong et al. 2013)

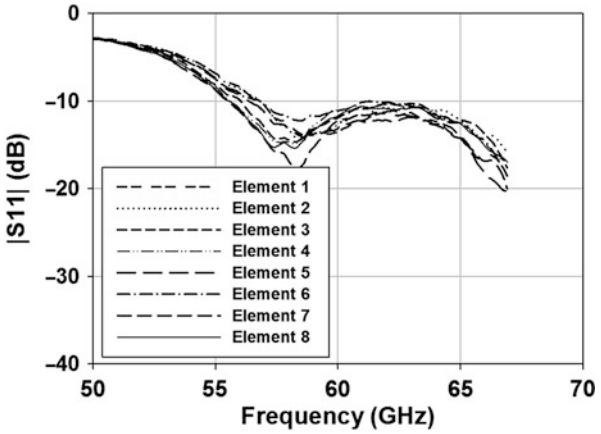


Fig. 44 Measured S11 amplitude of the antenna array (Hong et al. 2013)

from various active components that are implemented on the FR-4 carrier board. Simulation and measurement indicate the stand-alone model of the antenna array feedlines feature $|S_{11}| < -9.6$ dB and $|S_{21}| < 2.2$ dB between 57 and 66 GHz. The maximum measured phase deviation is determined to be less than 7.1°. Amplitudes of the input reflection coefficients of each of the antenna elements are measured one at a time. During measurement, the rest of the GSG ports are configured to be open. Measurement result of the antenna array is presented in Fig. 44. It can be observed that all eight antenna elements within the antenna array exhibit a −10 dB bandwidth of more than 9 GHz from 57 to 66 GHz, sufficiently meeting the bandwidth specification for WiGig/IEEE 802.11ad applications worldwide. Measurement errors occurred during the RF probing procedure and discrepancies between the antenna feedlines are identified as the primary cause for port-to-port variations.

Phased-array and beam steering characteristics are evaluated in the anechoic chamber using a far-field measurement setup. An enhanced serial interface (ESI) control board is connected to the antenna module through wired interface. Custom-made software program installed on the computer that is connected to the ESI board emulates the digital functions of the 60 GHz transceiver system. Beam steering is realized by modulating the phase distributions of equal-amplitude 60 GHz signals that feed each of the antenna elements. Eight 2-bit phase shifters within the RFIC are modulated to produce specific phase distributions for the corresponding beam steering scenario based on a predetermined beam table. A V-band standard

DOI 10.1007/978-981-4560-75-7_64-1

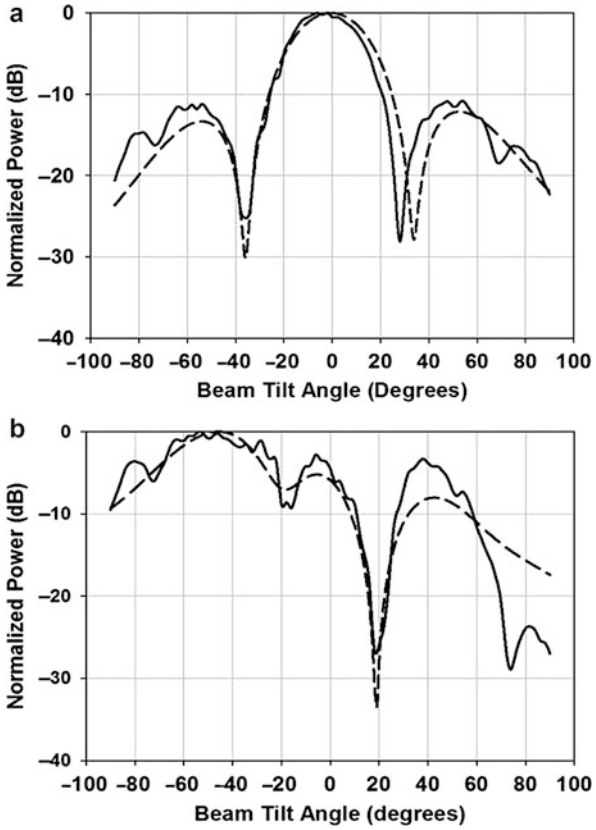


Fig. 45 Normalized far-field radiation pattern in the E-plane. (**a**) Boresight. (**b**) −45° beam tilt. *Solid*: measured. *Dash*: simulated (Hong et al. 2013)

gain horn antenna is used to transmit and receive a series of CW signals within the 57–66 GHz range to and from the DUT. The CW signals are up converted and down converted to 54 MHz intermediate frequency (IF) in both the transmitting and receiving end. The normalized measured and simulated far-field radiation pattern of the antenna module is presented in Figs. 45 and 46 in the E- and H-plane for both boresight and preset beam tilt angle of −45° and 30°, respectively. The designed 60 GHz antenna module exhibits 3 dB beamwidth of 28° and 50° at boresight in the E- and H-plane, respectively. The maximum scan range is determined to be approximately 50° in both planes. The measured gain values at preset beam tilt angle of −45° and 30 ° are 11.3 dBi and 11.6 dBi at 61.5 GHz, respectively.

mmWave Antennas for 5G Cellular Phones

Introduction and Background

Based on the remarkable technological advancements, the cellular market sector continues to maintain a healthy growth rate – according to the International Telecommunication Union (ITU), during 2013 Mobile World Congress (MWC), worldwide cellular phone subscriptions will exceed world population by 2014. Numerous approaches are being discussed and researched around the world to create a future cellular network which would one day succeed the incumbent 4G/LTE network. While the data capacity of the 4G/LTE service continues to improve quite remarkably, classical information theory established by Claude Shannon and Ralph Hartley points out that the bottleneck in achieving the desired cellular Gbps data rate mostly lies on the currently crammed spectrum bandwidth. In retrospect, it becomes possible to introduce an entirely new cellular technology by employing a higher frequency spectrum which contains a

DOI 10.1007/978-981-4560-75-7_64-1

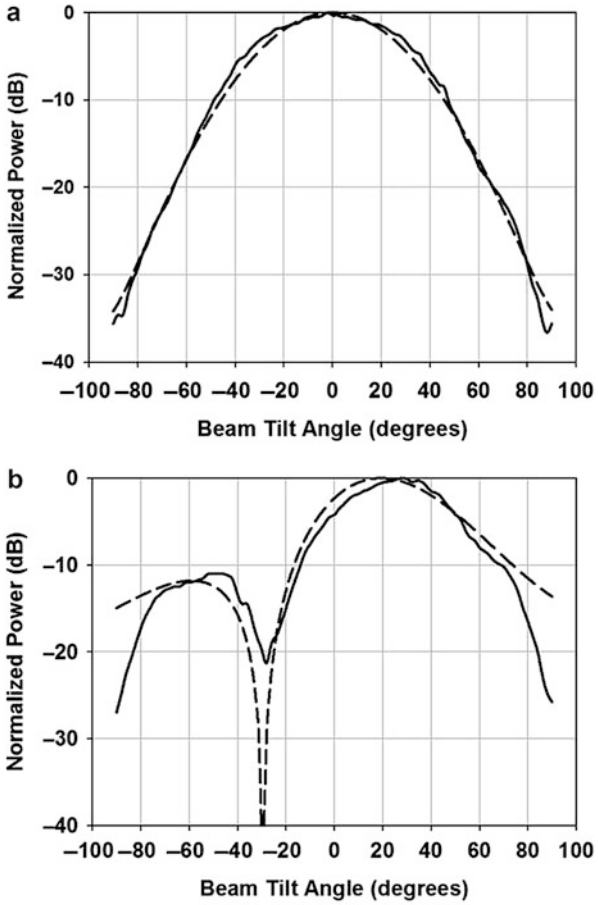


Fig. 46 Normalized far-field radiation pattern in the H-plane. (**a**) Boresight. (**b**) 30° beam tilt. *Solid*: measured. *Dash*: simulated (Hong et al. 2013)

much larger allocated signal bandwidth. The ongoing proliferation of small base stations denoted as microcells and picocells carried out by the cellular carriers becomes enormously instrumental in mitigating the relatively high free-space loss caused by the atmospheric conditions and penetration at extremely high frequencies. Adaptive beam steering technology discussed in section "Theory" can be utilized to get around obstructions, signal blind spots, and other NLOS propagation conditions. Theoretical study and measurement results of a 5G prototype operating at millimeter wave have been presented and discussed in detail in Roh et al. (2014). In the aforementioned study, a 28 GHz prototype exhibits a sharp radiation beam consisting of an array of multiple antennas which continually sweep the atmosphere to identify the strongest connection.

The implementation of a mmWave transceiver within a cellular phone introduces unprecedented challenges that are far from trivial. The enormity of this task is particularly exacerbated by the unique nature of user scenarios associated with cellular phones. In comparison to portable computers and wireless docking stations, cellular phones are much more compact in size and its user scenarios are highly diversified owing to the proliferation of smartphones. The 5G cellular phone must guarantee secure and reliable link accessibility of Gbps speed anywhere, anytime. Among key elements constituting the mmWave 5G radio, the antenna requires one of the most radical changes design wise. This is due to the relatively simple fact that every cellular standard up to present has operated below the 3 GHz spectrum. Present-day cellular antennas can be classified as electrically small antennas and are inherently omnidirectional. The antenna gain of these antennas typically falls in the range of −8 to 0 dBi. The much higher

DOI 10.1007/978-981-4560-75-7_64-1

gain that is needed to compensate the higher signal attenuation at mmWave frequencies introduces the concept of antenna arrays for cellular phones for the first time in the history of wireless communications. In this subsection, pragmatic methodologies to design and selectively utilize up to 32 antenna elements operating at 28 GHz within a future 5G cellular prototype are demonstrated. A number of key antenna design considerations that stem from the unique nature of mmWave bands are discussed in detail. Based on these considerations, a novel but practical 28 GHz phased array is designed and illustrated. While it is important to note that the 28 GHz carrier frequency has yet to be ratified for 5G communications, this design technique can be scaled to other mmWave frequencies in future applications.

Key Considerations for mmWave 5G Antenna

Determining the optimum placement of the 28 GHz antenna within a cellular handset is one of the most defining factors prior to the actual design. While it is difficult to accurately envision a future cellular handset device down to the detail, it is projected that radio functions will closely resemble its present-day counterpart. A conventional cellular phone chassis is examined and classified into a number of key segments as presented in Fig. 47. Major segments such as the rear case, battery cover, and the main PCB are predominantly based on polycarbonate and FR-4. For these materials, the loss tangent versus frequency relation is coherent with studies reported in (Hong et al. 2013). Therefore it would be valid to predict that the gain of the 28 GHz antenna will likely experience a non-negligible degradation when placed within the cellular handset device. A loss of 1.4 dB in gain of a rudimentary 28 GHz patch antenna is calculated when placed within in the top left corner of the presented chassis using electromagnetic numerical simulators. The metallic segments within the cellular handset such as the LCD bracket, LCD, and the battery are expected to be perceived as a finite conductive layer at 28 GHz. Consequently, the designed antenna must be implemented within the 5G cellular phone which excludes these regions.

The placement of the 28 GHz antenna is further complicated by the uncertainty of the user's hand. At 28 GHz, the user's body functions as a formidable absorber which is expected to have negative effects on the gain, radiation pattern, and return loss of the antenna. Figure 48 presents a simulation of the performance degradation of a conventional broadside 28 GHz antenna array implemented within the presented chassis. More than 9.5 dB decrease in gain is estimated when the user's hand fully encompasses

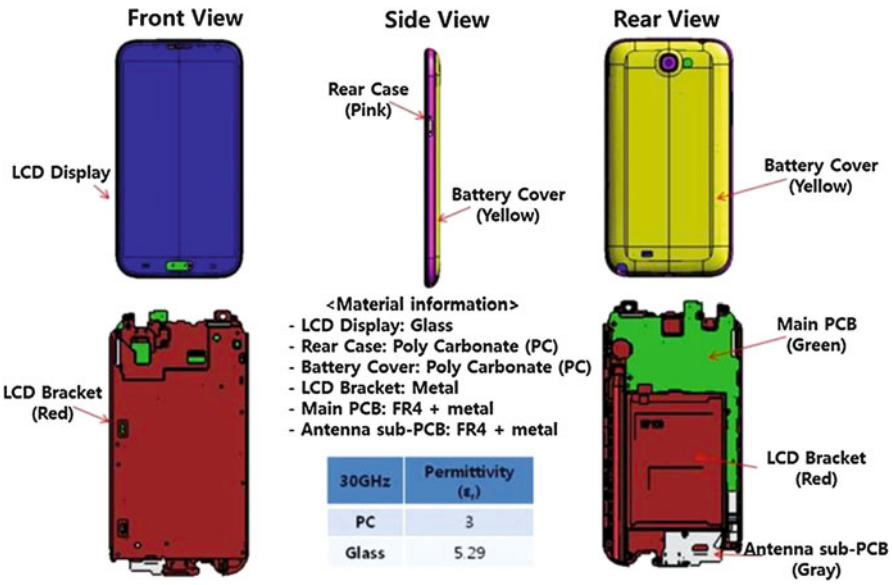


Fig. 47 Illustration of the material properties of the major segments of a conventional cellular handset (Hong et al. 2014a)

DOI 10.1007/978-981-4560-75-7_64-1

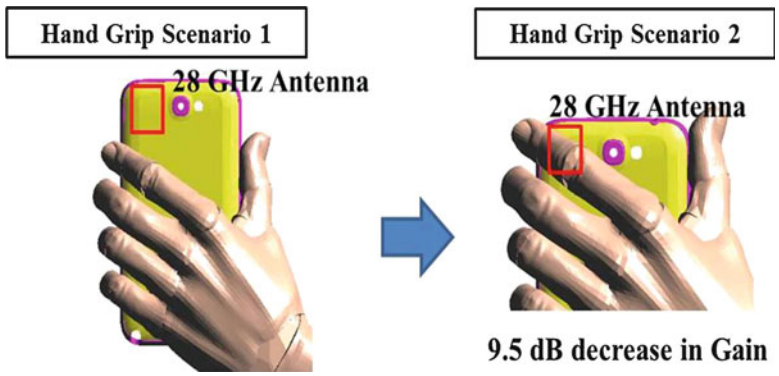


Fig. 48 Simulation of the effect of the user's hand on a rudimentary 28 GHz antenna array placed within a cellular handset (Hong et al. 2014a)

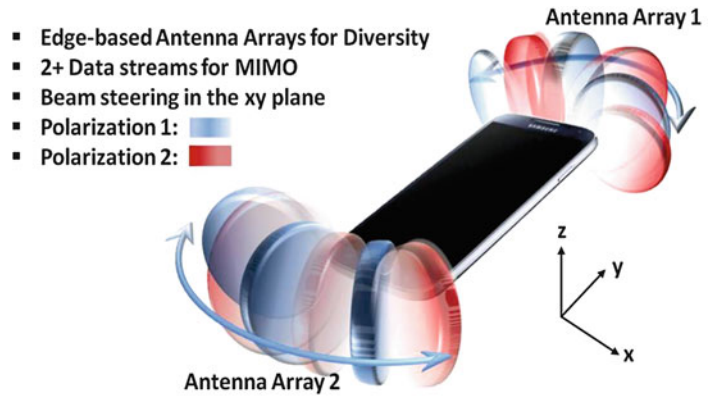
the region of the 28 GHz antenna array. One way of alleviating the potential performance risk caused by the user's hand would be to implement plural number of antennas within the cellular device.

Mobility requires cellular phones to operate virtually anywhere. It is statically and empirically impossible to predict the angle of arrival (AoA) of the strongest signal path. The omnidirectional patterns of 3G/4G cellular phone antennas today are advantageous in this regard. Likewise at mmWave frequencies, the incoming waves impinging on the antenna are predicted to be distributed across the entire sphere. This introduces a paradoxical situation for the antenna engineer that given a finite amount of energy, the antenna needs to radiate with extremely high intensity around the sphere. Since simultaneous radiation at all angles violates the classical laws of physics, phased-array beam steering as that employed in the previous subsection is applied.

In this subsection, Cartesian coordinates (x,y,z) and spherical coordinates θ (ranging from 0 to π and defined as the elevation angle) and φ(ranging from 0 to 2π and defined as the azimuth angle) will be used to describe the maximum allowable trajectory of the main lobe of the phased-array antenna. Based on the realistic physical limitations of antenna placement options within a cellular phone as described earlier, two-dimensional (planar) phased-array topology is deduced. With selected withstanding exceptions, the antenna elements are placed at least half wavelength apart to avoid spatial aliasing or also known as grating lobes.

For $1 \times N$ antenna array geometries, the scanning angle is a function of θ or φ. The scanning angle is a function of θ and φ for $N \times M$ antenna geometries. In addition, the scanning angle range of the main lobe is inversely related to the number of antenna elements (i.e., antenna array gain). Due to the high signal attenuation at 28 GHz, the antenna must be placed in very close proximity to the 28 GHz RFIC and the front-end module. Implementing the antenna array directly on the PCB of the 5G cellular device will therefore keep the insertion loss between the antenna and RFIC to be minimal. This implies that a placement of the RF blocks within the 5G architecture before the intermediate frequency (IF) stage will be dependent on the placement of the 28 GHz antenna array within the cellular phone. Collectively considering the aforementioned design environment, a minimum set of two 28 GHz antenna arrays are conceived for mmWave 5G cellular phones. As presented in Fig. 49, the two antenna arrays are implemented on the top and bottom portion of the cellular device. This is somewhat of an intuitive approach since similar to 3G/4G antennas, the large LCD display panel and the battery prohibit the antenna arrays from being placed in the center region of the cellular phone. This topology is applicable for both single-input single-out (SISO) and multiple-input multiple-output (MIMO) systems. Each 5G antenna array located on the top and bottom portion of the cellular phone is designed to respectively

DOI 10.1007/978-981-4560-75-7_64-1



Cellular Standards	4G	5G (This work)
Antenna Type	Sub-wavelength antennas	Phased array antennas
Radiation patterns	Omni-directional	Directional Fan-beam
Diversity & MIMO	Yes	Yes
Polarization	Single & Constant	Multiple & Reconfigurable

Fig. 49 The 28 GHz antenna array configuration for 5G cellular phones and its comparison with 4G standard (Hong et al. 2014b)

cover the corresponding hemisphere in the form of spatial diversity. For SISO, the secondary antenna array becomes instrumental when a secure link cannot be established using the primary antenna.

It is a rarity for thickness of cellular phone PCBs to exceed 1 mm. In the fiercely competitive consumer electronics arena, increasing the PCB thickness is directly related to the rise in production costs and bulkier designs. At 28 GHz, the free-space half wavelength is slightly larger than 10 mm. This discrepancy prohibits phased-array beam steering in the elevation plane (*zx* plane) and leads us to a one-dimensional beam steering approach using a 1 × N antenna array topology. Moreover, a very broad beamwidth is required in the elevation plane to compensate for the inability to beam steer.

mmWave 5G Mesh-Grid Antenna Array Design

In order to synthesize a highly directive fan beam, it is imperative to maximize the allowable number of antenna elements inside the cellular phone. Avoiding unnecessary loss in the antenna radiation is critical at mmWave since this is correlated to the wireless link budget and eventually the power consumption. Therefore each of the antenna elements is designed to be at least a quarter of λ_g large, which is the minimum threshold before the radiation efficiency starts to sharply deteriorate due to the rise of stored energy in the near-field region of the antenna. As mentioned in section "Key mmWave Antenna Design Techniques," mmWave planar dipoles are by far the most popular fan-beam antennas in the wireless industry. However, as it was also discussed in section "Key mmWave Antenna Design Techniques," the fan-beam radiation characteristics of planar dipole are preserved under the condition that the metallic patterns and tracelines are removed below the planar dipole topology. The elimination of the metallic patterns becomes problematic especially with the advent of smartphones due to the increased degree of integration of the PCB. Cellular phone PCBs which are less than 1 mm thick typically consists of 6–12 layers that are tens of μm in thickness respectfully. Within that space run various signal lines ranging from low-speed to high-speed interfaces and power distributed networks. Removal of the signal line traces for the implementation of the half-guided wavelength planar dipole structure puts a serious strain on the PCB layout. As an alternative, a low-profile antenna design which can coexist with the signal line traces and yet exhibit a fan-beam radiation characteristic is designed.

DOI 10.1007/978-981-4560-75-7_64-1

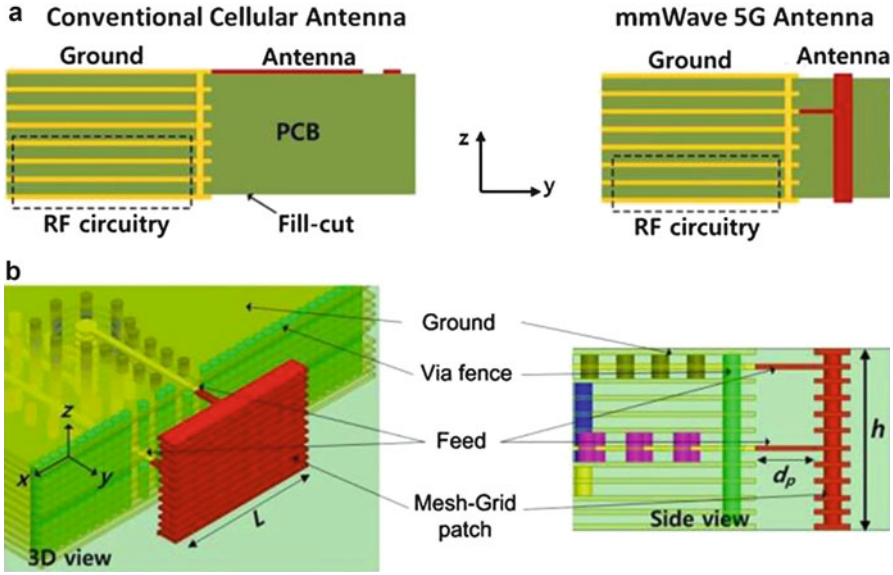


Fig. 50 (**a**) Simplified comparison between conventional cellular antenna and proposed mmWave 5G antenna topology. (**b**) Detailed view of the mesh-grid patch antenna (Hong et al. 2014b)

Illustrated in Fig. 50, in contrast to planar dipole antennas that spans across the *xy* plane, the newly devised structure is situated along the *zx* plane which can reduce the planar antenna footprint from the previous half λ_g to a fraction of a hundreds of a λ_g. The real estate in the *xy* plane is solely determined by the length of the dual antenna signal feed lines denoted as d_p. The entire radiating body of the antenna structure consists of tightly spaced via holes creating a mesh type grid as it can be observed from the *zx* plane. Hence, this antenna topology is named as the mesh-grid patch antenna by the authors. The physical dimension of the mesh-grid patch antenna ranges from quarter to half of the λ_g along the *x*-axis. The height denoted as *h* determines the number of PCB layers required for the specific mesh-grid patch antenna design and affects the elevation beamwidth in the θ angle. The mesh-grid antenna is designed to support multi-polarized MIMO based on any combination consisting of vertical, horizontal, and circular polarizations as illustrated in Fig. 51. Two antenna feedlines are configured to respectively enable horizontal and vertical polarizations. The polarization configurations are summarized in Table 2. Figure 52 presents the simulated and measured far-field radiation pattern of the mesh-grid patch antenna in the elevation plane. The measured −10 dB return loss bandwidth is approximately 1 GHz with a center frequency of 27.9 GHz.

The mesh-grid antenna elements are expanded and arranged in two sets of 1 × 16 phased arrays in the top and bottom portion of the cellular phone PCB as illustrated in Fig. 53. The high antenna gain realized by a reconfigurable and scalable array consisting of 16 antenna elements alleviates the linearity and power consumption requirements of the 5G amplifiers. The mesh-grid antenna elements situated at each corner of the PCB are arranged in slanted angles of approximately 50°. The conformal topology further maximizes the range of the beams steering scanning angles in the azimuth plane. Moreover, the slanted topology conforms to the cellular device and enables the designed mmWave antenna array to appear as an extremely low-profile metallic trace lines which encompasses around the edges of the PCB. The width of the trace lines is less than 0.2 mm which is even less than the 1 mm spacing required from the PCB edges for conventional surface-mount technologies (SMT). From the vantage point of the hardware layout, the inclusion of a total of 32 mmWave antenna elements requires negligible antenna footprint. Based upon this antenna solution, a truly massive MIMO antenna system may be realizable for mmWave 5G in the long term.

DOI 10.1007/978-981-4560-75-7_64-1

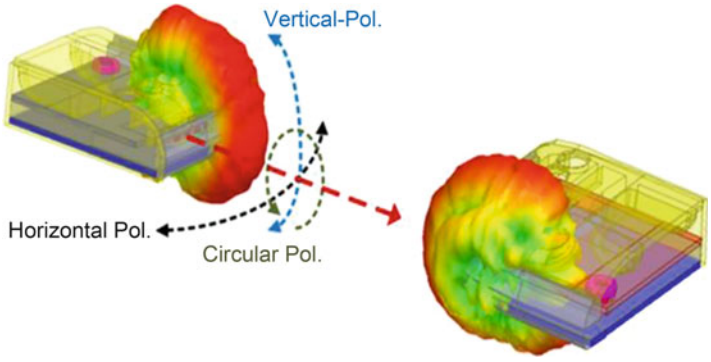


Fig. 51 Illustration of the effect of antenna alignment configuration on the polarization mismatch for mmWave-equipped cellular phones

Table 2 Polarization configuration of the dual-feed mesh-grid patch antenna

Antenna polarization	Horizontal polarization feedline state	Vertical polarization feedline state
Horizontal	On	Off
Vertical	Off	On
Diagonal	On	On
Circular	On	On+90°

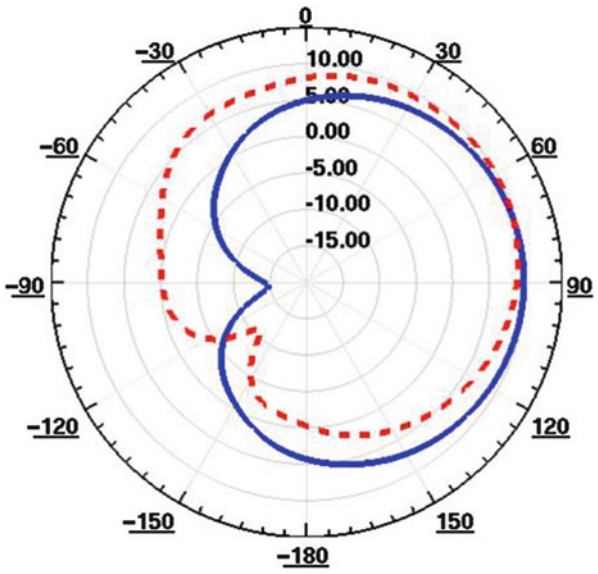


Fig. 52 Simulated and measured far-field radiation pattern of the mmWave mesh-grid patch antenna in the elevation plane. *Solid line*: simulation. *Dotted line*: measurement

The measured and normalized radiation patterns of the mesh-grid antenna array in the azimuth plane are plotted for different main beam scanning angles in Fig. 54. The antenna array prototype is first measured in a stand-alone state denoted as free space. Afterward, the measurement is repeated after the AUT is integrated within the cellular phone prototype. Measurements ascertain a scanning angle range of approximately $\pm 70^{o}$. By combining the two antenna arrays, nearly spherical radiation coverage can be

DOI 10.1007/978-981-4560-75-7_64-1

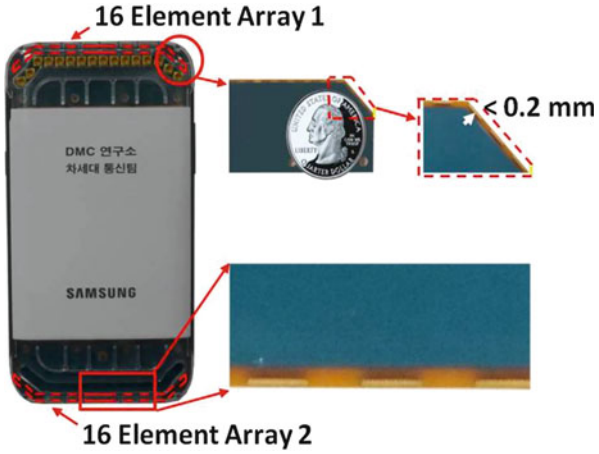


Fig. 53 Photographs of the mmWave 5G antenna system prototype integrated inside a cellular phone and zoomed in views of the mmWave antenna region (Hong et al. 2014b)

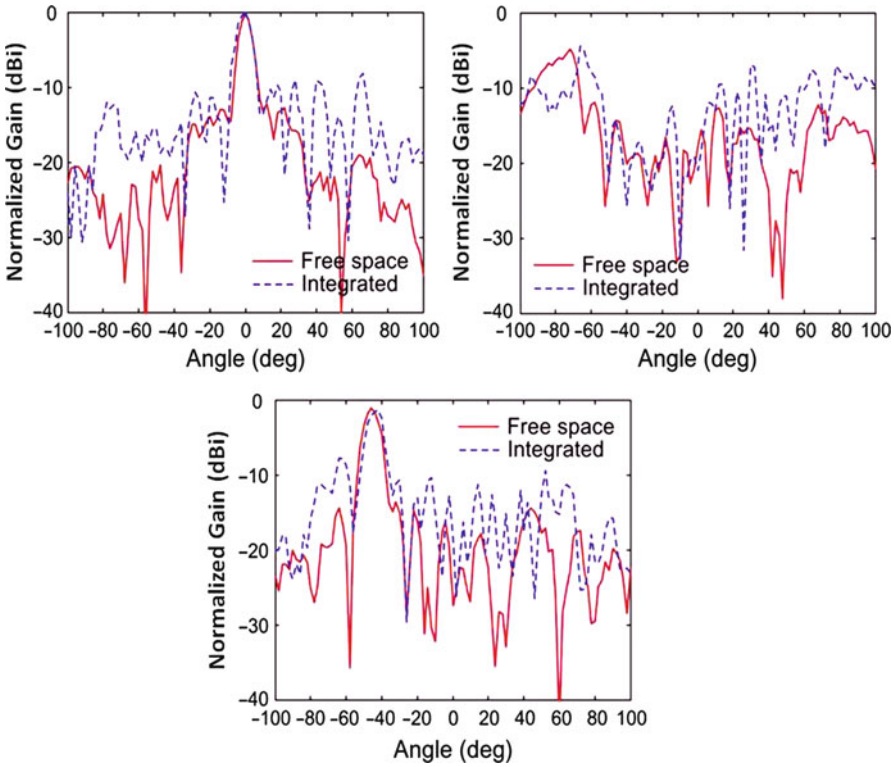


Fig. 54 Measured and normalized radiation patterns of the 16-element mesh-grid antenna array in the horizontal polarized configuration (Hong et al. 2014b)

attained. The radiation pattern discrepancies between the free space and integrated condition indicates the occurrence of pattern deformation caused by diffracted and refracted electromagnetic waves between the antenna elements and the cellular phone chassis. The discrepancy level is expected to vary as a function of the design and material composition of the cellular device. The 3 dB beamwidth in the azimuth plane is approximately 12° generating a null-free beam scanning coverage. The peak gain of the antenna array within the cellular device is calculated to be more than 10.5 dBi. Major loss factors include the 2.8 dB loss

DOI 10.1007/978-981-4560-75-7_64-1

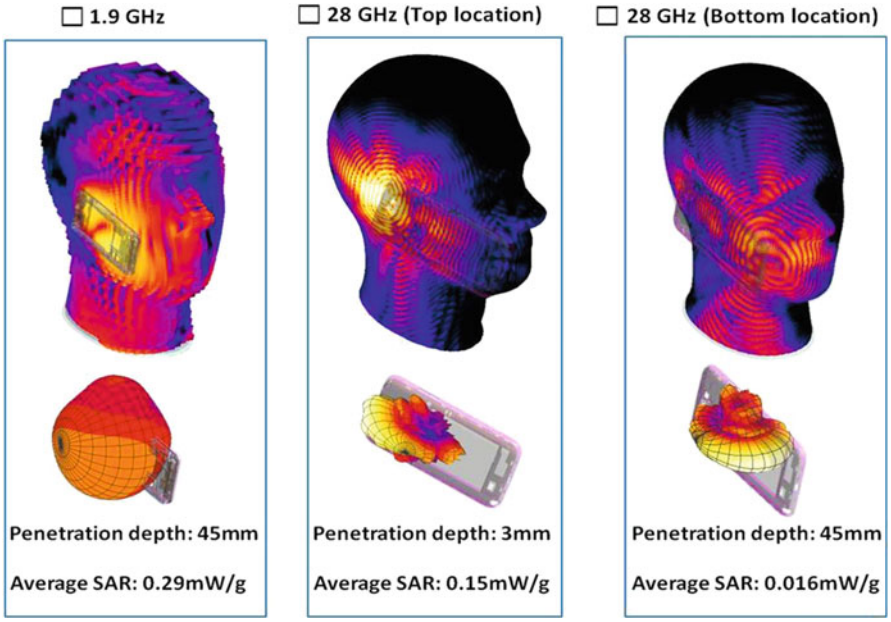


Fig. 55 Analysis and comparison of specific absorption rates of 4G and mmWave 5G antenna systems (Hong et al. 2014b)

from the mesh-grid antenna feed point to the K-type coaxial connector, 2.2 dB cable loss between the K-type coaxial connector and the RF unit, and the estimated 1.8 dB dielectric loss incurred by the polycarbonate chassis of the cellular phone prototype used in the measurement.

Biological Implications and Considerations

Assessing the biological implications on the user may be one of the most important parameters for mmWave 5G cellular devices in the future. The level of electromagnetic absorption by the human body is strictly and universally regulated by governmental bodies under the specific absorption rate (SAR) guideline. Every cellular device must pass a regulated SAR test prior to certification. Likewise although permissible SAR levels at mmWave frequencies do not exist at this point in time, it is possible to analyze and compare mmWave SAR levels with that of 3G/4G cellular phones. To investigate this, the measured radiation patterns of the two sets of mesh-grid antenna arrays are imported into a commercial numerical simulator.

With the inclusion of a human head phantom, the average SAR levels are calculated. As it can be observed in Fig. 55, the electromagnetic fields absorbed by the user at 28 GHz tend to be localized in comparison to that of a 3G/4G cellular frequency band. The penetrated skin depth of the absorbed 28 GHz signal is confirmed to be approximately 3 mm. In comparison the penetrated skin depth at 1.9 GHz exceeds 45 mm. This implies that majority of the absorbed energy is confined to the epidermis at mmWave. The average calculated SAR at 28 GHz for the antenna array situated in the bottom region of the cellular phone is found to be less than a factor of 18 compared to SAR levels retrieved at 1.9 GHz. The ability to beam steer the main lobe is projected to further reduce the SAR levels in the future.

Conclusion

As discussed in this chapter the exploitation of mmWave frequencies is projected to become the backbone of future mobile networks. mmWave wireless communication aims to enable unprecedented and intuitive

DOI 10.1007/978-981-4560-75-7_64-1

user experience with enhanced accessibility to the general public. However, it is vital to emphasize that mmWave spectrums will not render presently used microwave spectrum to become obsolete. The inherent limitations of the mmWave signals imply coexistence with legacy wireless architectures, and integrations will become an ever-important issue as widespread interest of mmWave communications continue to grow. In severe propagation conditions, mmWave radio systems will be designed to fall back to conventional radio systems to ensure the overall quality of service (QoS) of the wireless network. As the number of antennas required inside wireless mobile terminals such as cellular phones and small access points continue to increase on a yearly basis, the limited space constraints for mmWave antennas is projected to become an major issue. Researching novel mmWave antenna topologies which are easier to integrate with the rest of the legacy wireless RF system and the microwave frequency spectrum antennas will be an important area. In addition, the demand for three-dimensional, reconfigurable mmWave antennas is anticipated followed by the proliferation of wearable communication devices and the advancements of three-dimensional printing technologies. The mmWave antenna design philosophy must prioritize and focus on the wireless system performance such as link reliability and power consumption efficiency.

Another key area where mmWave antenna technologies are projected to be vital is associated with the emergence of multi-node/massive MIMO transmissions and receptions. Massive MIMO antenna architectures are composed of a very high number of antenna elements with the purpose of enhanced wireless capacity. They benefit from simplified MAC layers, enhanced latencies, and signal cell channel capacity. Devising efficient mmWave antenna topologies capable of three-dimensional, simultaneous multi-beam steering that is scalable remains a notable technical challenge. Future massive MIMO antenna technology must be investigated with full consideration of multi-heterogeneous networks (multi-HetNeT) which is composed of multiple carrier frequencies. The increasingly technical complexity related to interoperability and mutual dependency between the radio hardware and the wireless network researchers underlines the multidisciplinary nature of future mmWave antenna research. The anticipated Internet of Things (IoT) era will further introduce numerous wireless devices with diverse shapes and forms. The author predicts mmWave antennas will evolve from the current planar multilayer PCB types to conformal topologies based on novel materials and integration concepts.

Cross-References

- ▶ Advanced Antenna Fabrication Process (MEMS/LTCC/LCP/printing)
- ▶ Antennas in Hand-held Devices
- ▶ Beam-Scanning Leaky Wave Antennas
- ▶ Broadband and Multiband Planar Antennas
- ▶ Grid Antenna Arrays
- ▶ Lens Antennas
- ▶ MIMO Systems and Antennas for Terminals (Including Portable Devices Such as Handsets, iPad Laptops)
- ▶ Mm-Wave Sub-mm-Wave Antenna Measurement
- ▶ On-Chip Antennas
- ▶ Phased Arrays
- ▶ Radio Frequency Beamforming for Scanned and Multibeam Antenna Systems General
- ▶ Reconfigurable Antenna Arrays for Wireless Communications
- ▶ RF Material Characterization

DOI 10.1007/978-981-4560-75-7_64-1

► Substrate-Integrated Waveguide Antennas
► Taped Slot Antennas
► Waveguide Slot Antennas and Arrays

References

Abbaspour-Tamijani A, Sarabandi K (2003) An affordable millimeter-wave beam-steerable antenna using interleaved planar subarrays. IEEE Trans Antennas Propag 51:2193–2202

Al Henawy M, Schneider M (2011) Integrated antennas in eWLB packages for 77 GHz and 79 GHz automotive radar sensors. In: 41st European microwave conference (EuMC) 41st. Manchester, United Kingdom, pp 1312–1315

Alhalabi RA, Rebeiz GM (2009) High-gain Yagi-Uda antennas for millimeter-wave switched-beam systems. IEEE Trans Antennas Propag 57:3672–3676

Artemenko A, Maltsev A, Maslennikov R, Sevastyanov A, Ssorin V (2011) Beam steerable quartz integrated lens antenna for 60 GHz frequency band. In: Proceedings of the 5th European conference on antennas and propagation (EUCAP). Rome, Italy, pp 758–762

Baek KH, Hong W (2013) Highly miniaturized vertical end-fire antenna array for mmWave wireless communication. In: Antennas and propagation society international symposium (APSURSI). Orlando, Florida, USA, pp 992–993

Behdad N, Shi D, Hong W, Sarabandi K, Flynn MP (2007) A 0.3 mm^2 Miniaturized X-band on-chip slot antenna in 0.13um CMOS. In: IEEE radio frequency integrated circuits (RFIC) symposium. Honolulu, Hawaii, USA, pp 441–444

Bereka D, Gopinath A, Sainati B (2014) Design of a 60GHz integrated antenna on silicon substrate. In: IEEE antennas and propagation society international symposium (APSURSI). Memphis, Tennessee, USA, pp 1063–1064

Black DN, Wiltse JC (1987) Millimeter-wave characteristics of phase-correcting Fresnel zone plates. IEEE Trans Microwave Theory Tech 35:1122–1129

Chen J, Niknejad A (2011) A compact 1V 18.6dBm 60GHz power amplifier in 65nm CMOS. In: IEEE solid-state circuits conference (ISSCC). San Francisco, California, USA, pp 432–443

Chin WH, Fan Z, Haines R (2014) Emerging technologies and research challenges for 5G wireless networks. IEEE Wirel Commun Mag 21:106–112

Chu H, Guo Y, Lin F, Shi XQ (2009) Wideband 60GHz on-chip antenna with an artificial magnetic conductor. In: IEEE international symposium on radio-frequency integration technology, pp 307–310

Cohen E, Ruberto M, Cohen M, Pan H, Ravid S (2013) Antenna packaging of a 32 element TX/RX phased array RFIC for 60GHz communications. In: 2013 I.E. international conference on microwaves, communications, antennas and electronics systems (COMCAS). Tel Aviv, Israel, pp 1–5

El Sabbagh MA, Kermani MH, Omar MR (2004) Accurate broad-band measurement of complex permittivity using striplines. In: Antennas and Propagation Society international symposium, 2004, vol 1. IEEE, Monterey, pp 233–236

Emami S, Wiser R, Ali E, Forbes M, Gordon M, Guan X, Lo S, McElwee P, Parker J, Tani J, Gilbert J, Doan C (2011) A 60GHz CMOS phased-array transceiver pair for multi-Gb/s wireless communications. In: IEEE solid-state circuits conference (ISSCC). San Francisco, California, USA, pp 164–166

Fang X, Linton D, Walker C, Collins B (2004) A tunable split resonator method for nondestructive permittivity characterization. Instrum Meas IEEE Trans 53(6):1473–1478

Fratticcioli E, Dionigi M, Sorrentino R (2004) A simple and low-cost measurement system for the complex permittivity characterization of materials. Instrum Meas IEEE Trans 53(4):1071–1077

DOI 10.1007/978-981-4560-75-7_64-1

Ghosh A, Thomas T, Cudak M, Ratasuk R, Moorut P, Vook F, Rappaport T, MacCartney G, Sun S, Nie S (2014) Millimeter-wave enhanced local area systems: a high-data-rate approach for future wireless networks. IEEE J Sel Areas Commun 32:1152–1163

Harrington RF (1961) Time-harmonic electromagnetic fields McGraw-Hill

Hong W, Baek KH, Goudelev A (2012) Multilayer antenna package for IEEE 802.11 ad employing ultralow-cost FR4. IEEE Trans FR4 Antennas Propag 60:5932–5938

Hong W, Baek KH, Goudelev A (2013) Grid assembly-free 60-GHz antenna module embedded in FR-4 transceiver carrier board. IEEE Trans Antennas Propag 61:1573–1580

Hong W, Baek K, Lee Y, Kim YG (2014a) Design and analysis of a low-profile 28GHz beam steering antenna solution for Future 5G cellular applications. In: IEEE international microwave symposium (IMS). Tampa Bay, Florida, USA, pp 1–4

Hong W, Baek KH, Lee Y, Kim Y, Ko ST (2014b) Study and prototyping of practically large-scale mmWave antenna systems for 5G cellular devices. IEEE Commun Mag 52:63–69

Hristov HD, Herben MH (1995) Millimeter-wave Fresnel-zone plate lens and antenna. IEEE Trans Microwave Theory Tech 43:2779–2785

Jo O, Hong W, Choi S, Chang S, Kwon C, Oh J, Cheun K (2014) Holistic design considerations for realization of environmentally adaptive 60 GHz beamforming technology. IEEE Commun Mag 52:30–38

Katehi P, Alexopoulos N (1983) On the effect of substrate thickness and permittivity on printed circuit dipole properties. IEEE Trans Antennas Propag 31:34–39

Kong L, Seo D, Alon E (2013) A 50mW-TX 65mW-RX 60GHz 4-element phased-array transceiver with integrated antennas in 65nm CMOS. In: IEEE solid-state circuits conference (ISSCC), San Francisco, California, USA, pp 234–235

Lin F, Ooi BL (2009) Integrated millimeter-wave on-chip antenna design employing artificial magnetic conductor. In: IEEE international symposium on radio-frequency integration technology. Nanyang Executive Centre, Nanyang Technological University Singapore, Singapore, pp 174–177

Liu D, Pfeiffer U, Grzyb J, Gaucher B (2009) Advanced millimeter-wave technologies: antennas, packaging and circuits. Wiley, New York

Liu D, Akkermans JA, Chen HC, Floyd B (2011) Packages with integrated 60-GHz aperture-coupled patch antennas. IEEE Trans Antennas Propag 59:3607–3616

Marcus M, Pattan B (2005) Millimeter wave propagation; spectrum management implications. IEEE Microwave Mag 6:54–62

Marnat L, Carreno AA, Conchouso D, Martinez MG, Foulds IG, Shamim A (2013) New movable plate for efficient millimeter wave vertical on-chip antenna. IEEE Trans Antennas Propag 61:1608–1615

Natarajan AS (2007) Millimeter-wave phased arrays in silicon. California Institute of Technology. Pasadena, CA, USA

Natarajan A, Tsai MD, Floyd B (2009) 60GHz RF-path phase-shifting two-element phased-array front-end in silicon. In: 2009 symposium on VLSI circuits. Kyoto, Japan, pp 250–251

Okada K, Minami R, Tsukui Y, Kawai S, Seo Y, Sato S, Kondo S, Ueno T, Takeuchi Y, Yamaguchi T, Musa A, Wu R, Miyahara M, Matsuzawa A (2014) A 64-QAM 60GHz CMOS transceiver with 4-channel bonding. IEEE solid-state circuits conference (ISSCC). San Francisco, California, USA, pp 346–347

Pozar DM (1985) Microstrip antenna aperture-coupled to a microstripline. Electron Lett 21:49–50

Pozar DM (1992) Microstrip antennas. Proc IEEE 80:79–91

Roh W, Seol JY, Park J, Lee B, Lee J, Kim Y, Cho J, Cheun K, Aryanfar F (2014) Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype result. IEEE Commun Mag 52:106–113

Ryan CG, Chaharmir MR, Shaker J, Bray JR, Antar YM, Ittipiboon A (2010) A wideband transmitarray using dual-resonant double square rings. IEEE Trans Antennas Propag 58:1486–1493

Shannon CA (1948) Mathematical theory of communication. Bell Syst Tech J 27:379–423

Sun S, MacCartney G, Samimi M, Nie S, Rappaport T (2014) Millimeter wave multi-beam antenna combining for 5G cellular link improvement in New York City. In: 2014 I.E. international conference on communications (ICC). Sydney, Australia, pp 5468–5473

Talwar S, Choudhury D, Dimou K, Aryafar E, Bangerter B, Stewart K (2014) Enabling technologies and architectures for 5G wireless. In: IEEE international microwave symposium (IMS). Tampa Bay, Florida, USA, pp 1–4

Taori R, Sridharan A (2014) In-band, point to multi-point, mm-Wave backhaul for 5G networks. In: 2014 I.E. international conference on communications workshops (ICC). Sydney, Australia, pp 96–101

Tsukizawa T, Shirakata N, Morita T, Tanaka K, Sato J, Morishita Y, Kanemaru M, Kitamura R, Shima T, Nakatani T, Miyanaga K, Urushihara T, Yoshikawa H, Sakamoto T, Motozuka H, Shirakawa Y, Yosoku N, Yamamoto A, Shiozaki R, Saito N (2013) A fully integrated 60GHz CMOS transceiver chipset based on WiGig/IEEE802.11ad with built-in self calibration for mobile applications. In: IEEE international solid-state circuits conference (ISSCC). San Francisco, California, USA, pp 230–223

Valdes-Garcia A, Natarajan, A, Liu D, Sanduleanu M, Gu X, Ferriss M, Parker B, Baks C, Plouchart JO, Ainspan H, Sadhu B, Islam M, Reynolds S (2013) A fully-integrated dual-polarization 16-element W-band phased-array transceiver in SiGe BiCMOS. In: IEEE radio frequency integrated circuits symposium (RFIC). Kyoto, Japan, pp 375–378

Vogler LE (1982) An attenuation function for multiple knife-edge diffraction. Radio Sci 17:1541–1546, Wiley Online Library

Wang Y, Low K, Che F, Pang H, Yeo S (2003) Modeling and simulation of printed circuit board drop test. In: 2003 5th conference on electronics packaging technology, (EPTC 2003). Singapore, pp 263–268

Willmot R, Kim D, Peroulis D (2009) High-efficiency wire bond antennas for on-chip radios. IEEE international microwave symposium (IMS). Boston, MA, USA, pp 1561–1564

Yang XS, Wang BZ, Wu W (2005) Pattern reconfigurable patch antenna with two orthogonal quasi-Yagi arrays. IEEE Antennas Propag Soc Int Symp 2:617–620

Yngvesson KS, Korzeniowski T, Kim YS, Kollberg EL, Johansson JF (1989) The tapered slot antenna-a new integrated element for millimeter-wave applications. IEEE Trans Microw Theory Tech 37:365–374

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_65-1

Conformal Array Antennas

Lars Josefsson[a]* and Patrik Persson[b]
[a]Lars Microwave, Askim, Sweden
[b]Antenna System Technology, Ericsson Research, Gothenburg, Sweden

Abstract

The history, including the theory and the technology, leading to today's conformal array antennas (nonplanar array antennas) is discussed. Since the apertures of these antennas are curved (usually convex), the traditional methods for planar array analysis and design cannot be used. An important aspect is the mutual coupling between radiating elements due to the diffraction along the curved surface. Beamforming, phase modes, geodesics, and polarization are concepts that are dealt with.

Several applications of conformal array antennas are mentioned including wide angle (360°) coverage antennas for mobile base stations, antennas built into the skin of vehicles such as aircraft and satellites, spherical arrays for simultaneous tracking of signals from many directions, etc. Compared to planar arrays, the steering and control of conformal arrays is more complicated. However, the rapid growth in computer power (software and hardware) opens up for new solutions. In particular digital beamforming (DBF) looks promising for the future.

Keywords

Circular array; Commutating; DBF; Digital beamforming; Geodesic; Icosahedron; MIMO; Phase mode; Polarization; Spherical array; UTD; Wullenweber array

Introduction

Definition

A conformal array antenna conforms to a particular shape, e.g., a sphere, cylinder, paraboloid, etc. This is different from most array antennas that are typically planar and not curved.

The IEEE Standard Definitions of Terms for Antennas (IEEE Std 145-1993) gives the following definition:

2.74 conformal antenna (**conformal array**). An antenna (an array) that conforms to a surface whose shape is determined by considerations other than electromagnetic; for example, aerodynamic or hydrodynamic.

2.75 conformal array. *See*: **conformal antenna**.

An array antenna might have its shape determined by a particular electromagnetic requirement such as the angular coverage or beam shape characteristics, leading to a cylindrical or other particular shape. The term conformal is used also in such cases although it is not strictly according to the IEEE definition cited above. Technically it makes no difference.

*Email: lars.josefsson@ieee.org

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_65-1

Motivation

Why conformal array antennas? For several applications conformal antennas offer advantages over traditional planar antenna solutions. Some of them will be discussed briefly.

Aircraft antennas: Modern aircraft have antennas for radar, navigation, communication, instrument landing, altimeters, etc. There can be as many as 20–70 antennas on a typical aircraft, several of them protruding from the aircraft skin causing considerable drag and increased fuel consumption. Integrating these antennas into the curved aircraft skin is highly desirable (Hopkins et al. 1997).

Mobile base stations can benefit from a cylindrical array arrangement with steering of beams within 360° of coverage instead of using several planar arrays, each covering a fixed sector. For communication with aircraft and satellites from ground stations, even spherical or semispherical antenna arrangements may be required.

Also in *urban environments* – densely populated areas – more and more antennas are used. It is often desired to integrate these antennas into existing structures (lampposts, chimneys, flagpoles, etc.) to make them less obtrusive and more acceptable to the public eye.

Even *modern cars* can benefit from antenna functions integrated into the body (navigation, Internet, radio, television, collision avoidance, and so on).

Communication systems onboard *satellites* sometimes require large angular coverage leading to a need for curved arrays. One example is in low-orbiting, spin-stabilized satellites.

Conformal array antennas are also of interest for *defense applications* (radar, communication, and warning systems). These systems typically require large angular coverage and handling of many simultaneous targets as well as integration of the array antenna in the surface of various vehicles. With conformal arrays a separate radome might not be required.

History

A system of dipole elements arranged on a circle was proposed as an anti-fading broadcasting antenna in the 1930s (Chireix 1936). By sequential phase excitation, omnidirectional radiation was obtained in the horizontal plane, while signals reflected from the ionosphere at higher elevation angles (causing the interference and fading) were reduced. The first system was implemented in France, with only four radiators for cost reasons. Hence, the element phase excitation around the "circle" was 0°, 90°, 180°, and 270°. This is an example of a *phase mode* excitation (more about this later) (section "Array Theory").

Circular dipole arrays, so-called Wullenweber arrays, were developed in the 1950s for signal intelligence gathering and direction finding (Gething 1966). These arrays were quite large with diameters of several hundred meters. Most of them are now dismantled, however, and their functions have been replaced by satellite-borne systems.

Technical Progress

The circular (or ring) array is attractive for many reasons: it is relatively simple and proper excitation schemes enable a directive beam to be steered 360° in azimuth. Pattern synthesis methods based on phase mode theory and beam steering schemes have been developed (Knudsen 1953; Davies 1983). The applications are found in broadcasting, communication, navigation, direction finding, and radar. An interesting radar concept using circular arrays and beamforming on receive has been described by Dorey et al. (1989).

Several circular arrays can be placed on top of each other in order to increase the directivity, as in the electronically scanned TACAN (tactical air navigation) antenna. The antenna radiates a rotating phase-coded signal that helps aircraft to find their position in relation to, e.g., an airfield (Christopher 1974).

Many of the above circular arrays were made up of dipole elements. A dual structure of fundamental interest consists of aperture elements, or slots, in metallic cylinders (Wait 1959). This type has been

DOI 10.1007/978-981-4560-75-7_65-1

investigated using modal expansion techniques and high-frequency diffraction techniques, including mutual coupling effects (Hessel 1970; Pathak et al. 1980).

The step from cylindrical arrays to conical arrays may seem straightforward. Such arrays in the nose area of aircraft or missiles would be of great interest. However, arrays with elements near the pointed tip are difficult to manage (Munger et al. 1974). Another attractive arrangement for aircraft is an array conformal to the shape of the wing leading edge (Steyskal 2002). Also spherical arrays have been proposed (Liebman et al. 1975).

Array elements with associated electronics including feeding and cooling systems, etc. may be difficult to build on a curved surface, especially if the curvature is large and varying over the surface. However, good results have been obtained in many cases by approximating the curved surface with plane facets. Each facet forms a planar subarray with a small number of modular elements (Tomasic et al. 2003).

During the last decades, sophisticated software tools for analysis and optimization of complex electromagnetic structures have been developed. This development has been stimulated by the availability of advanced high-speed computers, without which many conformal arrays could not have been rigorously analyzed. Finite element methods as well as high-frequency diffraction methods have been developed. The uniform theory of diffraction (UTD) has been combined with the theory of geodesic lines on curved surfaces (see section "Analysis").

The steering and control of conformal arrays integrated into the skin of various structures must be matched by the proper processing of transmitted and received signals. Array calibration is another critical aspect. The beamforming function is probably most efficiently done by digital means. Hence, DBF, or digital beamforming, is an area very much related to conformal array technology, at least for more complex systems. Of course, analog beamforming will still find its place (see section "Beamforming in Conformal Array Antennas").

Conformal array antenna technology has been the focus of several workshops, conferences, and publications; see the reference list at the end of this chapter.

Array Theory

ULA Versus UCA

The *uniform linear array*, or ULA, can be seen as the fundamental building block of planar array antennas. Similarly, the *uniform circular array*, or UCA, is the fundamental building block of conformal rotationally symmetric array antennas, including cylindrical and conical shapes. The two arrays are illustrated in Fig. 1. Much of the theory for circular arrays applies to cylindrical arrays as well.

It should be observed that the term *circular array* is sometimes understood as a circularly shaped planar array (cf. rectangular array, elliptical array, etc.). However, this is not the meaning here. Another word for the UCA is *ring array*.

There is an important difference between ULAs and UCAs regarding the influence of the element patterns. In the linear array all elements point in the same direction. The ULA radiation pattern function can be written as the product of an element factor and the array factor:

$$E(\phi) = EL(\phi)\sum_{n} V_n e^{jnkd\sin\phi} \tag{1}$$

DOI 10.1007/978-981-4560-75-7_65-1

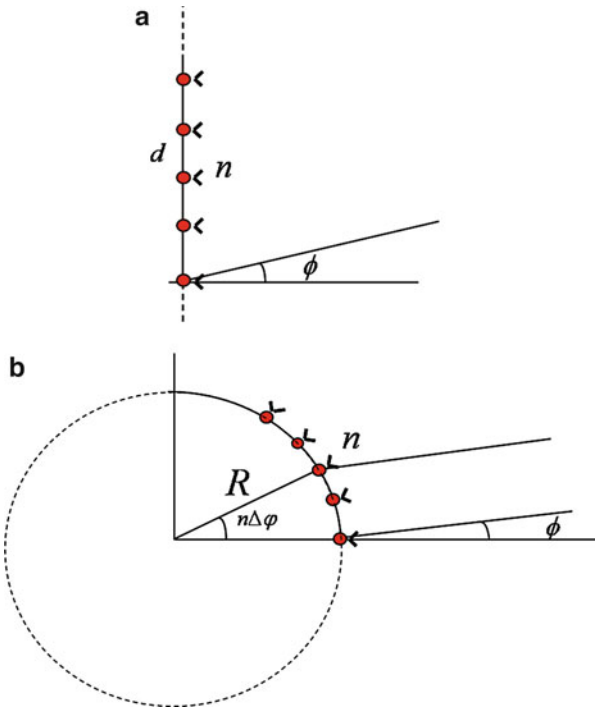


Fig. 1 Uniform linear array (**a**) and uniform circular array (**b**)

where

$EL(\phi)$ is the element pattern function (element factor)
V_n is the element excitation
d is the element spacing
$k = 2\pi/\lambda$

For the UCA an element factor can in general not be factored out; hence,

$$E(\phi) = \sum_n V_n EL(\phi - n\Delta\varphi) e^{jkR\cos(\phi - n\Delta\varphi)} \tag{2}$$

where $\Delta\varphi$ is the element angular spacing and R is the circular array radius.

Let us assume the elements to be isotropic, i.e., the element pattern $EL(\phi) = 1$ in all directions. If the element excitations are all equal, say, $V_n = 1$, the ULA would have a radiation pattern similar to a $\sin x/x$ function with a narrow beam, provided the element spacing d is less than a wavelength. The same assumptions for the UCA result in a more or less omnidirectional pattern in the plane of the array. Two problems may arise for both arrays: grating lobes for the ULA and pattern distortion (ripple) for the UCA. The common cause is a too large element spacing which results in the appearance of undesired higher modes.

In the following the focus will be on the UCA, the circular array antenna. Linear and planar phased array antennas are treated in other chapters.

Phase Mode Theory

An example of a practical UCA is shown in Fig. 2, where radiating elements in the form of vertical dipoles are suggested.

DOI 10.1007/978-981-4560-75-7_65-1

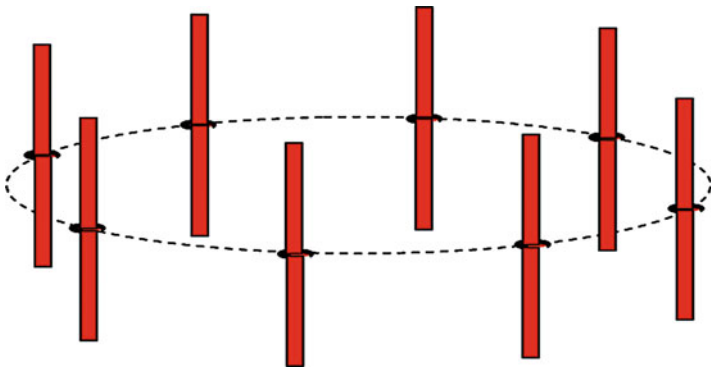


Fig. 2 A uniform circular array with vertical dipole elements

Table 1 Definition of the first few phase modes

Mode 0	Phase = constant	For one turn around the ring
Modes ± 1	Phase change ± 1 × 2π	"
Modes ± 2	Phase change ± 2 × 2π	"
Modes ± m	Phase change ± m × 2π	"
etc.		

The vertical dipoles in Fig. 2 are omnidirectional in the horizontal plane. Since the elements have identical patterns for all positions in the array, the element pattern can be factored out in the radiation function. More important is perhaps that the polarization remains vertical for all directions, and there is no cross polarization. This generic type of array is often used in radio communication and direction finding systems.

An idealized case is a UCA with a *continuous* ring excitation, thus approximating an array with densely spaced isotropic elements. The first phase mode excitations $V(\varphi)$ for this idealized ring are shown in Table 1.

A general excitation of the ring can be expressed as a summation of several phase modes with amplitudes C_m:

$$V(\varphi) = \sum_{-\infty}^{\infty} C_m e^{jm\varphi} \tag{3}$$

which is a Fourier series representation of the periodic function $V(\varphi)$. The period is 2π, one revolution around the ring. The Fourier coefficients C_m are

$$C_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} V(\varphi) e^{-jm\varphi} d\varphi \tag{4}$$

Also the far field must be periodic and can be written as

$$E(\phi) = \sum_{-\infty}^{\infty} A_m e^{jm\phi} \tag{5}$$

Alternatively, the far field can be written (cf. Eq. 2) as

DOI 10.1007/978-981-4560-75-7_65-1

$$E(\phi) = \frac{1}{2\pi} \int_{-\pi}^{\pi} V(\varphi) e^{jkR\cos(\phi-\varphi)} d\varphi \tag{6}$$

Combining Eqs. 3 and 6 gives

$$E(\phi) = \sum_{-\infty}^{\infty} \frac{1}{2\pi} C_m \int_{-\pi}^{\pi} e^{jm\varphi} e^{(jkR\cos(\phi-\varphi))} d\varphi \tag{7}$$

Comparing this with Eq. 5 gives a relation between the coefficients A_m and the coefficients C_m:

$$A_m = C_m \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{jm(\varphi-\phi)} e^{jkR\cos(\varphi-\phi)} d\varphi \tag{8}$$

This can also be written as

$$A_m = j^m C_m J_m(kR) \tag{9}$$

where the function $J_m(z)$ is the Bessel function of the first kind and order m. Thus, the Bessel function has a major impact on the relation between *radiation* and *excitation*, i.e., between the A_m and C_m coefficients. A graph of the Bessel function for the first few orders m is presented in Fig. 3.

From the magnitude of $J_m(kR)$, it is evident that for a given size kR only orders m up to about $m = kR$ will radiate efficiently. This "modal cutoff" is not very sharp for small sizes. However, it is more distinct for larger sizes, e.g., $kR = 100$. The limit $m = kR$ corresponds to a phase variation of 2π per wavelength along the periphery. Twice the number of modes can be used if also negative m values are included.

A second observation from Fig. 3 is that the Bessel function nulls will severely attenuate radiation for arguments kR close to these nulls. This problem can be studied in terms of the mode stability $|A_m/C_m|$. It will be shown that the mode stability improves considerably if directive elements are used.

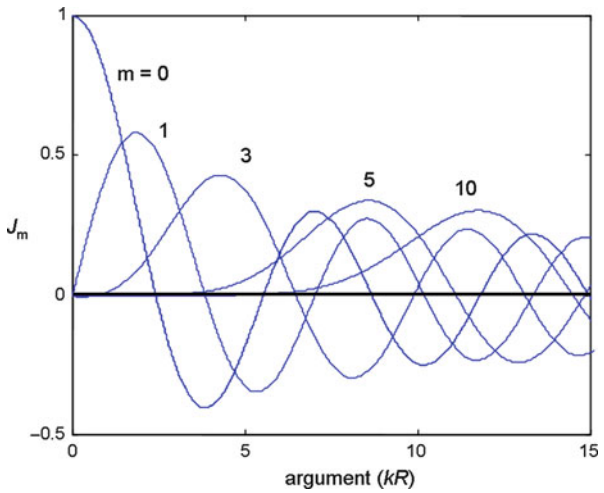


Fig. 3 The Bessel function $J_m(kR)$ for different orders m

DOI 10.1007/978-981-4560-75-7_65-1

Discrete Elements

The continuous ring array studied above approximates a real array with densely spaced radiating elements. A more realistic case is an array with evenly spaced discrete elements. To begin with it is still assumed that the radiating elements are isotropic radiators.

Consider the simplest case of a ring array with only the mode $m = 0$, i.e., a continuous and constant excitation, $V(\varphi) = V_0$. The Fourier expansion, Eq. 4, will have just one spectral term C_0, the "DC component." However, sampling this excitation at N points, corresponding to the N discrete element positions, will result in a line spectrum with harmonics at multiples of N/T, where T is the period 2π. If instead a phase slope $m > 0$ is considered, the same mode spectrum will result after sampling, but shifted by m/T.

A more complex excitation with several modes will form a spectrum of the modes used, repeated at intervals of N/T after sampling. These spectra should not overlap (no aliasing), i.e., it is required that $|m_{\max}| < N/2$. This implies that the number of elements should be large enough so that spectrum overlap is avoided. At the same time the array size should be large enough so that the desired number of modes is free to radiate. Finally, the "Bessel window" should block the undesired higher harmonics. These requirements are repeated below in mathematical form:

$$N > 2|m_{\max}| \tag{10}$$

$$kR > |m_{\max}| \tag{11}$$

$$\alpha kR < N - |m_{\max}| \tag{12}$$

Here it is assumed that a desired mode spectrum extending over a range of $\pm |m_{\max}|$ is used. Note the factor α introduced in Eq. 12, which nominally should be unity, but a factor up to perhaps 1.5 may be needed in some cases.

Figure 4 illustrates a situation with a desired spectrum extending over ± M modes. The related harmonics are not overlapping and the "Bessel window" admits most of the desired phase modes. In this case, assuming $M \approx kR$, an element spacing around the periphery of d/λ <0.5 is typically required.

A UCA with Omnidirectional Pattern

The following parameters are assumed: eight isotropic elements, uniform amplitude, one mode $m = 0$, and element spacing $d/\lambda = 0.65$. This gives $kR = Nd/\lambda = 5.2$. Summing up the element contributions (Eq. 2) gives the pattern below.

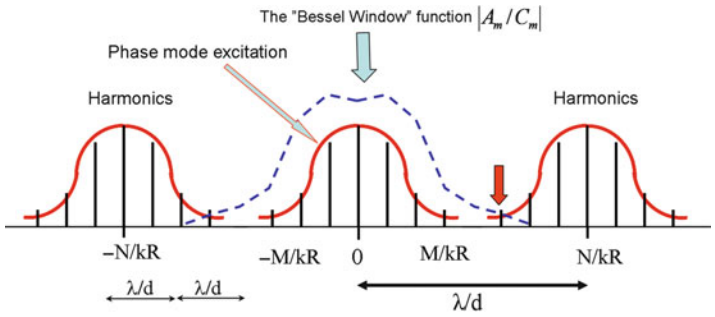


Fig. 4 Mode spectra: desired spectrum (*center arrow*) and part of the harmonic spectrum (see *right arrow*) causing some interference

DOI 10.1007/978-981-4560-75-7_65-1

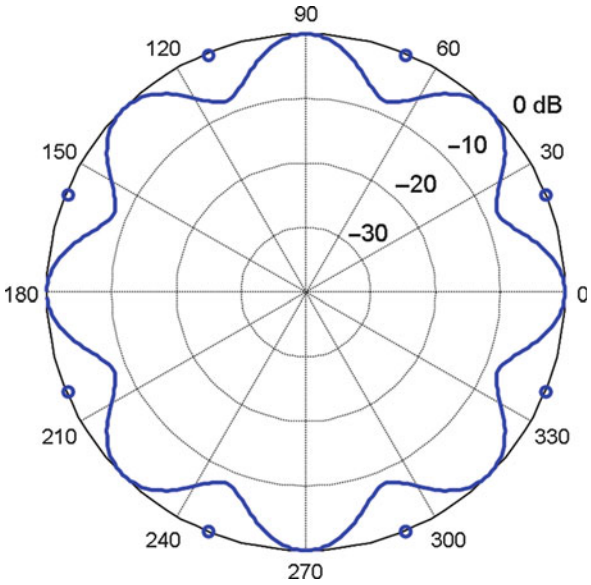


Fig. 5 Pattern from an eight-element UCA with $kR = 5.2$, isotropic elements (Josefsson and Persson 2006) (Copyright IEEE)

The "omnidirectional" pattern in Fig. 5 exhibits a rather severe ripple of about 8 dB. The reason is interference from two higher modes $m = \pm 8$ that are not sufficiently attenuated by the "Bessel window" (Fig. 4). The mode relative amplitudes (radiated) are found to be -0.215, 1.0, and 0.215.

Directive Beam

A second example is a UCA with isotropic elements, uniform amplitude, focused to radiate a narrow beam in the $\phi = 0$ direction. This implies that the elements are given a "cophasal" excitation. In the continuous case this results in an ideal pattern with -8 dB sidelobes:

$$E(\phi) = J_0\left(2kR\sin\frac{\phi}{2}\right) \tag{13}$$

This ideal pattern is shown in Fig. 6 together with a discrete example for the same kR value: 16 elements and $d/\lambda = 0.6$.

Directive Elements

With isotropic elements there was no element factor, as in Eq. 6. With the element factor inserted, the following far-field pattern is obtained:

$$E(\phi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} V(\varphi)EL(\phi-\varphi)\mathrm{e}^{jkR\cos(\phi-\varphi)}d\varphi \tag{14}$$

The element pattern is the same for all elements although the elements point in different directions. Since it is periodic, period 2π, the element pattern can be written as

$$EL(\alpha) = \sum_{p=-\infty}^{\infty} D_p e^{jp\alpha} \tag{15}$$

DOI 10.1007/978-981-4560-75-7_65-1

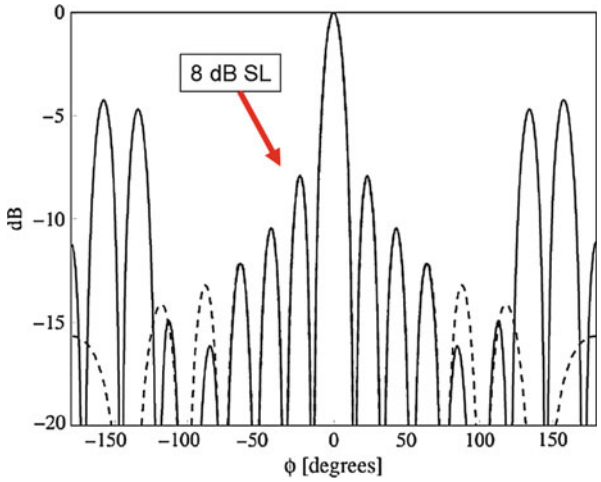


Fig. 6 Cophasal patterns for UCA with $kR = 9.6$. *Dashed curve*: Ideal case, continuous ring. *Solid line*: Discrete case with 16 elements and $d/\lambda = 0.6$ (Josefsson and Persson 2006) (Copyright IEEE)

Inserting this into Eq. 14 yields

$$E(\phi) = \frac{1}{2\pi} \int_{-\pi}^{\pi} V(\varphi) \sum_{p=-\infty}^{\infty} D_p e^{jp(\phi-\varphi)} e^{jkR\cos(\phi-\varphi)} d\varphi \tag{16}$$

Now, the excitation $V(\varphi)$ can be expanded in a Fourier series (Eq. 3). Taking out just the contribution from phase mode number m, the result is

$$E_m(\phi) = C_m e^{jm\phi} \sum_{p=-\infty}^{\infty} D_p \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j(m-p)(\varphi-\phi)} e^{kR\cos(\varphi-\phi)} d\varphi \tag{17}$$

or

$$E_m(\phi) = C_m \left[\sum_{p=-\infty}^{\infty} j^{m-p} J_{m-p}(kR) \right] e^{jm\phi} = A_m e^{jm\phi} \tag{18}$$

Thus, a new relation between C_m (excitation) and A_m (radiation) results (cf. Eq. 9). The mode stability ratio $|A_m/C_m|$ is much more favorable here since the expression within brackets in Eq. 18 has no nulls within the region of most interest, which makes the design less critical, and more bandwidth is attainable.

Much of the results for the UCA, the uniform circular array, apply also to cylindrical arrays.

Phase Mode Impedance

As in any phased array antenna, the mutual coupling among the elements plays an important role; several examples will be given in section "Design." In the UCA, the uniform circular array, there is a symmetric structure, and it is possible to define a phase mode impedance, i.e., the active impedance for a particular phase mode m excitation:

DOI 10.1007/978-981-4560-75-7_65-1

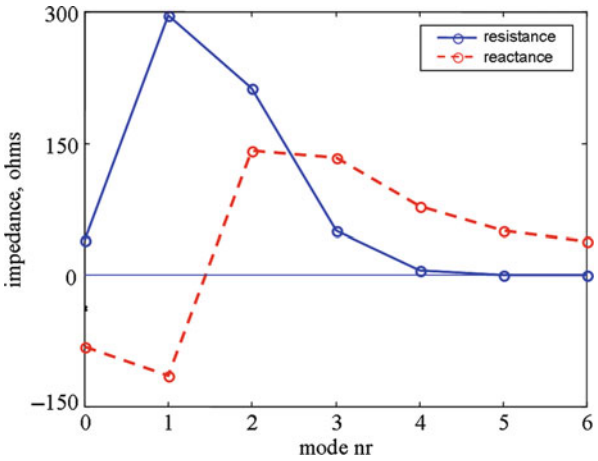


Fig. 7 Phase mode impedance vs mode order. UCA with 16 half-wavelength dipoles, $kR = 2.67$

$$Z^m = \sum_{k=1}^{N} Z_{ik} e^{2\pi j(k-1)m/N} \tag{19}$$

Here Z_{ik} is the mutual impedance between elements i and k. Because of the symmetry the impedance Z^m is independent of i and the same for all elements.

An example is shown in Fig. 7 for a 16-element circular array with parallel dipole elements spaced $d = \lambda/6$. The dipole mutual impedances Z_{ik} can be calculated from well-known formulas; see, e.g., Elliott (1981, p. 332) and Kraus (1988, p. 424). Since $kR = Nd/\lambda \approx 3$, in this case modes up to about $m = 3$ can be expected to radiate. This is also reflected in the figure; for higher modes the mismatch increases as the resistance approaches zero. Compare also with the Bessel function J_m (Fig. 3).

Doubly Curved Arrays

The ultimate conformal array antenna is a "smart skin" mounted on a part of an aircraft body or some other given structure. If the surface is only slightly curved, the design procedure might follow that of planar arrays (with some phase corrections due to curvature). The radiating elements would have to be directional, pointing outward from the surface. With increasing curvature it is obvious that all elements will not radiate efficiently in all scan directions. There would typically be an active sector to be moved (commutated) as the beam direction is changed.

Many systems require hemispherical coverage, for example, a ground-based surveillance radar or communication system, where the desired coverage is from zenith to horizon. Among the possible conformal array shapes are the half sphere, the cone, the paraboloid, and also the three- or four-sided pyramid. The latter, however, with planar surfaces, is not as efficient as the curved alternatives.

On the doubly curved arrays, it is not possible to lay out a perfectly periodic element grid. However, a reasonable quasiperiodic grid is possible, as the following element lattice on a paraboloid shows (Fig. 8).

The example shows circular elements laid out in rings (each ring is a UCA). A practical maximum look direction for each element is about $\pm$ 60° from the normal of the surface. For a beam direction 30° from zenith, this means that in this case only about half of the elements can contribute in this direction.

Another interesting shape is the half sphere, which can be thought to match quite well to the requirement for hemispherical coverage. A suitable configuration can be developed on the basis of one of the regular polyhedral shapes: the octahedron, the dodecahedron, and the icosahedron. These shapes approximate a sphere with 8, 12, and 20 equal plane facets, respectively. The last one with 20 triangular facets can be modified giving 12 pentagonal and 20 hexagonal surfaces (truncated icosahedron). The

DOI 10.1007/978-981-4560-75-7_65-1

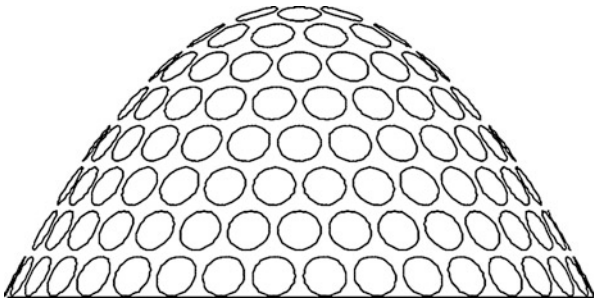


Fig. 8 An element lattice on a rotationally symmetric, parabolic surface (Josefsson and Lanne 2003)

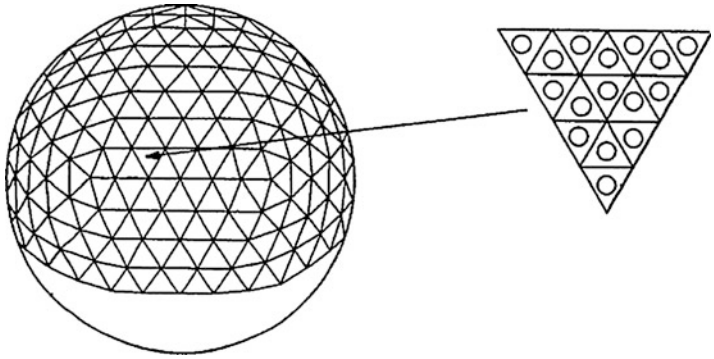


Fig. 9 A spherical surface built from triangular facets (Bondyopadhyay 2002) (Copyright IEEE)

facets can be subdivided into triangles projected onto the surrounding sphere where they can be further subdivided, etc. The result is a quasiperiodic lattice with some variations in element distances (see Fig. 9). All of these doubly curved arrays will have to use some kind of polarization control system or use circular polarization.

Analysis

Introduction

In the previous section UCA was introduced to present the fundamentals behind conformal antennas. So far the antenna elements considered have been positioned in free space, but in many applications there is a supporting structure as well. To fully predict the characteristics, the supporting structure should be included in the analysis.

This is not a trivial task since conformal array antennas can take many shapes. Furthermore, the surface is often electrically large (in terms of wavelengths) and it may be convex or concave or both. As illustrated in Fig. 9, faceted surfaces are also an alternative and this introduces edges or other discontinuities. Depending on the antenna design, dielectric layers may also need to be considered in addition to a PEC (perfect electric conductor) surface. Altogether, the surface can be very complex.

To some extent analytical methods can be used to derive the equations needed to solve the electromagnetic problem. But numerical methods will be needed to do the final analysis. Typically a combination is used, i.e., one uses analytical methods to take the first steps but then numerical methods need to be used. Many of the results presented in this chapter are based on such analyses. It should however be noted that today many of the commercially available codes such as HFSS and CST can analyze very complex and large surfaces. Hence, with the available computer power of today, a large surface is not as difficult to

DOI 10.1007/978-981-4560-75-7_65-1

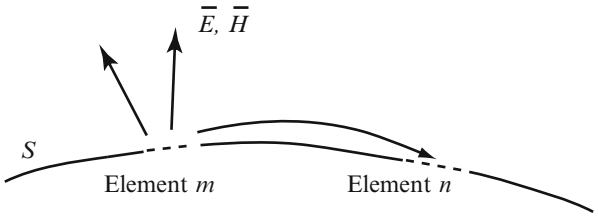


Fig. 10 The problem: find the field outside the curved surface *S* due to one or more radiating elements on the surface *S* (Josefsson and Persson 2006) (Copyright IEEE)

analyze due to memory requirements and time consumption as it was some years ago. Nevertheless, the focus here is to give a brief overview of some of the methods used traditionally when analyzing conformal antennas. Understanding the methods and theory behind is a very good starting point and makes it easier to also understand the results.

Note that the focus in this section is not to describe different analysis methods in detail, but to address some key aspects when analyzing conformal antennas.

The Electromagnetic Problem

As a start it is important to understand the problem to be solved and the associated electromagnetic problem. When having found a solution to the electromagnetic problem, it is possible to start analyzing the conformal (array) antenna. For the radiation problem, some approximations can be used when considering the far-zone region. A brief overview of circular array antenna theory was presented in the previous section to give an understanding about the fundamentals. Furthermore, the mutual coupling among the elements is essential in order to predict the characteristics of (conformal) array antennas. The electromagnetic solution to the problem also gives us the tool to study this, i.e., the surface field or the near field in order to find the mutual coupling among the radiating elements. This analysis is often difficult since no far-field approximations are possible.

Figure 10 illustrates the electromagnetic problem to be solved. It is a boundary value problem and now the field representation outside the surface *S* needs to be matched with the supposedly known field or current distribution at the antenna elements. When the solution has been found, the Green's function $\left(\overline{\overline{G}}\right)$ has been found for the antenna problem. Thus, the field outside *S* is related to the radiating element current $\left(\overline{J}\right)$ simply by the following general formula:

$$\overline{E} = \sum_{n=1}^{N} \int_{S_n} \overline{\overline{G}} \cdot \overline{J} dS \tag{20}$$

where the sum is taken over the number of elements considered. Unfortunately, the number of geometries where an explicit field representation outside *S* can be found is very limited. In many cases asymptotic expressions are needed, and in some cases a numerical representation is the only possible solution.

Calculating the mutual coupling between two radiators implies finding the field at element 2 due to the source at element 1 and then matching the boundary conditions at 2. This can typically be expressed as

$$Y_{12} = \frac{1}{V_1 V_2} \int_{S_2} \overline{E}_2 \times \overline{H}_2(1) d\overline{S} \tag{21}$$

DOI 10.1007/978-981-4560-75-7_65-1

By taking out two radiators like this out of a large array, with the other radiators short-circuited in case of apertures (or open-circuited for dipole-type elements), Eq. 21 represents an element in the mutual admittance (impedance) matrix. From this matrix the scattering matrix can be derived for the full array, i.e.,

$$\mathbf{S} = (\mathbf{I} - \mathbf{Y})(\mathbf{I} + \mathbf{Y})^{-1} \text{ or } \mathbf{S} = 2(\mathbf{I} + \mathbf{Y})^{-1} - \mathbf{I} \tag{22}$$

Modal Solutions: The Classical Solution

The classical way of analyzing conformal antennas is to use separation of variables and express the solution in a modal representation, i.e., in a vector harmonic series representation. This is possible only for special shapes, such as a perfectly conducting circular cylinder, elliptic cylinder, cone, or sphere. In the book *Electromagnetic Radiation from Cylindrical Structures* by J. R. Wait, a very good summary is found on the early work on cylinders (Wait 1959).

To briefly illustrate the technique of using separation of variables, an example will be given by using a circular cylinder. The cylinder is assumed to be infinitely long and perfectly conducting. In this example the antenna element is an aperture antenna. It is assumed that the tangential electric field distribution ($E_{\tan}$) on the surface is known as indicated in Fig. 11.

As a first step the problem is reduced to a two-dimensional problem by applying a Fourier transform with respect to the symmetry axis (here the cylinder or z-axis). Thus, define $\tilde{E}_{\tan}$ as a Fourier transform of the tangential components of $\bar{E}$ as

$$\tilde{E}_{\tan}(n, k_z) = \frac{1}{2\pi} \int_0^{2\pi} \int_{-\infty}^{\infty} E_{\tan}(r, \varphi, z) e^{-j(n\varphi + k_z z)} dz d\varphi \tag{23}$$

with the inverse given by

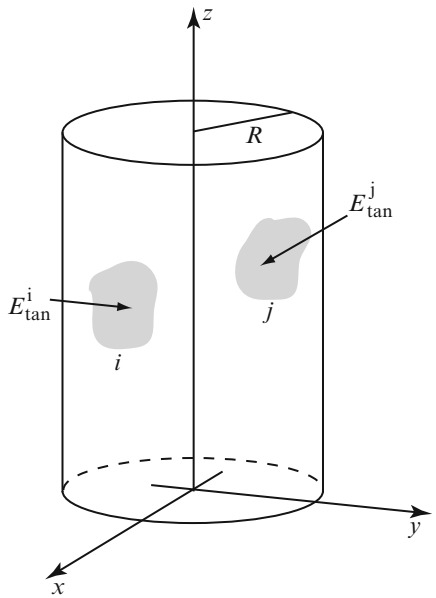


Fig. 11 Two apertures (i and j) in a conducting circular cylinder (Josefsson and Persson 2006) (Copyright IEEE)

DOI 10.1007/978-981-4560-75-7_65-1

$$E_{\tan}(r,\varphi,z) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{E}_{\tan}(n,k_z)e^{j(n\varphi+k_z z)}dk_z \tag{24}$$

The wave functions A_z and F_z are now constructed in a similar fashion as Eq. 24 but expressed in terms of the second-order Hankel function $H_n^{(2)}(k_r r)$ representing outward traveling waves. Using the boundary conditions, it is possible to find field expressions that are valid outside the cylinder $(R \le r < \infty)$ in terms of the known Fourier transform of the tangential electric field at the apertures. The electric field and the corresponding magnetic field are given in Eqs. 25–28:

$$E_\varphi(r,\varphi,z) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}\left[g_n(k_z)k_r H_n'^{(2)}(k_r r) + \frac{nk_z}{j\omega\varepsilon r}f_n(k_z)H_n^{(2)}(k_r r)\right]e^{-jk_z z+jn\varphi}dk_z \tag{25}$$

$$E_z(r,\varphi,z) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{k_r^2}{j\omega\varepsilon}f_n(k_z)H_n^{(2)}(k_r r)e^{-jk_z z+jn\varphi}dk_z \tag{26}$$

$$H_\varphi(r,\varphi,z) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}\left[-f_n(k_z)k_r H_n'^{(2)}(k_r r) + \frac{nk_z}{j\omega\mu r}g_n(k_z)H_n^{(2)}(k_r r)\right]e^{-jk_z z+jn\varphi}dk_z \tag{27}$$

$$H_z(r,\varphi,z) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}g_n(k_z)H_n^{(2)}(k_r r)e^{-jk_z z+jn\varphi}\frac{k_r^2}{j\omega\mu}dk_z \tag{28}$$

where

$$f_n(k_z) = \frac{j\omega\varepsilon\tilde{E}_z(n,k_z)}{k_r^2 H_n^{(2)}(k_r R)} \tag{29}$$

$$g_n(k_z) = \frac{1}{k_r H_n'^{(2)}(k_r R)}\left[\tilde{E}_\varphi(n,k_z) - \frac{k_z}{k_r^2}\frac{n}{R}\tilde{E}_z(n,k_z)\right] \tag{30}$$

and $k_r^2 = k^2 - k_z^2$.

The above solution has poor convergence if the radius of the cylinder is electrically large ($kR \gg 1$). This can be understood by recognizing the harmonic series as a series of standing waves around the cylinder. Therefore, the required number of terms in the summation over n is directly related to the size of the cylinder – many terms are necessary for an electrically large cylinder.

Now, the integral (in k_z) which represents the inversion to obtain the 3D solution from the spectrum of 2D solutions has to be solved. In general this must be done numerically, except for the far zone $(r \to \infty)$ where the radiating field can be found by using an asymptotic formula for the k_z-integral (Harrington 1961). Thus, the quite simple far-field expressions obtained from Eqs. 25 to 26 (when rewritten in spherical coordinates) are

DOI 10.1007/978-981-4560-75-7_65-1

$$\begin{pmatrix} E_\theta \\ E_\varphi \end{pmatrix} = \begin{pmatrix} j\omega\mu \\ -jk \end{pmatrix} \frac{e^{-jkr}}{\pi r} \sin\theta \sum_{n=-\infty}^{\infty} e^{jn\varphi} j^{n+1} \begin{pmatrix} f_n(-k\cos\theta) \\ g_n(-k\cos\theta) \end{pmatrix} \tag{31}$$

If the above (aperture) antenna is coated with one or more dielectric layers, the procedure can be extended to include these layers, although the complexity of the solution increases as well as the numerical burden. Explicit formulas for a single layer can be found in, e.g., Persson and Rojas (2003) and Thors and Rojas (2003), and multiple layers are treated in, e.g., Sipus et al. (1998).

High-Frequency Methods

As indicated earlier some special techniques may be needed when the surface becomes electrically large due to numerical problems. One way of avoiding these problems is to use a high-frequency approach. This means that asymptotic techniques are used to find an approximate solution. A well-known high-frequency method is the uniform theory of diffraction (UTD), which is a ray-based theory. Key references in the area are Pathak and Wang (1981), Pathak et al. (1980, 1981), and Pathak (1992). The reader is referred to these papers for explicit equations of the surface fields and radiated fields due to sources on smoothly shaped convex PEC surfaces.

Some of the approximations used for non-coated surfaces turn out to give non-satisfactory results when a coating is applied. Several attempts have been made, but some of the first ones to actually succeed in the analysis are presented in Ertürk and Rojas (2000) and Ertürk et al. (2004) where the microstrip antenna is considered, and the coated aperture antenna is solved in Persson and Rojas (2003) and Thors and Rojas (2003). Some later advances in the evolution of high-frequency methods can be found in Alisan and Ertürk (2010) and Akyuz et al. (2011).

Geodesics

Before concluding this part, it is important to highlight one of the key aspects of ray-based high-frequency methods: the geodesics. The geodesic is the key and will be needed in order to find how the electromagnetic field is propagating from one position to another on a curved surface. It should, however, be stated clearly that a geodesic is not just the shortest line, or any arbitrary line, connecting two points on a surface. A geodesic needs to fulfill certain requirements. To summarize, a definition of a geodesic only based on the length is not a useful one. A more general definition of a geodesic is given in Struik (1988) where a geodesic is defined as "a curve of zero geodesic curvature." This definition can also be expressed as: "the osculating plane of a geodesic contains the surface normal" or "along the geodesic the principal normal $(\hat{n})$ coincides with the surface normal $(\hat{N})$." Figure 12 illustrates a situation when a curve along the surface is not a geodesic.

Note that the normal property strictly fails in the case of a straight line on a (plane) surface, since the principle normal $\hat{n}$ is indeterminate. But, still, the straight line on flat surface is a geodesic. This fact, according to some authors writing about geodesics, has been taken as a reason for not using the above general definition of a geodesic. However, this definition will be adopted in this text since it gives an intuitive way of defining geodesics on curved surfaces.

By using differential geometry, it is possible to derive the geodesic equation that will solve the problem for any type of surface (Struik 1988). To solve this equation, extensive numerical analysis is needed, but fortunately it is possible to reduce the numerical complexity by restricting the analysis to certain types of surfaces such as cylinders, spheres, paraboloids, and hyperboloids. All of these surfaces are in fact very useful and can serve as valid approximations of a realistic surface to be considered. Assuming this, the geodesic equation can be reduced and in fact be solved analytically.

DOI 10.1007/978-981-4560-75-7_65-1

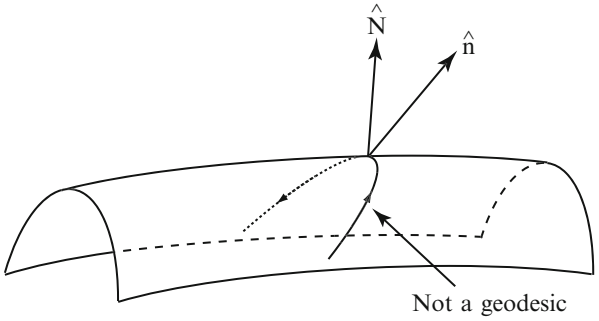


Fig. 12 A curve along a surface but not a geodesic surface ray (Josefsson and Persson 2006) (Copyright IEEE)

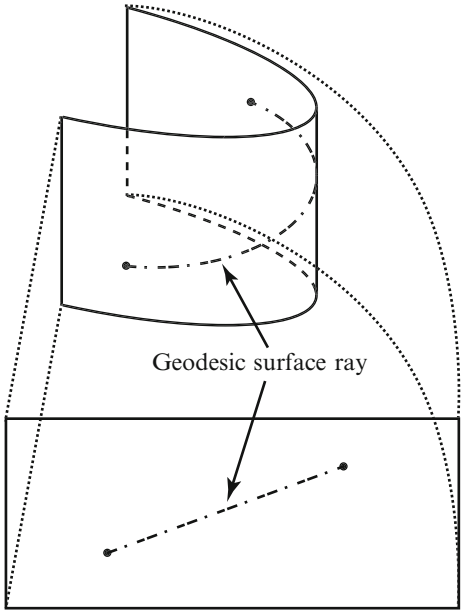


Fig. 13 A developed cylindrical surface (Josefsson and Persson 2006) (Copyright IEEE)

To better understand this, let us start with singly curved surfaces. Singly curved surfaces may also be denoted developable surfaces since they can be unfolded and analyzed as a flat surface and the key parameters are obtained from a two-dimensional analysis; see Fig. 13. As a consequence the geodesics on any singly curved surface can always be found, at least numerically. Thus, the singly curved surfaces can be studied without solving the geodesic differential equation.

Next, doubly curved surfaces will be considered, in particular rotationally symmetric surfaces. Solving the geodesic equation for doubly curved surfaces is of course more difficult than for singly curved surfaces. Often the equation has to be solved numerically resulting in a time-consuming analysis. However, as will be shown here, the geodesic equation for rotationally symmetric doubly curved surfaces can in many cases be solved in closed form. This simplifies the analysis of doubly curved surfaces and makes it possible to use a ray-based theory to analyze doubly curved conformal array antennas.

As a start, the surface needs to be represented as a parametric equation. For doubly curved rotationally symmetric surfaces, the parameterization is written as

$$x = f(u)\cos v, \quad y = f(u)\sin v, \quad z = g(u) \tag{32}$$

DOI 10.1007/978-981-4560-75-7_65-1

and the solution to the geodesic equation for rotational symmetric doubly curved surfaces then becomes

$$v(u) = \int \frac{\pm\alpha\sqrt{(f'(u))^2 + (g'(u))^2}}{f(u)\sqrt{(f(u))^2 - \alpha^2}} du + \beta = \mathcal{I}(u, \alpha) + \beta \tag{33}$$

The constants α and β are unique for a specific geodesic between the given points A and B. They can be obtained by solving

$$\begin{aligned} v_A &= \mathcal{I}(u_A, \alpha) + \beta \\ v_B &= \mathcal{I}(u_B, \alpha) + \beta \end{aligned} \tag{34}$$

The geodesics on any doubly curved rotationally symmetric surface can now be found by quadrature. However, analyzing doubly curved surfaces gets even more interesting when it turns out that the integral in Eq. 33 can be solved in closed form for many rotationally symmetric doubly curved surfaces, in particular those belonging to the Eisenhart coordinate system. In fact, Eq. 33 can actually be solved also for some non-Eisenhart surfaces as long as they are defined in the so-called geodesic coordinate system, i.e., the parameterization is such that the parameter lines are orthogonal (Jha and Wiesbeck 1995). One example of a non-Eisenhart surface is the ogive (Jha et al. 1989).

One thing to keep in mind is that for a closed, singly curved surface, the number of geodesics connecting two points is infinite. For doubly curved surfaces, on the other hand, the number of geodesics can be more than one, but it is bounded. Just illustrating the complexity of finding geodesics on a doubly curved surface, Fig. 14 shows a general paraboloid of revolution (GPOR) where one first-order geodesic is found in the positive φ-direction as well as a nontrivial second-order geodesic in the positive φ-direction.

Hybrid Methods

Before concluding this section, it is worth mentioning that a single method (separation of variables, a high-frequency method, etc.) is seldomly used alone. Often hybrid methods are used where one tries to combine the good things with several methods in order to analyze a complex problem. A very good combination is to use the numerical method called method of moments (MoM) together with a high-frequency method. A high-frequency method such as UTD can analyze electrically large objects; it is quite simple to use and can incorporate other solutions into its format. The MoM is useful for small objects since the unknowns become small in number; it can treat arbitrarily shaped objects and can give information about antenna parameters. The major disadvantage is that asymptotic formulations are still to be found for certain geometries. Nevertheless, combining UTD and MoM is probably the hybrid solution most often used when analyzing conformal antennas.

The main idea with a hybrid UTD-MoM method is to start with an exact formulation in terms of an integral equation formed at the boundary between the defined exterior and interior regions. The kernel of this integral equation is the exact Green's function, now replaced with an asymptotic version. Note that the asymptotic Green's function is not valid in the near field of the antenna (i.e., when the distances between the source and observation points are small). Often, the Green's function of the corresponding planar case is used to overcome this limitation. After this step the solution to the integral equation is not any different from solving traditional integral equations with the MoM procedure.

DOI 10.1007/978-981-4560-75-7_65-1

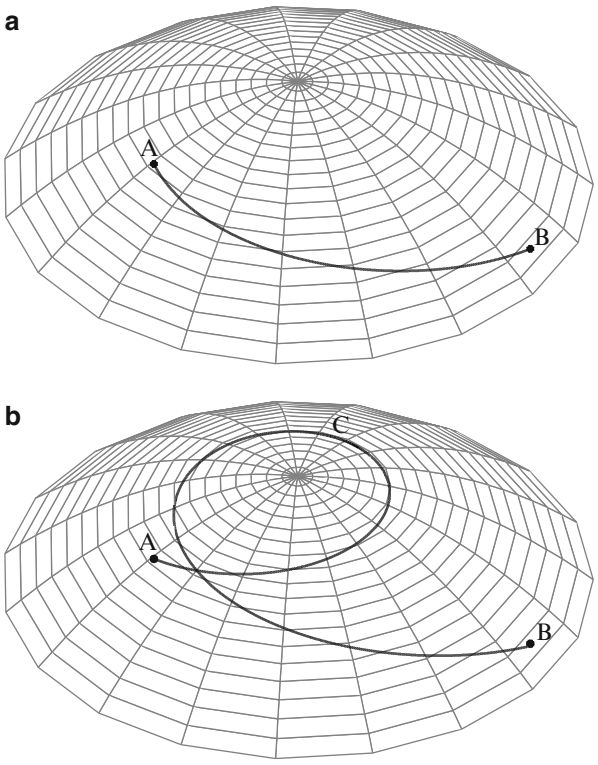


Fig. 14 A GPOR where $\alpha = 1.6$, ($v_A = 4$, $\varphi_A = 20°$) and ($v_B = 6.5$, $\varphi_B = 140°$). (**a**) The first-order geodesic in the positive φ-direction. (**b**) The nontrivial second-order geodesic in the positive φ-direction (Josefsson and Persson 2006) (Copyright IEEE)

Design

Introduction

In this section examples will be given to illustrate some of the characteristics of conformal antennas. The results have been obtained by using a hybrid UTD-MoM method unless otherwise stated. The purpose is to provide some fundamental knowledge about the characteristics of conformal antennas so that the reader can use this knowledge when advancing further into the topic, perhaps by analyzing more advanced conformal array antennas.

The examples to be considered are two conformal antennas with aperture-type radiating elements flush mounted on a conducting circular cylinder and a paraboloid of revolution, respectively. These two surfaces represent two important classes of conformal antennas. The analysis of such canonical problems provides insight into many of the characteristics of singly curved and doubly curved conformal antennas.

The Circular Cylinder

In the examples to follow, the circular cylinder is assumed to be infinitely long. The apertures have a rectangular shape with either axial or circumferential orientation. The aperture field is given by a combination of rectangular waveguide modes. In many cases a single-mode approximation is enough, but in the text several waveguide modes will be used to demonstrate their influence and importance if accurate results are required.

The geometry and all relevant parameters are defined in Fig. 15. The radius of the cylinder is denoted by R and equals 5λ; in most of the cases to be presented, the apertures are described by the width and height, w and h. s is the geodesic length between the centers of the apertures. The angle α is also defined in Fig. 15.

DOI 10.1007/978-981-4560-75-7_65-1

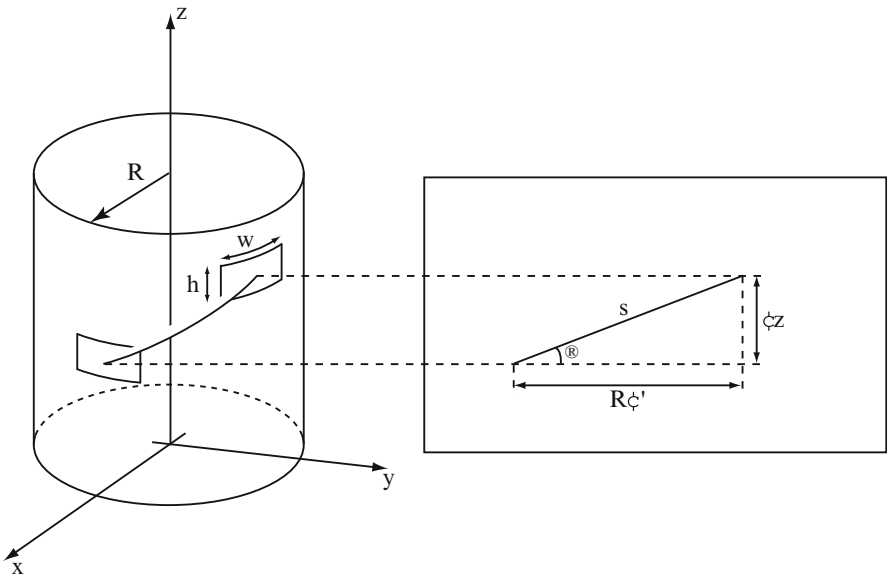


Fig. 15 Geometry for the study of the characteristics of azimuthally oriented aperture antenna elements on circular cylinders, i.e., vertical polarization (Josefsson and Persson 2006) (Copyright IEEE)

The apertures are linearly polarized (excited by the fundamental TE_{10} mode). Figure 15 shows a typical case with $w > h$ and axial (vertical) polarization. Axially oriented apertures, hence $w < h$, represent a case with circumferential (horizontal) polarization.

Isolated Mutual Coupling

The first example is about analyzing the isolated coupling, i.e., the mutual coupling between two elements only. The aperture dimensions are $0.27\lambda \times 0.65\lambda$. Figure 16 shows the *E*-plane coupling (i.e., $\alpha = 0°$) as a function of the cylinder radius. The following radii are used: $R = 2\lambda$, 5λ, 10λ, 20λ, and 50λ. The results are shown versus the geodesic length between the apertures. The corresponding planar result $(R \to \infty)$ is included for comparison.

As indicated in Fig. 16, the circular cylinder solution approaches the planar solution when the radius is increased, as expected. Another interesting result is that the planar solution seems to be surprisingly accurate for small distances (this also applies for H-plane coupling not shown here; see Josefsson and Persson (2006) for more results and details). This is a feature often used when analyzing arrays on curved surfaces. By using the planar solution, approximate values of the mutual coupling can be found easily for neighboring elements as well as between elements along the axis. However, the effect of the curvature of the surface is quite strong for larger distances. For a 50λ radius, the planar solution and the cylinder solution deviate considerably after a distance of approximately $6-8\lambda$.

Array Mutual Coupling

The next step is to study the *array* mutual coupling, i.e., the coupling among the elements in the array environment with all elements present and terminated in matched loads. At this point higher-order modes will be used in the calculations to increase the accuracy of the aperture field solution. The calculated results will be compared with measured data (Lindgren and Josefsson 1998) obtained from an experimental antenna built by Ericsson Microwave Systems AB in Mölndal, Sweden, as a test bed for singly curved conformal antennas. The surface is a circular cylinder with three rows of rectangular apertures (see Fig. 17). The main dimensions are the following:

DOI 10.1007/978-981-4560-75-7_65-1

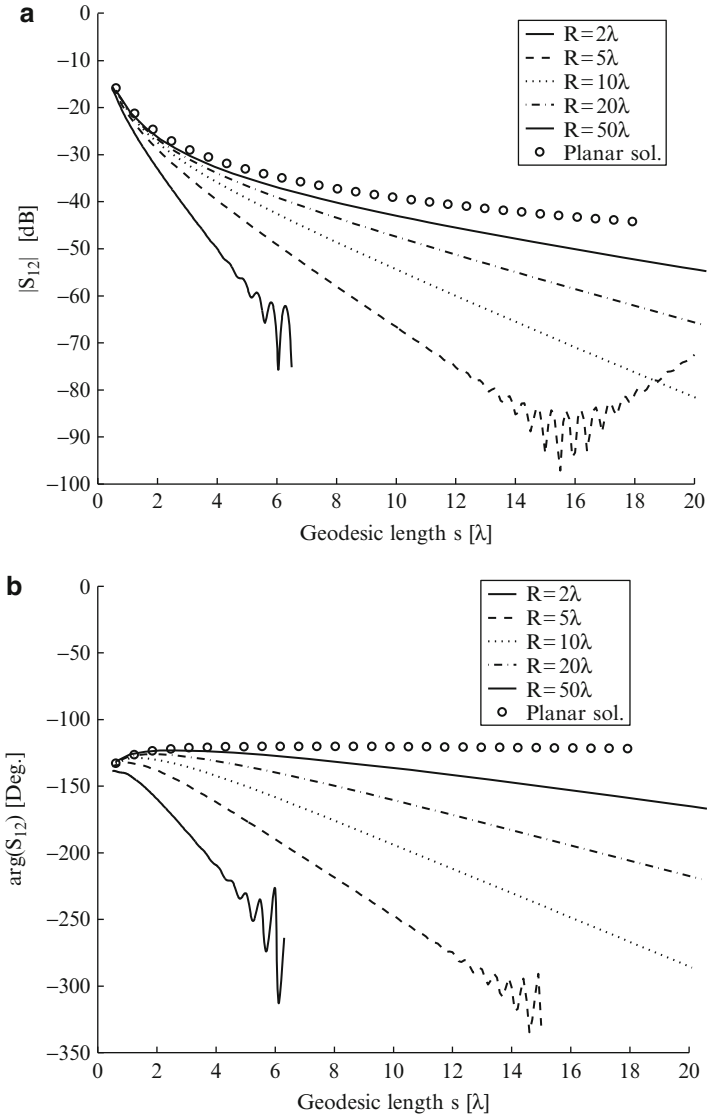


Fig. 16 The isolated *E*-plane mutual coupling along the circumferential direction of a cylinder with different radii. (**a**) The amplitude. (**b**) The phase, with the free space wave contribution subtracted (Josefsson and Persson 2006) (Copyright IEEE)

Element apertures	39 × 16 mm
Element grid	Rectangular, 3 × 18 elements
	Extending over about 120°
Element spacings	37.1 mm in azimuth
	41 mm in axial direction
Polarization	Circumferential
Cylinder radius	300 mm
Frequency interval	4.3 → 6.8 GHz
Cutoff freq. of second-order mode	$f_{c,20} \approx 7.6$ GHz
Frequency used here	$f = 5.65$ GHz

In the calculations no surface rays encircling the cylinder were taken into account since the cylinder is truncated at the rear due to practical reasons. Absorbers were placed on all edges to minimize edge effects.

DOI 10.1007/978-981-4560-75-7_65-1

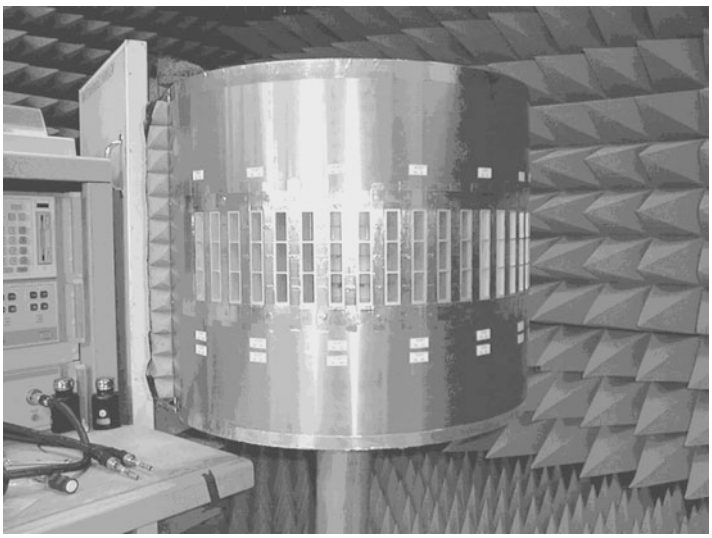


Fig. 17 The experimental antenna (Courtesy Saab AB)

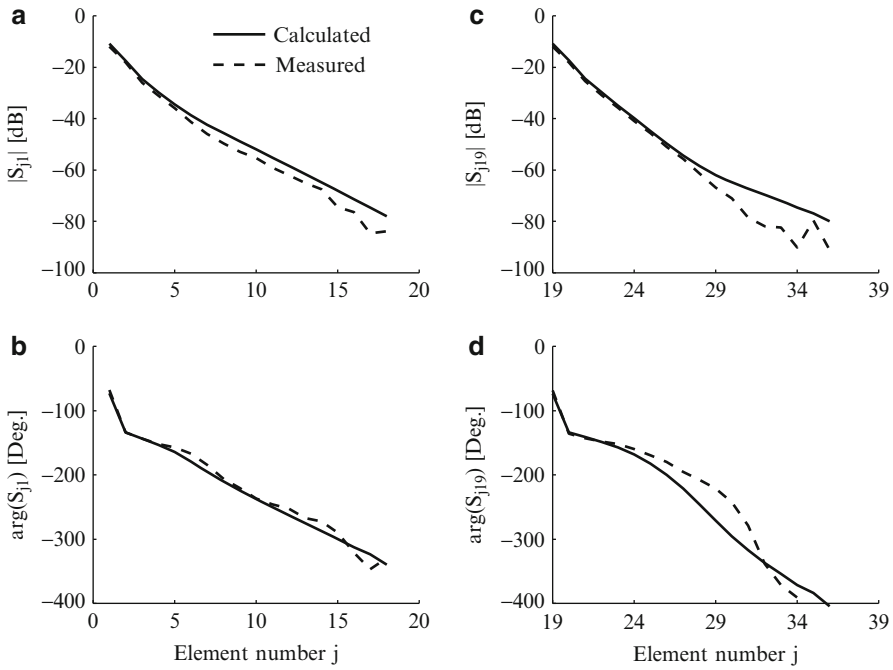


Fig. 18 The mutual coupling in the *E*-plane for the first row (elements 1 → 18) and the second row (elements 19 → 36). Single-mode approximation, $f = 5.65$ GHz. *Solid line*: Calculated. *Broken line*: Measured. (**a**) Amplitude, first row. (**b**) Phase, first row. (**c**) Amplitude, second row. (**d**) Phase, second row (Josefsson and Persson 2006) (Copyright IEEE)

The following numbering of the apertures is used: the aperture in the lower-left corner of Fig. 17 is aperture number 1; aperture number 54 is located in the upper-right corner.

In Fig. 18 the array mutual coupling is presented for $f = 5.65$ GHz. Additional results for other frequencies can be found in Persson and Josefsson (2001). Figure 18 shows the result for the mutual coupling in the *E*-plane for the first row of elements (1 → 18) and for the second row of elements (19 → 36) using a single-mode approximation. As can be seen, there are some differences between the first and second row depending on the different element surroundings. Furthermore, even with a

DOI 10.1007/978-981-4560-75-7_65-1

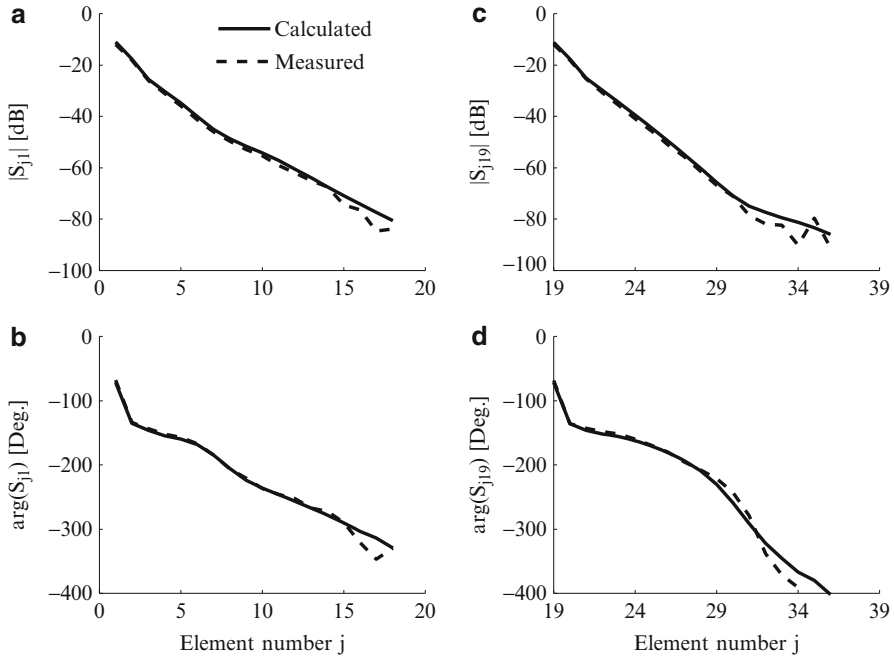


Fig. 19 The mutual coupling in the *E*-plane for the first row (elements 1 → 18) and the second row (elements 19 → 36). Five-mode approximation, $f = 5.65$ GHz. *Solid line*: Calculated. *Broken line*: Measured. (**a**) Amplitude, first row. (**b**) Phase, first row. (**c**) Amplitude, second row. (**d**) Phase, second row (Josefsson and Persson 2006) (Copyright IEEE)

single-mode approximation, the mutual coupling diagrams show good agreement between the measured and the calculated results, at least down to about –60 dB. For elements far away the agreement is not so good, but here the coupling levels are very low.

In order to improve the accuracy even further, higher-order waveguide modes were included in the analysis. With the four lowest TE modes (in increasing cutoff order, i.e., TE_{10}, TE_{20}, TE_{01}, and TE_{11}), only small differences were observed. But when taking also the TM_{11} mode into account, a significant improvement was obtained (see Fig. 19). Now the simulated results show good agreement with measurements down to coupling levels as low as –80 dB. The agreement of the phase is also improved. Using even more modes (up to 20 modes were tried) improved the results very little (Persson and Josefsson 2001).

The importance of the TM_{11} mode is probably explained by the fact that this mode is the first mode with a field component parallel with the direction of propagation. Obviously, this component is important for a precise modeling of the aperture field.

Embedded Element Patterns

As shown earlier the magnitude of the mutual coupling among the elements on a curved surface is less than for the corresponding planar case. However, the effects of the mutual coupling on the embedded element patterns can still be considerable for the curved case. The embedded element pattern gives an indication of the performance of the fully excited array and serves therefore as a valuable tool in the array design process. This will be illustrated by one example by considering a circular array with the elements uniformly spaced and covering the complete circumference of the cylinder. Thus, no edge effects are present in azimuth, and all azimuth element patterns are identical. This resembles the infinite planar array case, and the patterns for the two cases will be compared.

Figure 20 shows the gain of the *E*-plane embedded element pattern for a single ring of circumferentially polarized rectangular apertures for different cylinder radii. The element spacing is kept constant, equal to 0.6λ. The patterns are also compared with the corresponding planar case. As seen, there are no major

DOI 10.1007/978-981-4560-75-7_65-1

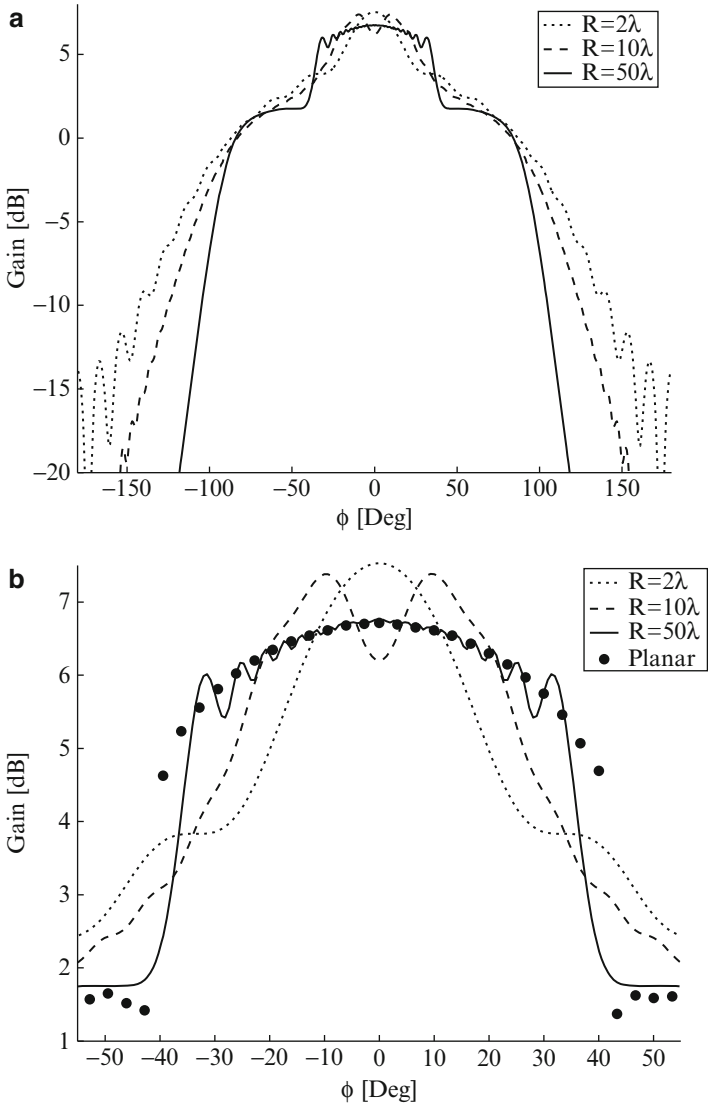


Fig. 20 The gain of the *E*-plane embedded element pattern for a single ring array in azimuth versus the radius of the cylinder for a fixed element spacing ($d = 0.6\lambda$). (**a**) Overview for $R = 2\lambda$, 10λ, and 50λ. (**b**) The corresponding planar case is shown in the close-up view (Josefsson and Persson 2006) (Copyright IEEE)

differences between the pattern for the planar case and the pattern for a cylinder with a large radius in the forward direction within the broadside region. Both patterns exhibit a clearly visible drop-off near $\phi = 42°$. This drop-off in the planar case is due to an end fire grating lobe at $\phi = 41.8°$. Its counterpart in the cylindrical array is a "cylindrical end fire grating lobe" (Mailloux 1994). It should be noted that the cylindrical grating lobe is not as distinct as in the planar case due to the curved surface. However, the phenomenon is caused by the same mechanisms in both cases.

Since the cylindrical grating lobe is more diffuse in the cylindrical case, the pattern dips are not as deep as in the (infinite) planar case. In the cylinder case there is a finite number of elements participating; the smaller the radius, the fewer the elements.

Another interesting observation is the ripple in the broadside region for the circular cylinder, not observed in the (infinite) planar case. The ripple is not caused by an edge effect since there is no discontinuity in the angular direction; instead it is due to the mutual coupling among the elements on this nonplanar geometry. The embedded element pattern is given by a superposition of the isolated

DOI 10.1007/978-981-4560-75-7_65-1

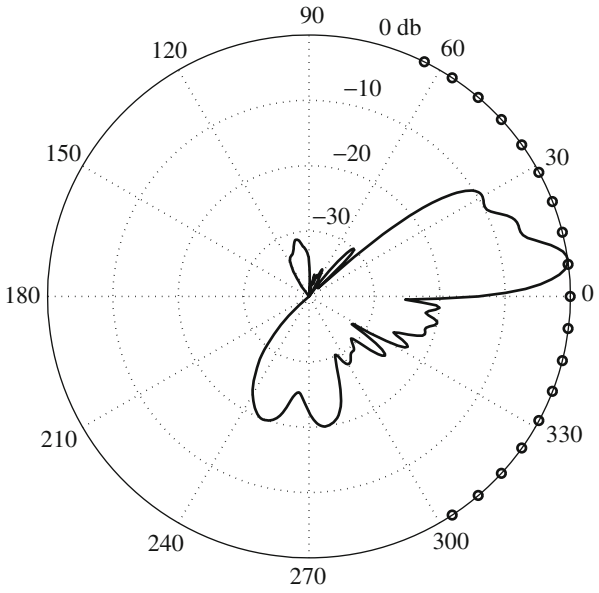


Fig. 21 Synthesized pattern with coupling after 100 iterations, also showing element locations (Josefsson and Persson 2006) (Copyright IEEE)

element patterns of all elements when excited through mutual coupling. When added together the interference between the different contributions will result in an oscillating behavior.

Array Synthesis

This section will be concluded by presenting a pattern synthesis exercise for a simple but typical application and shows the results in terms of pattern performance including mutual coupling effects. The geometry considered is the cylindrical array antenna that has been referred to several times (see Fig. 17), and the synthesis method used is the method of alternate projections (Bucci et al. 1990). The assumed requirements include both the sidelobe region and the shape of the main beam. The effect of the mutual coupling is included in the synthesis; neglecting the mutual coupling leads otherwise to poor results, especially in the shaped beam region (Josefsson and Persson 1999).

For simplicity, the elements are assumed to have a $\cos\phi$ isolated pattern, and only one row of waveguide aperture elements is studied. They are spaced 0.53λ on an arc of about 120° on the 4.3λ radius cylinder. The mutual coupling data is the results presented earlier in this section.

The following requirements are used in the synthesis procedure: a cosecant-squared main beam with unequal sidelobe envelopes. In more detail:

−180–0°	Maximum −20 dB sidelobes
0–35°	Cosec^2 within ± 1 dB
0–180°	Maximum −30 dB sidelobes

The pattern is iteratively modified in regions where the requirements are not met. A synthesized beam pattern after 100 iterations is shown in polar form in Fig. 21, where also the locations of the 18 radiating elements are indicated by small circles.

If the synthesis is repeated again, now assuming that the mutual coupling effect is neglected, there will be changes in the pattern. Figure 22b shows a close-up of the main beam analyzed with mutual coupling but synthesized without coupling. It is apparent that the mutual coupling plays an important role; it is essential to include the mutual coupling in the synthesis.

DOI 10.1007/978-981-4560-75-7_65-1

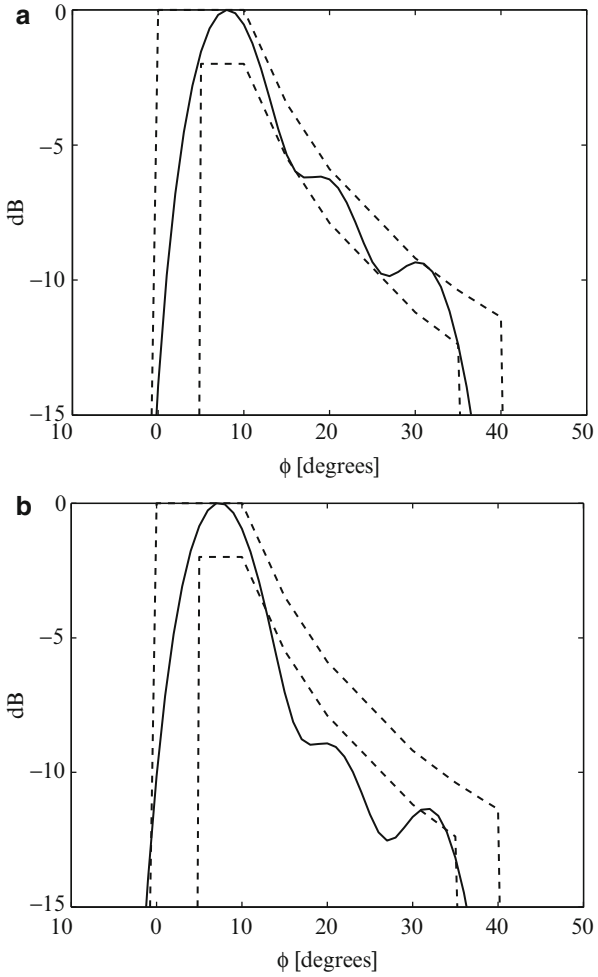


Fig. 22 Pattern synthesis with coupling (**a**). The pattern in (**b**) demonstrates the effect of the mutual coupling if not included in the synthesis (Josefsson and Persson 2006) (Copyright IEEE)

Doubly Curved Surfaces

It may have been understood from the discussion on geodesics that the analysis of doubly curved surfaces is a bit more challenging than the analysis of cylindrical surfaces but also more interesting. A doubly curved surface can more accurately model, e.g., realistic airborne or space-borne antennas than a singly curved surface can do. However, in many applications, a singly curved approach gives results accurate enough to be useful in the first stage of the design of a doubly curved antenna. Nevertheless, new features are introduced with doubly curved conformal antennas, and these features become important in a new way and must be accounted for. One such feature is the co-polar/cross-polar aspect of the radiation pattern since all elements on a doubly curved surface point in different directions.

In this section a few examples will be given to illustrate some properties of doubly curved array antennas. The analysis presented here is based on perfectly conducting general paraboloids of revolution (GPOR), and some of the results are also verified with measurements. Once again a hybrid UTD-MoM method is used. The antenna elements are circular waveguide-fed apertures; a single-mode approximation is used, i.e., only the dominant TE_{11} mode is accounted for in the simulations. Figure 23 illustrates the surface that is analyzed including an experimental antenna that was built.

The doubly curved experimental antenna shown in Fig. 23 was built at Ericsson Microwave Systems AB in Mölndal, Sweden, in 2000. The experimental antenna is shaped as a paraboloid of revolution with

DOI 10.1007/978-981-4560-75-7_65-1

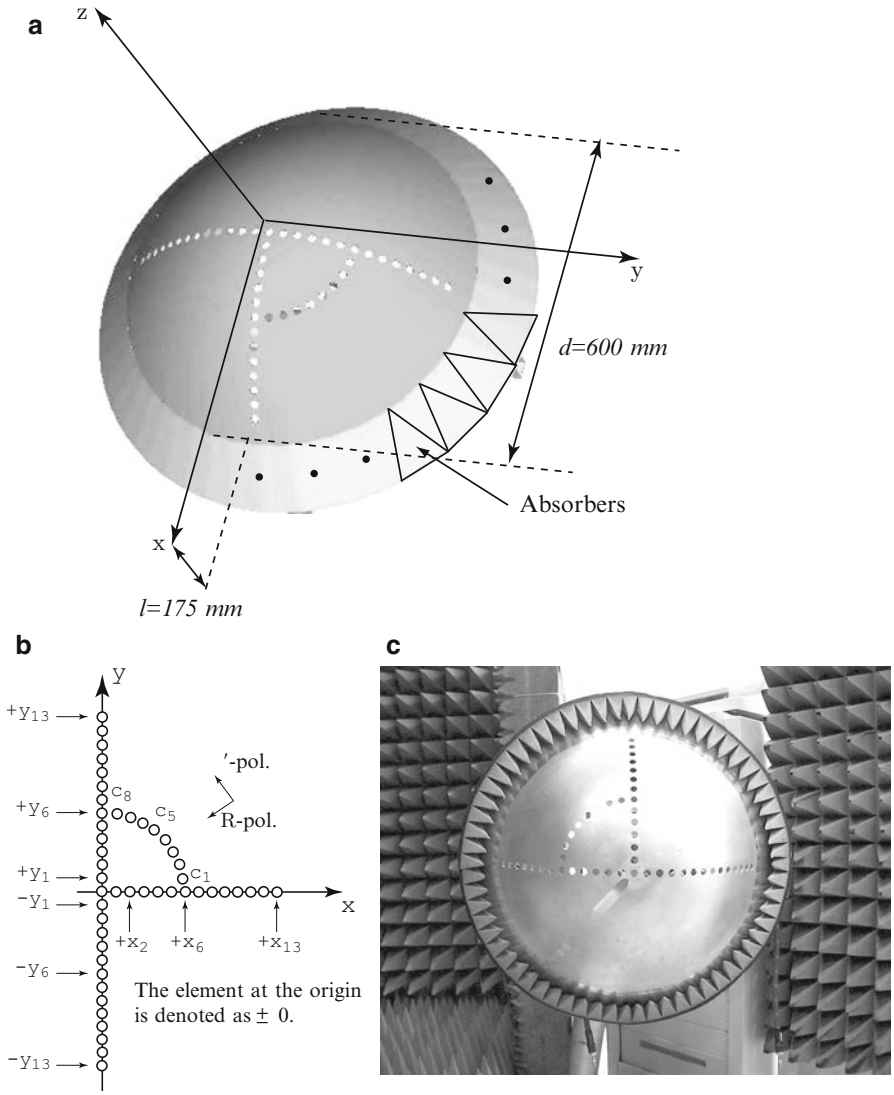


Fig. 23 (**a**) The geometry of the experimental doubly curved antenna with the coordinate system used in the analysis. (**b**) The numbering of the elements (Josefsson and Persson 2006) (Copyright IEEE). (**c**) A picture of the antenna on location for mutual coupling measurements (Courtesy Saab AB)

f/$d \approx 0.22$, and the diameter of the surface is approximately 600 mm with a depth of approximately 175 mm (see Fig. 23a). The surface is extended with a conical sheet of metal to gain extra space for the absorbers at the edge. The circular waveguide-fed apertures with diameter 14.40 mm are filled with Rexolite ($\varepsilon_r = 2.53$). The cutoff frequency for the dominant TE_{11} mode is 7.65 GHz. The nearest mode in cutoff order is TM_{01} with 9.99 GHz cutoff frequency. A single isolated element has a return loss (S_{11}) of about -8 dB.

Isolated Mutual Coupling

To illustrate the typical behavior of the mutual coupling on a doubly curved surface, one example is given to show the effect of the dependency on the polarization of the elements. Figure 24 shows the isolated mutual coupling between an element on the circular arc and elements along the positive *y*-axis. Here, the results are explained mainly as polarization effects. To illustrate this, the geodesic length is included in

DOI 10.1007/978-981-4560-75-7_65-1

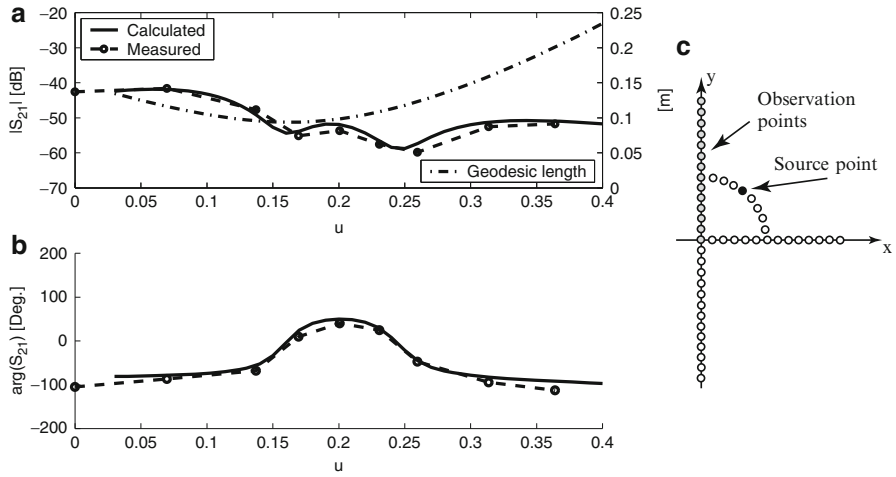


Fig. 24 Isolated coupling, [*R*, *R*]-polarization. The reference/source point is element c_5. (**a**) The amplitude. (**b**) The phase. (**c**) The geometry (Josefsson and Persson 2006) (Copyright IEEE)

Fig. 24 where the [*R*, *R*]-polarization case is considered; i.e., both the source and observation elements are *R*-polarized (see Fig. 23b).

More results can be found in Persson et al. (2003). Element patterns have also been measured and simulated (Josefsson and Persson 2006).

Polarization in Doubly Curved Arrays

As indicated in Fig. 24 the polarization issue is important and more complicated for doubly curved arrays compared to cylindrical arrays (Josefsson et al. 2002). This is because in general the radiating elements point in different directions, and they must be individually polarized in order to generate a desired polarization in the main beam direction. With circular polarization, however, rotation is not necessary since the rotation does not change the polarization. If not correctly oriented, there will instead be a phase error, but this can be compensated by the phase control that must be applied to each element in any case.

In order to simplify the modeling, it is assumed that the radiating elements are of type short dipoles, i.e., the element pattern is $E_0 \sin\theta$ with θ the angle from the dipole axis. The elements do not radiate "through" the parabolic surface, however. The dipoles are tangential to the curved surface and initially oriented in a vertical plane (see Fig. 25a). When considering curved array antennas, it is also of importance to introduce an active region since many elements may not radiate in the desired beam direction and should therefore be turned off in order to avoid sending unwanted signals into other parts of space. The active region is defined from the desired nominal beam direction $\hat{r}$ and the normal direction to the surface $\hat{n}$. The active region is obtained as the area encompassing all elements for which

$$\hat{r} \cdot \hat{n} \geq \cos\theta_{s\max} \tag{35}$$

where $\theta_{s\max}$ is a chosen maximum useful local scan angle, typically 45–60°. An example of an active region is shown in Fig. 25b.

In order to align the far-field contributions properly, each dipole is rotated around the surface normal until the far-field contribution in the main beam direction has the correct polarization. Additionally, the phase of each dipole is adjusted so that they add coherently to a plane phase front. However, it is generally not possible to obtain the same polarization throughout the pattern; thus, there will be some cross polarization. Figure 25c illustrates how the radiating elements are all aligned for the same (vertical) polarization in the main beam direction. When viewed from another angle (Fig. 25d), it is clear that a

DOI 10.1007/978-981-4560-75-7_65-1

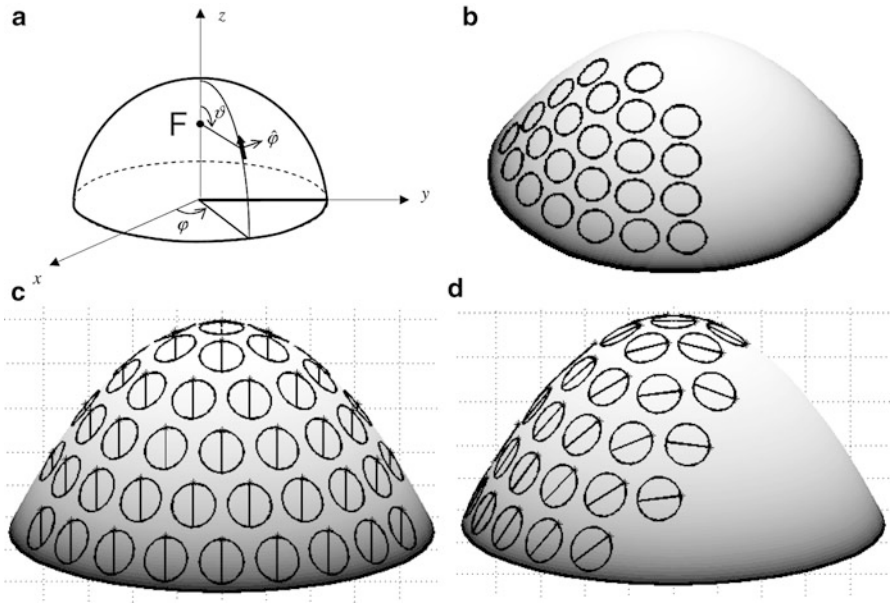


Fig. 25 (**a**) Illustration of the antenna element used. (**b**) An example of an active sector when scanning the beam. (**c**) Example of an array where the elements are aligned to give vertical polarization in the main beam direction. (**d**) As in (**c**), but now viewed from a different angle (Josefsson and Persson 2006) (Copyright IEEE)

certain degree of cross polarization will radiate in the sidelobe region. These are some aspects related to conformal array beamforming, to be discussed in the next section.

Beamforming in Conformal Array Antennas

Introduction

An array of radiating elements alone does not make a conformal antenna system. As a minimum, some kind of phase compensation for the array curvature is needed together with a summation network. In addition it may be required to move around the active sector of the array (commutating), to apply amplitude control, and to steer the polarization as well. All of this may sound complicated and expensive. However, there are many advantages that may be gained from the extra flexibility of conformal solutions such as multibeam capability, large angular coverage, large bandwidth, reduced RCS (radar cross section), etc. Some of the beamforming systems that have been developed will be described, in particular those that are used for circular and cylindrical arrays.

The principal function of an analog beamformer is illustrated in Fig. 26a, where the received signal is obtained after summing of the element signals with proper amplitude and phase weights.

The analog beamforming in Fig. 26a is illustrated for a receive situation. However, it can in principle also be designed for transmitting. Special analog networks have been used for generating simultaneous multiple beams, both constrained feed networks and optical type feeds (microwave lenses). The digital beamformer (Fig. 26b) is even more flexible. Several simultaneous beams can be generated from the digital processor. Corresponding digital solutions for the transmit case are also possible but not equally easy to implement.

DOI 10.1007/978-981-4560-75-7_65-1

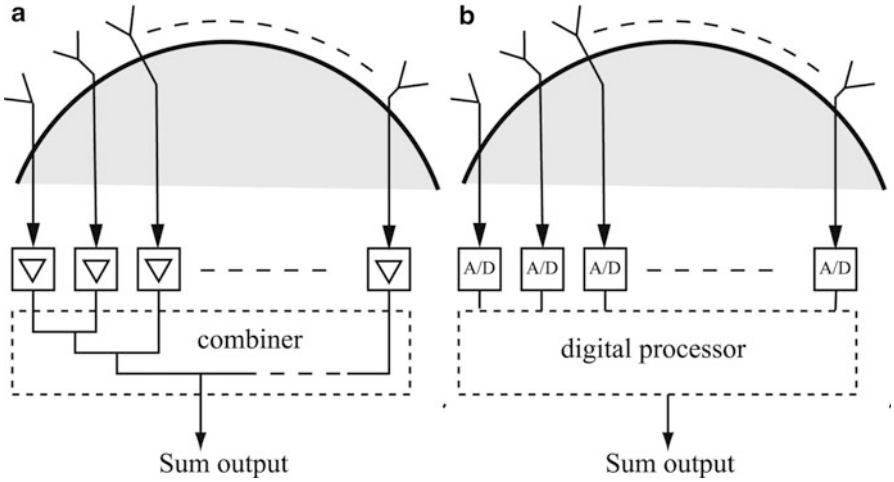


Fig. 26 (**a**) Analog beamforming system. (**b**) Digital beamforming system (Josefsson and Persson 2006) (Copyright IEEE)

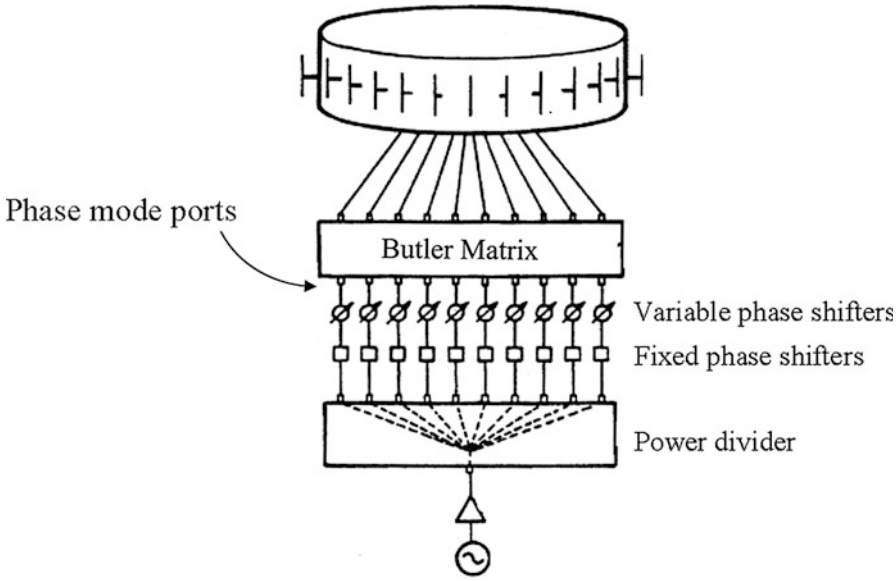


Fig. 27 The scanning multimode array (Sheleg 1968) (Copyright IEEE)

The Butler Matrix

The Butler matrix (Butler and Lowe 1961) is a passive network with N input ports and the same number of output ports. Feeding one of the input ports will generate equal amplitudes on the output ports but with a linearly varying phase from port to port. The phase slope will depend on which input port is excited. Thus, a set of (linear) phase mode can be generated. Connecting the output ports to a linear array antenna will produce sin *x*/*x* type patterns. Connecting to a circular array antenna (UCA) will produce the circular phase modes discussed in section "Phase Mode Theory."

A very interesting application of the Butler matrix is found in the scanning multimode array (Fig. 27).

In this system a beam can be steered 360° in azimuth by adding a linear phase shift to the phase mode ports, which simply shifts the excitation setup by the power divider and fixed phase shifters, i.e., a *commutating* function.

DOI 10.1007/978-981-4560-75-7_65-1

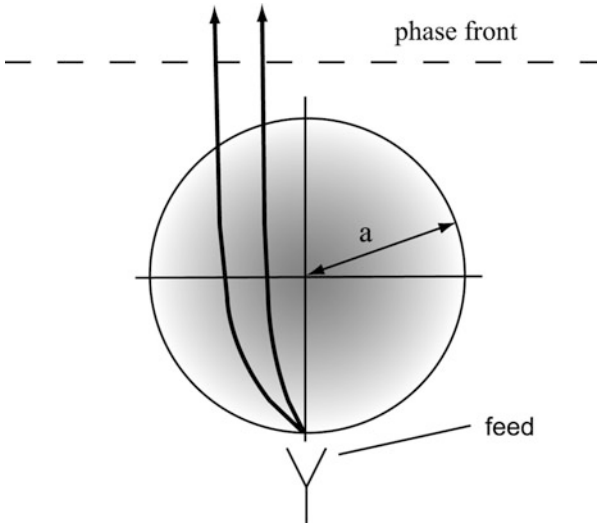


Fig. 28 Focusing by a Luneburg lens (Josefsson and Persson 2006) (Copyright IEEE)

A UCA with a Butler matrix feed can be used as a bearing discriminator (Rehnmark 1980). By extracting the phase values on a few phase mode ports, the DOA (direction of arrival) of received signals can be determined.

RF Lenses

RF lenses are based on optical analogies and can be used for focusing, beam shaping, and beam steering. In this way the amount of discrete components required in matrix-fed systems can be reduced significantly. An example is found in the Luneburg lens (Luneburg 1944) (Fig. 28).

In the Luneburg lens the spherical lens has a refractive index n that varies with radius as

$$n(r) = \sqrt{\varepsilon_r} = \sqrt{2 - (r/a)^2} \tag{36}$$

Several feed horns can be arranged to focus beams in different directions.

The R-2R lens and the R-kR lens are two lens types using parallel plate structures. The latter one requires a dielectric loading of the circular lens with a permittivity of $\varepsilon_r = 4$ or higher. An example of a compact R-kR lens-fed array on a high-permittivity substrate ($\varepsilon_r = 10.2$) is shown in Fig. 29. With this array ten beams separated by 10° were realized.

The Dome antenna makes half-sphere scanning possible from just one active planar array. The concept is illustrated in Fig. 30.

In the Dome antenna a planar array illuminates the concave inside of the passive spherical lens. Also other shapes such as ellipsoidal are possible. The phase gradient of the lens can be optimized for the desired elevation coverage. However, this revolutionary idea has not been put to much practical use because of its rather high cost and limited bandwidth.

Digital Beamforming

As already shown in Fig. 26, it seems that everything that can be done with analog beamforming can also be done digitally. The digital solution leads to strict requirements on data transmission, storage, and signal processing, especially for large arrays. However, these problems are more easily solved today than in the past, thanks to the rapid growth of computer power, both software and hardware. In analog receive beamforming, the element signals are combined with weights determined by feed networks

DOI 10.1007/978-981-4560-75-7_65-1

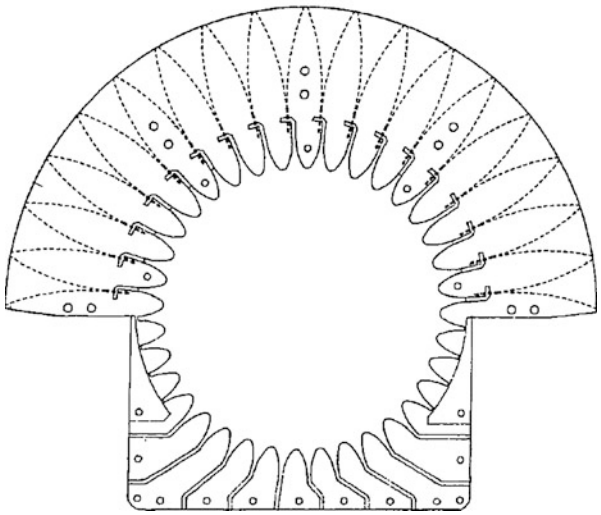


Fig. 29 An R-2R lens feed integrated with tapered slot elements in a high-permittivity substrate (Sierra and Bernal 1989) (Copyright IEEE)

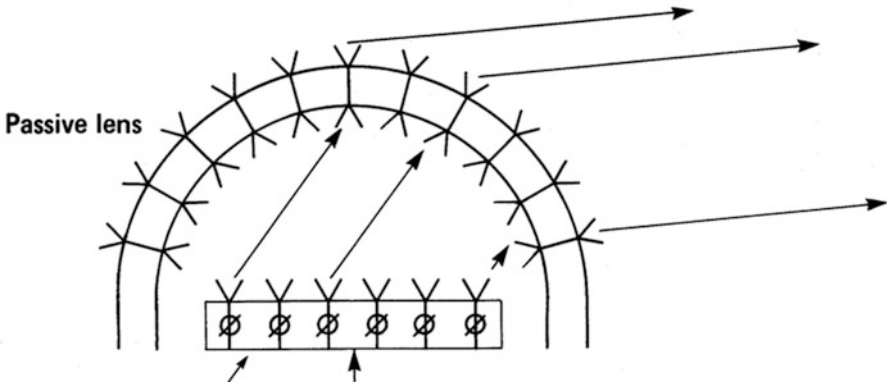


Fig. 30 The Dome antenna concept (Liebman et al. 1975) (Copyright IEEE)

and/or phase- and amplitude-controlled receive modules. The same operations can be performed in a digital computer on the element signals after analog-to-digital conversion. It is interesting to note that the Butler matrix performs the same transformation in the spatial domain as the FFT (fast Fourier transform) often used in the time/frequency domain. The digital approach allows the formation of simultaneous receive beams, even overlapping, without the feed losses that are incurred in an analog system. The element modules in the digital case have low-noise amplifiers (LNAs) preceding the analog-to-digital conversion. The signal-to-noise ratio is set by the LNAs, and it is therefore not affected by transmission losses in the further processing, hence "lossless beamforming."

Digital beamforming has been used mainly in linear and planar arrays. However, the following example shows a case applied to a circular 16 element array with 0.5λ element spacing. The elements are pointing radially, each with a cosα pattern. In this scenario it is assumed that there is one desired signal at 10 dB above noise level and two interference signals at 40 dB above noise level. For suppressing the interference, an adaptive nulling algorithm (SMI, sample matrix inversion) according to Applebaum (1976) is used. The geometry and the resulting patterns are shown in Fig. 31. The results show that the interfering signals are suppressed by at least 25 dB.

DOI 10.1007/978-981-4560-75-7_65-1

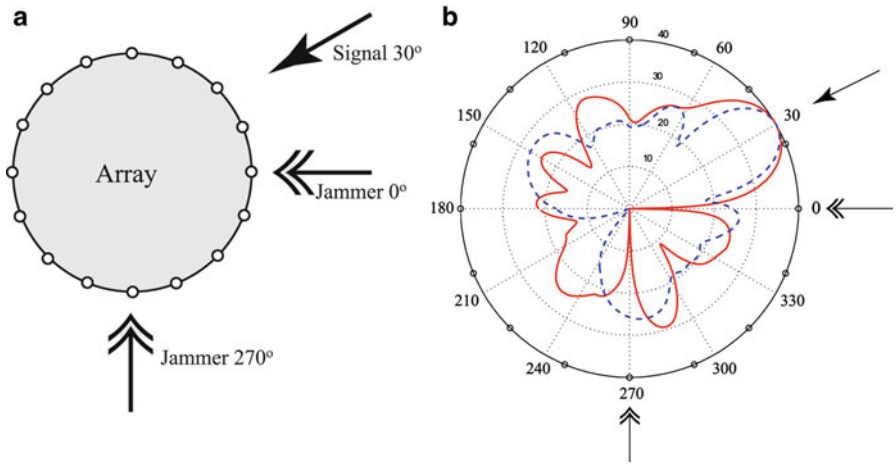


Fig. 31 (**a**) Scenario with one desired signal and two jamming signals. (**b**) Circular array patterns without interference suppression (*dashed line*) and with adaptation (*solid line*). Signal direction is from 30° and interfering signals from 0° and 270° (Josefsson and Persson 2006) (Copyright IEEE)

In Steyskal (1989) an interesting DBF experiment is described, using a 64-monopole element circular array with digitally formed low sidelobe patterns. Using in total $2M + 1 = 35$ phase modes in the synthesis, a sidelobe level of about -30 dB is successfully obtained over 40 % bandwidth (corresponding to $2kR$ values between 41 and 62). However, with more phase modes added in the synthesis, the pattern starts to deteriorate, e.g., $2M + 1 = 51$ is more than the "Bessel window" can accommodate at the low-frequency end (cf. Fig. 4).

Summary

The conformal array technology makes new antenna shapes possible, not necessarily limited to linear or planar configurations. Possibilities include structural integration in aircraft, very wide angular coverage, and reduced RCS (radar cross section); sometimes no radome is needed.

The electromagnetic problem of wave propagation on curved surfaces is now well understood, and the tools for analysis and design are available. Calibration, beam steering, and polarization control may still seem challenging in some cases. However, shifting the analog/digital interface closer to the radiating elements opens up new solutions, including DBF. New signal processing schemes show promise, e.g., sparse arrays, MIMO for radar (Dorey et al. 1989), compressive sensing (Migliore and Pinchera 2011), and so on. At the same time small systems with circular or cylindrical arrays will find increased use in communication and sensing systems in urban environments.

Cross-References

- Conformal Array Antennas
- Lens Antennas
- Multibeam Antenna Arrays
- Phased Arrays
- Waveguide Slot Antennas and Arrays
- Waveguide Slot Array Antennas

DOI 10.1007/978-981-4560-75-7_65-1

References

Akyuz MS, Ertürk VB, Kalfa M (2011) Closed-form Green's function representations for mutual coupling calculations between apertures on a perfect electric conductor circular cylinder covered with dielectric layers. IEEE Trans Antennas Propagat 59(8):3094–3098

Alisan B, Ertürk VB (2010) A high frequency based asymptotic solution for surface fields on a source excited sphere with an impedance boundary condition. Radio Sci 45:1–14, RS5008

Applebaum SP (1976) Adaptive arrays. IEEE Trans Antennas Propagat AP-24(5):584–598

Bondyopadhyay PK (2002) New phased array antenna architecture for the X-band radar. IEEE AP-S Int Symp 3:360–363, San Antonio, 2002

Bucci O, Franceschetti G, Mazzarella G, Panareillo G (1990) Intersection approach to array pattern synthesis. IEE Proc H 137:349–357

Butler J, Lowe R (1961) Beamforming matrix simplifies design of electronically scanned antennas. Electron Des 9:170–173

Chireix H (1936) Antennes a rayonnement zénithal réduit. L'Onde Êlectrique 15:440–456

Christopher EJ (1974) Electronically scanned TACAN antenna. IEEE Trans Antennas Propagat AP-22(1):12–16

Davies DEN (1983) Circular arrays. In: Rudge AW et al (eds) The handbook of antenna design, 2nd edn. Peter Peregrinus, London

Dorey J, Garnier G, Auvray G (1989) RIAS, synthetic impulse and antenna radar. In: Proceedings of international conference on Radar, Paris, 24–28 Apr 1989, p 556–562

Elliott RS (1981) Antenna theory and design. Prentice-Hall, Englewood Cliffs

Ertürk VB, Rojas RG (2000) Efficient computation of surface fields excited on a dielectric coated circular cylinder. IEEE Trans Antennas Propagat AP-48:1507–1516

Ertürk VB, Rojas RG, Lee KW (2004) Analysis of finite arrays of axially directed printed dipoles on electrically large circular cylinders. IEEE Trans Antennas Propagat 52(10):2586–2595

Gething PJD (1966) High-frequency direction finding. Proc IEE 113(1):49–61

Harrington R F (1961) Time harmonic electromagnetic fields. Prentice-Hall, New York

Hessel A (1970) Mutual coupling effects in circular arrays on cylindrical surfaces – aperture design implications and analysis. In: Proceedings of phased array symposium, Polytechnic Institute of Brooklyn, New York 1970

Hopkins MA, Tuss JM, Lockyer AJ, Alt K, Kinslow R, Kudva J N (1997) Smart skin conformal load-bearing antenna and other smart structures developments. American Institute of Aeronautics Astronaut (AIAA), structures, structural dynamics & materials conference, Kissimmee, FL USA, vol 1, pp 521–530

Jha RM, Wiesbeck W (1995) The geodesic constant method: a novel approach to analytical surface-ray tracing on convex conducting bodies. IEEE Antennas Propagat Mag 37(2):28–38

Jha R M, Bokhari SA, Sudhakar V, Mahapatra PR (1989) Closed form surface ray tracing on ogival surfaces. In: 1989 I.E. AP-S international symposium, San Jose, CA USA, 26–30 June 1989, pp 1294–1297

Josefsson L, Lanne M (2003) Shape optimization of doubly curved conformal array antennas. 3rd European workshop on conformal antennas, 22–23 Oct 2003, Bonn

Josefsson L, Persson P (1999) Conformal array synthesis including mutual coupling. Electron Lett 35(8):625–627

Josefsson L, Persson P (2006) Conformal array antenna theory and design. IEEE Press/Wiley, Hoboken, NJ USA

DOI 10.1007/978-981-4560-75-7_65-1

Josefsson L, Persson P, Lanne M (2002) The polarization problem in singly and doubly curved conformal array antennas. In: IEEE international symposium AP-S/URSI, San Antonio TX, 17–21 June 2002

Knudsen HL (1953) The field representation by a ring quasi-array of an infinite number of tangential or radial dipoles. Proc IRE 41(6):781–789

Kraus JD (1988) Antennas, 2nd edn. McGraw-Hill, New York

Liebman PM, Schwartzman L, Hylas AE (1975) Dome radar – a new phased array system. In: Proceedings of IEEE international Radar conference, Washington D.C. pp 349–353

Lindgren S, Josefsson L (1998) Konforma antenner, mätrapport. Technical report SR/R-1998:033, Ericsson microwave systems AB, Mölndal, Nov 1999. In Swedish

Luneburg RK (1944) Mathematical theory of optics. Brown University Press, Providence, RI USA

Mailloux RJ (1994) Phased array antenna handbook. Artech House, Boston, MA USA

Migliore M, Pinchera D (2011) Compressed sensing in electromagnetics: theory, applications, and perspectives. EuCAP, Rome

Munger AD, Vaughn G, Provencher JH, Gladman BR (1974) Conical array studies. IEEE Trans Antennas Propagat AP-22(1):35–43

Pathak PH (1992) High-frequency techniques for antenna analysis. Proc IEEE 80(1):44–65

Pathak PH, Wang N (1981) Ray analysis of mutual coupling between antennas on a convex surface. IEEE Trans Antennas Propagat AP-29(6):911–922

Pathak PH, Burnside WD, Marhefka RJ (1980) A uniform GTD analysis of the diffraction of electromagnetic waves by a smooth convex surface. IEEE Trans Antennas Propagat AP-28(5):631–642

Pathak PH, Wang N, Burnside WD, Kouyoumjian RG (1981) A uniform GTD solution for the radiation from sources on a convex surface. IEEE Trans Antennas Propagat AP-29(4):609–622

Persson P, Josefsson L (2001) Calculating the mutual coupling between apertures on a convex circular cylinder using a hybrid UTD-MoM method. IEEE Trans Antennas Propagat AP-49(4):672–677

Persson P, Rojas RG (2003) High-frequency approximation for mutual coupling calculations between apertures on a perfect electric conductor circular cylinder covered with a dielectric layer: nonparaxial region. Radio Sci 38(4):1079

Persson P, Josefsson L, Lanne M (2003) Investigation of the mutual coupling between apertures on doubly curved convex surfaces: theory and measurements. IEEE Trans Antennas Propagat 51(4):682–692

Rehnmark S (1980) Instantaneous bearing discriminators with omnidirectional coverage and high accuracy. Proc IEEE MTT Symp 120–122

Sheleg B, Juarez F, Alcaraz JA (1968) A matrix fed circular array for continuous scanning. IEEE AP-S Int Symp Dig 6:7–16

Sierra M, Bernal JC (1989) Multibeam antenna array. IEEE AP-S Int Symp Dig 1:150–153

Sipus Z, Kildal P-S, Leijon R, Johansson M (1998) An algorithm for calculating Green's functions of planar, circular cylindrical, and spherical multilayer substrates. Appl Comput Electromagn Soc J 13:243–254

Steyskal H (1989) Circular array with frequency-invariant pattern. IEEE AP-S Int Symp Dig 3:1477–1480

Steyskal H (2002) Pattern synthesis for a conformal wing array. In: IEEE aerospace conference, Big Sky, MT, Mar 2002

Struik DJ (1988) Lectures on classical differential geometry, 2nd edn. Dover, New York

DOI 10.1007/978-981-4560-75-7_65-1

Thors B, Rojas RG (2003) Uniform asymptotic solution for the radiation from a magnetic source on a large dielectric coated circular cylinder: non-paraxial region. Radio Sci 38(5):1–14

Tomasic B, Turtle J, Liu S, Schmier R, Bharj S, Oleski P (2003) The geodesic dome phased array antenna for satellite control and communication – subarray design, development and demonstration. In: IEEE international symposium on phased array systems and technology, Boston, 14–17 Oct 2003, pp 411–416

Wait JR (1959) Electromagnetic radiation from cylindrical structures. Pergamon Press, London

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_66-1

Multi-beam Antenna Arrays

Bybi P. Chacko*, Gijo Augustin and Tayeb A. Denidni
Centre for Energy, Materials and Telecommunication, National Institute of Scientific Research (INRS), Montreal, QC, Canada

Abstract

This chapter begins with a brief introduction of basic concepts in antenna arrays which is followed by slightly advanced concepts including smart antennas and related systems. This helps the readers to acquire basic knowledge necessary to understand the concepts of multi-beam antennas and various beam-forming networks outlined in the sessions that follow. The world of engineered materials which become an important research area is also outlined to broaden the knowledge of the readers. This chapter concludes with some recent multiple-beam antenna design examples with metamaterial technology.

Keywords

Antenna arrays; Multi-beam antenna arrays; Directive antenna arrays; Cylindrical antennas; Sectoral antennas; Beam scanning; Beam formers; Intelligent antenna system; Nimble radiation pattern; Multifunctional beams

Introduction

In layman's language, an antenna array is defined as a set of aerials arranged in a specific fashion in space so that the resulting radiation characteristics can be modified. As per IEEE standard definition of terms for antennas (IEEE Standard 145–1993), an array antenna can be defined as "an antenna comprised of a number of identical radiating elements in a regular arrangement that are excited to obtain a prescribed pattern." As explained by Randy L. Haupt (2010), the antenna array can be visualized as a system to collect rain. To collect more rain, it is necessary to use a very big bucket or employ large number of buckets and combine all of them. In the same way, in order to receive more electromagnetic radiation, it requires either a large aperture antenna, such as the Arecibo radio telescope in Puerto Rico (305 m wide), or an antenna array with more number of elements such as the square kilometer array as shown in Fig. 1.

The array can be formed by two or more individual elements which may or may not be identical in their radiation characteristics. However, in most of the simple antenna array designs, each individual element is designed to have identical radiation properties. The shape of the radiation pattern in an array with identical elements is determined by the following parameters (Balanis 2005):

- The shape of physical arrangement of each element in space (linear, rectangular, ring, etc.)
- The distance between each element
- The amplitude of the excitation signal to individual elements
- The phase of the excitation signal to individual elements
- The radiation pattern of the individual elements

*Email: bybichacko@emt.inrs.ca

DOI 10.1007/978-981-4560-75-7_66-1

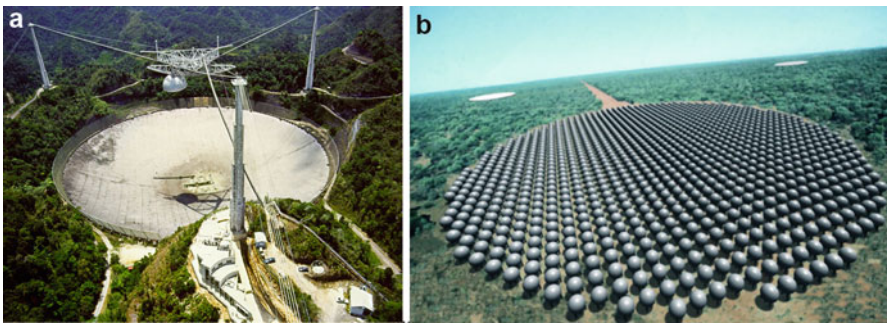


Fig. 1 (**a**) Arecibo radio telescope in Puerto Rico, (**b**) square kilometer array

Table 1 Application of antenna array

Objective	Application
Gain enhancement	Satellite antennas
Facilitate beam steering	Radar systems
Antenna diversity	Mobile phones
Improve the S/N ratio	Satellite receivers
Find direction of arrival	Radar systems

In modern world of wireless communication, this concept of antenna array is of high interest and is employed in various scenarios outlined above (Table 1).

Before proceeding further in this section, following definitions related to antenna array would be handy:

Radiation pattern: Radiation pattern of an antenna is the plot of radiated power variation around an antenna estimated in a constant distance from it.

Isotropic antenna: An antenna which radiates its energy equally in all directions. In reality, no such antenna physically exits. An isotropic antenna is used as the reference antenna to compare the characteristics of actual antennas.

Omni-directional antenna: Antennas having an isotropic pattern in one plane and directional pattern in an orthogonal plane.

Directional antenna: An antenna which radiates or concentrates its power to a particular direction or angular regions.

Bore sight: It is the direction at which the beam maximum of an antenna is pointed.

Beam forming: The process of combining signals from different elements of an antenna array is known as beam forming. The direction in which the array has maximum gain is said to be the beam-pointing direction. If the signals from the antenna elements are combined without any change in amplitude or phase, a broadside beam is produced. Beam forming can be generally categorized as analog beam forming and digital beam forming. As the name implies, in analog beam forming, the amplitude and phase of each antenna elements are controlled in the analog part of the system. Phase shifters are commonly used in the analog beam-forming-phased array configurations to control the phase of each antenna element and steering the main beam. On the other hand, in digital beam-forming method, the amplitude and phase of a signal are controlled in the digital part of the system.

DOI 10.1007/978-981-4560-75-7_66-1

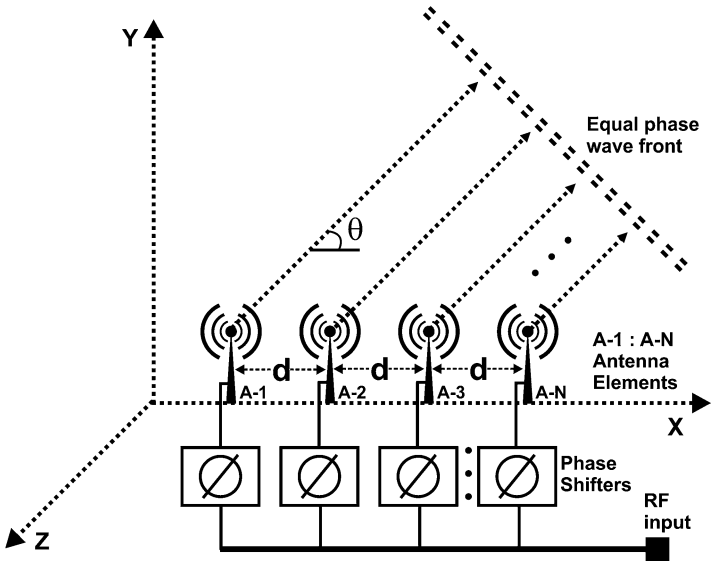


Fig. 2 Uniform N elements linear array

Beam steering: IEEE standard definition of terms (IEEE Standard 145–1993) defines beam steering as "changing the direction of the main beam of antenna." By mechanically moving, the main beam from any given array may be pointing to any direction; this is called mechanical steering. Electronic beam steering is accomplished by appropriately delaying the signals before combing. That is the beam-pointing direction, or the main beam direction can be controlled by changing the phase difference among different antenna elements.

Antenna Array Factor

The concept of an antenna array can be best understood through an example of a simple uniform array shown in Fig. 2. It consists of N identical antenna elements arranged in x-axis with an equal distance. Each antenna element possesses identical radiation characteristics and is fed through separate phase shifters so that each element is excited with progressive phase shift.

The radiation pattern of the linear array is formed based on the concept of pattern multiplication as expressed below:

$$\text{E Total} = \text{E (Single element)} * \text{Array Factor}$$

where E Total is the total field radiated from the antenna array and E (single element) represents the radiation pattern of the single element (A-1, A-2, etc.).

Array factor: It is a mathematical expression formed by the physical arrangement of an array and also with the properties of the excitation signal. The array factor varies from one type of array to another. This clearly indicate that the overall radiation pattern of an antenna array can be entirely different than that of the antenna element.

DOI 10.1007/978-981-4560-75-7_66-1

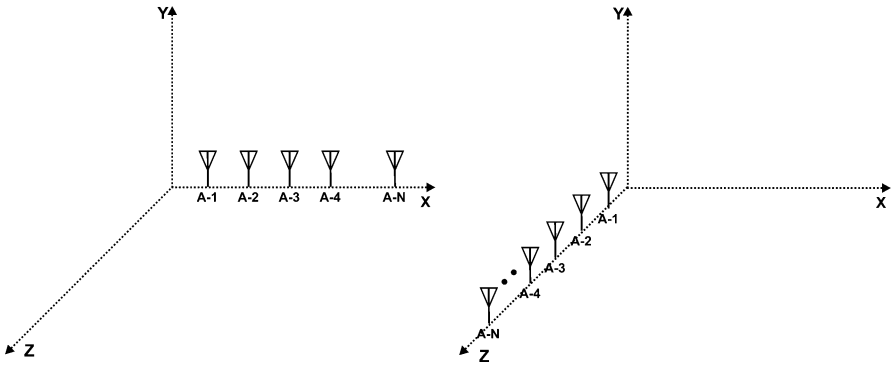


Fig. 3 Linear array with elements positioned in x-axis and z-axis

The array factor of the antenna described in Fig. 2 is defined as

$$AF = \frac{\mathrm{Sin}\left(\frac{N}{2}\ \varphi\right)}{\mathrm{Sin}\left(\frac{\varphi}{2}\right)} \tag{1}$$

$$\varphi = kd\cos\theta + \beta \tag{2}$$

where β is the difference in the phase excitation between the successive elements (progressive phase shift), d is the distance between the elements, k = 2*pi/ λ, and θ is the polar angle.

Configuration/Different Types of Arrays

There are various approaches in classifying antenna arrays. On the basis of the arrangement of the antenna elements in three-dimensional surfaces, the arrays are classified as:

1. **Linear or one-dimensional array:** In this configuration, each antenna element is arranged on a straight line in one dimension. Please note that it is not the shape of the individual antenna elements which needs to be in one dimension but the location of the individual elements. The elements can be simple dipoles to 3D conformal structures which are entirely based on the design objective of the communication system. Figure 3 illustrates some examples in which the antenna elements are distributed in a line parallel to x-axis and z-axis. During the design of the linear array, the expected radiation characteristics need to be taken care of. For example, when designing a linear array oriented along y-axis with array elements of omnidirectional radiation pattern, the broadside pattern will be directional and interferences from the neighboring elements will be minimum. This type of arrays is employed for application where a directional beam is desired and is useful when broad coverage in one plane and narrow beam in the orthogonal plane is required.
2. **Planar or two-dimensional arrays:** The planar array or two-dimensional array is designed with individual elements in a two dimensional plane as illustrated in Fig. 4. This configuration is widely employed for radar applications since it can facilitate highly versatile beam forming.
3. **Volumetric or three-dimensional arrays:** In this configuration, the array elements are designed to conform inside a three-dimensional space as depicted in Fig. 5. Conformal antenna arrays are also defined as a type of volumetric antenna arrays. Compared to the earlier antenna array designs, the volumetric arrays need relatively complex design strategies.

DOI 10.1007/978-981-4560-75-7_66-1

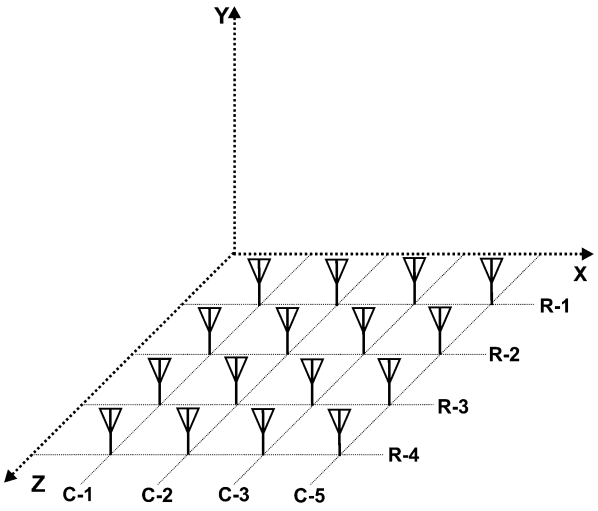


Fig. 4 Planar array arranged in XZ-plane

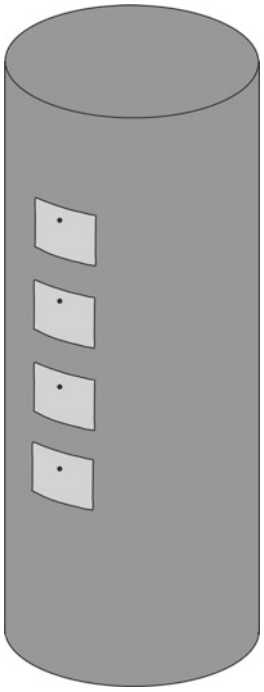


Fig. 5 Volumetric array

Another classification based on the radiation pattern is:

1. **Broadside Array:** In broadside arrays (Fig. 6), the direction of the main beam is perpendicular to the axis in which the array elements are arranged. The IEEE standard definition of terms states broadside antenna array as "A linear or planar array antenna whose direction of maximum radiation is

DOI 10.1007/978-981-4560-75-7_66-1

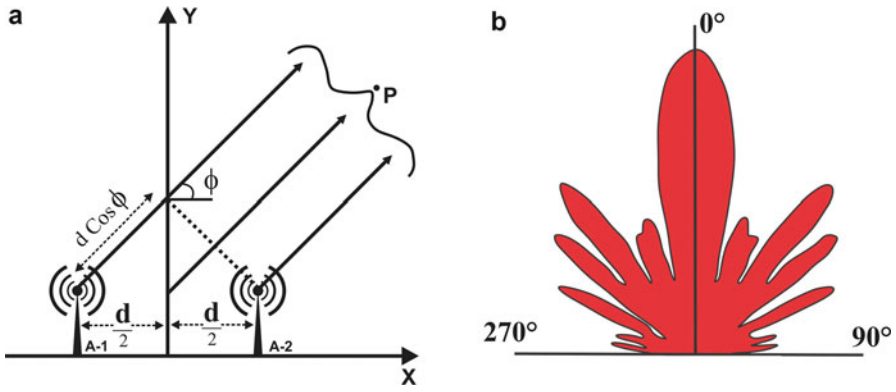


Fig. 6 Broadside array (**a**) configuration, (**b**) radiation pattern

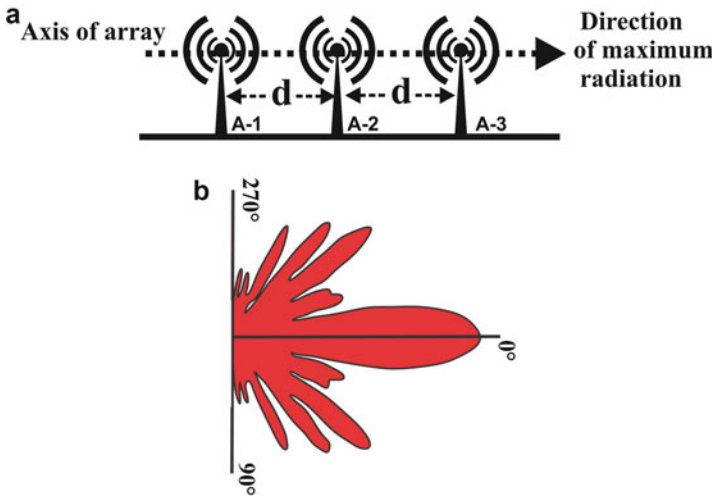


Fig. 7 End-fire array (**a**) configuration, (**b**) radiation pattern

perpendicular to the line or plane, respectively, of the array." The broadside pattern can be obtained in a linear array with identical elements by exciting them with equal magnitude and phase.

2. **End-Fire Array:** In these types of antenna array (Fig. 7), the beam maximum is focused along the axis of the array. Most of the planar or linear arrays can be configured to have a beam maximum that is pointing toward the array axis by exciting the individual antenna elements with equal magnitudes and progressive phase shift.
3. **Collinear Array:** The term collinear indicates that the individual antenna elements are connected end to end along a single line as shown in Fig. 8. In this configuration, the beam maximum is perpendicular to the axis of the array as in the case of broadside array. However, in this case, the fan beam is focused to every point perpendicular to the array axis resulting in a "donut"-shaped pattern as illustrated in Fig. 9. In order to get this pattern shape individual array, elements are excited with currents equal in magnitude and phase.

DOI 10.1007/978-981-4560-75-7_66-1

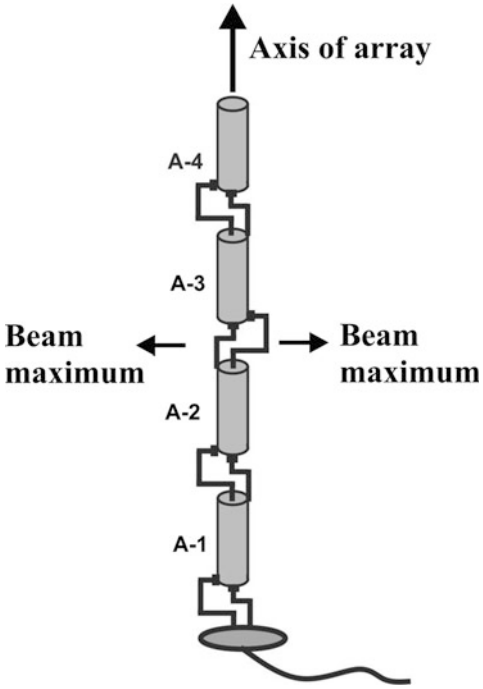


Fig. 8 Collinear array

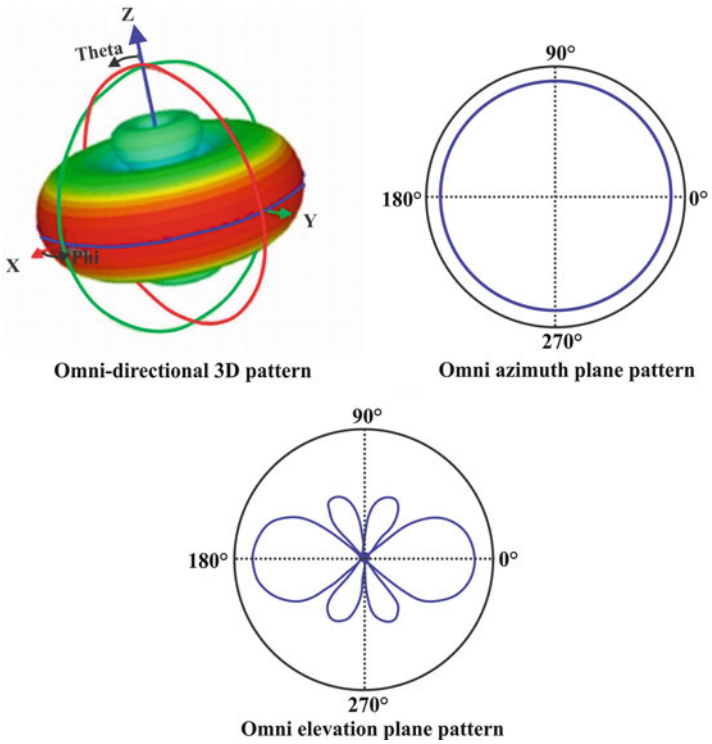


Fig. 9 Collinear array radiation pattern

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_66-1

Smart Antenna

In reality, antennas are not smart; an antenna along with its associated systems makes the antenna system smart. Basically, a smart antenna is a multielement antenna in which the signals received at each antenna element are intelligently combined to improve the performance of the wireless system. As illuminated by Balanis and Ioannides (2007), the working of a smart antenna can be better visualized by considering two persons carrying on a conversion inside an isolated room. Human's auditory system has the ability to distinguish and concentrate on a particular person's voice among a group of peoples talking simultaneously. In the same way, in a smart antenna, any unwanted signal is attenuated. Better range/coverage, increased wireless system capacity, multipath rejection, and reduced expenses are the main advantages of the smart antenna system.

The two basic configurations of smart antenna systems are:

Switched – multi-beam antenna: Either a number of fixed beams with one beam turned on toward the desired signal or a single beam (formed by phase adjustment only) that is steered toward the desired signal

Adaptive antenna array: An array of multiple antenna elements where the main beam is put in the direction of the desired signal, while nulls are in the direction of the interference

Multi-beam Antenna Arrays

In this section, various antenna systems that are employed to generate multiple beams are discussed. In a broad perspective, the multi-beam antennas are defined as antennas that generate either multiple numbers of beams simultaneously or switch between beams in multiple directions. The IEEE standard terms for antennas (IEEE Standard 145–1993) defines multi-beam antenna as "an antenna capable of creating a family of major lobes from a single non-moving aperture, through use of multiport feed, with one-to-one correspondence between input ports and member lobes, the latter characterized by having unique main beam pointing directions."

Some of the earlier designs utilized conventional lenses or modified reflectors to generate multiple number of radiating beams. However, the challenges such as shadows created by the cluster feeds led the researchers to focus on the development of phased array antennas that made the entire antenna system more "smart." The multiple-beam antenna systems facilitate the following advantages compared to single-beam antennas:

- Ability to focus the radiation to a desired direction
- Increase the capacity of the wireless system
- Combat signal fading by suppressing interfering signals
- Facilitate tracking of the objects in radar systems

These key benefits are obtained through relatively higher system complexities in contrast with less complex mechanically scanned or switched design concept. In the following section, the classical designs are first discussed that include designs such as a lens fed by various array feed horns, a reflector illuminated by different feed horns, and phased array concepts. In the later part of the section, relatively new concepts of using engineered materials for the formation of multi-beam antenna arrays are discussed.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_66-1

Historical Review of Multi-beam Antenna Arrays

Early History (1906–1970)

Even though the concepts behind the first antenna array design were well established by 1920s (Fleming 1919), the natural outgrowth of the array to multiple-beam antennas or beam scanning antennas took almost another decade to flourish into the level of successful experiments. In fact, it is reported that Marconi performed several experiments in the antenna of his wireless communication system to enhance the performance of his receiver in certain directions as early as 1906 (Marconi 1906). There were enormous amounts of research interest fuelled by the Second World War to develop an efficient electromechanically scanned antenna system for radar applications. One classical example is the SCR-270 (Fig. 10) developed during Second World War by the US Army installed in Hawaii that provided Japanese formation during the Pearl Harbor attack. The initial approaches to scan the beam using electromechanical systems were successful and widely developed by 1940s. The technological limitations made some of the key challenges such as mutual coupling between the array elements a major barrier in the design optimization in this period (Brown 1937).

In the 1950s, the special situations after Second World War powered considerable funding for the scientific research related to military applications. The invention of ferrite-based phase shifters (Button 1984) in between 1954 and 1955 provided the technological seed that led to the development in the field of electronically scanning antenna systems (Sarkar et al. 2006). This led to the initial theories and experimental validation of beam-steering antenna arrays by changing the phase of the signals that excite the antenna elements (Braun 1925; Friis et al. 1934). These initial discoveries of mechanically scanned arrays led to many advanced design variations including large ground-to-ship radars and complex airborne radar arrays. Due to various system level limitations, most of these initial discoveries were spread across HF band and then later to EHF range. One of the rich resources that outlines the

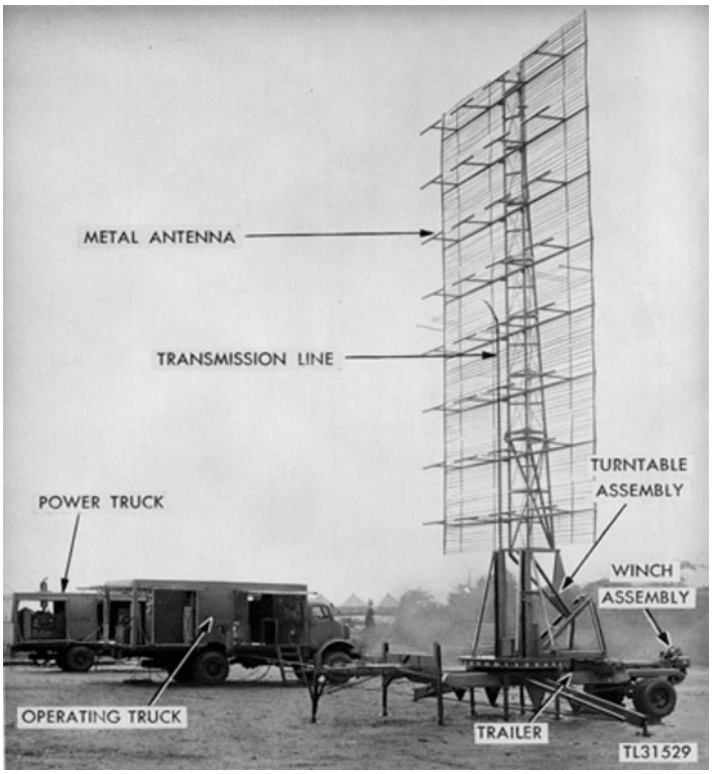


Fig. 10 SCR 270 antenna array with mechanical rotator

DOI 10.1007/978-981-4560-75-7_66-1

developments during this postwar period around the globe is entitled "Microwave Scanning Antennas" edited by Hansen (Hansen 1964, 1966). Until 1960s, the system cost was a major factor that prevents further developments on radar systems for nonmilitary applications. The evolution of printed circuit technology by late 1960s played a critical role in reducing the overall system expenses and thereby accelerating the research and development in broader areas of engineering physics. Even though the microstrip patch antennas were theoretically visualized by G.A. Deschamps in 1953 (Deschamps 1953), it took several years until early 1970, when Robert Eugene Munson succeeded in implementing the same in a sprint missile data link (Munson 1972).

Computational Challenges and Early Algorithms

While empathizing the challenges faced by researchers around the globe, it is important to note that by late 1950s even if the electromagnetic theory was quite established, the use of computer to solve these theories was completely nonexistent. It is quite interesting that, during the 1950s, the JPL in NASA used the word "computer" to represent an individual than a machine, where a specific group of people were designated to perform mathematical calculations by hand. By 1960s, the invention of mainframe computers made the complex electromagnetic calculations based on earlier algorithms (Mei 1965). Many of these initial computations were performed using point-matching solutions of Hallen's integral equation for metallic wires (an earlier version of method of moments) (Harrington and Harrington 1996). Hallen's equations were powerful enough to accurately model dipole antenna arrays. The introduction of algorithms based on geometrical theory of diffraction (GTD) was another breakthrough during the 1960s (Albertsen et al. 1975; Keller 1962). This enabled researchers to utilize well-established theories based on geometrical optics to accurately calculate the radiation properties of popular shapes such as edges and corners. Majority of the earlier computational techniques based on Green's functions were quite general, and most of them are in frequency domain compared to present-day tools which provide both frequency and time-domain solutions. The contributions from Schelkunoff and Dolph for the analysis of phased array antennas were another milestone. These contributions in the late 1940s utilized static weighting schemes (Dolph 1946; Schelkunoff 1943) and accelerated the development of antenna arrays with low side-lobe level. Another major work during the late 1960 by A.P Applebaum (1966) made key contributions to change the array weights which become one of the fundamental concepts behind the adaptive and reconfigurable antenna arrays. The concept of a retro-directive array that receives electromagnetic signals in one direction and retransmits in a mirror-image direction was originated in the late 1950s and credited with a US patent to Van Atta (1959).

Advanced era (1970–present): It is logical to classify the advanced era from a period in history when there were computational tools available, and many fundamental ideas were matured enough for practical implementation. During this period, the application spectrum of multi-beam antennas becomes much broader with the technological advancements. The availability of more efficient computational electromagnetic algorithms along with associate hardware made the research in this field of engineering rich. These scientific advancements enabled the researches to provide contributions with multi-beam antenna arrays with higher number of elements that are operating at the higher end of the spectrum. One of the earlier attempts for the implementation of passive electronically scanned array (PESA) took place by the early 1970s facilitating more efficient scanning than a mechanically rotating antenna. Another research focus during this era was on the development of conformal multi-beam antenna arrays for broadcasting, navigation, and directional finding. A notable development during the Cold War period (1962–1979) was the development of experimental Wullenweber arrays that are circular arrays designed for direction-finding applications. One of the prototypes, the 300 m diameter array with 120 radiating elements, was developed at the University of Illinois (Gething 1966). The initial ideas behind adaptive arrays begin with the motivation to control the receive pattern so that the interference of signals from unwanted directions

DOI 10.1007/978-981-4560-75-7_66-1

can be effectively minimized. One of the inventions during the late 1960s by P.W. Howells (1965) triggered the research across the globe leading to various key contributions in this exciting field of beam-scanning antenna arrays. The research during this period also diverted in various related areas including huge developments in many classical concepts in the field of multiple-beam antennas. Some of the earlier concepts to produce multiple beams were the use of digital/analog beam formers. In addition to these techniques, the antenna hardware concepts based on Butler matrices also evolved by this time (Butler 1961, 1965). The implementation of various lens implementations such as Rotman lens and Gent bootlace lens to produce multiple beams was also introduced in the 1960s (Gent 1957; Rotman and Turner 1963) which facilitated multiple focuses for one scanning plane. A decade later, some of these techniques were extended for various applications using microstrip technology (Archer 1984). The invention of P-I-N diodes and varactor diodes during this period attracted various researchers around the world to focus their research in the development of microwave phase shifters using this technology (White 1968, 1984). Even after a decade from the invention of P-I-N diodes and varactor diodes, the technology was not grown enough for high-power applications. At present, there are P-I-N diode-based microwave phase shifters that can handle kilowatts of peak power. Another important event in the mid-1970s is the development of microstrip antenna array by Munson. This development along with the cavity-based theory presented by Y.T.Lo et al. (1978) accelerated this exciting field of engineering. These developments along with the availability of low-cost fabrication technologies triggered the research world during the late 1980s with massive contributions in the development of multi-beam antenna arrays based on printed technology. The modern world of wireless communication is gifted with plenty of multidisciplinary technologies which accelerated this amazing field of engineering. Some of the exciting focus areas of multi-beam antenna research in this modern world include reconfigurable reflect arrays (Hum and Perruisseau-Carrier 2014), planar ultra-wideband modular arrays (Logan et al. 2013), and phased arrays with broad scanning capacity (Kavitha and Jacob 2013).

Types of Multi-beam Antennas

In general, the multi-beam antennas are classified as lens-based designs, reflector antennas, phased array antennas, and hybrid designs. These classifications are made based on the key system component that produces multi-beam. Each of these designs is detailed in the following sessions.

Lens Antennas

The basic working principle of lens antennas is based on the collimated action of any optical lens which forms a focused electromagnetic radiation. These antennas can be configured with single-feed or multi-feed elements. One of the generally used configurations of lens-based multi-beam antenna is outlined in Fig. 11. In a single-feed configuration, a single source illuminates the aperture and generates a beam based on the lens characteristics. The multiple-feed configuration or discrete lens array (DLA) excites the lens

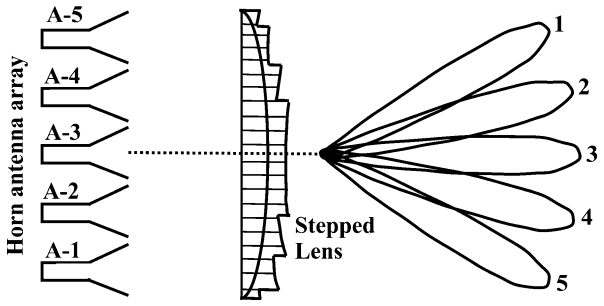


Fig. 11 Lens-based multi-beam antenna

DOI 10.1007/978-981-4560-75-7_66-1

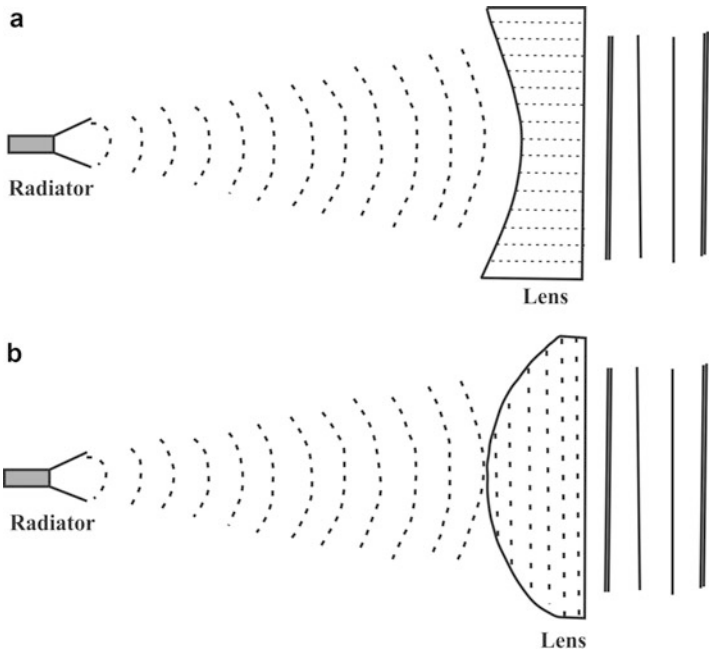


Fig. 12 (**a**) Fast-wave lens, (**b**) slow-wave lens

aperture and forms a constituent beam. By simultaneously exciting these beams, multiple beams can be formed based on the principle of superposition. Based on the properties of the lens, either constructive or destructive interactions can happen at a specific point in space, resulting in the formation of the beam pointing to a specific direction.

There are various types of lens antennas based on the principle of operation of the lens system; in a broad perspective, these are classified as:

(i) Fast-wave lenses (refractive index <1)
(ii) Slow-wave lenses (refractive index >1)

As the name indicates, these terms are related to the phase velocity of the wave that travels through the lens medium. These structures are illustrated in Fig. 12.

In fast-wave lenses, the electrical path length is decreased by the medium of the lens, and thus the phase velocity of the wave in the medium will be relatively higher. As shown in Fig. 12a, as the electromagnetic wave passes through the lens medium, the electric length is made shorter as the wave progresses. The fast lens antennas are also known as E-plane lens antennas. In contrast, a delay lens antenna is the one in which the wavelength in the medium will be increased, and this results in a reduced phase velocity as illustrated in Fig. 12b. The slow lens antennas are also known as dielectric lens antennas or H-plane lens antennas and further classified as:

- Lenses derived from conventional dielectrics such as lucite or polystyrene
- Lenses fabricated from artificial dielectric parts, ceramic, or metallic components

There are various lens topologies implemented for various applications including gain enhancement and beam forming. In the following section, various lens topologies that are employed for multi-beam antenna applications are being discussed. A detailed analysis of other lens architectures is detailed in Lee (1988).

DOI 10.1007/978-981-4560-75-7_66-1

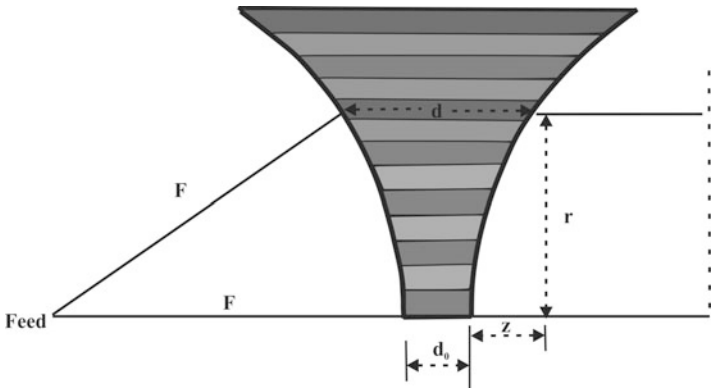


Fig. 13 The topology of a waveguide lens

The key lens topologies employed in multi-beam antennas are waveguide lens, bootlace lens, Rotman lens, and Luneburg lens.

Waveguide Lens The basic topology of a waveguide lens is shown in Fig. 13. It is primarily made of stacking waveguides of different dimensions. The concept is based on the fact that the phase velocity of electromagnetic waves exceeds the speed of light inside a waveguide. The shape of the waveguide is designed in order to provide corresponding phase velocity. More specifically, the distance travelled by the center ray from the source located on one side of the lens to a perpendicular plane on the other side of the lens is made equal to the path travelled by a general ray through other waveguide pieces.

This condition is satisfied from Eq. 3 to Eq. 5:

$$d = d_0 + z/(1 - n) \tag{3}$$

where d_0 is the smallest dimension of the lens.

$$n = \left[1 - (\lambda/\lambda c)^2\right]^{1/2} \tag{4}$$

where n is the refractive index of the lens based on the free space wavelength λ and waveguide cutoff wavelength λc.

$$z = F - \left(F^2 - r^2\right)^{1/2} \simeq r^2/2F \tag{5}$$

where F is the focal length of the lens and r is the distance from the element to the lens axis.

One of the classical applications for a waveguide lens antenna is shown in Fig. 14, formed by an array of waveguides with either circular or spherical cross sections for satellite communication systems. The length of the waveguides is made so as to produce corresponding phase delays when electromagnetic signals propagate through them.

One of the design challenges for the waveguide-based antenna array technology is the limited bandwidth. This is mainly because of the dispersive nature of the waveguide media. That is, the phase velocity of electromagnetic waves inside the waveguide varies with frequency. One of the classical solutions for this challenge is the use of a stepped design as shown in Fig. 15, in which a step is made in the lens. As a general rule, this is done when the lens thickness exceeds [λ/(1−μ)]. Even though the stepping

DOI 10.1007/978-981-4560-75-7_66-1

Fig. 14 Classical application of waveguide lens array antenna used in DSCS-III satellite

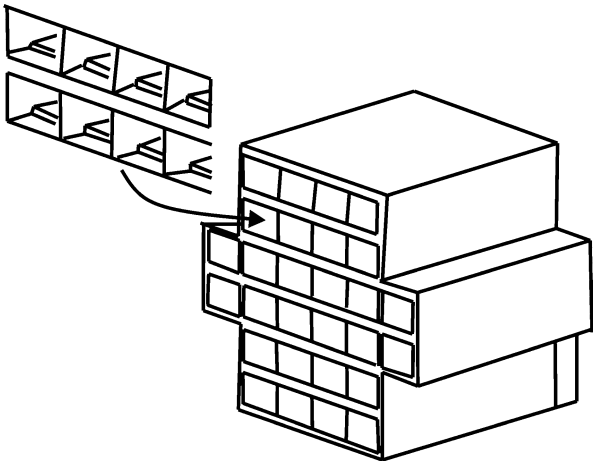


Fig. 15 Stepped waveguide array

can be made on either one or both inner sides of the waveguide, it is suggested to avoid the stepping on the feeding side of the waveguide array to avoid shadow effects.

Bootlace Lens The bootlace lenses are known for its capability to have more than one focal point and thereby provide wide-angle scanning capabilities. It consists of two array geometries that are arranged back to back to form a combined lens system. The array elements on one side are connected to the other side through a processing circuitry that consists of amplifiers and transmission lines. The beam characteristics of the bootlace lens depend on the length of the transmission lines along with the amplitude and phase between the receiving and transmitting elements. The transmission lines that connect the radiators on the inner and outer surfaces are connected through transmission lines mostly with coaxial cables. This gives an appearance of untightened bootlaces which gives the name bootlace lens. The geometry of a bootlace lens is shown below (Fig. 16).

A multi-beam antenna can be formed by employing bootlace lens by placing more than one source antennas at various focus points. Even though the initial designs were only capable of providing two-dimensional scanning (Kales 1964), later designs provided multifocal three-dimensional scanning

DOI 10.1007/978-981-4560-75-7_66-1

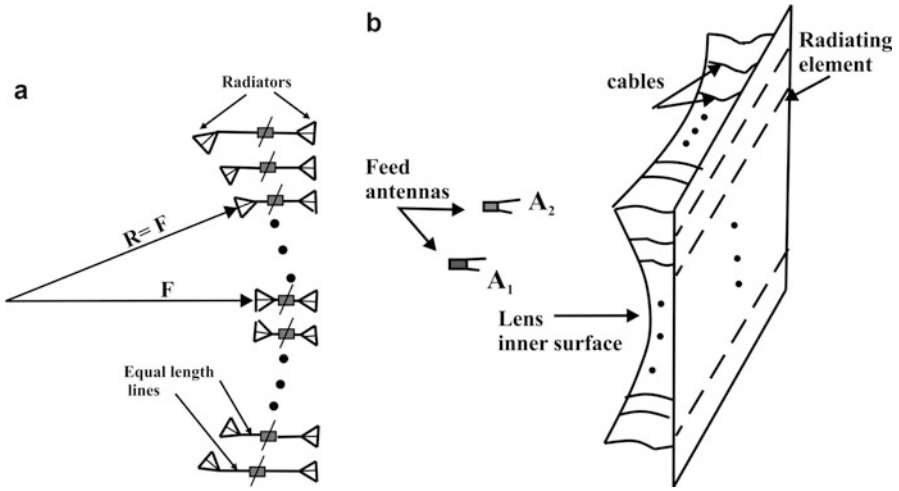


Fig. 16 (**a**) Diagram of bootlace lens, (**b**) geometry of one of the earlier versions of multifocal bootlace lens with three-dimensional scanning proposed by Jaganmohan B. L. Rao (1982)

facility (Rao 1982). One of the general theories developed by Gent in 1957 states that if the lens has a radius of 2R, then its focal points lie on a circle with radius R (Gent 1957).

In general, the bootlace lenses provide the following key features:

- Three-dimensional degrees of freedom for the design.
- The embedded microwave components in between the two lens surfaces facilitate illumination control of the outer surface and thereby provide beam scanning.
- The ability to implement stepping without the issues of shadow effect.

Rotman Lens Rotman lens is capable of producing simultaneous multiple beams when implemented on an array configuration. The diagram shown in Fig. 17 represents a Rotman lens which is a modified version of Gent bootlace lens (Gent 1957). Apart from the bootlace lens, the Rotman lens can provide much wider coverage exceeding scan angles of greater than 50° (Rotman and Turner 1963).

In one of its simplest configurations shown in Fig. 17, the transmitting or receiving antenna elements are mounted on a curved arc which either radiates or receives from the other curved portion of the lens. The array elements are arranged in a straight line at a geometrical plane perpendicular to the lens axis outside the lens. This generates multiple beams that are capable of steering toward the preprogrammed directions without any physical movement of the surface.

In comparison with other lens topologies, Rotman lens is relatively broadband in nature. The feed portion shown in the right-hand side can be implemented either using microstrip or stripline; meanwhile, the bootlace portion represented in the left-hand side can be implemented using coaxial cables. The key design parameters are outlined in Fig. 17b. This include:

- Focal angle, α
- Focal length, f1
- Ratio of beam angle and ray angle, γ
- Focal ratio, β
- Element spacing, d
- Peak beam angle, ψm

DOI 10.1007/978-981-4560-75-7_66-1

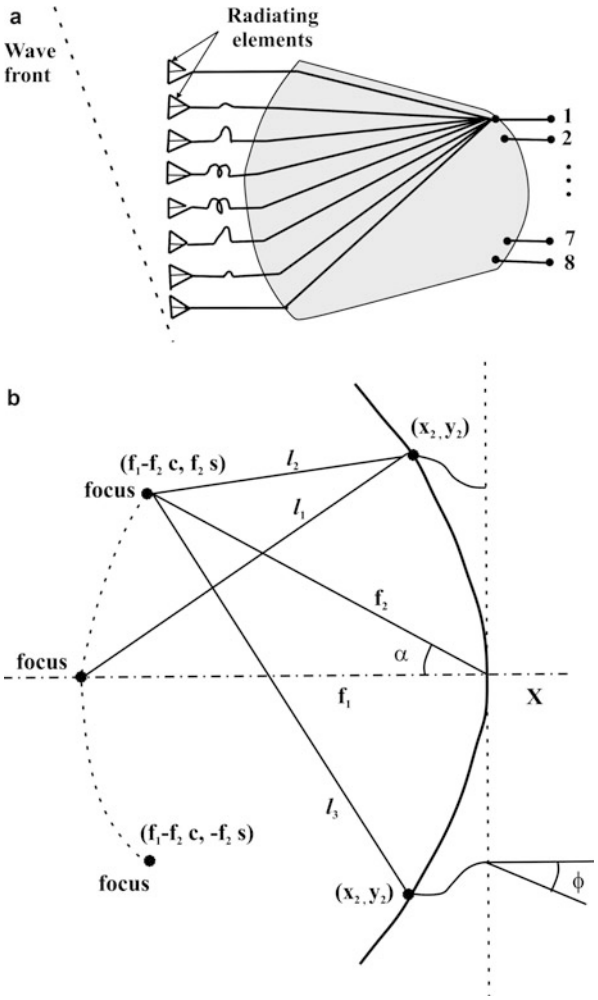


Fig. 17 (**a**) Rotman lens, (**b**) ray geometry of the Rotman lens

The design process is a trade-off between mutual coupling reduction and fabrication limits. Based on the detailed analysis provided in Hansen (2009), the design guidelines are given below:

- The focal ratio β can be calculated from the upper and lower focal length f_2 and f_1 using

$$\beta = \frac{f_2}{f_1} \tag{6}$$

- The beam angle ratio γ is given by

$$\gamma = \frac{\text{Sin}\ \psi}{\text{Sin}\ \alpha} \tag{7}$$

  where ψ is the beam angle and α is the focal angle.
- Another parameter **ζ** relates the distance γ_3 with f_1 and controls a percentage of the phase and amplitude error the lens possess can be calculated from

DOI 10.1007/978-981-4560-75-7_66-1

$$\zeta = \frac{\gamma_3 \gamma}{f_1} \tag{8}$$

- The maximum beam angle, ψm, is a critical design parameter that can be derived from

$$\zeta_{max} = \frac{(NE - 1)\gamma d}{2f_1} \tag{9}$$

where NE represents the number of elements.

- The quadratic lens equation is given by

$$a\left(\frac{w}{f_1}\right)^2 + \frac{bw}{f_1} + c = 0 \tag{10}$$

where the coefficients can be calculated from

$$a = 1 - \frac{(1-\beta)^2}{(1-\beta C)^2} - \frac{\zeta^2}{\beta^2} \tag{11}$$

$$b = -2 + \frac{2\zeta^2}{\beta^2} + \frac{2(1-\beta)}{1-\beta C} - \frac{\zeta^2 S^2 (1-\beta)}{(1-\beta C)^2} \tag{12}$$

$$c = -\zeta^2 + \frac{\zeta^2 S^2}{1-\beta\ C} - \frac{\zeta^4 S^4 (1-\beta)}{4(1-\beta C)^2} \tag{13}$$

$$\text{and} \quad C = \cos\alpha, S = \sin\alpha$$

The element spacing is critical since it has high impact on the grating lobes which can be calculated from

$$\frac{d}{\lambda} = \frac{1}{1 + \sin\psi_m} \tag{14}$$

Luneburg Lens In contrast to the earlier lens topologies, Luneburg lens is a spherically symmetric gradient-index lens whose refractive index decreases radially from center to the outer surface (Fig. 18), theoretically proposed by Luneburg (1944). The geometry of this lens includes two spheres in which one has an infinite radius and the other sphere is the lens surface.

The refractive index, n, of a unit length radius Luneburg lens is defined as

$$n = \left(2 - r^2\right)^{1/2} = \sqrt{\varepsilon_r} \tag{15}$$

where r is the radius of the lens.

The Luneburg lens produces a plane wave when the feed point is at one of the surface regions of the sphere. This in turn enables the Luneburg lens to produce multiple beams either by moving the feed point through the sphere surface or by exciting more than one feed locations around the sphere surface.

DOI 10.1007/978-981-4560-75-7_66-1

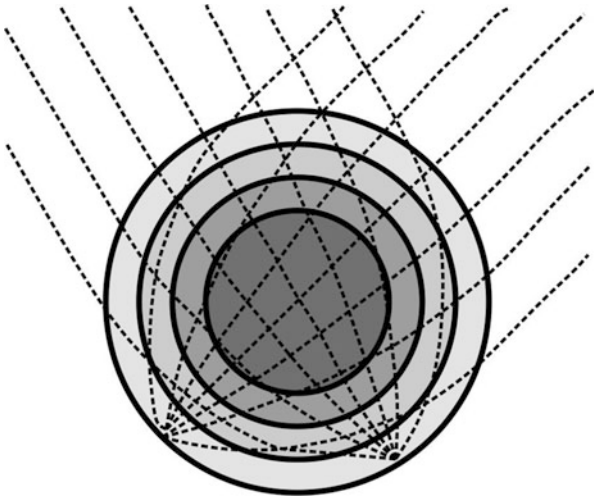


Fig. 18 The cross-sectional view of a Luneburg lens in which the gray shading is proportional to the refractive index, n

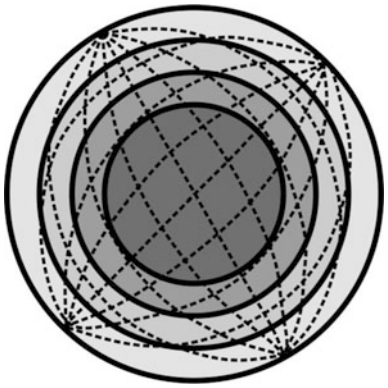


Fig. 19 Cross section of Maxwell's fish-eye lens, with gray shading representing increasing refractive index

Maxwell's fish-eye lens (Fig. 19) is a classical example of generalized Luneburg lens in which the refractive index varies according to

$$n = \sqrt{\varepsilon_r} = \frac{n_0}{1 + \left(\frac{r}{R}\right)^2} \tag{16}$$

Reflector Antenna

In a simple form, a reflector antenna consists of a reflector and an antenna feed. These types of antennas are widely used in long-distance high-gain applications like communication satellites, military satellites, high-resolution radars, etc. In the design, the reflector surface can take any general shape like paraboloid, hyperboloid, spheroid, ellipsoid, cylindrical, etc. The basic working principle of the reflector antenna is that the reflector converts the spherical wave to a plane wave in transmission mode and vice versa in reception mode. One of the most commonly used focal point-fed parabolic reflector antennas is shown in Fig. 20. On receiving mode, as a beam of parallel rays incident up on the parabolic reflector, the reflector

DOI 10.1007/978-981-4560-75-7_66-1

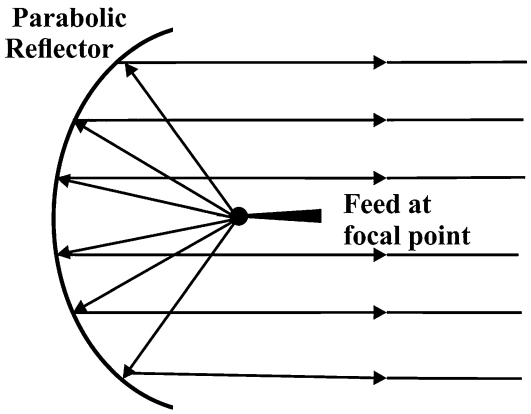


Fig. 20 Reflector antenna

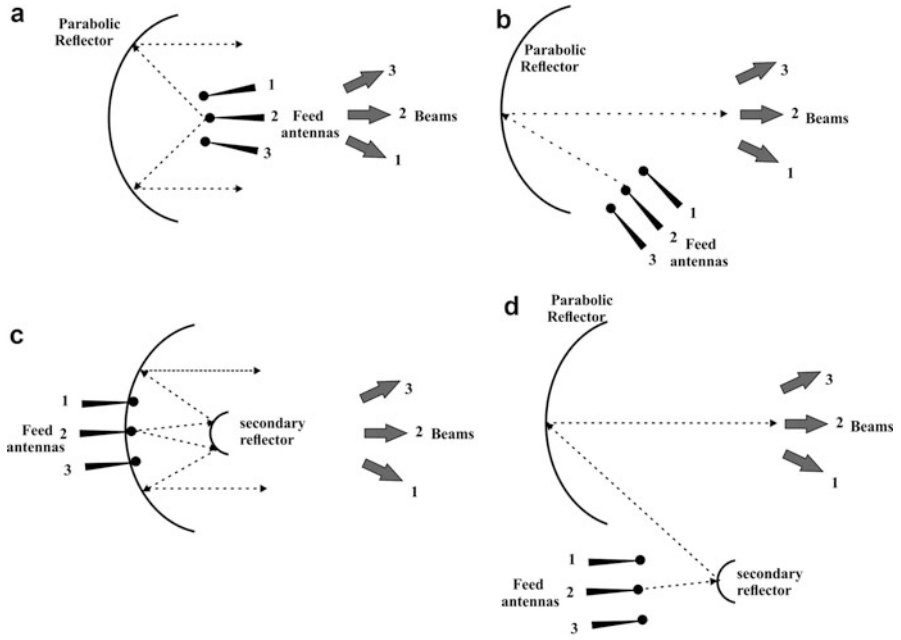


Fig. 21 (**a**) Parabolic reflector multi-beam antenna, (**b**) offset-fed parabolic reflector antenna, (**c**) focal point-fed Cassegrain reflector antenna, (**d**) offset-fed Cassegrain reflector antenna

focuses the entire ray to the focal point. Similarly on transmission mode, the beam is produced by the antenna feed, placed at the focal point, and emerges as parallel rays.

By using antenna array as the primary radiator, these antennas can also produce multiple beams. The performance of the reflector multi-beam antenna is similar to that of a lens antenna. Some of the most commonly used multi-beam antenna configurations based on this concept are shown in Fig. 21.

In a reflector antenna, aperture size of the reflector, type of the reflector, focal length, offset distance, and surface tolerance are the main design parameters to get a desired radiation pattern. The aperture size is a key design parameter that determines the gain and bandwidth of the antenna. In multiple-beam antenna applications, a large aperture size reflector is quite commonly used.

Aperture blockage by the feed and its support is the main disadvantage of a focal point-fed reflector antenna. One of the solutions to reduce this blockage which causes side lobes is the offset-fed reflector antenna. Cassegrain feeding is another solution which helps to overcome both the aperture blockage and

DOI 10.1007/978-981-4560-75-7_66-1

lack of control over the main reflector illumination. As illustrated in Fig. 21b, the reflected wave from a secondary hyperboloidal reflector illuminates the paraboloidal primary reflector. In this design mechanism, the reflectors and the feed antenna are positioned in such a way that the focus of the main reflector and one of the foci of the secondary reflector is coinciding in such a way that the outline of the secondary reflector controls the illumination of the main reflector. Electrically, such a system is same as that of a single reflector antenna with a longer focal length. Therefore, it provides better scanning performance than a single reflector type antenna. The Cassegrain focal point-fed reflector antenna also suffers from the aperture blockage. It can be minimized by choosing the diameter of the secondary reflector same as that of the feed. Another solution to minimize the aperture blockage is to employ offset Cassegrain feeding as depicted in Fig. 21d.

Phased Array Antenna

A phased array antenna is a directive antenna composed of an array of equally spaced individual antennas or elements. In such a design, the shape and direction of the radiation pattern is determined by the relative amplitude and phase of the electromagnetic signal at the individual elements. Similar to reflector or lens antennas, a phased array antenna with a beam-forming network can also be used to produce multiple beams. This configuration is normally preferred when the number of antenna elements ranges between a few to a few hundred. The beam-forming network in a phased-antenna system mainly combines the signals from all the elements in an array to form a receiving beam or conversely distributes the transmitted signal to different elements in an array to form a transmitting beam. Beam forming can be generally categorized as analog beam forming and digital beam forming. As the name implies, in analog beam forming, the amplitude and phase of each antenna element are controlled by an analog circuitry in the system. This includes analog phase shifters that provide smooth beam forming by controlling the phase of excitation of each antenna elements. Most of the modern-phased array systems use much advanced techniques that include digital beam-forming circuits in which the amplitude and phase of excitation of array elements can be software controlled with digital hardware.

Butler Matrix Butler matrix is one of the most popular multiple beam-forming networks employed to feed a uniformly spaced antenna array which was first introduced by Jesse Butler and Ralph Lowe in 1961. It is a feed network with $N = 2^n$ number of input ports and $N = 2^n$ number of output ports, where n is an integer. Thus, a Butler matrix network can provide linear phase distributions when the number of antenna elements is a power of two. In other words, if a signal is introduced at one input port, the matrix will produce equal amplitude excitations with constant phase difference at all output ports, resulting radiation in a particular direction in space. A signal introduced at another port will similarly create another beam pointing to different direction. Regarding the structure of the matrix, hybrid couplers and phase shifters are the basic building blocks of a Butler matrix network. This includes hybrid couplers with 90° and 180° designs along with Butler matrices. For example, the design to produce a broadside beam can employ a 180° hybrid coupler.

A Butler matrix with $N = 2^n$ number of input and output ports comprises Nn/2 number of hybrid couplers and (n-1)N/2 phase shifters. The phase difference between radiating elements in this matrix for the *i*th beam location is given by $(2i-1)\ 180^\circ/N$.

Similarly, the angle of beam title when a signal is introduced at the *i*th input of the matrix can be calculated by the equation

$$\theta_i = \mathrm{Sin}\left(\frac{N+1-2i}{2N}\ \frac{\lambda}{d}\right), \quad i = 1, 2, 3 \ldots .N \tag{17}$$

DOI 10.1007/978-981-4560-75-7_66-1

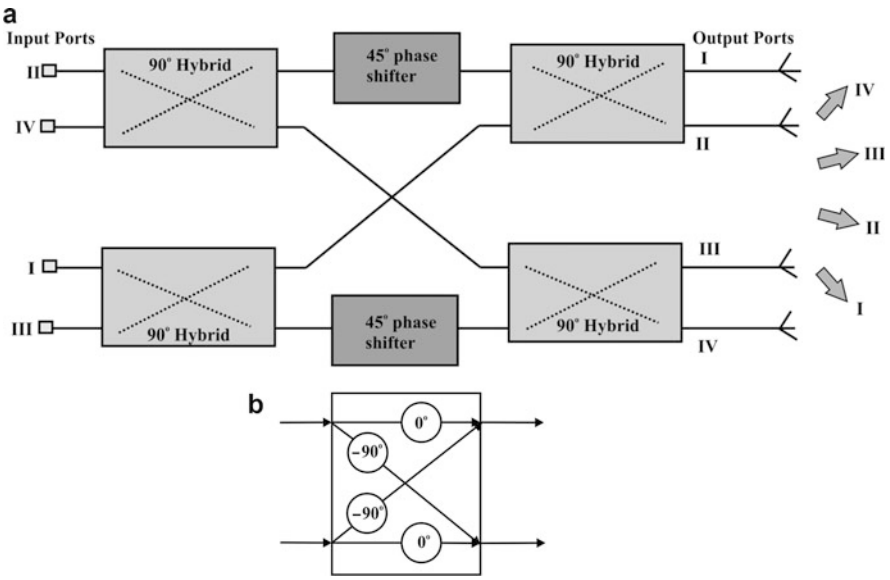


Fig. 22 (**a**) 4 × 4 Butler matrix, (**b**) 90° hybrid coupler

Table 2 Phase distribution and beam direction of a 4 × 4 matrix

Input port		I	II	III	IV
Output port	I	−90°	−45°	−180°	−135°
	II	−225°	−90°	−135°	0°
	III	0°	−135°	−90°	−225°
	IV	135°	−180°	−45°	−90°
Beam direction		48.59°	14.47°	−14.47°	−48.59°

Figure 22 demonstrates the basic functional block diagram of a 4 × 4 Butler matrix using 90° hybrid couplers. As shown, total of four hybrid couplers are arranged in two columns along with two phase shifters in the network topology. Table 2 shows the phase distributions produced by the matrix when a signal is introduced at a particular input port. The corresponding main beam directions from the antenna array are also given.

Another exciting feature of this network is that the antennas can produce multiple beams by simultaneously exciting two input ports. For example, if port 1 and port 4 are excited at the same time, the phase array will produce two beams simultaneously with a beam tilt of 48.59° and −48.59°. One of the limitations of this network is the lack of capability to produce two adjacent beams since there is high tendency to form a merged single beam.

Blass and Nolen Matrix *Blass matrix:* Another interesting matrix design is depicted in Fig. 23, called Blass matrix. It is a multi-beam feed network, in which "N" numbers of antenna array element feeders are connected to M number of beam port lines through directional couplers. In a receiving configuration, each beam-port couple signals from each antenna with the help of series feed lines. In this feed network, the couplers are equally spaced along the transmission line to produce a constant phase shift between elements that results in beam steering in a desired angle. In order to prevent reflections, the beam and element transmission lines are terminated with matched loads. One of the consequences of this

DOI 10.1007/978-981-4560-75-7_66-1

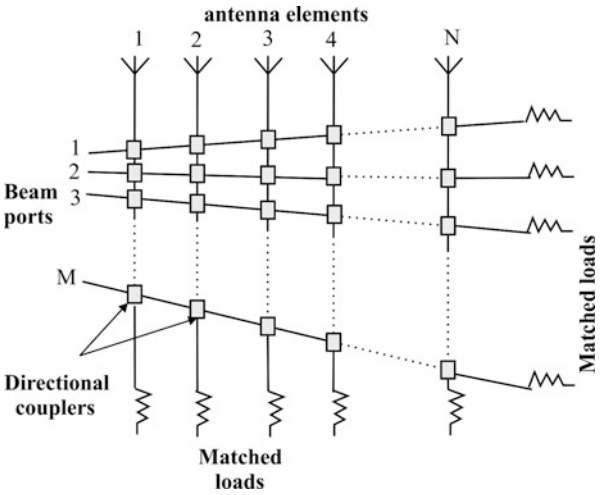


Fig. 23 Block diagram representation of Blass matrix

termination is the reduced efficiency of the entire system. If $\psi_{m,n}$ is the phase length from the m^{th} port to n^{th} element, then the phase difference of any two adjacent elements is

$$\begin{aligned} \Delta\psi_m &= \psi_{m,n+1} - \psi_{m,n} \\ &= \psi_{m,n} - \psi_{m,n-1} \end{aligned} \tag{18}$$

The corresponding output beam m is steered to an angle of

$$u_s = -\frac{\Delta\psi_m}{k\ d} \tag{19}$$

Nolen matrix: This is a combined design of both Butler matrix and Blass matrix in which the design utilizes a unitary coupling matrix circuit as shown in Fig. 24. As in a Blass matrix, the Nolen matrix also has different numbers of antenna elements and beam ports. In the basic configuration, the M^{th} port is coupled from the N^{th} antenna element as demonstrated in Fig. 25. When the number of antenna elements is a power of two, the Nolen matrix can be reduced to a simple Butler matrix.

Metamaterials and Its Application in Multi-beam Antenna Arrays

For the last two decades, metamaterials have been an interesting research area in the scientific community. The terminology metamaterials originated from the Greek term "meta" meaning "after" or "beyond" which implies metamaterials possess some special characteristics that are beyond the natural materials existing in nature. Thus, metamaterials are artificial structures that exhibit electromagnetic properties not usually found. The term metamaterials was first formed by Rodger M. Walser, University of Austin, in 1999. It was originally defined as "macroscopic composites having a synthetic, three dimensional, periodic cellular architecture designed to produce an optimized, not available in nature, of two or more responses to specific excitations." Some of the other definitions that have been found in the literatures are "materials whose permeability and permittivity derived from their structure" and "artificial effectively homogeneous electromagnetic structures with unusual properties not readily available in nature" (Caloz and Itoh 2006). Based on the way in which these materials treat the incident fields of an electromagnetic radiation, they are classified as double-negative materials (DNG), electromagnetic band gap (EBG) or

DOI 10.1007/978-981-4560-75-7_66-1

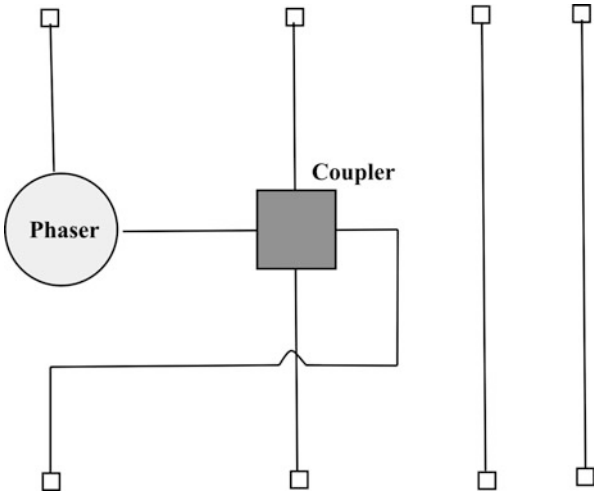


Fig. 24 Unitary coupling matrix

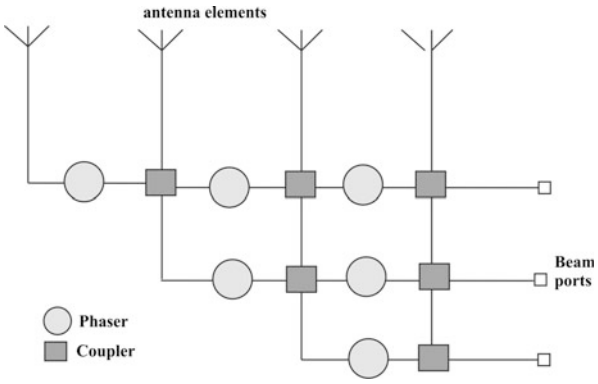


Fig. 25 Nolen matrix

photonic band gap materials (PBG), and frequency-selective surfaces (FSS). In general, they are composed of an array of dielectric or metallic elements in one, two, or three dimensions. The electromagnetic properties of these materials are characterized by various design factors including the element shape, array texture, electrical parameters of the material, and the distance between elements in the array.

As an electromagnetic wave incident up on a DNG material, it provides negative permittivity and negative permeability. These types of materials have found applications in beam-steering antennas, lens antennas, and microwave filters. However, metamaterials like PBG or EBG prohibit the propagation of incident waves across a range of frequency band. As a general rule, the periodicity in these materials is in the range of half wavelength. These materials also found great applications in antenna engineering to enhance the gain and suppress the surface wave, in microwave filters to create band-stop response, and in electromagnetic shielding applications. FSS is a kind of planar metamaterials, which is made of an array of metallic patches or perforated conductors. As the name implies, these materials are transparent to some frequency bands while reflective, absorbing, or redirecting to others. These materials found significant applications in antenna, filters, polarizers, radomes, and electromagnetic shielding. FSS-based EBG surfaces formed by cascading some FSS layers are also used for antenna applications.

As shown in Fig. 26, in terms of permittivity and permeability, materials are generally classified into four categories. A material with permittivity and permeability greater than zero is known as double-

DOI 10.1007/978-981-4560-75-7_66-1

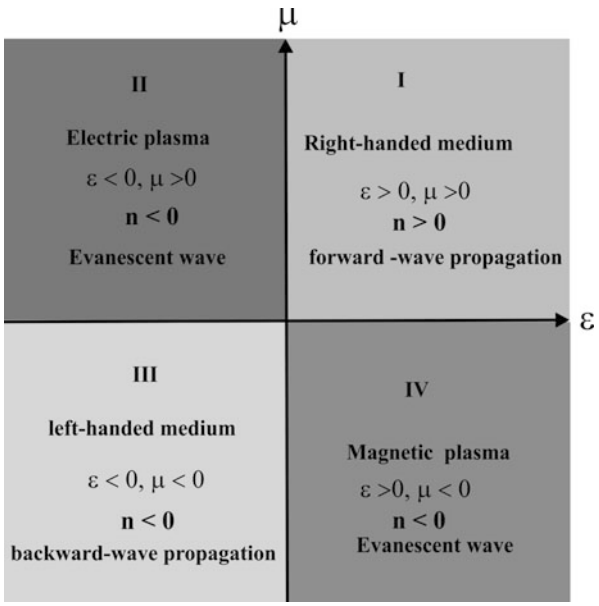


Fig. 26 Material classifications in terms of permittivity and permeability

positive metamaterial or right-handed materials which support forward propagating wave. A material with permittivity less than zero and permeability greater than zero will be designated as epsilon-negative medium. Similarly, a material with permeability less than zero and permittivity positive is known as mu-negative metamaterial. A material with permittivity and permeability less than zero will be designated as double-negative material or LHM which supports backward propagation of waves.

Applications in Antenna Engineering

Short history: The history of metamaterials started in 1968 when Vaselago (1968) proposed the existence of a material with simultaneous negative electric permittivity and magnetic permeability. However, due to the lack of experimental verifications, his invention was ignored for more than three decades. The first revolution of metamaterial occurred in the late 1990s (Pendry et al. 1996) when Sir Pendry demonstrated a wire medium whose permittivity is negative. Then, in 1999, Sir Pendry discovered the split ring resonators having negative permeability (Pendry et al. 1999). Another important landmark was the first artificial left-handed material using a combination of wires and SRRs by Dr. Smith in 2001 (Shelby et al. 2001). The year following this event, LHM was presented almost simultaneously using transmission line approach by three research groups (Eleftheriades, Oliner, and Caloz-Ltoh) (Caloz and Itoh 2002; Iyer and Eleftheriades 2002; Oliner 2003). Any one-dimensional RHM can be resented or equivalent to a conventional transmission line composed of series inductance and shunt capacitance, whereas an LHM can be represented as a transmission line consisting of series capacitance and shunt inductance. In fact, the scenario with a current flow through this structure can be represented as a series capacitance accompanied by an inductance, and a shunt capacitance is accompanied by a capacitance. Thus, eventually, a composite right-left-handed transmission line model has been evolved to represent the RHM and LHM (Caloz and Itoh 2006).

The second revolution of metamaterial has happened with the concept of electromagnetic wave bending using gradient refractive index medium in 2005 (Smith et al. 2005) and the invisible cloaks in 2006 (Leonhardt 2006; Pendry et al. 2006). After these realizations, the interest toward the metamaterial has been considerably increased, and many scientific papers have been published. Due to the extraordinary features, metamaterials have been found and are potential candidates for various engineering

DOI 10.1007/978-981-4560-75-7_66-1

applications. Especially in antenna engineering, metamaterials are widely being used to improve its performances. Examples include antenna substrates for miniaturization, bandwidth enhancement, gain enhancement, and controlling the direction of radiation. Metamaterials as antenna superstrates can be utilized to improve impedance and directivity bandwidth, change the polarization, increase gain, and beam tilting. Instead of conventional transmission lines, metamaterial-based transmission lines can be employed in relatively more efficient feed networks. Compared to conventional design of transmission line-based feed networks, the metamaterial-based feed networks have the advantage of low profile and broad bandwidth. In phased array antennas, metamaterials are employed to improve the impedance matching and beam-scanning range. Another area of interest especially in military applications is in the field of Radom design where the ground plane regions can be realized with metamaterials. In addition, the design flexibility of these engineered materials possesses good potential to be embedded in applications such as Luneburg lens antennas and reflector antennas.

Design Examples

This session outlines some of the recent developments in metamaterial-based multi-beam antenna arrays. The objective of this session is to provide an experimental level demonstration of this magnificent field of engineering.

High-Gain Reconfigurable Sectoral Antenna Using an Active Cylindrical FSS Structure

Sectoral antennas have been widely reported and are popular in various classical applications such as base station antennas and point-to-multipoint communications to cover a specific area. It is a type of directional antenna with a narrow beam width in one plane and a wide beam width in the other plane. A sectoral antenna with radiation pattern reconfigurability provides more design flexibility and in turn enhances the overall system performance. In addition, since the beam is directed toward the desired user, it provides energy saving and thereby higher system efficiency.

One of the practical realizations of metamaterial-based beam-scanning antenna is shown in Fig. 27. It composed of active unit cell-based cylindrical FSS structure around an omnidirectional electromagnetically coupled coaxial dipole array. An array of discontinuous strips and PIN diodes inserted into their discontinuities forms the active FSS unit cell. They are placed cylindrically with an angular periodicity of $\theta_{FSS} = 30^\circ$ and radius of 50 mm on a substrate with permittivity 3, thickness 0.254, and loss tangent 0.0013. High-frequency PIN diodes GMP-4201 are used in the prototype. Two narrow DC feeding lines are used at both sides of the strips for supplying diodes. To produce same amount of current in all the diodes, high-value resistors are also used at the top of each line between the strip and DC feeding line. As shown in Fig. 27b, the ECCD array was also fabricated with a UT-141B semirigid coaxial cable from Micro-Coax.

In order to produce reconfigurable radiation pattern, the cylindrical FSS structure is divided into two semicylinder sectors, one with diodes on and the other with diodes off. The semicylinder with the diodes on operates as an array of continuous strips with high reflection, whereas the other semicylinder with the diodes off operates as an FSS with an array of printed dipole below resonance frequency and allows the transmission of the incident electromagnetic wave. Therefore, the FSS structure converts the omnidirectional radiation pattern of the source into directive pattern. By switching the diode states, the direction of the high and low reflective sectors changes; therefore, the radiation pattern of the antenna can be swept in the entire 360° azimuth plane.

DOI 10.1007/978-981-4560-75-7_66-1

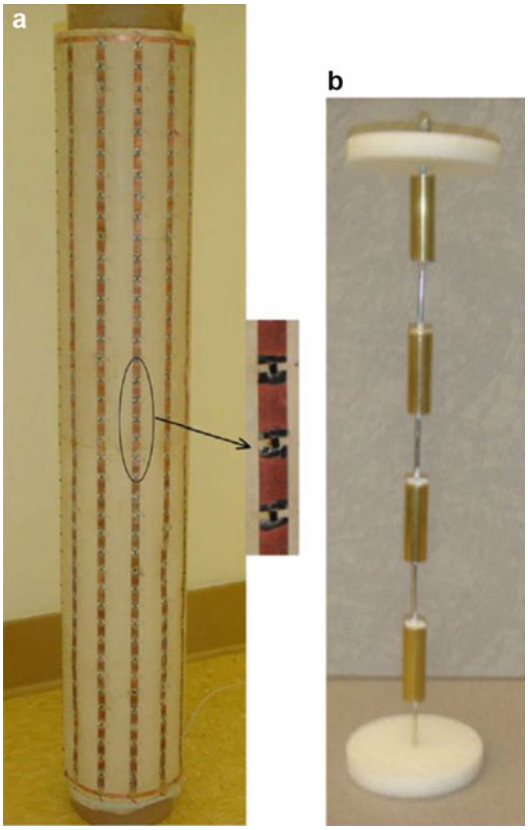


Fig. 27 Photograph of a metamaterial-based beam scanning antenna prototype, (**a**) antenna system, (**b**) source antenna array. (© The Institute of Electrical and Electronics Engineers, Published with permission from (Edalati and Denidni 2011))

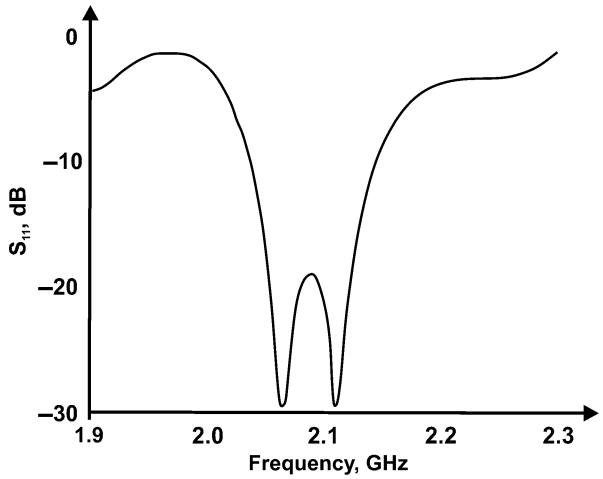


Fig. 28 Reflection coefficient of the FSS-based beam-scanning antenna

During measurement, five columns of diodes were in the off state and seven columns of diodes in the on state. A supply voltage of 0 V is given to the diode for off state, whereas 37 V DC is used for the on state. The S_{11} (Fig. 28) of the reconfigurable antenna indicates ~6 % of 2:1 VSWR bandwidth around the center frequency.

DOI 10.1007/978-981-4560-75-7_66-1

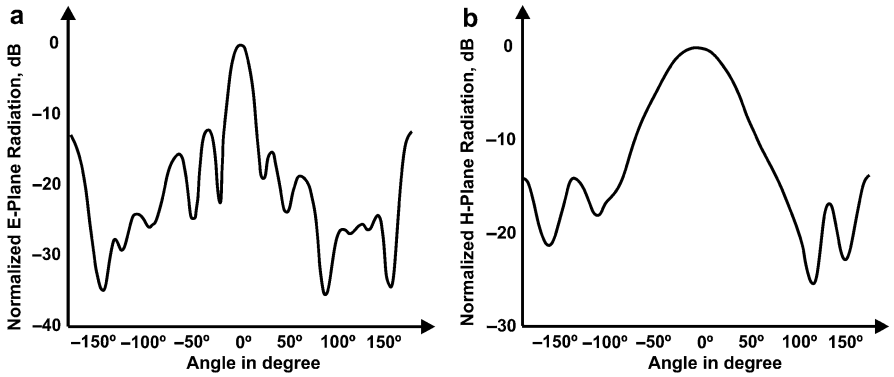


Fig. 29 Normalized radiation patterns of the antenna in two principle planes

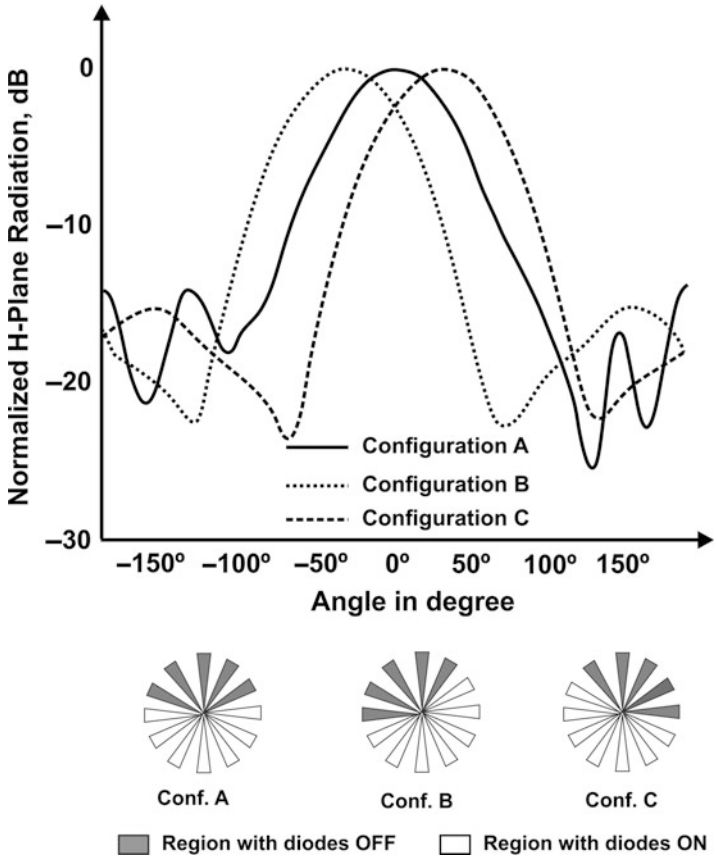


Fig. 30 Beam scanning in H-plane with various diode reconfiguration scenarios

The normalized radiation patterns of the antenna in the E- and H-planes are shown in Fig. 29. The antenna provides a directional radiation pattern with half power beam width of 20° and 70° in the E- and H-plane respectably with a side-lobe level of −13 dB. The H-plane radiation pattern of the antenna, for switching the beam toward 0°,−30° with different reconfiguration scenarios, is demonstrated in Fig. 30. The radiation characteristics show a gain of around 13dBi.

DOI 10.1007/978-981-4560-75-7_66-1

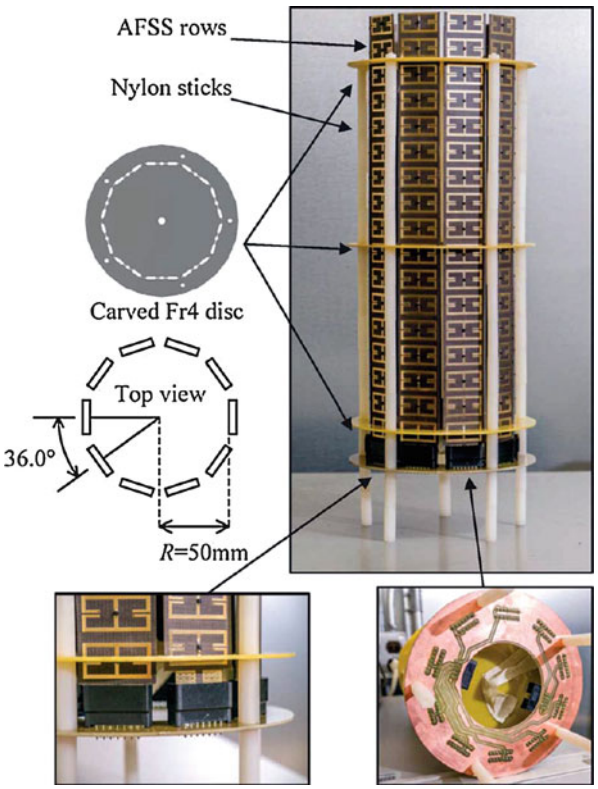


Fig. 31 Photograph of an electronically beam reconfigurable antenna (© The Institute of Electrical and Electronics Engineers, Published with permission from (Liang et al. 2013))

Electronically Radiation Pattern Steerable Antennas Using Active Frequency-Selective Surfaces

Compared to conventional smart antenna systems with phase shifters, the demand for electronically steerable antenna solutions is quite demanding in this modern telecommunication era. The advantages include more flexible design and omnidirectional beam formation. One of the interesting design examples capable of electronic steering of radiation pattern is shown in Fig. 31.

It consists of ten columns of active FSS unit cell-based cylindrical array with a radius of 50 mm. Each column of cylindrical array has 17 unit cells. Special design schemes are employed for biasing through the bottom of the antenna system. Finally, the pair of bias cables is connected with a programmable multichannel voltage controller. The structure of the omnidirectional coaxial collinear source antenna at the center is made of LMR-195 coaxial cable sections.

The antenna operates in three single mode configurations, namely, case-1, case-2, and case-3, as illustrated in Fig. 32. In the case-1 scenario, five columns of the cylindrical FSS are set into reflective mode, and the other five columns are transparent mode. The reflective FSS arrays block the electromagnetic energy, and the transparent FSS array allows passing the beam. The case-2 also works the same as that of the case-1 mode with an additional opaque layer resulting steering of the beam. In case-3, one column of the cylindrical FSS is set to partially opaque using appropriately biased varactor diode, resulting a beam in between 0° and 18°. By changing the different sectors, this particular antenna can provide 360° steerability of the main beam in H-plane. With the same beam control method, multi-beams (two beams) are also generated.

Figure 33 shows the layout of the active FSS unit cell realized on a high-frequency laminate with permittivity 3.5, loss tangent 0.0007, and thickness 0.8 mm. The top side of the unit cell consists of a pair

DOI 10.1007/978-981-4560-75-7_66-1

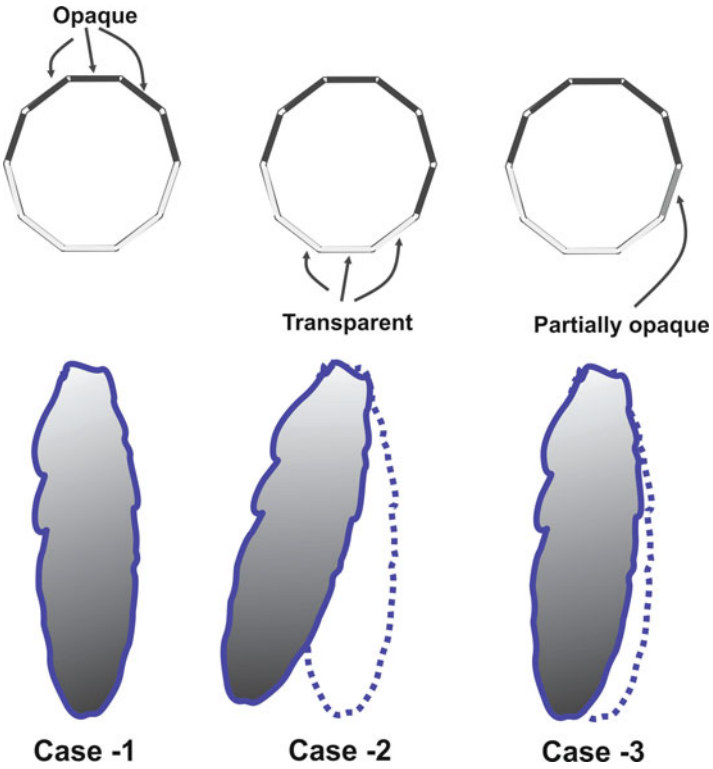


Fig. 32 Single-beam configurations

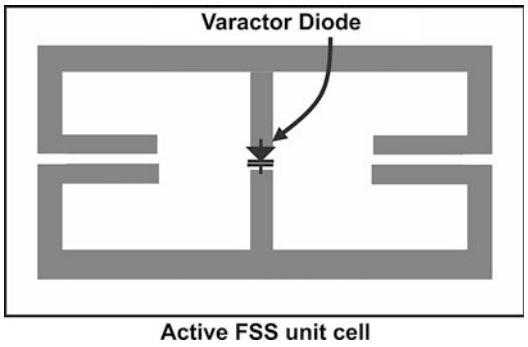


Fig. 33 Unit cell layout

of vertical anchor- shaped metal strips and a varactor mounted in the middle. The unit cells repeat along vertical direction to form an array. The bottom side comprises the DC biasing network. Vias are used to connect the front side strip to the biasing network. RF chokes are also applied between two unit cells and near the vias. BB857 (by Infineon) and 22nH high Q inductor (by Murata) varactor diodes are used. At 1.78 GHz, 6 V and 30 V are set as the voltage for reflection and transmission mode of the cylindrical FSS, respectively. And the voltage for transparent mode is 12 V.

Figure 34 shows the H-plane radiation pattern of the antenna for various reconfigurations. In case-1, the main beam direction is 178° and 3 dB width is 77°. For case-2 configuration, the main beam direction is 196° and 3 dB width is 84°. Meanwhile, with case-3 mode, the main beam direction is 192° and 3 dB width is 88°. Moreover, the case-1 also indicates a null point 51.9 dB smaller than the main lobe, which is an interesting characteristic to isolate from noise source, especially in anti-jammer systems. The antenna

DOI 10.1007/978-981-4560-75-7_66-1

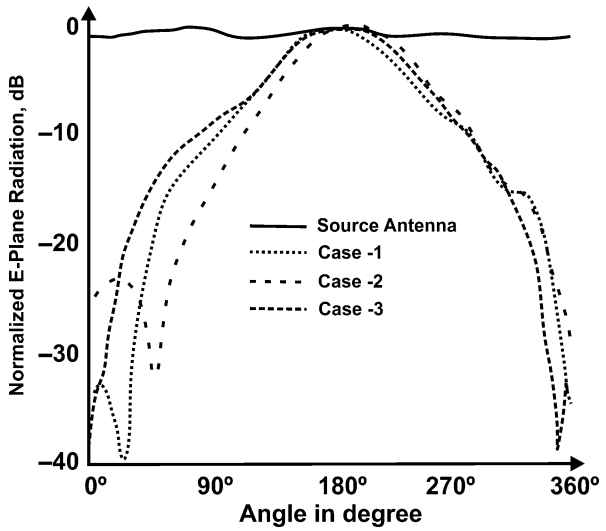


Fig. 34 H-plane radiation pattern of the antenna in various pin diode configurations

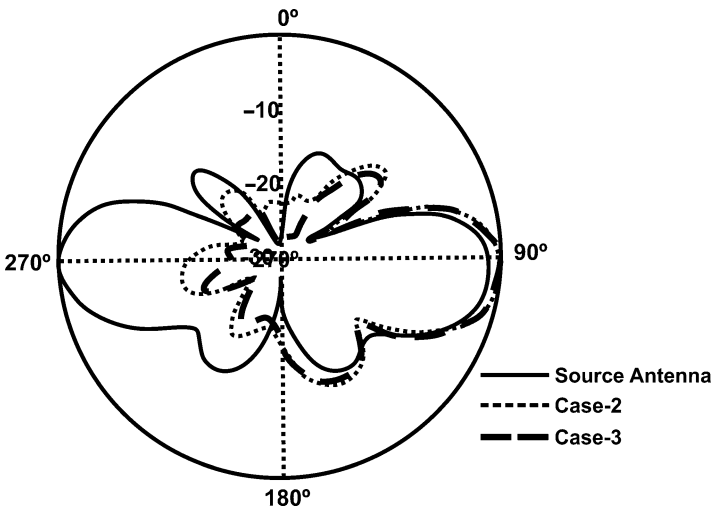


Fig. 35 Radiation pattern in the E-plane for the reconfigurable antenna in various reconfiguration states

also exhibits good cross polar level of 20 dB. The gain of the source antenna is 3.6 dBi. The designed AFSS antenna provides a gain of 7dBi for case-1, 6.9 dBi for case-2, and 6.6 dBi for case-3 mode. The E-plane radiation pattern of the antenna is presented in Fig. 35. The E-plane pattern is mainly decided by the inner omnidirectional source antenna.

The dual beam configurations are shown in Fig. 36, and the corresponding radiation patterns are illustrated in Fig. 37. The two beams of configuration-C point to 54° and 206°. As expected, they are supposed to be 54° and 120°. Moreover, there is less lobe pointing at about 120° that is due to the two columns that are not wide enough to stop the signal, and signal from the other two lobes also contribute to this radiation. Both radiation direction and amplitudes can be controlled.

DOI 10.1007/978-981-4560-75-7_66-1

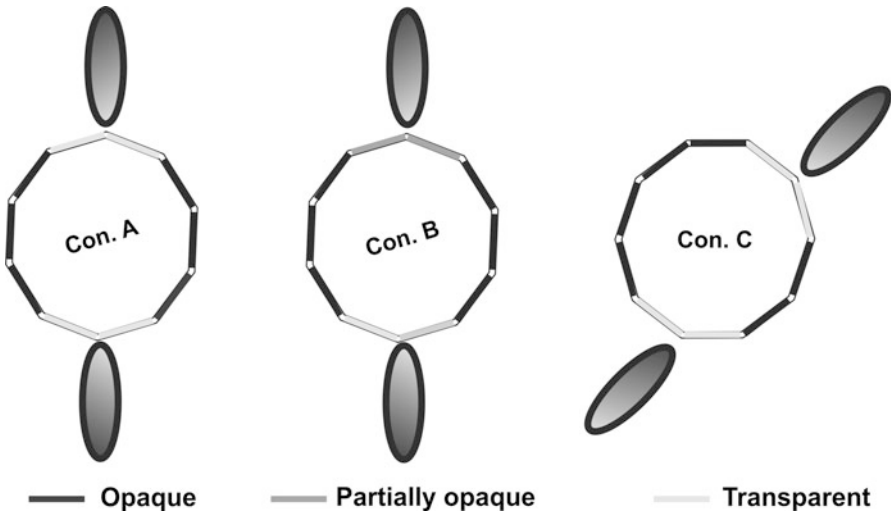


Fig. 36 Dual-beam mode configurations

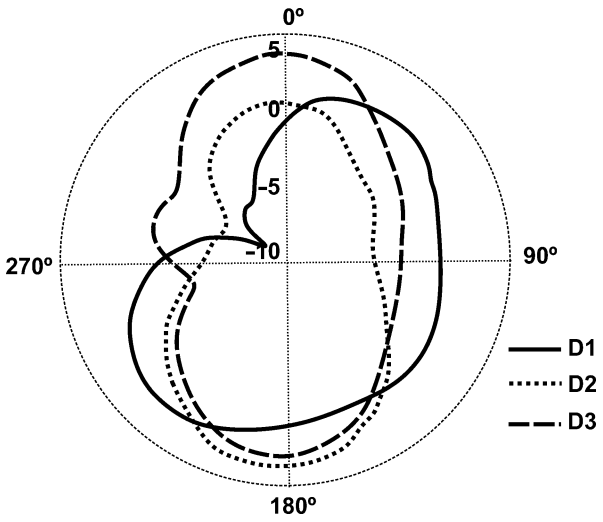


Fig. 37 The gain patterns in the H-plane for various beam-steering configurations

Conclusion

In this technological era, wireless communication plays a key role in every phase of life. The modern wireless communication systems and related technologies are synergistic products of interdisciplinary areas of engineering. The main objective of this chapter was to introduce the concept of classical and most recent design concepts of multi-beam antenna arrays. Toward this goal, basics of antenna arrays and smart antenna concepts were first presented. A considerable emphasis was given to classical multiband antenna designs and their beam-forming circuits. In order to provide the most recent advancements in multi-beam antenna technology, the idea of metamaterial and its applications were then introduced. Finally, some metamaterial-based multi-beam antenna designs were also provided.

DOI 10.1007/978-981-4560-75-7_66-1

Cross-References

- ▶ Antennas in Automobile Radar
- ▶ Antennas in Radio Telescope Systems
- ▶ Applications of Phased Array Feeders in Reflector Antennas
- ▶ Beam-Scanning Leaky Wave Antennas
- ▶ Conformal Array Antennas
- ▶ Phased Arrays
- ▶ Reconfigurable Antenna Arrays for Wireless Communications
- ▶ Reflectarray Antennas

References

Albertsen N, Bach H, Balling P, Jensen F, Pontoppidan K, Solymar L, . . . Jensen N (1975) A study on radiation pattern prediction for high frequency satellite antennas. Final report. Technical University of Denmark, Lyngby Institute of Electromagnetics 1

Applebaum S (1966) Adaptive arrays, Syracuse University Research Corporation. SPL TR66-1, August

Archer D (1984) Lens-fed multiple beam arrays. Microw J 27:171

Balanis CA (2005) Antenna theory: analysis and design, 3rd edn. Wiley, Hoboken

Balanis CA, Ioannides P (2007) Introduction to smart antennas. Morgan & Claypool Publishers, San Rafael

Braun F (1925) Directed wireless telegraphy. Electrician 57:222

Brown G (1937) Directional antennas. Radio Eng, Proc Inst 25(1):78–145

Butler J (1961) Beam-forming matrix simplifies design of electronically scanned antennas. Electron Des 9(8):170–173

Butler J (1965) Digital, matrix and intermediate-frequency scanning. Paper presented at the antennas and propagation society international symposium, New York

Button KJ (1984) Microwave ferrite devices: the first ten years. Microw Theory Tech, IEEE Trans 32(9):1088–1096

Caloz C, Itoh T (2002) Application of the transmission line theory of left-handed (LH) materials to the realization of a microstrip "LH line". Paper presented at the 2002 I.E. antennas and propagation society international symposium, San Antonio

Caloz C, Itoh T (2006) Electromagnetic metamaterials: transmission line theory and microwave applications. Wiley, Hoboken

Deschamps GA (1953) Microstrip microwave antennas. Paper presented at the third USAF symposium on antennas, Illinois

Dolph C (1946) A current distribution for broadside arrays which optimizes the relationship between beam width and side-lobe level. Proc IRE 34(6):335–348

Edalati A, Denidni TA (2011) High-gain reconfigurable sectoral antenna using an active cylindrical FSS structure. Antennas Propag IEEE Transac 59:2464–2472

Fleming JA (1919) The principles of electric wave telegraphy and telephony. Longmans, Green, London

Friis H, Feldman C, Sharpless W (1934) The determination of the direction of arrival of short radio waves. Radio Eng, Proc Inst 22(1):47–78

Gent H (1957) The bootlace aerial. Royal Radar Establishment J (40):47–58

Gething PJD (1966) High-frequency direction finding. Paper presented at the proceedings of the institution of electrical engineers, London 113:49–61

Hansen RC (1964) Microwave scanning antennas: apertures, vol 1. Academic, New York
Hansen RC (1966) Microwave scanning antennas: array systems, vol 3. Academic, New York
Hansen RC (2009) Phased array antennas, vol 213. Wiley, Hoboken
Harrington RF, Harrington JL (1996) Field computation by moment methods. Oxford University Press, New York, NY, USA
Haupt RL (2010) Antenna arrays: a computational approach. Wiley, Hoboken
Howelss PW (1965) Intermediate frequency side-lobe canceller: Google Patents
Hum SV, Perruisseau-Carrier J (2014) Reconfigurable reflect arrays and array lenses for dynamic antenna beam control: a review. IEEE Trans Antennas Propag 62(1):183–198. doi:10.1109/TAP.2013.2287296
IEEE Standard 145-1993 (R2004) IEEE standard definitions of terms for antennas
Iyer AK, Eleftheriades GV (2002) Negative refractive index metamaterials supporting 2-D waves. Paper presented at the IEEE MSS-S international microwave symposium digest, Seattle
Kales M (1964) Design considerations for two-dimensional symmetric bootlace lenses: DTIC document
Kavitha BS, Raglend IJ (2013) A wide-scan phased array antenna for a small active electronically scanned array: a review. Paper presented at the Circuits, Power and Computing Technologies (ICCPCT), India, pp.1008–1016
Keller JB (1962) Geometrical theory of diffraction. JOSA 52(2):116–130
Lee JJ (1988) Lens antennas. In: Antenna handbook. Springer, New York, pp 1073–1131
Leonhardt U (2006) Optical conformal mapping. Science 312(5781):1777–1780
Liang Z, Qun W, Denidni TA (2013) Electronically radiation pattern steerable antennas using active frequency selective surfaces. Antennas Propag, IEEE Trans 61:6000–6007
Lo YT, Solomon D, Richards WF (1978) Theory and experiment on microstrip antennas. In: Proceedings of the 1978 antenna applications symposium, Illinois
Logan JT, Holland SS, Schaubert DH, Kindt RW, Vouvakis MN (2013) A review of Planar Ultrawideband Modular Antenna (PUMA) arrays. Paper presented at the Electromagnetic Theory (EMTS), Proceedings of 2013 URSI international symposium on, Orlando, Florida.
Luneberg R (1944) Mathematical theory of optics, Brown Univ, Providence, RI, USA
Marconi G (1906) On methods whereby the radiation of electric waves may be mainly confined to certain directions, and whereby the receptivity of a receiver may be restricted to electric waves emanating from certain directions. Proc R Soc Lond, Ser A 77(518):413–421
Mei K (1965) On the integral equations of thin wire antennas. Antennas Propag, IEEE Trans 13(3):374–378
Munson R (1972) Microstrip phased array antennas. Paper presented at the proceedings of twenty-second symposium on USAF antenna research and development program
Oliner AA (2003) Invited – a planar negative-refractive-index medium without resonant elements. Paper presented at the 2003 I.E. MTT-S international microwave symposium digest, Philadelphia
Pendry JB, Holden AJ, Stewart WJ, Youngs I (1996) Extremely low frequency plasmons in metallic mesostructures. Phys Rev Lett 76(25):4773–4776
Pendry JB, Holden AJ, Robbins DJ, Stewart WJ (1999) Magnetism from conductors and enhanced nonlinear phenomena. IEEE Trans Microw Theory Tech 47(11):2075–2084
Pendry JB, Schurig D, Smith DR (2006) Controlling electromagnetic fields. Science 312(5781):1780–1782
Rao J (1982) Multifocal three-dimensional bootlace lenses. Antennas Propag, IEEE Trans 30(6):1050–1056
Rotman W, Turner R (1963) Wide-angle microwave lens for line source applications. Antennas Propag, IEEE Trans 11(6):623–632

Sarkar TK, Mailloux R, Oliner AA, Salazar-Palma M, Sengupta DL (2006) History of wireless, vol 177. Wiley, Hoboken
Schelkunoff SA (1943) A mathematical theory of linear arrays. Bell Syst Tech J 22(1):80–107
Shelby RA, Smith DR, Schultz S (2001) Experimental verification of a negative index of refraction. Science 292(5514):77–79
Smith DR, Mock JJ, Starr AF, Schurig D (2005) Gradient index metamaterials. Phys Rev E 71:036609
Van Atta LC (1959) Electromagnetic reflector: Google Patents
Veselago VG (1968) The electrodynamics of substances with simultaneously negative values of ε and υ. Sov Phys Usp 10:509
White JF (1968) Review of semiconductor microwave phase shifters. Proc IEEE 56(11):1924–1931
White JF (1984) Origins of high-power diode switching. IEEE Trans MTT 32(9):1105–1117

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_68-1

Antenna Measurement Setups: Introduction

Xianming Qing[a]* and Zhi Ning Chen[a,b]
[a]Institute for Infocomm Research (I^2R), Singapore, Singapore
[b]Department of Electrical and Computer Engineering, National University of Singapore, Singapore, Singapore

Abstract

Measurement is an important and necessary way to characterize the performance of an antenna and verify the design of an antenna. This chapter aims to give a brief overview on the basic setups of antenna measurement instead of detailed discussion of measurement techniques for each antenna parameter. The typical setups of far-field and near-field antenna measurements as well as the different antenna test ranges are reviewed. In particular, the setups for the impedance measurement of differential antennas, the on-wafer measurement of antennas, and the measurement of special absorption rate (SAR) are introduced.

Keywords

Antenna measurement; Antenna test range; Near-field; Far-field; Impedance; Anechoic chamber; Free space; On-wafer antenna measurement; Special absorption rate (SAR)

Introduction

In antenna engineering, measurement is basic and crucial for characterizing the performance of an antenna to ensure the antenna to meet specifications, to validate theoretical simulation or design, and to verify the construction and fabrication process of the designed antenna. Many of the basic measurement methods for characterizing antennas had been developed even before World War II. However, new approaches and technologies of measurement have continuously emerged with the rapid growth of modern mobile communications in the last decades (Hollis and Lyon 1972; Kummer and Gillespie 1978; IEEE 1979; Balanis and Birtcher (2008)). Antenna metrology requires not only sound background knowledge in both antenna theory and ample experience in antenna engineering but also sophisticated equipment capable of providing the acceptable accuracy and purity of measured data. Commercial systems specifically designed for a variety of antenna measurements have become available since the 1960s, in particular, due to the strong demand of aerospace, space, and defense industries.

In antenna engineering, impedance, gain, radiation pattern, efficiency, and polarization are certainly the most commonly used parameters to be measured for evaluating the performance of an antenna. For specific applications, some other parameters may be of more importance and interest, for example, antenna temperature is crucial for the antennas in telescope systems, and special absorption rate (SAR) is critical for the antennas in consumer electronics.

*Email: qingxm@i2r.a-star.edu.sg

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_68-1

Impedance Measurement

To attain the maximum power transfer between a source or a source-transmission line and an antenna (or between an antenna and a receiver or receiver-transmission line), a conjugate impedance matching is essentially desired. In most of wireless systems, the feeding transmission line features a characteristic impedance of 50 ohms so that most of the practical antennas are designed with an input impedance of 50 ohms as well for a good impedance matching.

In a mismatched system, the degree of the mismatch determines the ratio of reflected power to incident or available power at the input antenna terminals. The degree of the mismatch is a function of the input impedance of the antenna under test (AUT) and the characteristic impedance of the feed line. They are directly related to the input reflection coefficient and the input voltage standing wave ratio (VSWR) at the input terminal of the antenna by the standard transmission-line relationships of

$$\frac{P_{\text{refl}}}{P_{\text{in}}} = |\Gamma|^2 = \frac{|Z_{\text{ant}} - Z_{\text{c}}|^2}{|Z_{\text{ant}} + Z_{\text{c}}|^2} = \left|\frac{\text{VSWR} - 1}{\text{VSWR} + 1}\right|^2 \tag{1}$$

where P_{refl} is the reflected power from the antenna, P_{in} is the input power to the antenna, Γ is the voltage reflection coefficient at the antenna input terminal, Z_{ant} is the input impedance of the antenna, and Z_{c} is the characteristic impedance of the feed line.

Rewrite (Eq. 1); the input impedance of the antenna can be expressed as

$$Z_{\text{ant}} = Z_{\text{c}} \frac{1 + \Gamma}{1 - \Gamma} \tag{2}$$

For a single-ended antenna or unbalanced antenna, namely, the antenna is with a ground plane, such as microstrip antenna and monopole antenna, the antenna can be presented using a single port network with a loading impedance of Z_{ant} as shown in Fig. 1. The reflection coefficient Γ is equal to the S_{11} of the equivalent single-port network, which can be easily measured using a vector network analyzer (VNA) as shown in Fig. 2, where the test ports are normally set as the interfaces of coaxial lines.

The impedance of a balanced or differential antenna cannot be measured directly using a network analyzer with unbalanced ports. As shown in Fig. 3, when a balanced antenna is connected to an unbalanced test port, the currents flowing into the two radiators of the antenna are unequal because of

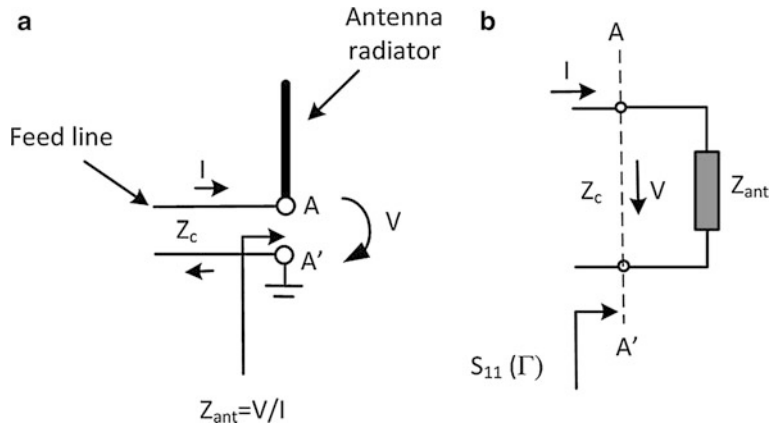


Fig. 1 Single-ended antenna and the equivalent single-port network; (**a**) antenna schematic view and (**b**) single-port equivalent network

DOI 10.1007/978-981-4560-75-7_68-1

Fig. 2 Vector network analyzer (VNA) with single-ended coaxial test ports (Courtesy of Keysight/Agilent Technologies)

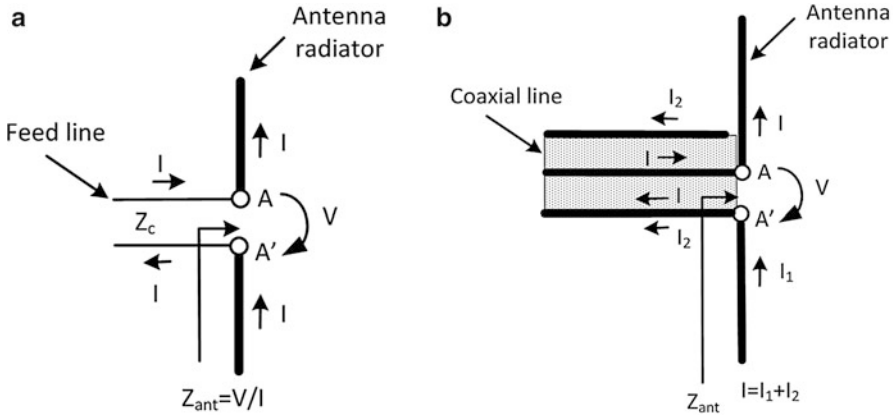


Fig. 3 Current distribution on balanced antenna; (**a**) ideal case and (**b**) fed by single-ended coaxial line

part of the current flowing on the outer conductor of the feed cable so that the impedance of the balanced antenna cannot be characterized correctly. Conventionally, the impedance of a balanced antenna can be measured by adding a balun which forces the current equally to flow in each part of the antenna radiator. However, the accuracy of such measurement results relies on the performance of the balun, especially over a wide testing frequency band. The impedance of a symmetrical balanced antenna can also be measured using imaging method, i.e., mounting half part of the radiators of the antenna on a ground plane which provides a mirror image to configure the other parts of radiators. The impedance of the antenna can be obtained by multiplying a factor of 2 of the measured impedance of the equivalent monopole. However, the accuracy of the measurement results is much dependent on the size of the ground plane used, and this method is valid just for symmetrical balanced antennas (Meier and Summers 1949).

DOI 10.1007/978-981-4560-75-7_68-1

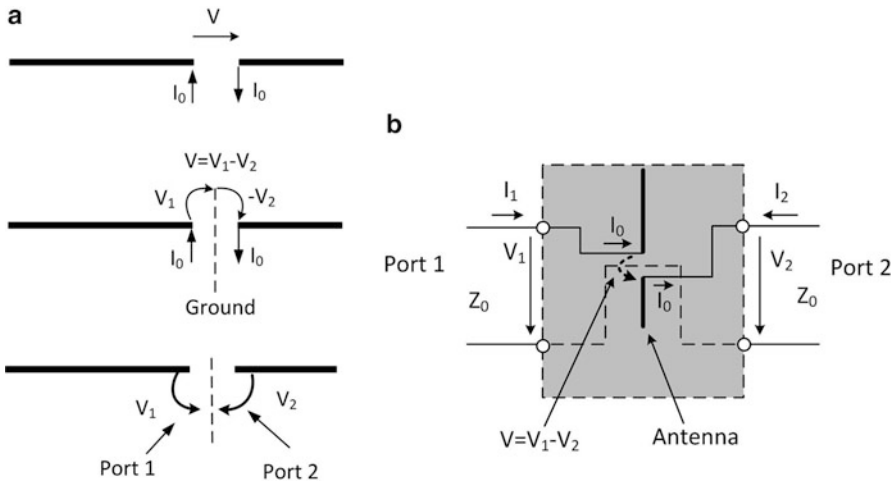


Fig. 4 Balanced antenna and equivalent two-port network; (**a**) excitation and port definition and (**b**) network representation of the balanced antenna

The impedance of a balanced antenna can be measured using a single-ended two-port network analyzer, wherein the full set of the S-parameters of a two-port network instead of S_{11} is utilized (Meys and Janssens 1998; Qing et al. 2009). As shown in Fig. 4a, the positive and negative ports of the source with a voltage of V are connected to the input terminals of the radiators of the balanced antenna, respectively; the driven voltage can be split as V_1 and V_2 with a virtual ground plane without any disturbance of the current distribution on the antenna. Therefore, each terminal of the antenna radiators and the ground plane can be considered as a "port." The antenna can be considered to be equivalent to a "two-port" network as shown in Fig. 4b. The impedance of the antenna, Z_d, is thus related to the network parameters of the equivalent two-port network and can be characterized by the S-parameters of the two-port network.

$$Z_d = \frac{2Z_0(1 - S_{11}S_{22} + S_{12}S_{21} - S_{12} - S_{21})}{(1 - S_{11})(1 - S_{22}) - S_{21}S_{12}} \tag{3}$$

where Z_0 is the characteristic impedance of the feed transmission line, which is 50 Ω for most of the measurement systems.

For a symmetrical balanced antenna, $S_{11} = S_{22}$, and $S_{12} = S_{21}$, and Eq. 3 can be simplified to

$$Z_d = \frac{2Z_0\left(1 - S_{11}^2 + S_{21}^2 - 2S_{12}\right)}{(1 - S_{11})^2 - S_{21}^2} \tag{4}$$

A configuration of the measurement setup is illustrated in Fig. 5. The measurement is conducted using a two-port VNA (Keysight/Agilent N5230A) and a test fixture. Any other VNA which features the port-extension function is suitable for the measurement as well. The test fixture is constructed using two semi-rigid coaxial cables which are soldered together on their outer conductors. One end of the fixture is with two subminiature version A (SMA) connectors and connected to the VNA through the test cables. The other end of the fixture is kept open with the small extensions of inner conductors to form the two tips to connect the AUT.

The measurement is recommended to conduct in an anechoic chamber or a place with controllable environment reflection. The main measurement procedures are as follows:

DOI 10.1007/978-981-4560-75-7_68-1

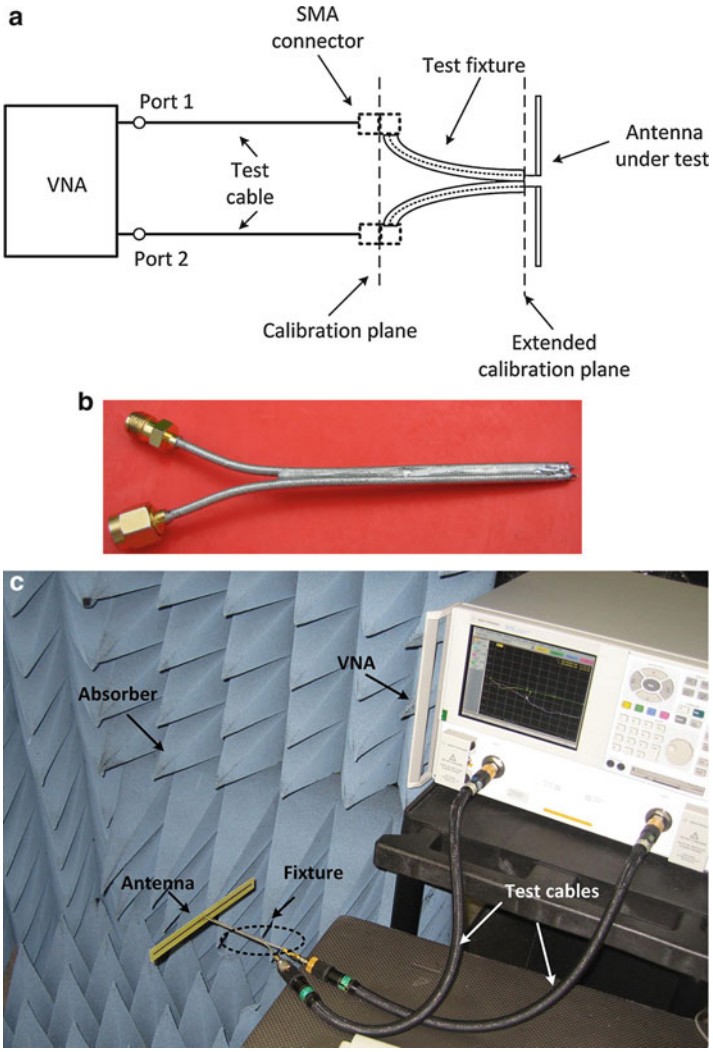


Fig. 5 Measurement setup for input impedance of a balanced antenna; (**a**) schematic configuration, (**b**) fixture prototype, and (**c**) measurement setup using Keysight/Agilent N5230A network analyzer

- Carry out standard VNA parameter setting and calibration.
- Shift the calibration plane through the port-extension technique to de-embed the influence of the text fixture by connecting the test fixture to the test cables and then carrying out the port extension to shift the calibration plane to the tips of the fixture. The fixture is required to be open or short circuited when performing the port extension. Short-circuited configuration (soldering the tips of the fixture and outer conductors of the coaxial cables together) is recommended since perfect "open" condition is very difficult to achieve, especially at higher frequencies.
- Connect the AUT to the fixture and record the S-parameters.
- Calculate the impedance of the AUT using Eqs. 3 or 4.

To validate the measurement methodology, two types of balanced antennas, an asymmetrical dipole antenna and a symmetrical meander line dipole antenna, are prototyped and measured.

Figure 6 shows the antenna prototypes and the detailed dimensions. All the antenna prototypes are printed onto an FR4 print circuit board (PCB) with a dielectric constant of 4.4, a loss tangent of 0.02, and a thickness of 0.504 mm.

DOI 10.1007/978-981-4560-75-7_68-1

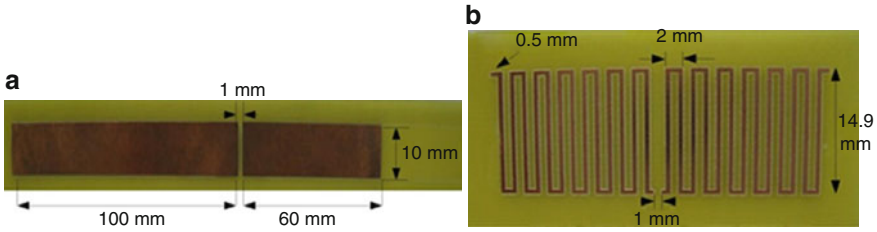


Fig. 6 Antennas under test; (**a**) asymmetrical dipole antenna and (**b**) symmetrical meander line dipole antenna

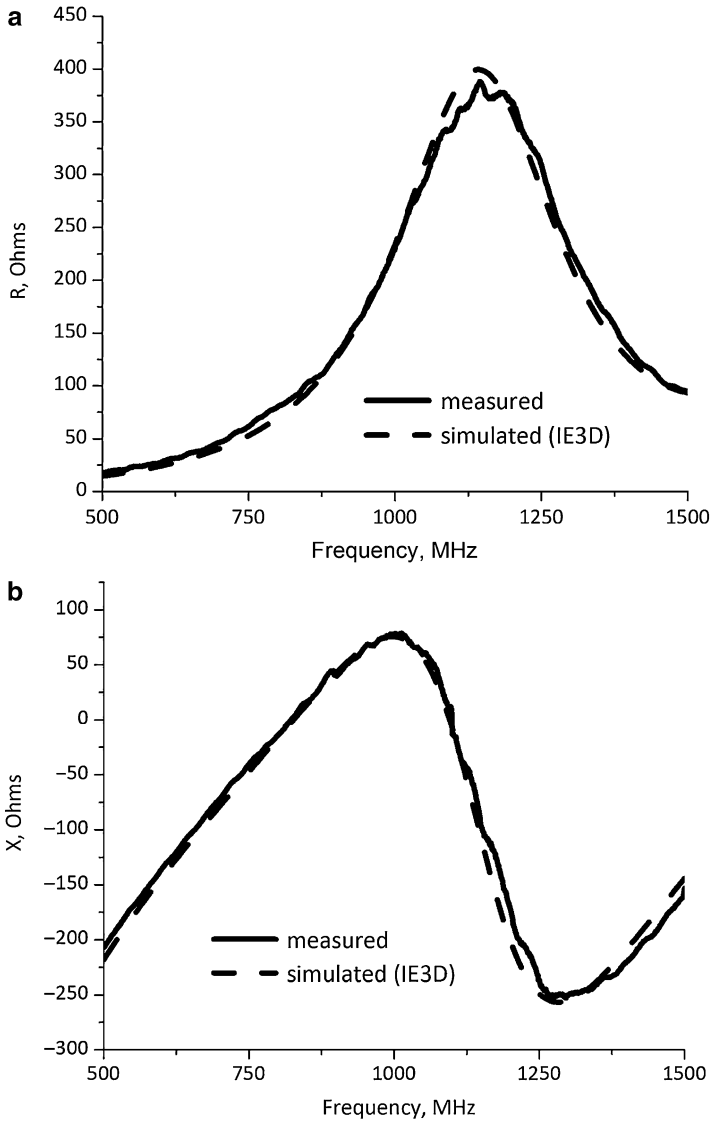


Fig. 7 Measured and simulated impedance of the asymmetrical dipole antenna; (**a**) real part and (**b**) imaginary part

Figure 7 illustrates the measured and simulated impedance of the asymmetrical dipole antenna over a frequency range of 500−1500 MHz. One of the dipole antenna arms is 100 mm in length, and the other 60 mm. The width of the arms is 10 mm. The simulation was conducted with IE3D software, which is based on the method of moments. Very good agreement has been observed for both real and imaginary

DOI 10.1007/978-981-4560-75-7_68-1

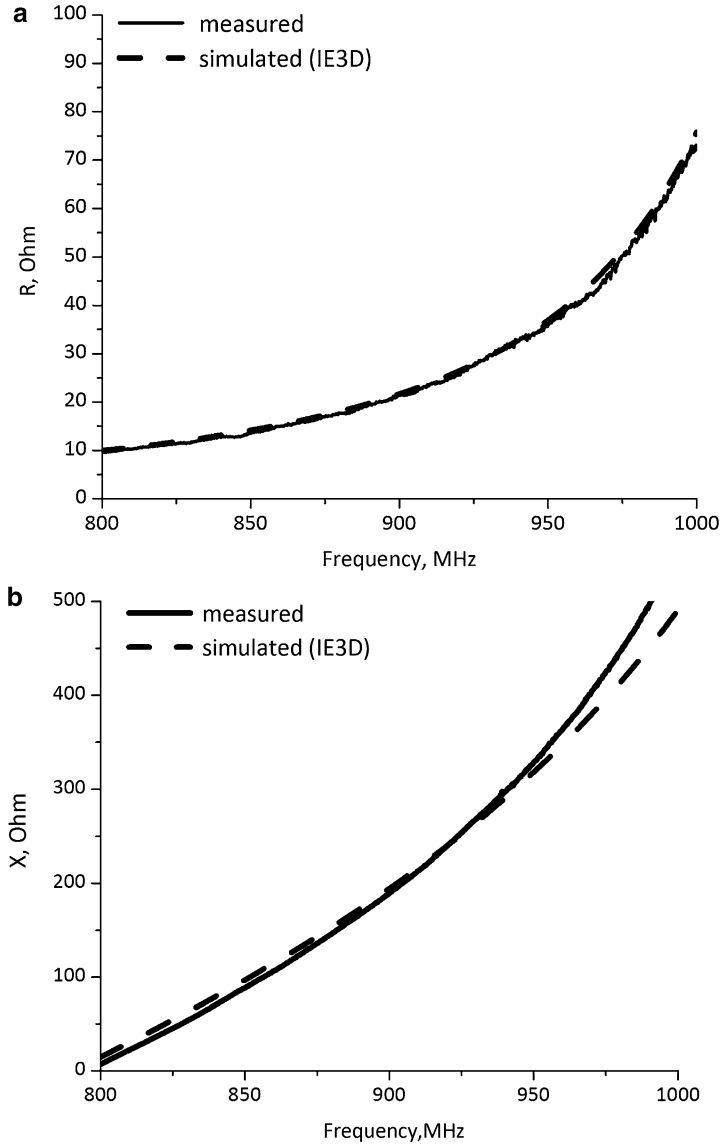


Fig. 8 Measured and simulated impedance of the symmetrical meander line dipole antenna; (**a**) real part and (**b**) imaginary part

parts of the impedance. Figure 8 compares the measured and simulated impedance of the symmetrical meander line dipole antenna as shown in Fig. 6b. Both measured real and imaginary parts of the impedance show good agreement with the simulation across the frequency range of 800−1000 MHz.

Radiation Measurement

The distribution of the fields radiated by an antenna does not assume the same amplitude (or phase) pattern at all radial distances; it evolves with the increasing distance from the antenna. Three regions surrounding an antenna are recognized and classified according to the development of the field distribution (Balanis 2005) as shown in Fig. 9.

Reactive near-field region: In the space immediately surrounding the antenna closer than approximately 0.62 $\sqrt{D^3/\lambda}$ (or $\lambda/2\pi$ for electrically small antennas), where D is the largest dimension of the

DOI 10.1007/978-981-4560-75-7_68-1

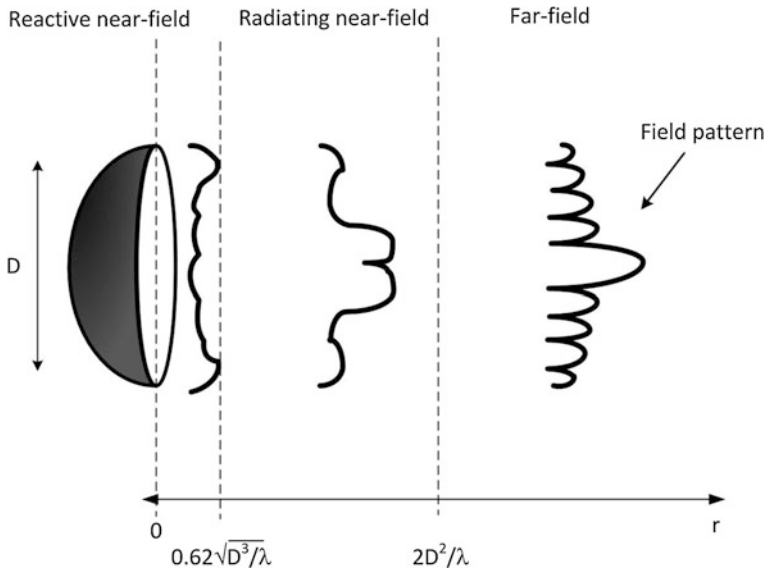


Fig. 9 The field patterns of an antenna as a function of a radial distance

antenna and λ is the wavelength. In this region, the electric or magnetic field does not propagate radially, and the energy is predominantly reactive and decays more rapidly than $1/r$.

Radiating near-field region: Beyond the immediate neighborhood of the reactive field, the radiating field begins to dominate. The extent of this region is $0.62\sqrt{D^3/\lambda} < r < 2D^2/\lambda$. This region can be divided into two subregions. For $0.62\sqrt{D^3/\lambda} < r < D^2/4\lambda$, the fields decay more rapidly than $1/r$ and the radiation pattern is dependent on r. For $D^2/4\lambda < r < 2D^2/\lambda$, the fields decay as $1/r$, but the radiation pattern is dependent on r. The main lobe of the radiation pattern in this radiating near-field region is frequently broadened and bifurcated.

Radiating far-field region: Beyond the radiating near-field region, $r > 2D^2/\lambda$ (or $> 10\lambda$ for electrically small antennas). The fields decay as $1/r$ and the radiation pattern is independent of r. The radiation pattern approaches the same shape that would basically be observed at an infinite distance.

Generally, the antenna measurement techniques can be categorized as far-field antenna measurement and near-field antenna measurement. In the far-field antenna measurement system, the antenna is illuminated by a uniform plane wave and the antenna parameters are measured directly. In a near-field antenna measurement system, the near-field distribution of the antenna is measured using a probe and the far-field characteristics are obtained through near-field to far-field transforms.

Far-Field Antenna Measurement Setup

For far-field antenna measurement, the AUT is required to be illuminated by a uniform plane wave, so that a signal source and a transmit antenna with desired radiation characteristics are needed. The block diagrams of the far-field antenna measurement systems are shown in Fig. 10; the main parts of the systems include:

Signal source: The signal source must be selected so that it has frequency control, frequency stability, spectral purity, controllable power level, and possible modulation.

Transmitting antenna: The antenna is used to generate the plane wave and illuminate the AUT; it should be with a known pattern and controlled polarization. Some frequently used transmitting antennas are shown in Fig. 11.

Receiver: It determines how much power is received by the AUT, where a receiver with great sensitivity and large dynamic range is preferable.

DOI 10.1007/978-981-4560-75-7_68-1

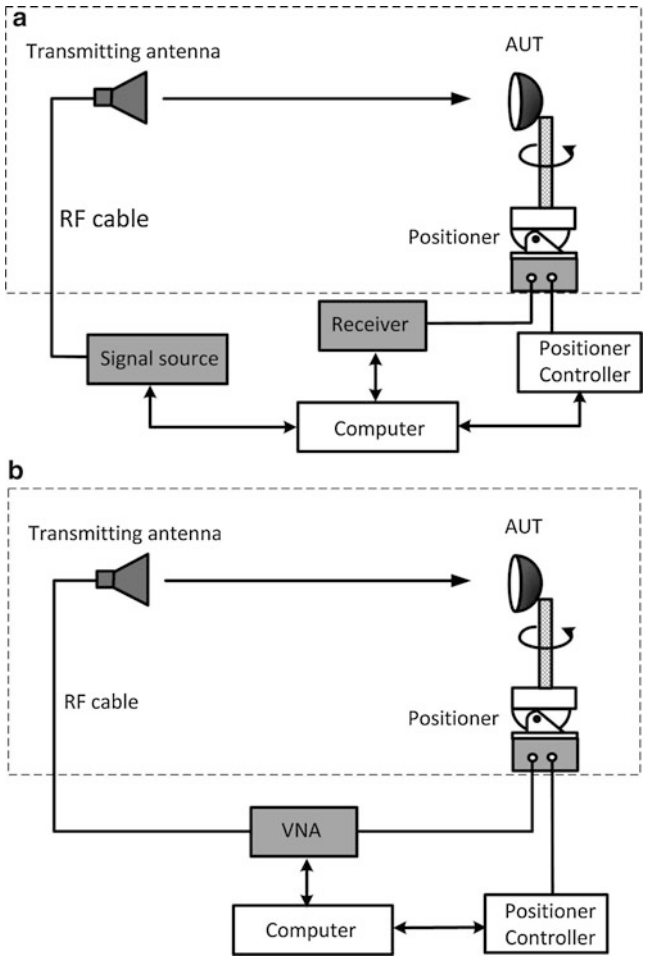


Fig. 10 Block diagrams of far-field antenna measurement setup; (**a**) measurement setup with separate signal source and receiver and (**b**) measurement setup with a VNA

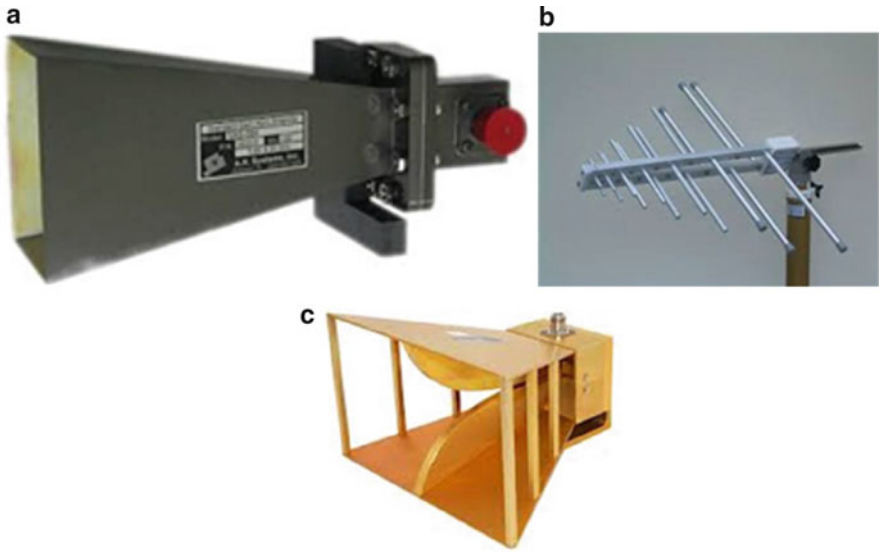


Fig. 11 Commonly used transmitting antennas; (**a**) horn antenna, (**b**) log-periodic antenna, and (**c**) double-ridge horn antenna

DOI 10.1007/978-981-4560-75-7_68-1

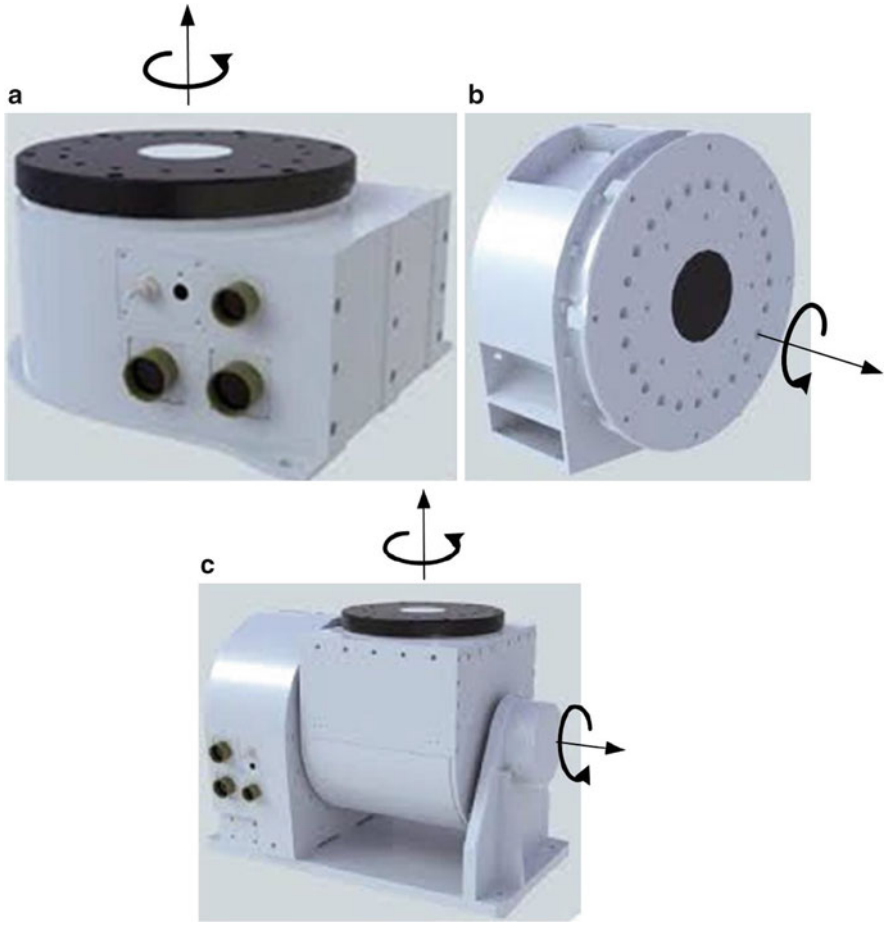


Fig. 12 Positioners (turntable) for antenna measurement; (**a**) one-dimensional positioner, (**b**) polarization positioner, and (**c**) two-dimensional positioner

Positioning system (turntable): This system is used to rotate the AUT relative to the transmit antenna, to measure the radiation pattern as a function of angle. Some positioning systems are exhibited in Fig. 12.

Recording system/data processing system: Modern antenna measurement facilities utilizing a computer and software to coordinate the actions of the positioning and measuring systems, to store and process the acquired data, and to display or plot the resulting measurements. A communication bus connects the computer to the positioner controller and instruments. The IEEE-488 GPIB (general purpose instrument bus) is frequently used.

Apart from the separate signal source and receiver configuration, the VNA has been applied in antenna measurement system for many years since an antenna pattern measurement can be conceptually viewed as the S_{21} between the input connectors of the transmit antenna and the AUT. VNAs exhibit excellent phase stability, spectral purity, sensitivity, and dynamic range (well in excess of 100 dB), which makes it well suitable for antenna measurement.

Far-Field Antenna Test Ranges

The far-field measurement can be performed in outdoor or indoor ranges. In general, there are two basic types of far-field antenna ranges: reflection range and free-space range. In the ground reflection range, the specular reflection from the ground is used to obtain a uniform phase and amplitude distribution over

DOI 10.1007/978-981-4560-75-7_68-1

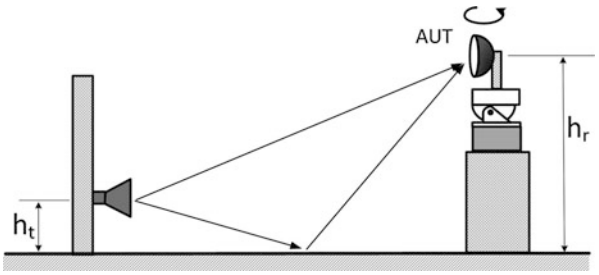


Fig. 13 Configuration of the ground reflection range

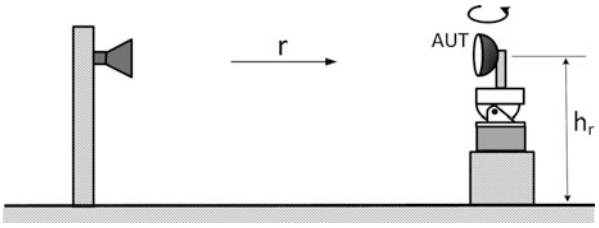


Fig. 14 Illustration of the outdoor elevated range

the AUT. In the free-space ranges, the reflections from the environment are minimized. The typical free-space ranges are elevated range, slant range, anechoic chamber, and compact range.

Ground Reflection Range

In the ground reflection range as shown in Fig. 13, the specular reflection from the ground is used to obtain a uniform phase and amplitude distribution in the test zone, or quiet zone, of the antenna (Hemming and Heaton 1973). This is accomplished by designing the range so that the specular reflections from the ground combine constructively with the direct illumination. The length of the range is designed to meet the far-field criterion of $r > 2D^2/\lambda$; the heights of the transmitting antenna and the AUT are chosen so that the AUT is centered on the first interference lobe of the direct and specular reflected rays. The outdoor type ground reflection ranges are usually employed in the ultrahigh-frequency (UHF) region for measuring antennas with moderately broad beam; they can also be used for the AUTs operating at higher frequencies.

Elevated Range

Elevated range is an outdoor range. In this setup, the transmitting antenna and the AUT are mounted above the ground. These antennas can be on mountains, towers, buildings, or wherever one finds that is suitable. This is often done for very large antennas or at low frequencies (VHF and below, <100 MHz) where indoor measurements would be intractable. The basic diagram of an elevated range is shown in Fig. 14. The length of the range, r, is designed to meet the far-field criterion $r > 2D^2/\lambda$. The height of the AUT, h_r, is determined by two criteria related to the transmitting antenna. The transmitting antenna should be chosen so that the amplitude taper over the AUT is typically no greater than 0.25 dB, to ensure the uniform illumination. In addition, to minimize the range reflections, its first null should point toward the base of the test tower, ensuring that the range surface intercepts only sidelobe energy. These two criteria determine the height of the AUT to be $h_r > 4D$. In the elevated range, diffraction fences are somehow used to further minimize the reflections from the ground. There are two cautions that should be emphasized. First, the diffraction fence should not intercept the main beam of the source antenna. Second, the top edge of the fence should not be straight knife edge but rather serrated to reduce the edge diffraction.

DOI 10.1007/978-981-4560-75-7_68-1

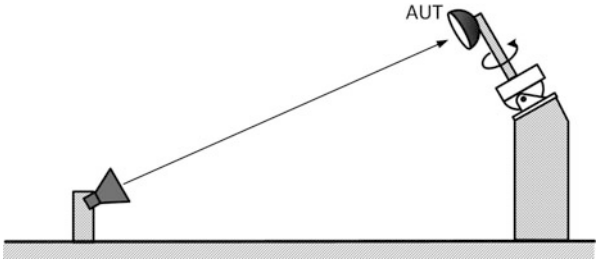


Fig. 15 Geometry arrangement of the slant range

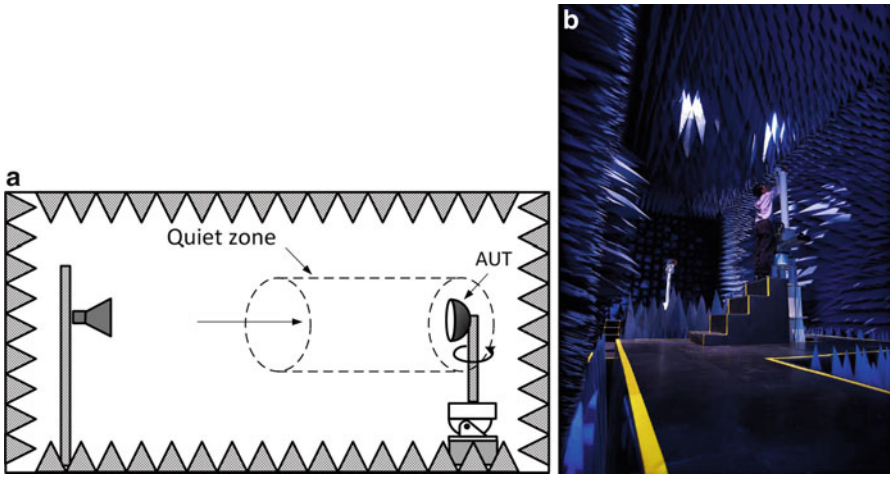


Fig. 16 Anechoic chamber; (**a**) geometry of the rectangular anechoic chamber and (**b**) photo of an anechoic chamber (Courtesy of Institute for Infocomm Research, Singapore)

Slant Range

A slant range (Arnold 1966) is one in which the transmit antenna is located close to the ground and the AUT is mounted on a tower as shown in Fig. 15. The transmit antenna is positioned so that the pattern maximum is oriented toward the center of the test antenna. The first null is usually directed toward the ground specular reflection point to suppress reflected signals. It is desirable that the tower of the AUT be constructed of nonconducting materials to reduce reflections. Slant ranges, in general, require less real estate than the elevated ranges.

Anechoic Chamber

The anechoic chambers are the most popular antenna measurement sites. Compared to the outdoor ranges, the anechoic chamber is able to provide a controlled environment, all-weather capability, and security and to minimize possible electromagnetic interference for antenna measurement. As shown in Fig. 16, an anechoic chamber is a large room, where the walls, floor, and ceiling of which are lined with metallic sheets (normally steel sheet) and all the inner surfaces are lined with RF/microwave absorbers (Hemming 2002). The test zone (quite zone) is able to be generated by arranging the absorbers properly. In the test zone, the amplitude and phase of the incident plane wave from the transmitting antenna are with very little variation such as 0.2 dB and 0.5°, respectively. Anechoic chambers are mostly utilized in microwave frequency regions, but materials have been developed which provide a normal-incidence reflection

DOI 10.1007/978-981-4560-75-7_68-1

coefficient of −40 dB at frequencies as low as 100 MHz. In general, as the operating frequency is lowered, the thickness of RF absorbing material must be increased to maintain a given level of reflectivity performance (Emerson 1973). An RF absorber that meets the minimum electrical requirements at the lower frequencies usually possesses improved performance at higher frequencies.

There are two types of anechoic chambers: the rectangular and tapered chambers. The design of both chamber types is based on the principles of geometrical optics. The goal is to minimize the specular reflections in the quiet zone. The rectangular chamber is usually designed to simulate free-space conditions and maximize the volume of the quiet zone. The tapered chamber has the advantage of tuning by moving the transmit antenna closer to (at higher frequencies) or farther from (at lower frequencies) the apex of the taper. Thus, the reflected rays are adjusted to produce nearly constructive interference with the direct rays at the test location.

The anechoic chamber techniques are discussed in much greater detail in Chapter Anechoic Chamber Design of this Handbook.

Compact Range

The main drawback of the anechoic chamber is the size limitation; the chamber is required to be large enough to meet the far-field conditions so that it is desired to have an anechoic chamber as large as possible. However, the cost and practical constraints often limit their size, especially for electrically large antennas. One solution to this problem is via a compact range as shown in Fig. 14. A compact antenna test range (Johnson, et al. 1969; Johnson 1986) is a collimating setup that generates nearly planar wave fronts in a very short distance (typically 10–20 m) compared to the $2D^2/\lambda$ distance required to produce the same size test zone using the standard system configuration of testing shown in Fig. 16.

As exhibited in Fig. 17, in the compact test range, a transmitting antenna is used as an offset primary feed to illuminate a paraboloidal reflector that converts the impinging spherical waves into plane waves. This is very similar to the principle upon which a dish antenna operates. Compact range configurations are often designated according to their analogous reflector antenna configurations: parabolic, Cassegrain, Gregorian, and so on (Burnside et al. 1987; Pistorius and Burnside 1987; Sanad and Shafai 1990; Gupta et al. 1991). A perfect plane wave is produced by the compact test range if the paraboloidal reflector has a perfect surface, infinite size, and if the feed is a point source with a pattern which compensates for the space attenuation. Of course, such ideal conditions cannot be achieved, and the field distribution in a real compact test range deviates from the uniform plane wave. In general, the deviation is within acceptable parameters in the quite zone which is typically 50 to 60 % the aperture of the reflector. The imperfections of the field in the quiet zone are measured in terms of phase errors, ripple-amplitude deviations, and taper-amplitude deviations. Acceptable deviations for most compact test ranges are less than 10 % phase error and less than 1 dB ripple- and taper-amplitude deviations.

In short, the ground reflection range uses the surface reflections to obtain uniform field distribution along the AUT aperture. However, this imposes very strict requirements on the range smoothness (related to the highest usable frequency), which has cost and maintenance impact. The measurement errors due to the ground reflections in the slant range are lower than that of a comparable elevated range; however, its receiving tower is usually higher than that in the elevated range. This fact puts a limit on the maximum antenna size to be installed at the top of the tower. The elevated range is commonly used in microwave measurements for medium-size antennas. The ground reflection errors encountered in this type of range are minimized using diffraction fences or measurement techniques to determine precisely the reflection level and subtract it from the measurement. Among the far-field antenna ranges, the most accurate one is the indoor compact range because all the reflections can be controlled and minimized. Its size is also minimal compared to the other ranges.

DOI 10.1007/978-981-4560-75-7_68-1

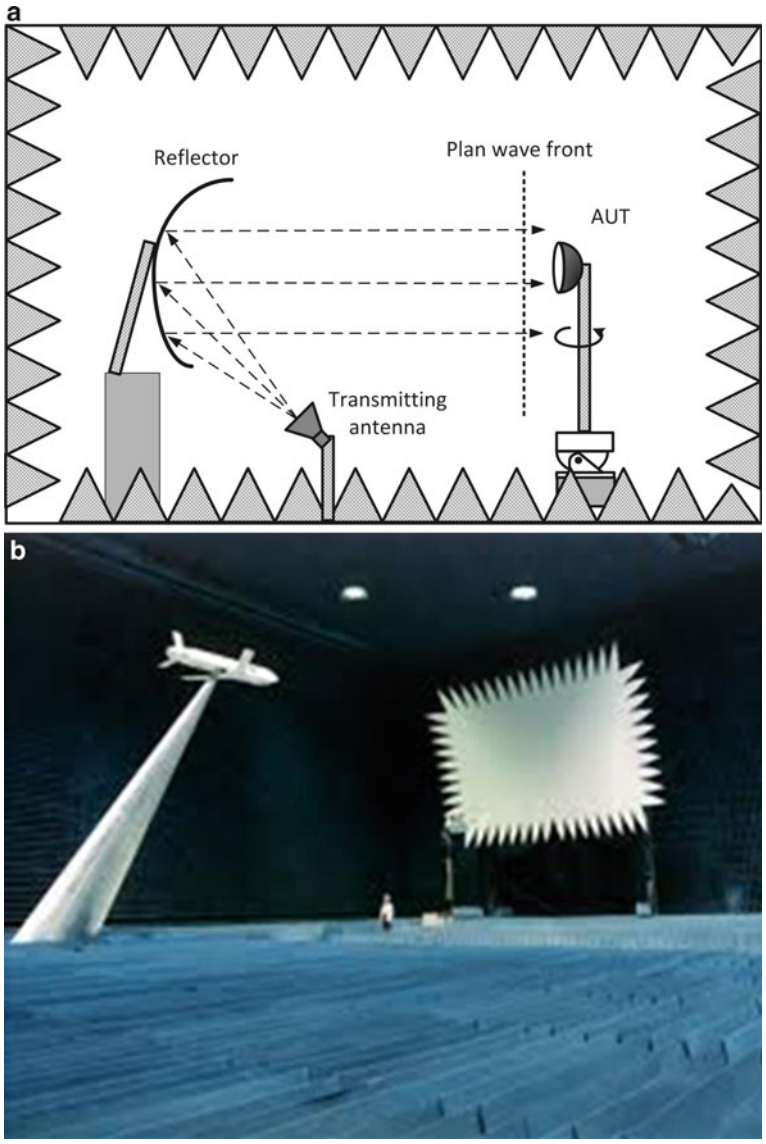


Fig. 17 Compact antenna test range; (**a**) geometry of the configuration and (**b**) photo of a compact test range

Near-Field Antenna Measurement Setup

The dimensions of a conventional antenna test range can be reduced by making measurements in the near-field of the AUT and then using analytical methods to transform the measured near-field data to far-field radiation characteristics (Johnson et al. 1973; Paris et al. 1978; Joy et al. 1978; Yaghjian 1986). This type of range uses a small RF probe antenna that is scanned over a surface surrounding the AUT. Typically, the separation between the probe and the antenna structure is in the order of 4 to 10 wavelengths. During the measurement, the information about near-field phase and amplitude is collected over a discrete matrix of points. The data are then transformed to the far-field using the Fourier transform. In near-field testing, the AUT is usually aligned to the scanner's coordinate system and then either the probe or the test antenna is moved. In practice, it is easier and more cost effective to scan the RF probe over linear axes or the AUT over angular axes. In principle, the measurement can be done over a surface, which can be defined in any of the six orthogonal coordinate systems: rectangular, cylindrical (circular cylindrical), spherical, elliptic

DOI 10.1007/978-981-4560-75-7_68-1

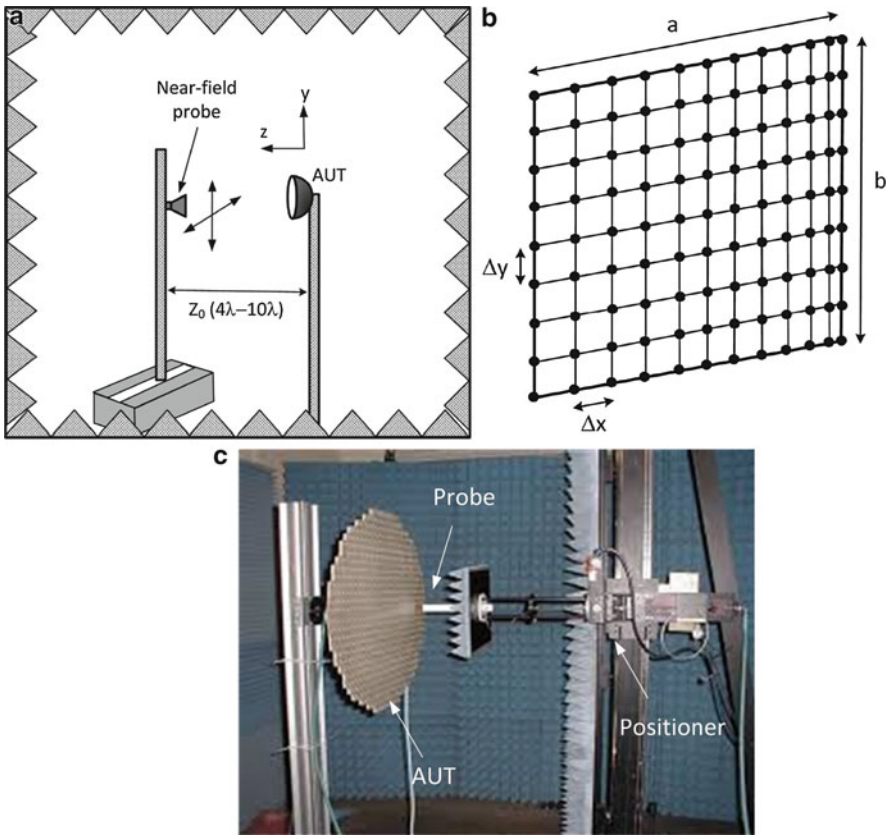


Fig. 18 Planar near-field antenna measurement setup; (**a**) block diagram, (**b**) rectangular scanning surface, and (**c**) photo of a planar scanning near-field antenna measurement system

cylindrical, parabolic cylindrical, and conical. However, only the first three are deemed convenient for data acquisition, and of them, the complexity of the analytical transformations increases from the planar to the cylindrical and then to the spherical surfaces.

Planar Near-Field Antenna Measurement Setup

With the planar near-field scanning setup as shown in Fig. 18, a probe antenna is moved in a plane situated in front of the AUT and the received signal (amplitude and phase) is recorded. The position of the probe is characterized by the coordinates (x, y, z_0) in the xyz coordinate system of the AUT. During the scanning, z_0 is kept constant, while both x and y are varied. The distance z_0 is approximately $4\lambda - 10\lambda$ to avoid sampling of the reactive energy of the AUT. The dimensions of the near-field scanning aperture ($a \times b$) must be large enough to accept all significant energy from the AUT.

Cylindrical Near-Field Antenna Measurement Setup

In the cylindrical scanning technique, the AUT is rotated around the z-axis in $\Delta\varphi$ step, while the probe is moved on the cylindrical surface at various heights in Δz step as shown in Fig. 19. The probe is located at a distance, a, which is the available smallest cylinder radius enclosing the AUT. The cylindrical scanning enables obtaining the exact azimuth pattern but limited elevation pattern due to the truncation of the scanning aperture in z-direction. In accordance with the sampling theory, the sampling spacing is determined from $\Delta\varphi = \lambda/2a$ and $\Delta z = \lambda/2$. A two-dimensional Fourier transform of the near-field data gives the cylindrical vector wave functions that determine the test radiation pattern.

DOI 10.1007/978-981-4560-75-7_68-1

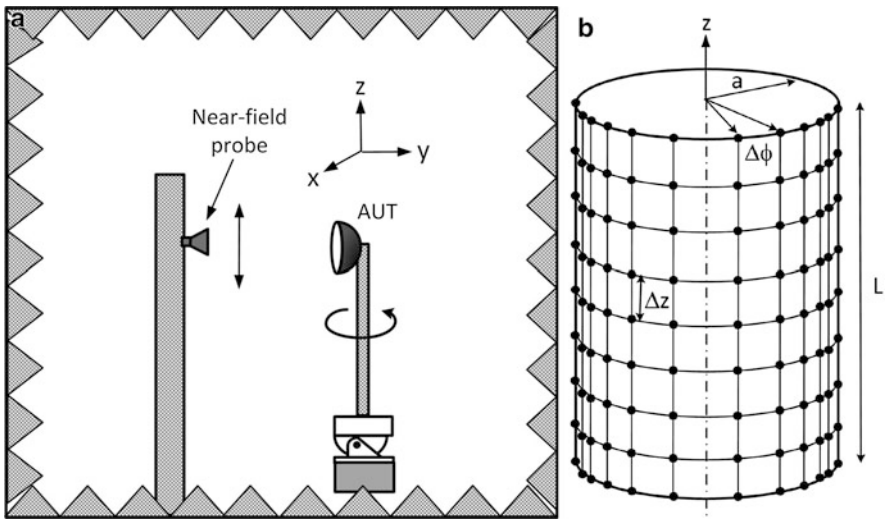


Fig. 19 Cylindrical near-field antenna measurement setup; (**a**) block diagram and (**b**) cylindrical scanning surface

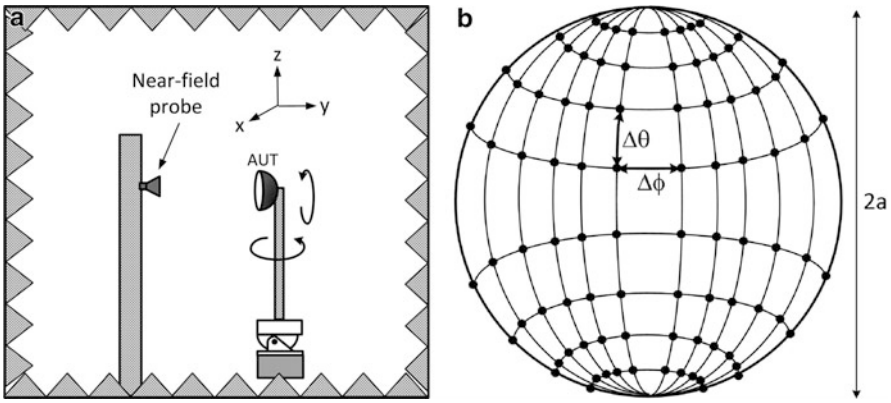


Fig. 20 Spherical near-field antenna measurement setup; (**a**) block diagram and (**b**) spherical scanning surface

Spherical Near-Field Antenna Measurement Setup

In the spherical near-field antenna measurement system as exhibited in Fig. 20, the probe is normally stationary and the AUT is moved in $\Delta\varphi$ and $\Delta\theta$ steps with a spherical scanning positioner. The resulting data are collected over a spherical envelope surrounding the test antenna. The radius of the rotation is "a" which is the available smallest radius enclosing the AUT. Full or nearly full coverage of the test antenna's radiating field can be evaluated with this type of near-field system, so that it is able to provide a three-dimensional pattern. The sampling spacing is determined by the sampling theory to be $\Delta\varphi = \Delta\theta = \lambda/2a$.

Multi-Probe Spherical Near-Field Antenna Measurement Setup

Different from conventional near-field antenna measurement systems using one probe to collect the amplitude and phase responses at the test point, a multi-probe spherical near-field system has been developed for "real-time" measurement (Iversen et al. 2001; Duchesne et al. 2004). A simplified block diagram of the multi-probe spherical near-field system is shown in Fig. 21. The measurement concept is based on a technique that identifies the signal from each probe in an array of probes by perturbing the

DOI 10.1007/978-981-4560-75-7_68-1

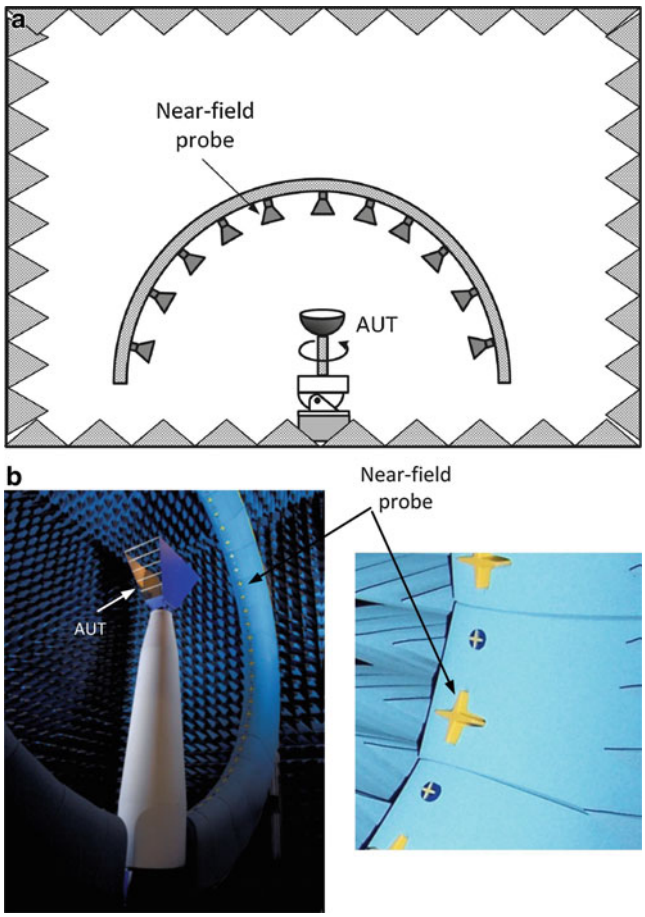


Fig. 21 Multi-probe spherical near-field antenna measurement setup; (**a**) block diagram and (**b**) photo of a multi-probe spherical near-field antenna measurement system (Foged et al. 2010)

electromagnetic properties of the probe (Lasserre et al. 1995; Bolomey and Gardiol 2001). The result of this perturbation is a modulation-frequency component in the output signal, which is directly related to the amplitude and phase of the incident field at the location of the probe. By sequentially modulating each probe, it is possible to measure in "real time" the amplitude and phase at each probe location, without the need for an expensive high-frequency multiplexing network. The near-field measurement data is transformed to a far-field radiation pattern using numerical techniques based on spherical wave expansion.

The near-field antenna measurement techniques are discussed in much greater detail in Chapter Near-Field Antenna Measurement Techniques of this Handbook.

Special Antenna Measurement Setups

On-Wafer Antenna Measurement

To achieve high-performance system with integration at a high level, miniaturize the overall device size, and low manufacturing cost, much effort has been devoted into the development of system-on-chip (SoC) or system-on-board (SoB) and system-in-package (SiP). The realization of antennas on chip offers tremendous advantages as it reduces the chip-to-board design complexity and the associated cost.

DOI 10.1007/978-981-4560-75-7_68-1

However, the adaptation of on-chip antennas in systems is hindered by the lack of accurate characterization methodologies. To design a robust wireless link using fully integrated transceivers with embedded antennas, a systematic method for determining the on-chip antenna gain and radiation pattern is critical.

A number of on-wafer antenna measurement setups have been reported (Zwick et al. 2004; Farinelli et al. 2006; Van Caekenberghe et al. 2008; Tsai et al. 2013; Qing and Chen 2014). Figure 22a shows a typical VNA based on-wafer antenna measurement setup. The probe station includes an anti-vibration table and accessories for on-wafer measurement, which enables a stable and reliable contact between the probe and the AUT. The probe positioner with three-dimensional movement structure offers great convenient for probe contacting with the AUT. The horn antenna on the rotating arm is used for signal transmitting/receiving and connected to the other port of the VNA. The AUT is positioned on a platform on the anti-vibration table. The distance between the horn antenna and the AUT is required to be greater than the far-field criteria; the half-space (180°) patterns in the main cuts are measured. The probe shown in Fig. 22c is another key component; one end of the probe with coaxial or waveguide structure is connected to one port of the VNA, and the other with coplanar waveguide (CPW) tips is connected to the AUT. The typical configurations of the CPW tips are signal-ground (SG), ground-signal-ground (GSG), and ground-signal-ground-signal-ground (GSGSG).

Specific Absorption Rate (SAR) Measurement

Specific Absorption Rate (SAR) is a measure of how transmitted RF energy is absorbed by human tissue. SAR is a function of the electrical conductivity, the induced E-field from the radiated source, and the mass density of the tissue. The SAR is calculated by averaging (or integrating) over a specific volume (typically a 1-g or 10-g tissue):

$$SAR = \iiint_V \frac{\sigma(v)|E(v)|^2}{\rho(v)}\,dv \tag{5}$$

where σ is in siemens/meter, E in the electric field voltage/meter, and ρ in kg/cubic meter.

The units of SAR are W/kg or equivalently mW/g. Various governments have defined maximum SAR levels for RF energy emitted by mobile devices. For example, the SAR limit in the USA for mobile phones is 1.6 W/kg, averaged over 1 g of tissue. In Europe, the SAR limit is 2.0 W/kg averaged over 10 g of tissue.

The value of SAR depends heavily on the geometry of the part of the body that is exposed to the antenna and on the exact location and geometry of the antenna. Thus, SAR measurement must be made with each specific source, such as a specific mobile phone model, and at the intended position of use. The DASYS system from SEGAG (Speag 2014) is the most often used measurement setup for SAR measurement. As exhibited in Fig. 23, the system consists of a human body-shaped hollow tub, a near-field probe, a controlling robot, a receiver, and a computer. The human body-shaped hollow tub is filled with a fluid that has similar properties to human tissue, and the AUT is positioned directly underneath the tub near the head. The robot-controlled near-field probe is immersed into the fluid in the tub to collect the magnitude and phase of the electric field in a specific region. The measured data are then sent to the computer for post processing and SAR calculation.

DOI 10.1007/978-981-4560-75-7_68-1

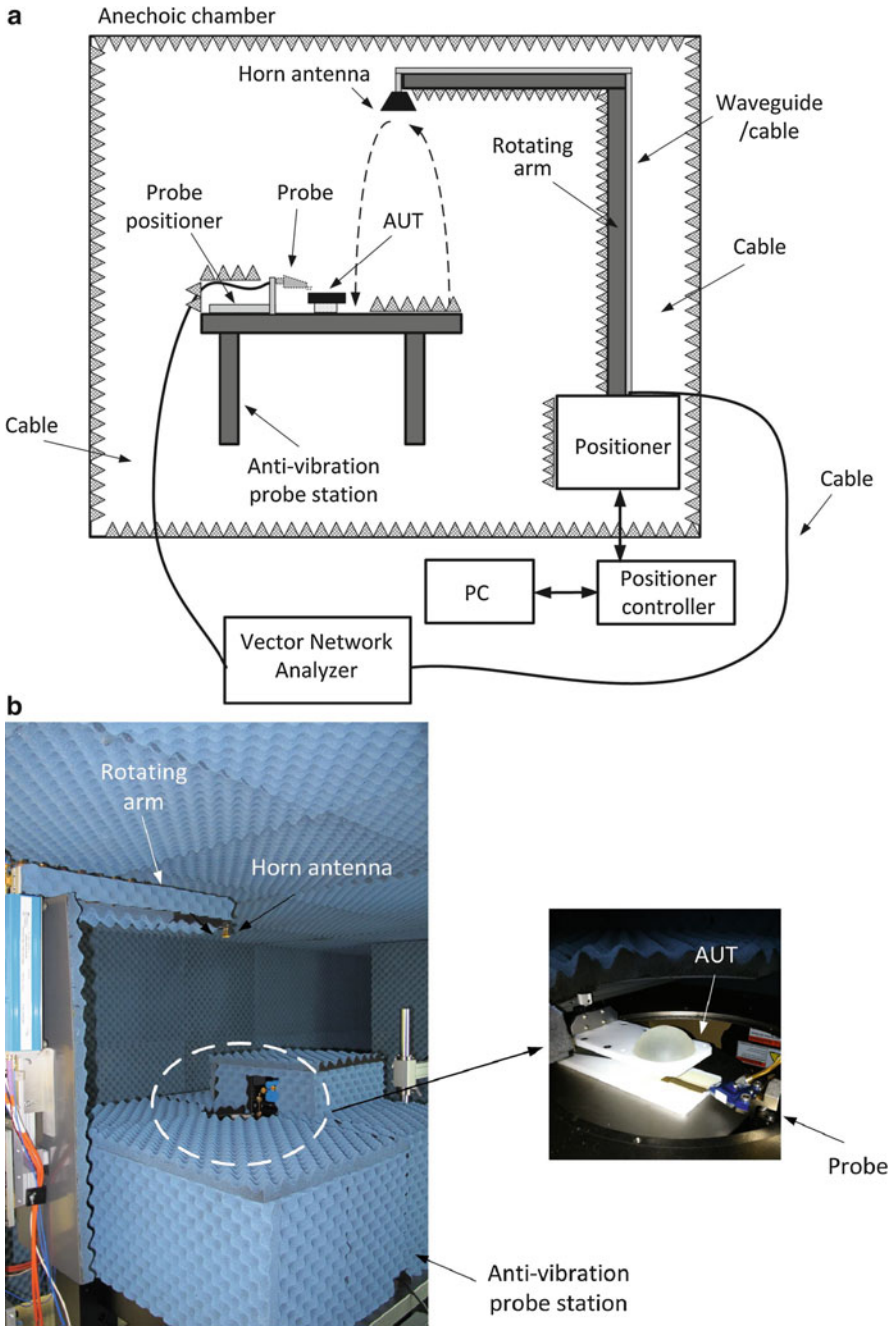


Fig. 22 (continued)

DOI 10.1007/978-981-4560-75-7_68-1

Fig. 22 On-wafer antenna measurement setup; (**a**) block diagram, (**b**) photo of an on-wafer measurement setup (Courtesy of Institute for Infocomm Research, Singapore), and (**c**) probes

Fig. 23 A SAR measurement system

Conclusion

Antenna measurement is essential to ensure the antenna to meet the desired specifications, to validate theoretical simulation, or to verify antenna construction and design. Among the different antenna measurement setups and test ranges, a proper selection is dependent on many factors, such as the operating frequency range, the directivity of the AUT, the mechanical features of the antenna (size, weight, and volume), and so forth. Furthermore, as the technology progresses, the requirements placed upon wireless communication systems and the antennas will continue to be more stringent, which will push the development of the antenna measurement technique and setup with higher and higher accuracy and more and more powerful measurement capability to meet the needs of the coming advanced antenna systems.

DOI 10.1007/978-981-4560-75-7_68-1

Cross-References

- ▶ Anechoic Chamber Design
- ▶ EMI/EMC Chamber Design, Measurement, and Instrument
- ▶ Mm-Wave Sub-mm-Wave Antenna Measurement
- ▶ Near-Field Antenna Measurement Techniques
- ▶ Radiation Efficiency Measurements of Small Antennas

References

Arnold PW (1966) The "slant" antenna range. IEEE Trans Antennas Propag 14(5):658–659

Balanis CA (2005) Antenna theory analysis and design. John Wiley & Sons, Hoboken

Balanis CA, Birtcher C (2008). Modern antenna handbook, Chapter 20. John Wiley & Sons, Hoboken

Bolomey JC, Gardiol FE (2001) Engineering applications of the modulated scatterer technique. Artech House, Boston

Burnside WD, Gilreath M, Kent BM, Clerici G (1987) Curved edge modification of compact range reflector. IEEE Trans Antennas Propag 35(2):176–182

Duchesne L, Garreau P, Robic N, Gandois A, Iversen PO, Barone G (2004) Compact multiprobe antenna test station for rapid testing of antennas and wireless terminals. The 5th international symposium on multi-dimensional mobile communications, pp 553–557

Emerson WH (1973) Electromagnetic wave absorbers and anechoic chambers through the years. IEEE Trans Antennas Propag 21(4):484–489

Farinelli P, Vaha-Heikkila T, Saily J, Sorrentino R, Tuovinen J (2006) A Novel Method for Millimeter-Wave On-Wafer Characterization of Reflect Patch Antennas. IEEE Antennas Wireless Propag Letters 5(1):556–558

Foged LJ, Duchesne L, Rosa A, Belot D, Lopez JM (2010) High gain P-band antenna measurements in multi probe spherical near field range. 14th international symposium on antenna technology and applied electromagnetics & the American electromagnetics conference, 1–4

Gupta IJ, Brown DG, Burnside WD, Lin W (1991) A serrated edge Gregorian subreflector for dual chamber compact range systems. IEEE Trans Antennas Propag 39(8):1258–1261

Hemming LH (2002) Electromagnetic anechoic chambers: a fundamental design and specifications guide. IEEE Press/Wiley, Piscataway/New York

Hemming LH, Heaton RA (1973) Antenna gain calibration on a ground reflection range. IEEE Trans Antennas Propag 21(4):532–537

Hollis JS, Lyon TJ (1972) Microwave antenna measurements. Scientific-Atlanta, Atlanta

IEEE (1979) IEEE standard test procedures for antennas. IEEE Std:149

Iversen PO, Garreau P, Burrell D (2001) Real-time spherical near-field handset antenna measurements. IEEE Antennas PropagMag 43(3):90–94

Johnson R (1986) Some design parameters for point-source compact ranges. IEEE Trans Antennas Propag 34(6):845–847

Johnson RC, Ecker HA, Moore RA (1969) Compact range techniques and measurements. IEEE Trans Antennas Propag 17(5):568–576

Johnson RC, Ecker HA, Hollis JS (1973) Determination of far-field antenna patterns from near-field measurements. Proc IEEE 61(12):1668–1694

Joy EB, Leach WM Jr, Rodrigue GP, Paris DT (1978) Applications of probe compensated near-field measurements. IEEE Trans Antennas Propag 26(3):379–389

Kummer WH, Gillespie ES (1978) Antenna measurements. Proc IEEE 66(4):483–507

Lasserre, J. L., Serafin, D., Bolomey, J. C., Lucas, F. & Therond, F. (1995). Characterization of portable communication systems by means of near-field techniques. *Proc. SPIE 2556, Wireless Communications*, 50 (September 18, 1995); doi:10.1117/12.220888.

Meier AS, Summers WP (1949) Measured impedance of vertical antennas over finite ground planes. Proc IRE 37(6):609–616

Meys R, Janssens F (1998) Measuring the impedance of balanced antennas by an S-parameter method. IEEE Antennas Propag Mag 40(6):62–65

Paris DT, Leach WM Jr, Joy EB (1978) Basic theory of probe compensated near-field measurements. IEEE Trans Antennas Propag 26(3):373–379

Pistorius CWI, Burnside WD (1987) An improved main reflector design for compact range applications. IEEE Trans Antennas Propag 35(3):342–347

Qing X, Chen ZN (2014) Measurement setups for millimeter-wave antennas at 60/140/270 GHz bands. IEEE Intl Work Ant Technol (iWAT):281–284

Qing X, Goh CK, Chen ZN (2009) Impedance characterization of asymmetrical balanced antennas and application in RFID tag design. IEEE Trans Microwave Theory Techniq 57(5):1268–1274

Sanad MSA, Shafai L (1990) Dual parabolic cylindrical reflectors employed as a compact range. IEEE Trans Antennas Propag 38(6):814–822

Speag (2014) http://www.speag.com/products/dasy/dasy-systems/. Accessed 5 July 2015

Tsai ZM, Wu YC, Chen SY, Lee T, Wang H (2013) A V-band on-wafer near-field antenna measurement system using an IC probe station. IEEE Trans Antennas Propag 61(4):2058–2067

Van Caekenberghe K, Brakora KM, Hong W, Jumani K, Liao D, Rangwala M, Wee YZ, Zhu X, Sarabandi K (2008) A 2–40 GHz probe station based setup for on-wafer antenna measurements. IEEE Trans Antennas Propag 56(10):3241–3247

Yaghjian AD (1986) An overview of near-field antenna measurements. IEEE Trans Antennas Propag 34(1):30–45

Zwick T, Baks C, Pfeiffer UR, Liu D, Gaucher BP (2004) Probe based MMW antenna measurement setup. IEEE Int Symp Ant Propag:747–750

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_69-1

Anechoic Chamber Design

Boon Kuan Chung*
Faculty of Engineering and Sciences, Universiti Tunku Abdul Rahman, Kuala Lumpur, Malaysia

Abstract

An anechoic chamber usually involves a substantial investment both financially and in building space. Hence, there is much interest to attain the required technical performance with lowest possible investment. The screened room must be designed to provide an environment free of extraneous signals. The suitable type of RF absorber must be chosen to line the entire inner surface of the shielded room in order to simulate a free-space environment with no reflection from the walls, ceiling, and floor. Using a suitable computer simulation tool together with the appropriate model which characterizes the absorber scattering behavior, the chamber geometry may be optimized to achieve the cost-effectiveness target. Upon completion of the construction, the performance of the anechoic chamber should be evaluated using the standard acceptance test methods.

Keywords

Anechoic chamber; Electromagnetic; Radio frequency; Antenna; Radiation pattern; Quiet zone; EMC; Radar cross section

Introduction

Free-space electromagnetic wave typically has very low signal levels in the order of −100 dBm or 3 μV. The signal can be easily masked off by extraneous energy in the environment. Hence, measurement of electromagnetic wave must be performed in laboratory facilities which provide high isolation from external electromagnetic environment. Shielded enclosures can provide isolation performance in excess of 100 dB. However, electromagnetic waves propagating in a shielded enclosure will be reflected by the walls, ceiling, and floor back and forth. Consequently, a complex wave front will be formed at the test region where the test antenna is to be placed, as opposed to the desired uniform plane wave. In order to create a free-space environment without reflection so that electromagnetic wave propagates from the source to the receiver in a simple and well-defined manner, suitable absorbing material must be installed over the entire inner surfaces of the shielded enclosure to prevent reflection of the electromagnetic waves. Such laboratory facility is known as anechoic chamber. With no echo and no extraneous energy, the anechoic chamber provides a "quiet" environment for electromagnetic wave measurements.

The World War II period gave rise to increasing needs to test radio frequency (RF) devices, antennas, and radar signature. It created the need for anechoic chambers as a means to improve the accuracy of indoor RF measurements. The facilities provide controlled environment, with all-weather capability and minimum electromagnetic interference. The proliferation of mobile and satellite communication industries since the 1980s brought about the design of many chambers with the capabilities to perform a wide

*Email: chungbk@utar.edu.my
*Email: bkchung1429@gmail.com

DOI 10.1007/978-981-4560-75-7_69-1

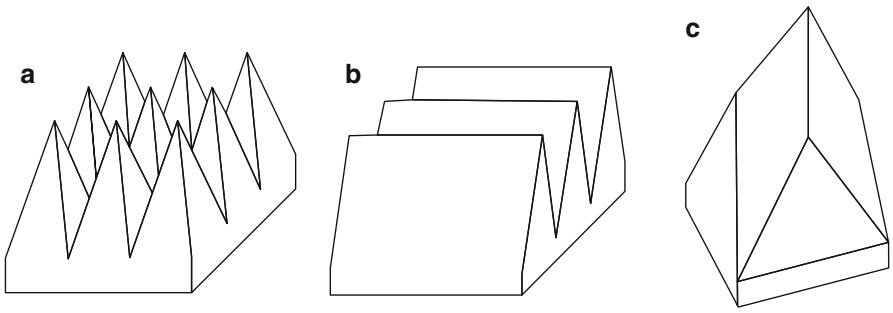


Fig. 1 Absorbers of the shape of (**a**) pyramid, (**b**) wedge, and (**c**) twisted pyramid

variety of measurements which include antenna gain, radiation pattern, polarization, radar cross section (RCS), radiation hazard, electromagnetic compatibility (EMC), etc. Others have been designed for a specific type of measurement.

Wave Absorbers

The operating frequency range of an anechoic chamber is determined by the wave absorbing material used for lining the chamber walls, ceiling, and floor. Wave absorber with narrowband performance will limit the capability of the chamber. Modern anechoic chambers typically use wave absorbers with superior absorbing properties such as low normal incidence reflection, low forward-scatter and backscatter at wide-angle incidence, lower minimum operating frequency, wide bandwidth, and reduced absorber thickness in the course of improving the "quietness" of anechoic chambers, cost efficiency, and required room size. Wave absorber with flat surface and lossy property (low impedance) will reflect a substantial amount of incident wave and absorb the remaining electromagnetic energy. Impedance discontinuity at the interface between the free space and the absorbing medium causes more reflection and less transmission. When material with higher impedance is used such that there is lesser discontinuity at the air-absorber interface, more energy can be transmitted into the absorbing medium, but a much thicker material will be required to fully dissipate the electromagnetic energy. A method to reduce the required material thickness is to use material property which gradually tapers from free-space impedance (377 ohms) at the front surface to that of a lossy medium at the back surface. It means multiple layers of carbon-impregnated material with increasing loss tangent from the front layer toward the bottom layer must be stacked up. Alternatively, lossy material can be cut into pyramid shape (see Fig. 1a) in contrast to flat front surface. With lower filling factor of lossy material with respect to air at the tip and higher filling factor at the base of the pyramid, the desired impedance taper can be achieved. It can yield a good absorption or low reflectivity level over a very wide frequency range. However, the height of the pyramid must be greater than a quarter wavelength for the absorber to function efficiently. Hence, shorter pyramidal absorber can be used if the anechoic chamber is targeted for higher microwave frequencies. The upper limit of the frequency range is determined by the surface quality and carbon loading of the absorber. A 24-inch pyramidal foam absorber may offer a reflectivity level below −40 dB over the frequency range of 1–40 GHz. If the desired lowest operating frequency of the chamber is set at 30 MHz, the taper length must be greater than 8 ft. The bulky size gives rise to high cost of material, difficulty of installation, and requirement for large room space and thus limits its use to few applications where low-frequency measurements are required.

DOI 10.1007/978-981-4560-75-7_69-1

Wedge-shaped absorber shown in Fig. 1b may give slightly poorer reflectivity level compared to pyramid-shaped absorber for wave at normal incident. However, it can give better performance at wide-angle incident for wave direction nearly parallel to the ridge of the wedge. Hence, it can be strategically used at certain areas of a chamber wall to improve the quietness of the anechoic chamber.

Long pyramid absorber used for low-frequency anechoic chamber tends to droop with age. Twisted pyramid structure has been used to improve the mechanical strength. The pyramid is "twisted" by 45° with respect to the base (see Fig. 1c). The trade-off is a poorer reflectivity performance compared to the standard pyramid.

Hollow pyramid-shaped substrate (some made of paper) can be coated with special resistive material to make lightweight broadband wave absorber. Microscopic particles of metal are dispersed in a nonconducting matrix to adjust the electrical conductivity of the resistive material.

Taper length less than a quarter wavelength may provide the required low reflectivity level using a suitable carbon loading for the pyramidal foam absorber. If the carbon loading is too high, the impedance taper will be too abrupt, and incident wave will be reflected from the region near the pyramid tip instead of propagating toward the pyramid base (hence, degrade the usable upper frequency limit). On the other hand, if the carbon loading is too low, the incident wave may be able to propagate toward the pyramid base, but it may not be sufficiently dissipated (hence, lower the absorber reflectivity performance). With a carbon loading which is optimized for the 200–1000 MHz range (for EMC test chamber) with reflectivity of −20 dB, the short pyramid absorber may not provide adequate performance at higher frequencies where the typical applications in antenna and radar cross-section measurements require absorber reflectivity below −40 dB in the anechoic chamber design.

In the design of EMC test chamber where the lowest operating frequency is 30 MHz, thin absorber with wideband performance is desirable for reduction of chamber size. Ferrite tile absorber with thickness of 0.25 inch may provide a reflectivity level below −15 dB over the 30–600 MHz range. Pyramidal foam absorber with carbon loading optimized to operate at 200–1000 MHz can be combined with ferrite tiles to cover the required frequency range of 30–1000 MHz. This type of absorber is known as hybrid absorber.

As far as the designer of anechoic chamber is concerned, the most important technical specifications of wave absorber are the usable frequency range, normal incidence reflectivity, and oblique incidence reflectivity. To qualify the wave absorber, the simplest measurement method that is currently in use is the arch method which uses a transmit antenna and a receive antenna mounted on a circular arch above the RF absorber surface. The transmit antenna illuminates the RF absorber, and the receive antenna picks up the reflection. The reflection coefficient is obtained by comparing the reflections from the RF absorber to the reflection from a conducting surface placed in the same position as the RF absorber. Different incident/reflection angles are obtained by moving the transmit antenna and the receive antenna along the arch. Parallel and perpendicular polarizations can be evaluated by the appropriate orientation of the linear-polarized antennas. The arch method is usually used above 1 GHz. Since the sample has to be in the far field, a large arch radius would be required. The sample size must be large enough, and the antenna beamwidth must not be too narrow in order to obtain an accurate result. Edge diffraction effects from the sample, direct coupling between transmit and receive antennas, and extraneous reflections from the environment must be reduced to the level below the typical reflectivity of the absorber under test.

In the selection of wave absorber to design an anechoic chamber, the first consideration should be the intended use of the facility. As mentioned above, anechoic chamber for antenna and radar cross-section measurements requires absorber reflectivity below −40 dB. This performance is typically achievable at frequencies above 1 GHz. In fact, the reflectivity performance increases with frequency at approximately 6 dB per octave. Hence, an engineer should make sure the required minimum reflectivity level is achieved at the expected lowest operating frequency of the anechoic chamber. At a much higher frequency when the reflectivity performance reaches its peak, it will start to drop off rapidly as frequency increases.

DOI 10.1007/978-981-4560-75-7_69-1

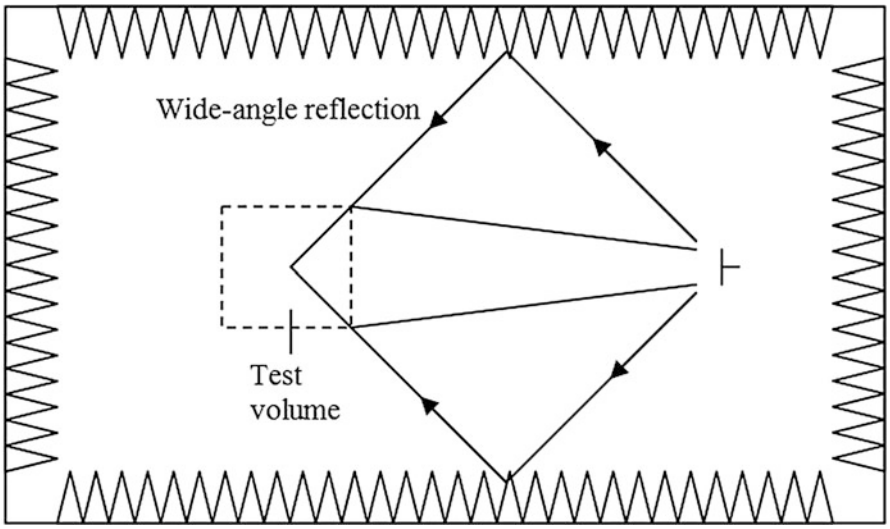


Fig. 2 Plan view of a simple rectangular anechoic chamber

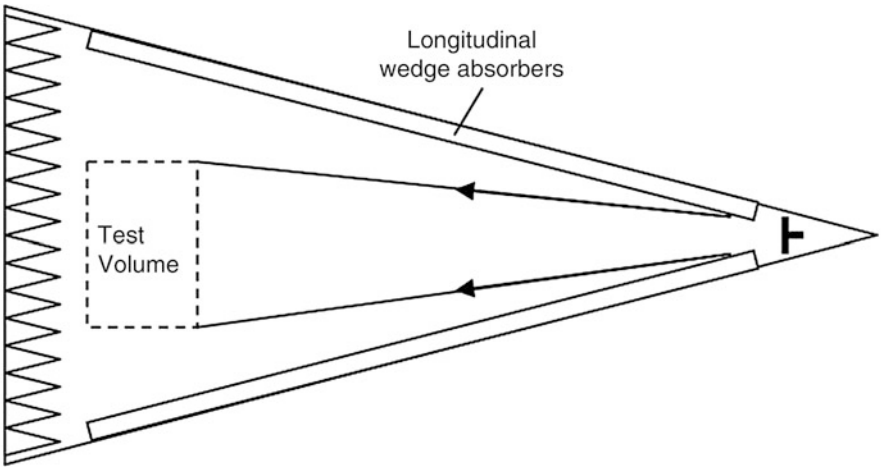


Fig. 3 Diverging geometry of a tapered chamber

This is because when the absorber size becomes large in term of wavelengths, the incident wave front starts to scatter. An optimum carbon loading and aspect ratio will be required to extend the operating frequency of the wave absorber.

Geometry of Anechoic Chamber

A rectangular-shaped chamber is simple to construct and it is also easy for lining of absorbers. One shortcoming of this simple geometry, however, is that the absorbers on the side walls and ceiling may reflect a considerable amount of wave toward the test region. The absorber performance is inadequate when the incident wave approaches it at wide angle of incidence as shown in Fig. 2. As a result, a complex wave front with sharp peaks and nulls is formed at the test region due to constructive and destructive interferences of the reflected waves and the direct-path signal. The nonuniform field distribution will cause errors in various electromagnetic wave measurements.

DOI 10.1007/978-981-4560-75-7_69-1

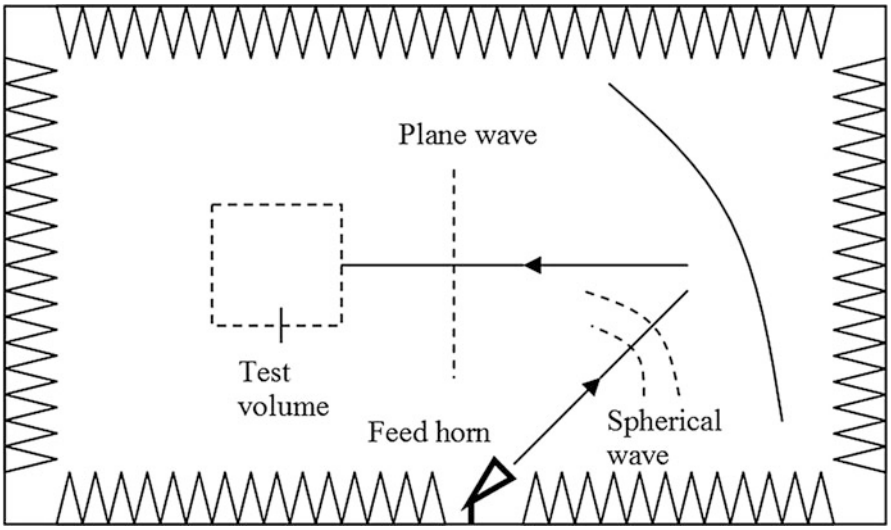


Fig. 4 Compact range anechoic chamber configuration

The problem of wide-angle reflection can be eliminated using a tapered-shaped chamber shown in Fig. 3. The concomitant benefit is the cost saving from the use of less expensive absorbers at the tapered section. The reduced total surface area of the tapered chamber means less number of wave absorber is required, and this provide further cost saving. However, there are a few shortcomings associated with it. Firstly, only a single-source antenna can be accommodated at the taper apex. Measurements involving multiple sources, moving sources, or bistatic radar cross section cannot be performed in the tapered chamber. Secondly, the path loss is different from that of free space due to coupling between the electromagnetic field and the diverging chamber walls. Hence, the chamber is not suitable for measurement involving absolute field strength. At higher frequencies, tapered chamber does not offer any advantage over rectangular chamber because wide-angle reflection can be overcome using antenna with highly directional radiation pattern (i.e., less radiation toward the walls). In addition, absorbers typically have better absorbing property at these frequencies; thus, the reflectivity level may be sufficiently low albeit wide-angle reflection.

The shortcomings of tapered chamber can be overcome using a quasi-tapered geometry. It consists of a small rectangular room, connected to a tapered section, and then followed by a larger rectangular room where the test region is located. This design incorporates the performance advantages of a fully tapered chamber at low frequencies and provides some flexibilities of a rectangular chamber. Wedge-shaped absorbers can be used at the tapered section to improve the chamber quietness performance since the wave transmitted from the smaller rectangular room toward the larger rectangular room will propagate nearly parallel to the ridge of the wedge of which a better absorbing characteristic is offered compared to pyramid absorber.

Microwave antenna and RCS measurements usually require the target under test to be illuminated by a uniform plane wave. Uniform wave front in terms of amplitude and phase can only be achieved in the far-field region and often dictates a very large distance. The distance can be reduced using an offset-feed parabolic reflector to transform spherical wave front into plane wave as shown in Fig. 4. The reduced room size is known as compact range anechoic chamber. The resulting plane wave has very flat phase front and small amplitude taper over the test zone, making the compact range equivalent to a much larger conventional antenna test range. The offset focal point eliminates scattering and shadowing from the feed and feed support.

DOI 10.1007/978-981-4560-75-7_69-1

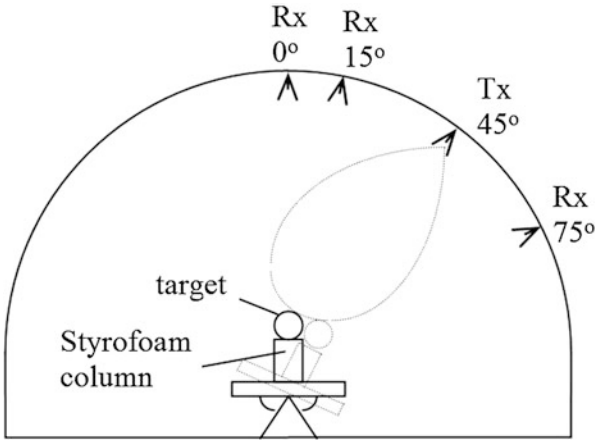


Fig. 5 Dome-shaped anechoic chamber for monostatic and bistatic RCS measurements

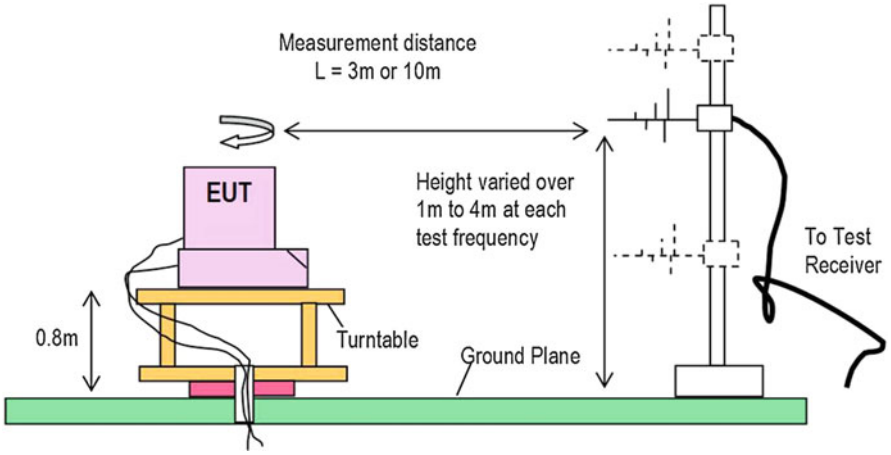


Fig. 6 EMC RF emission test configuration in an OATS

Special-purpose dome-shaped chambers are used at a few research institutions (Chung et al. 1997). The chamber facilitates a variety of monostatic as well as bistatic RCS measurements at a constant range as shown in Fig. 5. A number of antennas are mounted on the dome surface inside the chamber to receive scattered electromagnetic waves from a radar target placed at the center of the spherical space.

A large number of low-frequency (30–1000 MHz) electromagnetic anechoic and semi-anechoic test chambers have been manufactured since the 1980s due to the rapid growth in multimedia and telecommunication industries. Electromagnetic compatibility (EMC) regulation mandates that both intentional and unintentional radiators must meet the EMC requirements before they can be marketed. A full anechoic chamber is commonly used to determine the immunity of products to impinging electromagnetic field. As for the emission limit, electromagnetic interference (EMI) emissions from electronic products are referenced to measurements performed on an ideal open-area test site (OATS) having a perfectly conducting infinite ground plane. The typical test configuration is shown in Fig. 6. A semi-anechoic chamber in which the conductive floor is not covered with wave absorbing material can be used as the alternative test site for this measurement provided it fulfills the normalized site attenuation (NSA) requirements defined in CISPR 16-1-4 and ANSI C63.4. The site attenuation is defined as the ratio of the voltage input to a matched and balanced lossless tuned dipole transmit antenna to that at the output of a similarly matched balanced lossless tuned dipole receive antenna. The instrumentation configuration is

DOI 10.1007/978-981-4560-75-7_69-1

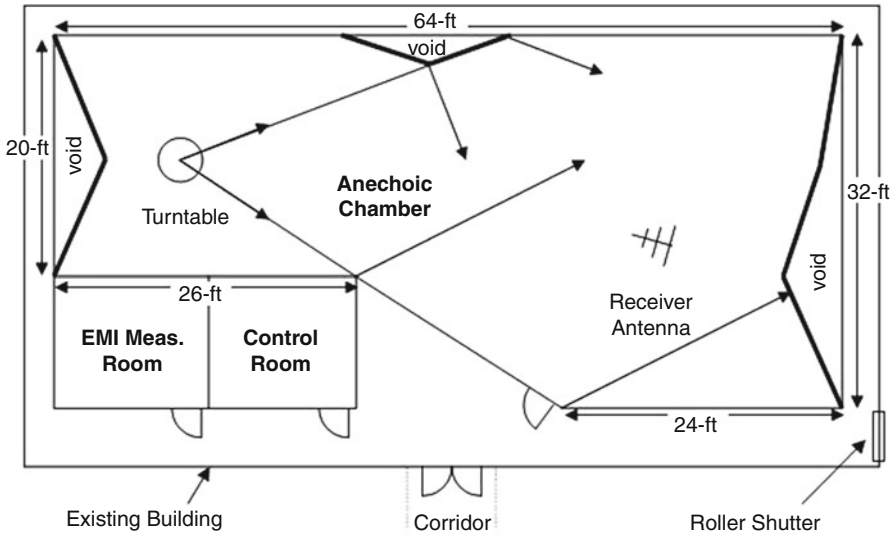


Fig. 7 Floor plan of an asymmetrical-shaped chamber

the same as the setup for product EMC compliance testing except the equipment under test (EUT) is replaced by a known transmitting source. The measured NSA data for both vertical and horizontal polarizations shall be compared to that calculated for an ideal site. A test site is considered acceptable for radiated emission measurements if the measured results are within ±4 dB of the theoretical NSA for an ideal site. The ±4 dB tolerance includes instrumentation calibration errors, measurement technique errors, and errors due to site anomalies. If the measuring distance is 3 m, the mutual impedance correction factors for tuned dipole antennas must be applied to the measurement results.

EMC radiated emission measurement above 1 GHz complying with EN55022:2010 standard is also required for information technology equipment (ITE) shipped into Europe since October 2011. Full anechoic chamber is specified for the measurement up to 400 GHz. For the US market, measurement of radiated emission above 1 GHz up to 40 GHz is also required by the FCC. Before the chamber can be used for EMC compliance measurement, the test site must be validated according to the site voltage standing wave ratio (SVSWR) method defined in CISPR 16-1-4 and/or time-domain reflectivity (TDR) method proposed for ANSI C63.4.

Anechoic chambers that are built specifically for EMC tests are usually not suitable for radar cross section and antenna measurements. Similarly, a microwave anechoic chamber using pyramidal foam absorbers with thickness less than a meter is usually not suitable for EMC tests below 1 GHz. The technical background and the required site performances for different measurement needs are described in the following sections. A multipurpose anechoic chamber was designed and built at Multimedia University (Chung and Chuah 2003). An innovative design technique has been developed, giving rise to the construction of an asymmetrical-shaped anechoic chamber which fulfills the different requirements for various measurement needs using lower-cost materials. The facility can be used for EMC tests, antenna measurements, monostatic and bistatic RCS measurements, RF transceiver testing, calibration of scatterometer, and other electromagnetic research experiments. Figure 7 shows the floor plan of the anechoic chamber. Figure 8 shows the interior view of the anechoic chamber. RF absorbers are glued onto the walls and ceiling to absorb the waves transmitted in the chamber. For treatment of the corners between two walls and between the wall and the ceiling, 12-inch pyramidal absorber and flat foam absorber are cut to fit the angle of the corners as shown in Fig. 9. Wide-angle reflection is reduced due to the asymmetrical chamber geometry in which the plane of some of the walls is parallel to the direction of wave propagation.

DOI 10.1007/978-981-4560-75-7_69-1

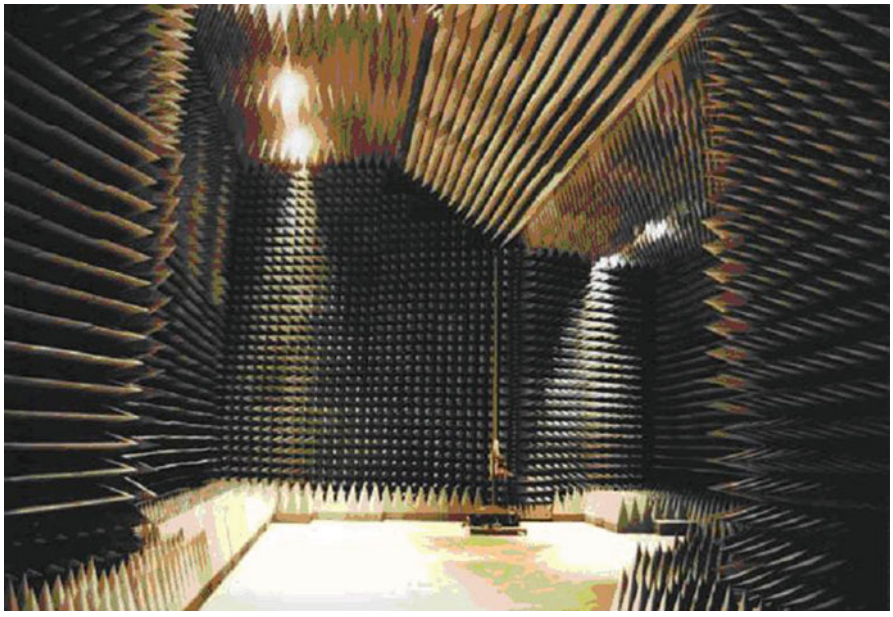


Fig. 8 The interior view of the asymmetrical-shaped chamber

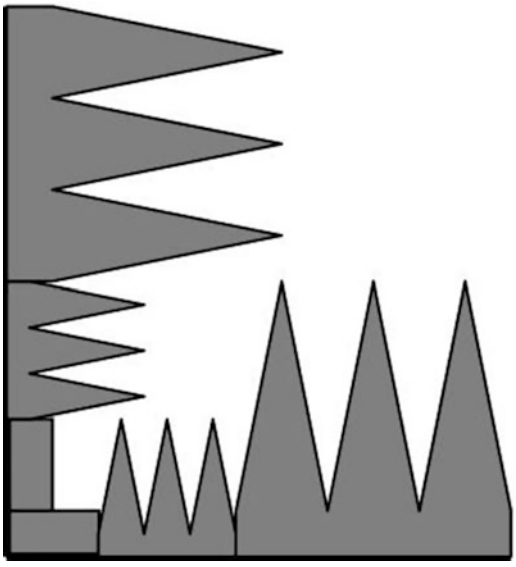


Fig. 9 Treatment of corners in the anechoic chamber

At two ends of the chamber, supplementary wall panels are installed such that the incident waves are reflected away from the test region. The walls and ceiling of the asymmetrical-shaped chamber are designed such that no first-order and second-order reflections (except those reflected from the floor) will propagate into the receiver region. When the electromagnetic rays finally reach the test region after a few bounces off the absorbing walls and ceiling, the wave energy will be sufficiently attenuated. Hence, the required quietness performance can be extended to the low-frequency range even though the absorber reflectivity is poorer at lower frequencies (since the height of pyramid becomes less than a quarter wavelength at frequency below 100 MHz). The required NSA performance can be achieved without utilizing bulky and expensive absorber.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_69-1

Size of Anechoic Chamber

In a typical antenna measurement, the separation distance between the transmit antenna and receive antenna is defined by the Rayleigh range criterion or Fraunhofer distance. Assuming the transmit antenna is a point source, and the receive antenna has a maximum transverse dimension D, the separation distance must be greater than $2D^2/\lambda$ where λ is the wavelength. This is to ensure the field at the edge and the field at the center of the receive antenna have a phase difference of less than 22.5°, thus providing a near-uniform plane wave front. The same criterion is used for radar cross-section measurement. If D is in the order of half wavelength, the Fraunhofer distance is only $\lambda/2$. This is usually not considered as a far-field distance. However, this separation distance may be sufficient to avoid undesirable near-field coupling or reactive coupling (capacitive or inductive) between transmit antenna and receive antenna, supporting only the desired radiative coupling (since the field strength falls off by $1/r$ vs. distance r). However, if D is several wavelengths such as the cases of antenna array or large radar target, the required Fraunhofer distance increases parabolically. When the separation distance is more than 2λ, it is considered as the far-field distance if the Rayleigh range criterion is fulfilled.

In the design of an anechoic chamber, the measurement facility must be sufficiently large to accommodate the transmit antenna and the receive antennas according to the Rayleigh range criterion. At lower frequency such as 30 MHz or wavelength of 10 m, separation distance of 5 m may be acceptable when the transverse dimension of the intended test volume is less than half wavelength. If the separation distance is only 3 m, such as the case of 3-m range EMC chamber, some degree of near-field reactive coupling may set in. At microwave frequency of 30 GHz or wavelength of 10 mm, and the transverse dimension of the intended test volume is 1 m, the required Fraunhofer distance will be 200 m. Anechoic chamber of this size will be prohibitive in most cases. If the Rayleigh range criterion is to be fulfilled at 10-m range, the dimension of the test volume must be reduced to 22 cm. Alternatively, the upper operating frequency must be limited to 1.5 GHz (for test volume width of 1 m). Higher frequencies can be used if the test volume width or object under test is smaller. Hence, in the specification of an anechoic chamber, specifying the size of test volume does not necessarily mean the full size of quiet zone is usable for antenna or radar cross-section measurement needs.

For antennas with large aperture (large D) operating at high microwave frequency (small λ), the required far-field distance will be very large. Near-field measurement techniques can be used to test the large antennas. A small RF probe antenna is moved over a virtual surface surrounding the antenna under test (AUT) at a typical distance of 4–10 wavelengths. Near-field amplitude and phase are measured over a discrete matrix of points. These data are then transformed to the far-field values using Fourier transform techniques. The far-field antenna pattern can then be presented in the conventional format.

Wave absorber is designed to function at the Fresnel radiative region of radiating source. It is designed to absorb electromagnetic wave which has intrinsic wave impedance (E/H) of 377 ohms. The near-field wave impedance of a dipole-type source is higher than 377 ohms, while the near-field wave impedance of a current-loop-type source is lower than 377 ohms. Hence, the distance between the radiating source and the absorber must be greater than about $\lambda_{max}/2\pi$ (where λ_{max} is the maximum wavelength at the lowest operating frequency) for the specified absorber reflectivity to be achieved. Since the thickness of pyramidal foam absorber must be greater than $\lambda_{max}/4$ to be effective, the length of an anechoic chamber must be longer than the measurement range distance by λ_{max}, and the width must be wider than the test volume size by λ_{max} as well. For a 10-m range with test volume size of 2 m and lowest operating frequency of 30 MHz, the anechoic chamber may have to be larger than 20 m long $\times$ 12 m width $\times$ 10 m height, if 2.5-m long pyramidal foam absorber is used for lining the chamber walls and ceiling. Of course, the chamber size can be greatly reduced if thin ferrite tile or hybrid absorbers are utilized, but the upper operating frequency of the chamber may be limited to 1000 MHz or so.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_69-1

Chamber Design

RF wave measurement should be performed in an environment free from electromagnetic energy not directly associated with the measurement. The isolation is normally provided by RF shielded enclosure to prevent external extraneous energy from getting into the test region and masking the low-level signal of interest. At the same time, the isolation also protects nearby systems from being interfered by the EM energy generated by radiation source in the shielded enclosure used in the measurement setup.

RF shielded enclosure is a room completely enclosed by metal plates. However, ventilation holes and access door to the shielded enclosure, as well as cables penetrating the shielded enclosure for signal/power interconnection purposes, will cause leakages of electromagnetic waves. The shielding effectiveness is often degraded by these imperfections and limited by the shielding performance of ventilators and access door. Generally, any discontinuity in the metallic surfaces, including metal-to-metal joints and overlapping metal contacts, will allow EM fields to penetrate. High-conductivity metallic gaskets (such as wire knit mesh or beryllium copper "finger stock") are commonly used to close the gaps around the access door. Specialized "honeycomb" vent, comprising a large number of small hollow waveguide tubes welded together in a "honeycomb" fashion, is commonly installed to provide ventilation in the screened room. In most practical cases, shielding effectiveness better than 120 dB is hard to achieve. Doors, vents, filters, and piping must be carefully designed and constructed to maintain the shielding integrity.

Modular clamp-up structure is commonly used to construct smaller screened room. Prefabricated wooden panels laminated with sheet metals are tied together to form the six-sided free-standing room by a framework made of plated steel. Welded construction is normally used when the required room size exceeds 240 ft^2. It provides the required structural strength and long-term high-performance shielding.

The choice of RF absorber to be used is a function of the purpose of the anechoic chamber. The required electrical performances as well as the absorber shape, thickness, weight, mounting method, cost, and logistic must be considered. For antenna and radar measurements above 1-GHz, a 24-inch thick pyramid absorber may be suitable. However, it would not be able to provide sufficiently low reflectivity at lower frequencies. Standing wave will be reinforced by reflections between the opposing walls and between the ceiling and the floor, causing cavity resonance to be established in the shielded enclosure. The absorber may not give sufficient damping of Q-factors of the resonant modes. However, the geometry of the anechoic chamber may be optimized to extend the usable range to lower frequencies.

An anechoic chamber with a floor area of 20 m $\times$ 10 m and 8-m height requires a substantial investment both financially and in building space. The important considerations in an anechoic chamber project may include the specifications, procurement process for various components and instrumentation, logistical arrangements (contract, shipping, insurance, haulage, time to delivery, storage space), construction process, and performance evaluation.

The geometry of the chamber structure must not be too complex so that the screened room can be easily constructed, the shielding effectiveness can be under control, and the RF absorbers can be properly lined on the chamber walls, ceiling, and floor. In order to optimize the chamber design in terms of geometry and combination of RF absorbers, wave propagation simulation is often conducted to compute the vector sum of direct-path signal and the reflected signals. The objective of the simulation is to evaluate the quietness level within the test volume which is indicated by the magnitude of amplitude ripple along the longitudinal and transverse axes.

The calculations of wave components due to reflections from absorber-lined walls and ceiling require a proper model that characterizes the absorber scattering behavior as a function of incident angle, frequency, and polarization. The magnitude and phase must be known in order to calculate the total field due to the sum of the direct wave and the reflected waves.

DOI 10.1007/978-981-4560-75-7_69-1

Wave energy incidents on the absorber will be scattered in virtually all directions much like a rough surface. A geometrical optics modeling method that traces the rays of electromagnetic waves propagating from the transmitter to the receiver assumes specular reflection at the walls, ceiling, and floor. The calculated field quantities may not compare well with the measurement results. The only reliable modeling method is to compute a complete 3D solution of Maxwell's equations over the entire interior of the chamber. Finite-difference time-domain (FDTD) methods have been used in the analysis of compact anechoic chamber by Luebbers et al. (1991). This method is very laborious in terms of specifying the exact dimensions and material properties of the chamber structure, fixtures, and absorber configuration, and it involves extremely long computation time. In all practical cases, the measured results will be different from the predicted performance due to the simplified models of the chamber fixtures (such as antenna mast, turntable, lighting, CCTV camera, "honeycomb" waveguide vents, penetration panels, etc.) and neglecting the imperfection in lining the absorbers. Hence, the use of geometrical optics method together with a reasonable model that represents the absorber-lined surfaces will suffice to predict the approximate chamber performance prior to the chamber construction.

Holloway and Kuester (1996) have used a geometrical optics (GO) method to predict the site attenuation of a semi-anechoic chamber (SAC). The method of homogenization (Holloway and Kuester 1994) has been used to model the RF absorber. A rectangular chamber shape was considered, and only a single bounce of the rays off the absorbing walls and ceiling was accounted in the calculation of the total field strength. Reflections due to waves that make more than one bounce off the absorbing surfaces were assumed negligible. This assumption may only be valid if a good quality absorber has been used. Site attenuation measured at two chambers lined with 4-ft and 5-ft pyramidal absorbers shows poor correlation with the calculated values with deviations more than 4 dB for both horizontal and vertical polarizations. Larger deviations were found for horizontal polarization and for frequencies below 80 MHz (Mayer et al. 1998). One of the reasons for these deviations is due to the variation in material properties over the bulk of the pyramidal absorber. Average permittivity and conductivity from a few material samples are typically used in the theoretical analysis. The other error factors are mutual impedance between the antennas, near-field effects (between each antenna and the conductive floor), poor low-frequency performance of the absorber (that more than one bounce must be considered), and room resonance.

Geometrical optics (GO) is the most popular technique for modeling of wave propagation behavior. Ray representations of electromagnetic waves are traced from the transmitter, changing ray direction as they undergo reflection and refraction, and the field quantities of those rays that arrive at the receiver are determined. The geometrical theory of diffraction (GTD) and uniform theory of diffraction (UTD) methods are extension to GO rays to include the effects of diffraction. The behavior of the diffracted wave at edges and corners can be determined from an asymptotic form of the exact solution for simpler canonical problems. For example, the diffraction around a sharp edge is found by considering the asymptotic form of the solution for an infinite wedge. The geometrical optics methods are only accurate when the dimensions of objects being analyzed are large relative to the wavelength of the field.

There are two basic categories of ray-tracing propagation models: brute-force shoot and bounce ray (SBR) (Yang et al. 1998) and image-based methods (Kimpe et al. 1999). Conventional SBR methods suffer large kinematics error due to the use of reception sphere. Image-based methods do not have such problem. However, without an efficient image generation algorithm, they are not manageable for complex scene. A hybrid SBR-image model that improves the efficiency of image generation has been developed by Chen and Jeng (1997). The choice of SBR technique would govern the performance of the hybrid model. The basic classes of SBR techniques that can be used are ray, tube, and beam approaches.

In the use of ray approach, individual rays are launched and traced. The separation between various rays launched from the transmitter is critical. It must be small enough that two adjacent rays never straddle and miss a potential scatterer or receiver. On the other hand, it must not be too small that the computation time

DOI 10.1007/978-981-4560-75-7_69-1

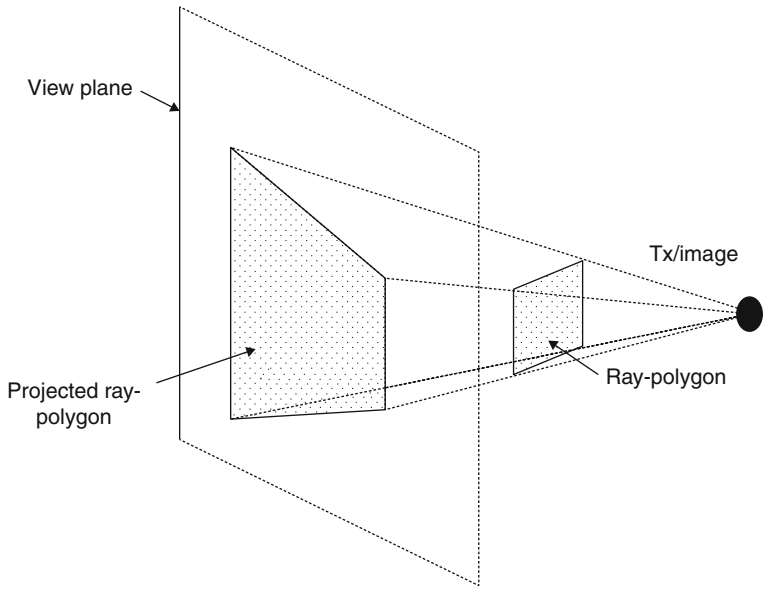


Fig. 10 Projection of ray polygon to a view plane

is excessive due to the huge number of rays. The optimum ray separation is difficult to determine because the ray separation increases with distance. Different ray separations must be used for different scenes. Reception has to be validated by a backward-tracing process. A significant number of backward-tracing efforts may end in invalid reception and discarded. This kind of trivial effort grows with the number of receivers or observation points.

A ray-tube approach where several rays are grouped to form a tube with simple polygon cross section (e.g., triangle or rectangular) has been developed by Chen and Jeng (1997). It enhances the reception validation process by eliminating backward ray tracing of invalid ray paths. A more complicated tube-tracing algorithm is required. Ray separation or tube size remains critical, and the concerns of optimum ray separation also apply here.

Improvement using large ray tubes is provided by a beam-tracing method (Fortune 1998). Beams with irregular cross section, representing infinite number of rays, are used. Ray separation is not a concern, and the beam size is not critical. Ideally, one beam represents one ray path, in contrast to the large number of rays or ray tubes required for keeping track of one ray path in the ray or ray-tube approach. The beams also assist in reception test by eliminating trivial backward tracing of invalid ray paths. However, the beam tracing is also the most complicated to code among the SBR techniques.

A variant of beam-tracing technique has been developed by Teh and Chuah (2003). Generic polygons are traced instead of the beams. Thus, the technique is called polygon-tracing method. A generic polygon may consist of one or more polygons which may be concave or convex, is self-intersecting, or contains holes. Scatterers are modeled as objects with polygonal facets. Reception sphere is not required, thus giving rise to very low kinematics errors. In a problem where a large number of receivers are involved, a minimum number of beams can cover all the receiver locations. Hence, the required computation time can be dramatically reduced. This feature is particularly useful in calculating the field strength at different positions of the receiving antenna in typical RF measurement site validation methods.

The polygon-tracing method shoots and bounces polygonal projections from the transmitter and its images. The ray polygon may be transformed to a different polygonal shape when projected onto an arbitrary view plane as illustrated in Fig. 10. Each view plane is selected to coincide with a plane object at the scene. The footprint on the view plane represents the region illuminated by the transmitter or the image source.

DOI 10.1007/978-981-4560-75-7_69-1

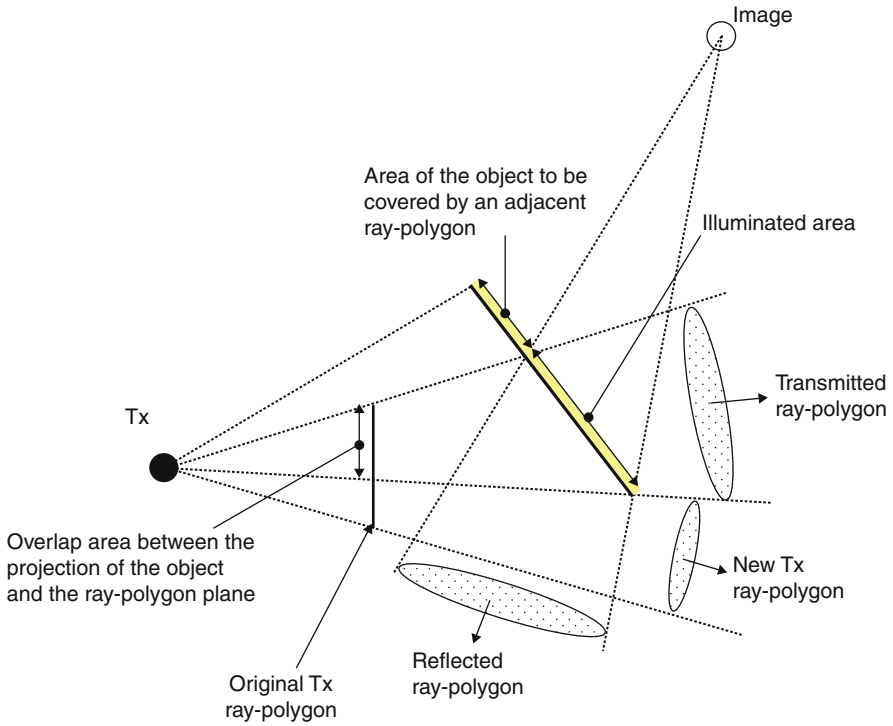


Fig. 11 Creation of transmitted and reflected ray polygons

To cover rays in all directions, the original transmitter ray polygons may be represented by the six faces of a rectangular box enclosing the transmitter. To begin tracing a ray polygon, the entity (scatterer or receiver) nearest to the source is projected onto the plane containing the ray polygon. The center of projection (COP) is the position of the source. If the projection of a receiver lies within the ray polygon, the receiver will receive a ray from the source. Backward ray tracing is then performed to determine the exact path information for electromagnetic field calculations based on GO and UTD. In the case of a scatterer entity, the overlap area between the projection of the object and the ray-polygon plane indicates the area where the beam energy will be intercepted. Reflection and transmission ray polygons that take the shape of the illuminated area of the object are created and queued for further tracing to the subsequent scatterer or receiver at the scene. Other areas of the object will be illuminated by an adjacent ray polygon. An illustration of the polygon-tracing mechanism is given in Fig. 11.

To include diffraction, edges of the polygonal object plane that fall within the projection of the ray polygon are considered. Diffraction ray polygons are created and traced for their subsequent reflection and transmission. However, the projection of the ray polygon based on the diffraction law is rather complicated. The projection for a straight line may become a curved line. Since the diffracted wave amplitude is usually very weak, the subsequent reflection, transmission, and diffraction may be negligible.

In a GO or polygon-tracing method, the electric field strength $|E_r|$ in free space (with no ground reflection) at a distance r from the transmitter antenna is given by:

$$|E_r| = \frac{|E_1|}{r} \tag{1}$$

where $|E_1|$ is the radiated electric field strength at a distance of 1 m from the transmitter antenna. At the quiet zone or test volume of an anechoic chamber, the electric field strength $|E_{obs}|$ observed at a receiver position will be given by the vector sum of the direct-path signal and all the reflected signals.

DOI 10.1007/978-981-4560-75-7_69-1

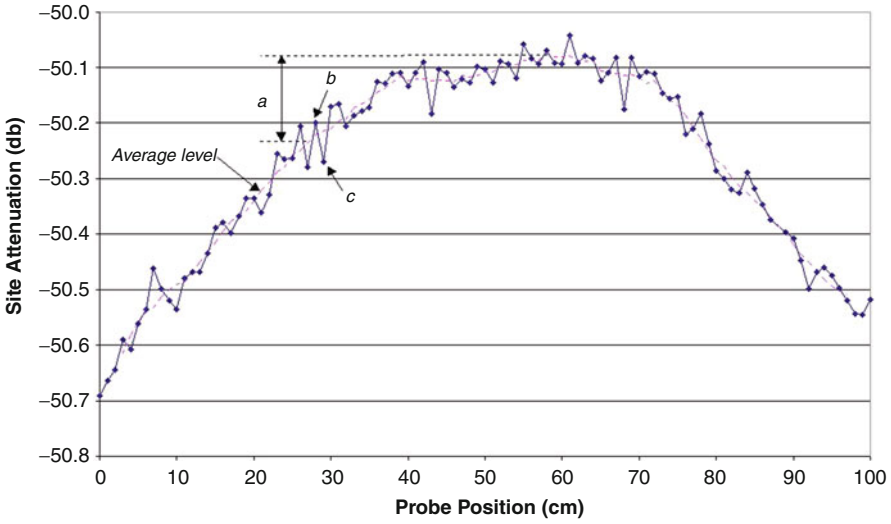


Fig. 12 Amplitude variations within the quiet zone

The magnitude of $|E_{obs}|$ at various positions within the test volume is to be found. In the computer simulation program, an impulse is transmitted from a dipole antenna, and it propagates in all directions in the form of ray polygons. For every bounce from the absorbing wall, the electric field strength is multiplied with the reflection coefficient. The degradation of the absorber reflectivity at oblique incidence is taken care of by the appropriate absorber model. Instead of calculating $|E_{obs}|$ for a large number of discrete frequency points, the received impulses resulting from various multipath propagations (with different time delays) are combined to give a time-domain impulse response. A fast Fourier transform (FFT) is performed to obtain the desired frequency-domain results.

If the chamber has a rectangular shape, multiple first-order reflections and second-order reflections would interfere with the direct wave at the receiving antenna as illustrated in Fig. 12. To improve the quietness performance, the chamber geometry can be designed such that no first-order and second-order reflections will propagate into the test volume. When the electromagnetic rays finally reach the receiver antenna after a few bounces off the absorbing walls, the wave energy will be sufficiently attenuated.

An asymmetric anechoic chamber shown in Figs. 7 and 8 has been designed and constructed at Multimedia University (Chung and Chuah 2003). With no symmetry, waves reflected from the walls will have different path lengths, and therefore the possibility of two waves with about the same amplitude to add in phase is very low. The required quietness performance can be achieved without the need to utilize bulky and expensive absorber.

Field Uniformity of Quiet Zone

Full anechoic chambers are commonly used for antenna and radar cross-section (RCS) measurements at microwave frequencies above 1 GHz. The performance of the chamber is defined by the uniformity of the illuminating field over the *quiet zone*. It is the region in which the antenna under test or radar target is to be placed. The size of the quiet zone may be specified in terms of the amplitude and phase variations of the field from the center of the quiet zone to the edge of the zone. The illuminating field over the quiet zone will deviate from that calculated based upon ideal range geometry because of reflections from various mounting structures, cables, and obstacles on or near the range surface and from irregularities in the range surface. Amplitude ripple and cross-polarization may result.

DOI 10.1007/978-981-4560-75-7_69-1

In an anechoic chamber, the size of the usable quiet zone is dictated by the Rayleigh range criterion. The phase variation of the direct wave from the center of the quiet zone to the edge of the zone must be less than 22.5° (equivalent to λ/16). Larger phase variation may cause the measured antenna pattern to depart appreciably from the true far-field pattern. For a fixed range distance R of 10 m, the diameter of the quiet zone decreases with frequency from 1.22 m at 1 GHz to 28.9 cm at 18 GHz.

The amplitude taper from the center to the edge of the quiet zone must be minimized so that it will not adversely affect the electromagnetic wave measurement accuracy. This can be achieved when the 3-dB beamwidth of the transmit antenna is much wider than the quiet zone. Amplitude ripple may result due to interference of the direct wave with reflections from the walls, ceiling, and floor of the anechoic chamber. The effect of wide-angle reflection may be reduced by the use of directional transmit and receive antennas. Hence, the amplitude ripple depends on the radiation pattern of the antennas, and standardized figure of merit cannot be defined for an anechoic chamber.

The quietness of the quiet zone may be defined as the average ratio of reflected energy to that of the direct wave. The "free-space voltage standing wave ratio" field probe technique (IEEE Std. 149–1979) is usually used to evaluate the quietness level. An antenna with stated directivity is used as the probe. A probe carriage is utilized for the probe to move continuously along a transverse line. The interference pattern of the field is recorded along with the linear motion of the probe. Measurements are made for horizontal and vertical transverse directions with both horizontal and vertical polarizations. Standard gain horn antenna is usually used as the probe.

Figure 12 shows an example of signal strength variation similar to a standing wave pattern measured over a quiet zone. The amplitude of the ripples varies across the graph due to the complicated variation in reflected field amplitude with position. At the same time, the average level of the standing wave curve varies slowly due to the change in "eyesight" angle between the transmit antenna and receive antenna as well as the distance variation between them. One can see this as the composite variation in pattern level of the antennas.

Referring to Fig. 12, the pattern level at probe position of 28 cm is a, and the detected signal level varies from b to c when the direct wave E_d and the reflected wave E_r are detected in phase and out of phase, respectively. Then,

$$b(dB) = 20 \cdot \log \frac{E_d \cdot 10^{a/20} + E_r}{E_d} \tag{2}$$

$$c(dB) = 20 \cdot \log \frac{E_d \cdot 10^{a/20} - E_r}{E_d} \tag{3}$$

$$R_c(dB) = 20 \cdot \log \frac{E_r}{E_d} = a + 20 \log \frac{10^{(b-c)/20} - 1}{10^{(b-c)/20} + 1} \tag{4}$$

where R_c is known as the reflectivity level of the quiet zone. As there are more than one reflected field received by the probe at any one time, what is really established by Eq. 4 is an equivalent reflected wave which is the aggregate effect of all reflected waves incident upon the probe antenna. The result obtained will depend on the directivity of the antenna. For example, an omnidirectional antenna will yield a higher reflectivity level (i.e., poorer apparent performance) since the equivalent reflected wave is composed of all the reflected waves from all surfaces of the chamber. If directional source antenna and/or directional probe antenna are used, less number of reflected waves will be received by the probe. Hence, a lower reflectivity level (i.e., better apparent performance) will be obtained, although both results refer to the same anechoic

DOI 10.1007/978-981-4560-75-7_69-1

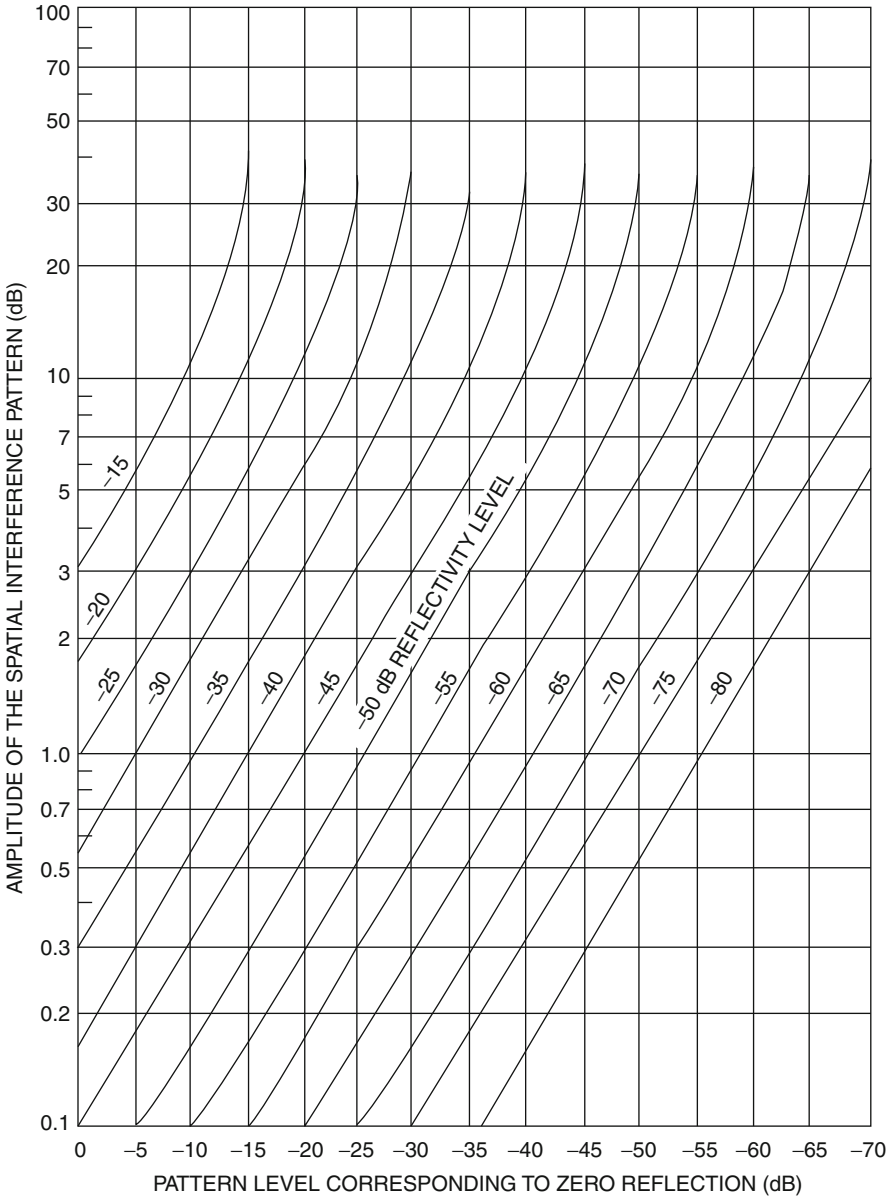


Fig. 13 Tabulated graph for determining the reflectivity level of an anechoic chamber

chamber. Anyway, a standardized reflectivity level has little usefulness other than for benchmarking purpose because "poorer reflectivity" may not appreciably affect the measurement accuracy if the AUT directivity is high.

In the analysis of the standing wave curve, the maximum peak-to-peak interference amplitude (b–c) must be determined. The average level of the interference pattern is used to determine the pattern level, a. The reflectivity level R_c can be calculated using Eq. 4 or using the tabulated graph in Fig. 13. It is apparent that R_c depends upon the pattern level a at which (b–c) is found. For the same value of (b–c), the reflectivity level R_c increases with the pattern level. For the pattern level $a = -0.15$ dB and peak-to-peak interference amplitude (b–c) = 0.1 dB, the reflectivity level of the anechoic chamber is −45 dB. If the same (b–c) = 0.1 dB is obtained at pattern level $a = -5$ dB, it indicates the anechoic chamber has a better reflectivity of −50 dB.

DOI 10.1007/978-981-4560-75-7_69-1

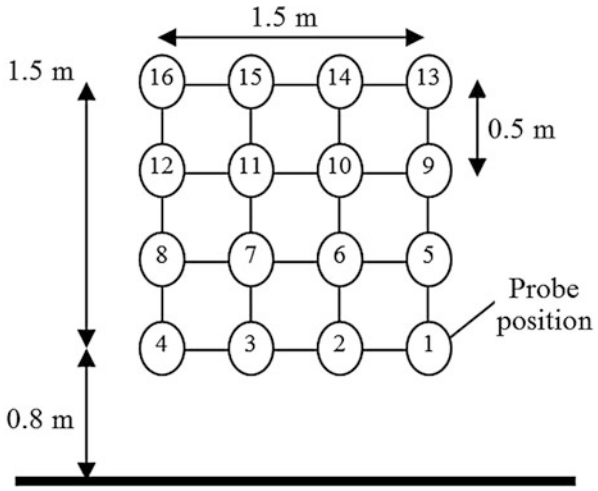


Fig. 14 Positions for field uniformity test

Low-frequency full anechoic chambers are used for radiated RF field immunity tests required by European EMC standard EN61000-4-3. In this test, a strong RF field is radiated from a transmit antenna toward the equipment under test (EUT). The standard field strengths for severity levels 1, 2, and 3 are 1 V/m, 3 V/m, and 10 V/m, respectively. Illumination of the electromagnetic field is applied at four faces of the EUT with both vertical and horizontal polarizations. The most susceptible configuration and operating mode of the EUT are to be set up. A set of parameters that represent the operation of the EUT are monitored from the control room. Any degradation of the operating parameters is recorded. The standard requires the field at the EUT's position to have uniform amplitude distribution from 80 to 2700 MHz (EN61000-4-3:2006 + A2:2010). To validate the chamber for RF field immunity tests, field uniformity test is made at 16 points over a grid covering a plane area of 1.5 m × 1.5 m with the lowest point at 0.8 m above the floor as illustrated in Fig. 14. The test site is considered acceptable for radiated RF immunity test if the field strength at 12 of the 16 points is within the tolerance of −0 dB/+6 dB of the test level. The large tolerance is allowed because the standard takes into practical consideration that wave absorbers typically have poor reflectivity performance at lower frequencies.

RCS Chamber Acceptance Test

Radar cross section (RCS), σ, is a parameter used in characterizing the radar signature of an object. Given the actual target echo at the receiving system, σ is the effective area which would intercept sufficient power out of the transmitted field to produce the given echo at the receiving system by isotropic reradiation. This parameter is a function of the material properties (such as conductivity, electric permittivity, and magnetic permeability) and the shape of the target and is unique for a particular frequency and polarization. It is evident that σ can be greater than the physical cross section of the target if the object is highly reflective and/or has a preference direction of scattering which is in the direction toward the receiving system. Similarly, σ can otherwise be less than the physical cross section. The measured value of σ may be influenced by the transmitted waveform (CW, pulse, or FM) and wave front (plane or spherical).

Radar technology has advanced into the spectrum of millimeter and micrometer wavelengths. The application of radar equipment is no longer confined to looking up at aircraft or out at ships. Airborne and spaceborne radars, capable of producing images of ground, have been developed and extensively used in

DOI 10.1007/978-981-4560-75-7_69-1

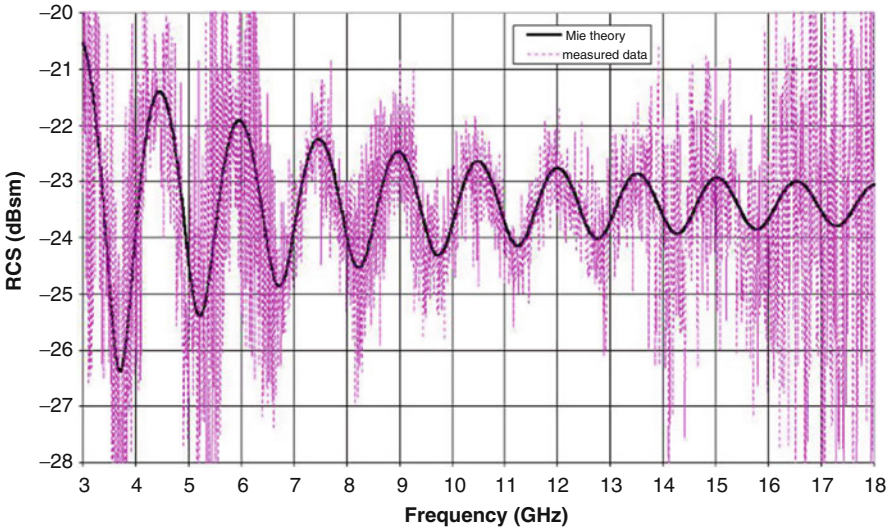


Fig. 15 Comparison between measured and theoretical monostatic RCS of a 3-inch sphere, without time gating

remote sensing applications. Clutter due to backscattering from ground in the early imaging radar has been utilized as useful information associated to the vegetation cover and underground materials. RCS control has also progressed from RCS reduction using radar-absorbing materials to body shaping and impedance loading (loading the body surface with distributed or lumped impedance). A considerable amount of researches have been directed to RCS enhancement when, in some applications, it is desirable to enhance the radar cross section above the physical cross section of the body, over a specific range of aspect angles. One example of these applications is in the tracking of civilian aircraft where a strong and relatively steady return is desired for good tracking accuracy by ground radar. Methods for this purpose include the use of impedance loading, corner reflectors, dielectric lenses, and retrodirective arrays. Dihedral and trihedral corner reflectors have been extensively used in the calibration of polarimetric radars.

RCS of deterministic target is usually measured in an anechoic chamber. Chamber for this measurement has an additional requirement to provide low background level. The performance criterion is therefore defined by its apparent RCS measured in the absence of any target placed at the test region. The background RCS is effectively the total system sensitivity of the measurement system which includes the anechoic chamber as well as the RCS measurement instrumentation. It determines the smallest RCS that can be measured.

Vector network analyzer with time-gating capability is typically used in the instrumentation. A known target such as a flat reflecting metal sheet or a metallic sphere is placed at the target pedestal. The scattered signals are measured in a series of step-frequency continuous wave transmission. Figure 15 shows an example of co-polarized RCS measurement result of a 3-in. metal sphere. Note that the measured RCS (expressed in dB square meter or dBsm) is oscillating around the theoretical value. This is primarily due to the coupling effect between the target and the pedestal. The electromagnetic wave scattered from the pedestal to the target, and vice versa, follows a longer path and interferes with that scattered directly from the target surface. This results in constructive or destructive interference, depending on the wavelength.

The result is converted to its time-domain equivalent in Fig. 16 using inverse discrete Fourier transform computation. Errors due to target-pedestal coupling appear as secondary impulses after the “arrival” of the main impulse. The noise level in the vicinity of the target indicates the minimum detectable RCS or the apparent RCS of the chamber. A time window can be applied to remove the unwanted signal. When the time-gated transient RCS is converted back to the frequency domain using a discrete Fourier transform computation, the RCS of the target in frequency domain can be obtained.

DOI 10.1007/978-981-4560-75-7_69-1

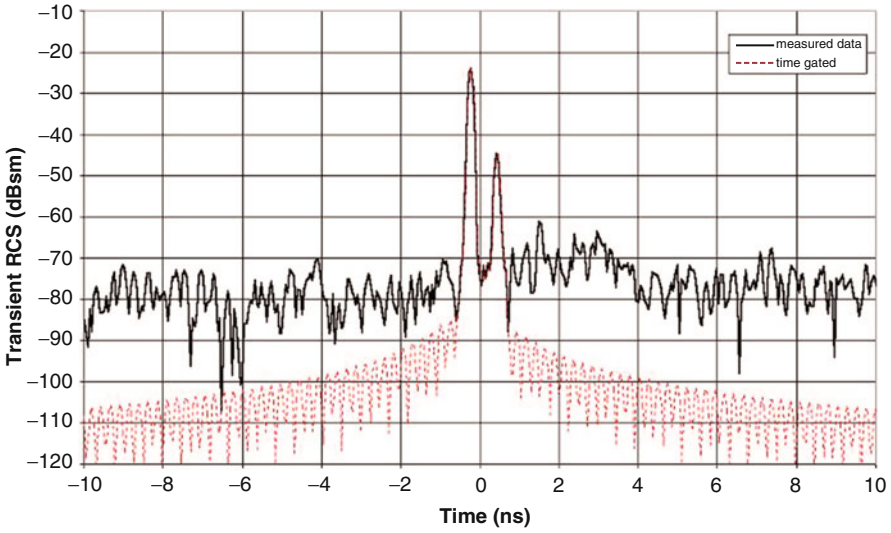


Fig. 16 Transient RCS of a 3-inch sphere with 3–18 GHz bandwidth, before and after time gating

In the acceptance test of an RCS chamber, the metallic sphere is moved through the quiet zone. The radar returns may vary depending on the ripple in the illuminating field. The peak-to-peak ripple in the radar returns can be related to the extraneous signal level in the test region.

Conclusion

Anechoic chamber may be designed for different measurement needs. The performance criteria employed to evaluate the anechoic chamber depend on the type of measurements to be made in the facility. Accordingly, the chamber geometry and wave absorber to be used should be carefully considered in the design. In addition, the design must include the necessary features such as access doors, ventilation, lighting, electrical power supply, filters, connection panels, cables, pneumatic air supply, turntable, motorized antenna mast, etc. Wave propagation simulation should be conducted to compute the vector sum of direct-path signal and the reflected signals. The objective is to evaluate the quietness level within the test volume which is indicated by the magnitude of amplitude ripple along the longitudinal and transverse axes. The proper model that characterizes the absorber scattering behavior as a function of incident angle, frequency, and polarization should be used in the calculations of wave components due to reflections from absorber-lined walls and ceiling. After the construction of the anechoic chamber, acceptance tests should be performed according to the available standards.

Cross-References

- ▶ Anechoic Chamber Design
- ▶ Antenna Measurement Setups-Introduction
- ▶ EMI/EMC Chamber Design, Measurement and Instrument
- ▶ mm-Wave Sub-mm-Wave Antenna Measurement
- ▶ Near-Field Antenna Measurement Techniques
- ▶ Radio Wave Propagation Models

DOI 10.1007/978-981-4560-75-7_69-1

References

Chen SH, Jeng SK (1997) An SBR/Image approach for radio wave propagation in indoor environments with metallic furniture. IEEE Trans Antennas Propag 45(1):98–106

Chung BK, Chuah HT (2003) Design and construction of a multipurpose anechoic chamber. IEEE Antennas Propag Mag 45(6):41–47

Chung BK, Chuah HT, Bredow JW (1997) A microwave anechoic chamber for radar cross-section measurement. IEEE Antennas Propag Mag 39(3):21–26

Fortune S (1998) Efficient algorithms for prediction of indoor radio propagation. 48th IEEE Veh Technol Conf 1:572–576

Holloway CL, Kuester EF (1994) A low-frequency model for wedge or pyramid absorber arrays – II: computed and measured results. IEEE Trans Electromagn Compat 36(4):307–313

Holloway CL, Kuester EF (1996) Modeling semi-anechoic electromagnetic measurement chambers. IEEE Trans Electromagn Compat 38(1):79–84

Kimpe M, Leib H, Maquelin O, Szymanski TH (1999) Fast computational techniques for indoor radio channel estimation. Comput Sci Eng 1(1):31–41

Luebbers R, Steich D, Ryan D, Kunz K (1991) Analysis of compact electromagnetic anechoic chamber performance using finite difference time domain methods. Southeastcon 1:4–7

Mayer F, Ellam T, Cohn Z (1998) High frequency broadband absorption structures. IEEE Int Symp EMC 2:894–899

Teh CH, Chuah HT (2003) An improved image-based propagation model for indoor and outdoor communication channels. J Electromagn Waves Appl 17(1):31–5

Yang CF, Wu BC, Ko CJ (1998) A ray-tracing method for modelling indoor wave propagation and penetration. IEEE Trans Antennas Propag 46(6):907–919

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_71-1

Radiation Efficiency Measurements of Small Antennas

Yi Huang*
Department of Electrical Engineering and Electronics, The University of Liverpool, Liverpool, UK

Abstract

This chapter deals with a range of antenna radiation efficiency measurement methods and techniques: from the traditional to the state-of-the-art. The most widely used methods are discussed in details, along with their backgrounds and scientific foundations. The relevant measurement setup, procedure, and data processing techniques are also given. A comparison of these methods is made at the end of this chapter, and their advantages and limitations are identified.

Keywords

Antenna efficiency measurement; Radiation efficiency measurement; Antenna measurements; Small antennas

Introduction

The antenna *radiation efficiency* (η) is one of the most important parameters since it is the most useful and informative measure of how efficient an antenna is to radiate or receive the radio wave at a desired frequency. It is defined as the ratio of the total power radiated by the antenna (P_r) to the total power accepted by the antenna (P_a) at its input terminal.

$$\eta = \frac{P_r}{P_a} \tag{1}$$

It is normally expressed in percentage although some people prefer to express it in dB. A good and efficient antenna should have very high radiation efficiency, ideally 100 % (i.e., 0 dB). It should be pointed out that the radiation efficiency is a function of the frequency but not a function of the radiation direction. It is determined by the antenna loss but does not include the antenna impedance mismatch. The mismatch between the feedline/connector and the antenna is taken into account by the total efficiency of the antenna (η_T) which is defined as the ratio of the total power radiated by the antenna (P_r) to the total power supplied to the antenna (P_s) and can be expressed as

$$\eta_T = \frac{P_r}{P_s} = \frac{P_a}{P_s}\frac{P_r}{P_a} = \left(1 - |\Gamma|^2\right) \cdot \eta \tag{2}$$

where Γ is the reflection coefficient (the same as the S-parameter S_{11} or S_{22}) which is a complex number (note: some published papers did not use the magnitude in this equation which is not correct) and can be easily measured using a vector network analyzer. When the reflection coefficient is zero (the perfect match

*Email: Yi.Huang@IEEE.org

DOI 10.1007/978-981-4560-75-7_71-1

case), the power accepted by the antenna is the same as the power supplied to the antenna, and there is no mismatch; thus, the total efficiency is the same as the radiation efficiency – this is an ideal case. In reality, the radiation efficiency is always larger than the total efficiency of the antenna.

The term "antenna efficiency" can cause confusion in reality since it could mean the radiation efficiency or the total efficiency. Although they are closely linked as shown in Eq. 2, they can be very different: an antenna may have a radiation efficiency of 90 % but a total efficiency of just 9 %! Thus, it should be made clear and specific especially when a value is mentioned. In this chapter, the radiation efficiency will be used since the total efficiency can be obtained easily using Eq. 2 if it is required.

The radiation efficiency is a very important figure of merit and needed for the computation of the radiated or received power once the input power is known. For some applications, the antenna efficiency is a much more important parameter than the radiation pattern. Take mobile phones as an example; the battery power drain is a big and bottleneck problem and significantly affected by the antenna efficiency, while the antenna radiation pattern (very important for many applications) is not important in this case since the field around a mobile phone is relatively random (especially in the urban environment) and the way we use the phone is also relatively arbitrary. Thus, an efficient antenna can be considered as a power amplifier without the need for a power supplier – it is the best solution to make an efficiency radio system. In this example, an efficient antenna can improve the mobile system performance and extend the battery life. The antenna efficiency is therefore of huge interest to a wide range of applications. However, this parameter is actually not easy to obtain: the theoretical or numerical value is normally not accurate or reliable because the antenna material electromagnetic properties may not be ideal or known, and sometimes complex supporting and/or housing structures are involved. Just like the antenna gain, the measured antenna efficiency is preferred in practice. There has been a lot of interest in how to make accurate measurements on the antenna efficiency. A number of measurement methods and techniques have been proposed and studied over the years (Wheeler 1959; Newman et al. 1975; Pozar and Kaufman 1988; Johnson and McRory 1998; Schantz 2002; Huang et al. 2003, 2010, 2011; Huynh 2004; Rosengren and Kildal 2005; Schroeder and Gapski 2006; Fur et al. 2009; Holloway et al. 2012). Some have already gained wide recognition and have been adopted in practice, but some are still subject to further investigation and/or improvement for real-world applications. In this chapter, mainly the well-developed and matured methods and techniques will be introduced. Some newly developed promising methods will also be introduced and discussed. The antennas of interest are physically small and can be measured in a standard laboratory.

Measurement Methods

Many radiation efficiency measurement methods have been developed, and it is very hard to say which method is the best since each method has its advantages and disadvantages – it is also linked to the specific antenna. Some methods are particularly suitable for certain types of antennas. The intention is to provide a relatively comprehensive review of the radiation efficiency measurement methods in this section. A comparison of these methods will be provided at the end of the chapter.

Pattern Integration Method

Background and Theory

This is probably the oldest conventional method to measure the antenna radiation efficiency, and it is based on the definition (i.e., Eq. 1) which means that the power accepted by the antenna and the power radiated by the antenna are required. The accepted power can be calculated using

DOI 10.1007/978-981-4560-75-7_71-1

$$P_a = \left(1 - |\Gamma|^2\right) \cdot P_s \tag{3}$$

The power supplied to the antenna (P_s) can be measured using a power meter (or a receiver), and the reflection coefficient can be measured using a vector network analyzer. The difficulty is how to measure the total radiated power. In theory, it is

$$P_r = \oint U \cdot d\Omega = \int_0^{2\pi} \int_0^{\pi} U(\theta, \phi) \cdot \sin\theta d\theta d\phi \tag{4}$$

where U is the radiation intensity in W/unit solid angle and linked to the radiated electric field E in the far field by

$$U = r^2 \cdot |E \times H| = r^2 \cdot \left|E^2\right| / 120\pi \tag{5}$$

where r is the distance from the antenna to the observation/measurement point. Thus, one needs to measure the field E in the three-dimensional (3D) space at a fixed distance r (which should be large enough to be considered in the antenna far field). The integration over the whole spherical space for the antenna field (pattern) has to be conducted, which is why this method is called the pattern integration method. In reality, the E field can be decomposed to two orthogonal components (E_θ and E_ϕ) and measured at discrete sampling points (i, j), thus the total radiated power:

$$P_r = \frac{r^2}{120\pi} \sum_{i,j} \left(|E_\theta(i,j)|^2 + \left|E_\phi(i,j)\right|^2\right) \sin\theta_i \Delta\theta_i \Delta\phi_j \tag{6}$$

The accurate measurements of these fields in the 3D space could be very time-consuming and difficult.

Facility and Equipment

Typically, the measurement should be conducted in an anechoic chamber, which is a large room lined with radio-absorbing materials (RAMs) on the walls, the ceiling, and the floor. It is an indoor facility to simulate the free space. Alternative facilities such as the open area test site (OATS) and near-field probe system are also fine. A power meter or receiver is required to measure the power supplied to the antenna and the E field at the selected sampling points. A calibrated field probe or antenna is required for the E field measurement as well. A vector network analyzer (VNA) is needed to measure the reflection coefficient of the antenna. Since only the magnitude of the reflection coefficient is needed, other equipment (such as a network analyzer) or methods could be used to replace the VNA to determine the reflection coefficient magnitude (Fig. 1).

Measurement Setup and Procedure

For the E field measurement, a typical setup in an anechoic chamber is shown in Fig. 2.

The procedure to obtain the radiation efficiency can be summarized by the following six steps:

1. Measure the power supplied to the antenna using a power meter or a receiver.
2. Measure the antenna reflection coefficient using a VNA.
3. Calculate the power accepted by the antenna using Eq. 3.

DOI 10.1007/978-981-4560-75-7_71-1

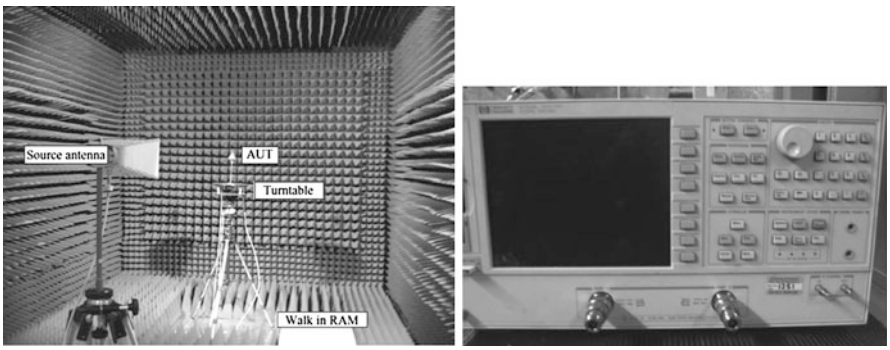


Fig. 1 An anechoic chamber and a vector network analyzer

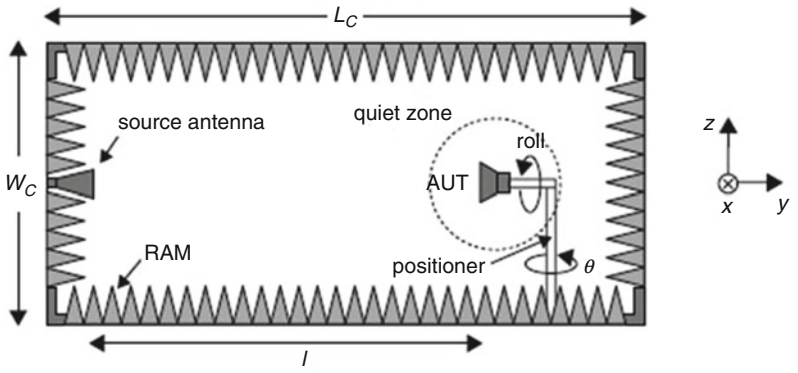


Fig. 2 Antenna radiated field measurement inside an anechoic chamber

4. Measure the *E* field in 3D at a fixed distance.
5. Calculate the total radiated power using Eq. 6.
6. Calculate the antenna radiation efficiency using Eq. 1.

Discussion and Conclusions

The beauty of this measurement method is that it is closely linked to other antenna characterization measurements such as the antenna impedance (via the reflection coefficient measurement) and radiation pattern measurements. The same facilities and equipment are used. However, it needs the integration/summation of the measured radiated field to obtain the total radiated power. This could be a very time-consuming process since the antenna field/pattern measurement should be done in 3D; the results are of high uncertainty (>10 %) (Pozar and Kaufman 1988). Furthermore, when the operational frequency is relatively low, it may not be possible to obtain its radiated field pattern in a standard laboratory (either a large test site would be required to meet the far-field condition or a near-field facility should be used).

Directivity/Gain Method

Background and Theory

From the antenna theory, we know that the antenna radiation efficiency is the ratio of the gain *G* to the directivity *D*:

$$\eta = G/D \tag{7}$$

DOI 10.1007/978-981-4560-75-7_71-1

The directivity does not take the antenna loss into account but the gain does. Thus, the gain is normally smaller than the directivity and can be measured using a number of well-known measurement methods. The most straightforward approach is to use the comparison method where a reference antenna with a known antenna gain is employed; the gain is obtained by comparing the received powers with the antenna under test (AUT) and the reference antenna. Another popular approach is to employ two identical antennas as the transmitting and receiving antennas. Using the Friis transmission formula, we can obtain the antenna gain:

$$G = \left(\frac{4\pi r}{\lambda}\right)\sqrt{\frac{P_R}{P_a}} \tag{8}$$

where λ is the wavelength, P_R is the received power in watts, and P_a is the power accepted by the transmitting antenna in watts. It is assumed that the polarizations of the antennas are matched and the peaks are aligned with the line of measurement. The measurement should be conducted under the far-field condition.

The directivity of the antenna cannot be measured directly, but it can be computed from its definition formula:

$$D = \frac{4\pi U}{\int_0^{2\pi}\int_0^{\pi} U(\theta,\phi)\cdot \sin\theta d\theta d\phi} \tag{9}$$

Just like the pattern integration method, the E field measurement has to be conducted in 3D at discrete sampling points. Thus, this process can be very time-consuming and expensive. However, here we can use the normalized field or power radiation pattern, the distance information is not required, and there is no need to obtain the absolute field value; thus, the computed directivity can be more accurate, and the resulted radiation efficiency can therefore be more reliable than the previous method.

If the antenna is directional and the radiation is in the main lobe, the directivity is closely linked to the beamwidths (θ_{HP}, ϕ_{HP}) in the E- and H-planes; thus, the directivity can be estimated using (Huang and Boyle 2008)

$$D \approx \frac{4\pi}{\theta_{HP}\phi_{HP}} \tag{10}$$

This method is much quicker and simpler than the pattern integration method by using Eq. 9, but the accuracy is very much dependent on the directivity value obtained using Eq. 10.

Due to the significant advancement in computational electromagnetics (CEM), it is now relatively easy to obtain the antenna directivity using a numerical simulation tool such as CST Microwave Studio or HFSS. Since the directivity does not include any loss of the antennas, the simulated directivity can be very close to the measured directivity. Thus, instead of measurement, it is also acceptable to use the simulated directivity (but care has to be taken to ensure the simulation is conducted properly. It is easy to obtain inaccurate results using a very good simulation tool). This approach is also simple, accurate, and fast.

Facility and Equipment

Again, the measurement should be conducted in an anechoic chamber. Alternative facilities such as an open area test site (OATS) and a near-field measurement system are also acceptable.

DOI 10.1007/978-981-4560-75-7_71-1

Unlike the pattern integration method, here we just need to measure and compare the relative gain and *E* field values; thus, there is no need to use a power or receiver to obtain the absolute values. A VNA is sufficient since it can be used to measure the field (power) difference of two antennas, the reflection coefficient, and the radiation pattern of the antenna. If the antenna directivity is to be obtained via the computer simulation, relevant simulation software and detailed antenna configurations are also required. If the antenna gain is to be obtained using comparison method, a standard reference antenna with a known antenna gain is required.

Measurement Setup and Procedure

For the antenna gain and directivity/pattern measurements, a typical setup in an anechoic chamber is basically the same as the pattern integration method which is shown in Fig. 2.

The procedure to obtain the radiation efficiency can be summarized by the following three major steps:

1. Measure the antenna gain using one of the recommended methods (such as the comparison method).
2. Obtain the antenna directivity using one of the recommended methods (using Eq. 9 or Eq. 10 or via computer simulation).
3. Calculate the radiation efficiency using Eq. 7.

Discussion and Conclusions

Like the pattern integration method, standard antenna measurement equipment and facility are required. But only comparative measurements are to be conducted, and no calibration is required. This measurement method may provide more accurate results than the pattern integration method.

Wheeler Cap Method

Background and Theory

As we have seen that the determination of the radiation efficiency, one of the most important antenna parameters, was not easy, more efficient and convenient methods were required. In 1959, Wheeler introduced a theory about the radiansphere (Wheeler 1959), which has a radius of $\lambda/2\pi$ and is defined as the boundary of the reactive field and radiating near-field regions. The field inside the radiansphere is mainly involved in the loss power and stored energy, while the field outside the radiansphere is dominated by the radiation power, and the ratio of the electric field to the magnetic field equals to the intrinsic impedance (about 377 Ω in free space). "For purposes of measurement, it may be desired to remove the radiation resistance of a small antenna while retaining its other properties (loss resistance, capacitance, inductance). This can be accomplished to a close approximation by enclosing the antenna in a radiation shield which ideally is a perfectly conducting spherical" (Wheeler 1959). This method has now become the well-known Wheeler cap method which is the most popular radiation efficiency measurement method for small antennas due to its simplicity, speed, repeatability, and accuracy.

Basically, when the antenna is capped by a conducting cap, the power consumed by the antenna in this case is considered as the antenna loss power P_{cap}. While the antenna is placed in free space (or an anechoic chamber), the power accepted by the antenna P_{free} is viewed as the summation of the loss and radiation powers (i.e., $P_{\text{free}} = P_{\text{cap}} + P_r$). Linking the power and resistance of the antenna, we have

$$\eta = \frac{P_r}{P_a} = \frac{P_{\text{free}} - P_{\text{cap}}}{P_{\text{free}}} = \frac{I^2 R_{\text{free}} - I^2 R_{\text{cap}}}{I^2 R_{\text{free}}} = \frac{R_{\text{free}} - R_{\text{cap}}}{R_{\text{free}}} \tag{11}$$

DOI 10.1007/978-981-4560-75-7_71-1

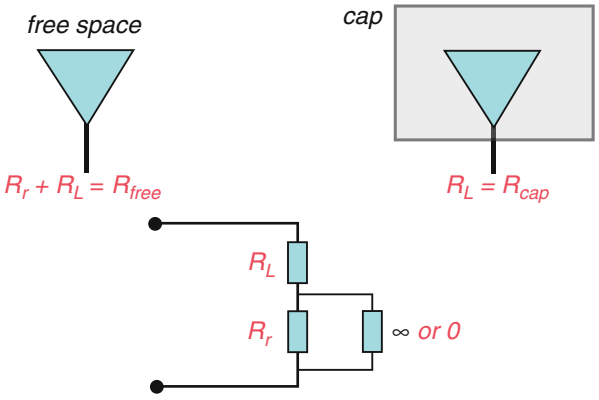


Fig. 3 The equivalent circuit for the Wheeler cap method

We therefore need to measure the resistance of the antenna in free space ($R_{\text{free}} = \textit{antenna resistance} + \textit{loss resistance}$) and its resistance inside the cap ($R_{\text{cap}} = \textit{loss resistance}$) separately and then use Eq. 11 to calculate the antenna radiation efficiency. This process can also be explained using the equivalent circuit theory as shown in Fig. 3 where the reactance of the antenna impedance is ignored since we are only interested in the resistance of the antenna. When the antenna is placed in free space, the resistance is the summation of the radiation resistance and the loss resistance which is in parallel with an open circuit (the impedance is infinite). When the antenna is placed inside the cap which acts as a short circuit (the impedance is zero) in parallel with the radiation resistance, thus, the antenna resistance is just the loss resistance.

The conditions here are:

- The antenna is small (electrically).
- The cap is made of a very good conductor (the loss is negligible compared with the antenna loss).
- The radius of the cap is small, typically about $\lambda/2\pi$ which is the boundary between the reactive and the radiating near fields of a small antenna.

The current on the antenna is also assumed unchanged during the measurement when it is inside and outside the cap. This is reasonable since the current distribution along an electrically small antenna is not sensitive to the environment (following a sine function for a dipole antenna).

Facility and Equipment

The Wheeler cap method is very straightforward. To obtain the radiation efficiency, one has to measure the antenna impedance (resistance more specifically) when the antenna is placed inside and outside the conducting cap. Thus, the essential equipment and facilities required are:

- A suitable sized conducting cap
- A VNA for impedance measurement

Ideally, the impedance measurement should be conducted inside an anechoic chamber. If the antenna is not very directive (all small antennas are not very directive), the measurement can be conducted in an OATS or just a big room.

It should be pointed out that, in reality, the cap does not have to be a sphere or hemisphere in shape. Other shapes (such as a rectangular shape) are also fine as long as the conditions above are met.

DOI 10.1007/978-981-4560-75-7_71-1

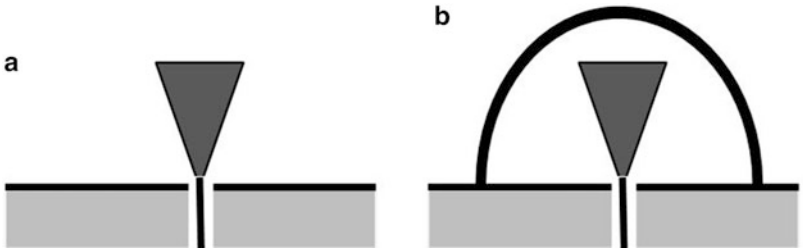


Fig. 4 Antenna measurement (**a**) in free space and (**b**) in a Wheeler cap

Measurement Setup and Procedure

A typical measurement setup is shown in Fig. 4. A calibration should be conducted before the measurement, and the cap size should be large enough to accommodate the antenna.

The procedure to obtain the radiation efficiency is to:

1. Measure the antenna impedance (resistance) when it is in free space (anechoic chamber).
2. Measure the antenna impedance (resistance) when it is in a Wheeler cap.
3. Calculate the antenna radiation efficiency using Eq. 11.

Discussion and Conclusions

To use the Wheeler cap method, one just needs a VNA and a conducting cap to obtain the radiation efficiency with a minimum of data processing. It is simple, fast, cost-effective, and accurate. However, it also has some drawbacks and challenges:

(a) It is only suitable for electrically small antennas. This means that the method cannot be applied to the antennas which are comparable or larger than the wavelength; thus, a large number of antennas cannot benefit from this wonderful method, which include broadband antennas (they are not electrically small).

(b) The cap cannot be too small or too large; although $\lambda/2\pi$ is just a guideline, it can be slightly larger or smaller in practice. The real condition is that no resonant modes can be generated inside the cap as we will see later. This means that for different antenna sizes, we may need different caps.

(c) For some antennas, such as a short dipole or monopole, the reflection coefficient is very close to 1, and the impedance solution is not unique. Because the reflection coefficient for an antenna with input impedance Z_a is

$$\Gamma = \frac{Z_a - 50}{Z_a + 50} = \frac{R_a - 50 + jX_a}{R_a + 50 + jX_a} \tag{12}$$

when the reflection coefficient is close to 1, the solutions could be (i) any resistance R_a but large reactance X_a or (ii) large resistance and any reactance. Thus, it is actually very hard to obtain the impedance accurately in practice. Any small errors in the measured reflection coefficient can result in large errors in the impedance. How to measure the impedance of an electrically small antenna accurately is actually a challenging task.

(d) The assumption that the antenna impedance can be modeled by either a series or parallel RLC circuit may not hold for some antennas (such as circularly polarized antennas or complex antennas) and result in errors (McKinzie 1997; Cho et al. 2014).

DOI 10.1007/978-981-4560-75-7_71-1

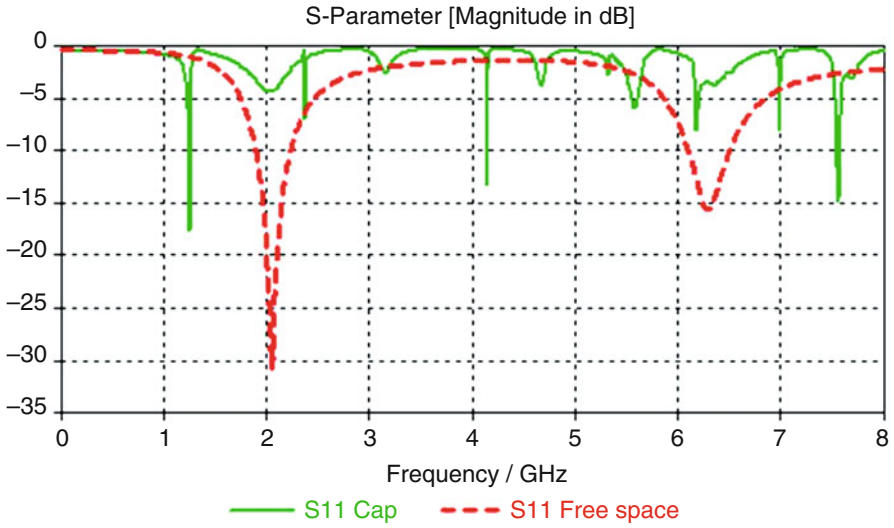


Fig. 5 Typical dipole antenna reflection coefficients inside and outside (free space) the Wheeler cap when the cap size is comparable with or larger than the wavelength

Modified Wheeler Cap Methods

Background

As compared with other antenna efficiency measurements, the Wheeler cap method offers the most attractive features but only suitable for electrically small antennas. Thus, there have been many attempts to modify the original Wheeler cap method in order to measure the antenna which is not electrically small.

When an antenna which is not electrically small is placed inside a cap, the cap will not be electrically small (at least comparable with the wavelength) – this means that resonance can be generated inside the cap. In this case, the real part of antenna impedance includes not just the loss resistance of the antenna but also the loss of the cap. At a resonant frequency, the loss of the cap can be significant which results in a small reflection coefficient. Take a lossy dipole antenna with a resonant frequency around 2 GHz as an example; a typical reflection coefficient in free space is shown in Fig. 5 as a broken line. When this antenna is placed inside a conducting cap which is larger than this half-wave dipole, the reflection coefficient in this case is shown as a solid line with many dips/nulls which are caused by the cap resonance, and the corresponding "antenna impedance" is no longer just determined by the antenna. Thus, Eq. 11 is no longer valid for calculating the antenna efficiency, and the Wheeler cap method cannot be used. To resolve this problem, a number of modified methods have been proposed and investigated.

Sliding Wall Cavity Method To remove the resonance, a sliding wall cap/cavity was employed to conduct the measurement as shown in Johnson and McRory (1998) and Huang et al. (2003). A typical measurement setup and a measurement result are shown in Fig. 6. The cap/cavity is made of a good conductor, but its size can be much larger than $\lambda/2\pi$. The input resistance of the AUT inside and outside the cap/cavity should be measured as in the standard Wheeler cap method. If a cavity mode appears near the operation frequency at which the measurement is taken, the cavity resonance is tuned by moving the sliding walls until the resonance peak is shifted away from the operation frequency. Thus, a range of measurements at different sliding wall positions may have to be undertaken. The antenna efficiency can be calculated using Eq. 11. Johnson and McRory (1998) described a method to obtain the S-parameter by measuring the same antenna with several dimensions, using a sliding wall. These measurement results form a circle in the Smith chart. If the maximum ($\Delta_{S,\max}$) and minimum ($\Delta_{S,\min}$) S_{11} differences

DOI 10.1007/978-981-4560-75-7_71-1

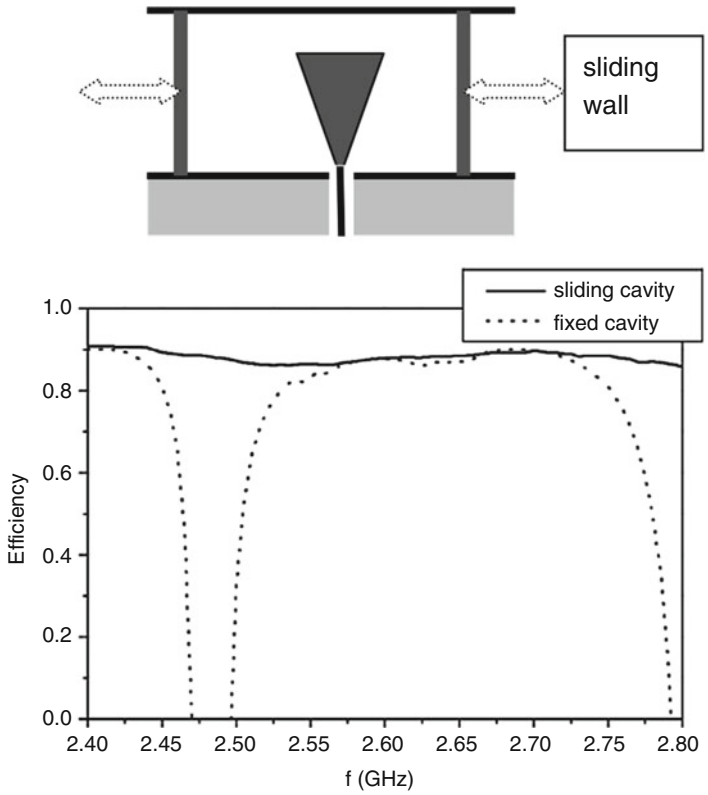


Fig. 6 A typical setup for the sliding wall cavity method and a measurement result (Huang et al. 2003)

(in absolute values) from the free space reflection coefficient S_{11} are obtained, the antenna radiation efficiency can be calculated using the following equation:

$$\eta = \frac{2}{\left(\Delta_{S,\max}\right)^{-1} + \left(\Delta_{S,\min}\right)^{-1}} \cdot \frac{1}{1 - |S_{11}|^2} \tag{13}$$

The data processing for this approach is more complicated than the Wheeler cap method.

The sliding wall cavity method is an effective way to remove the resonance and has kept most advantages of the original Wheeler cap method. The results are reasonably accurate for most antennas. The main drawback is that a conducting sliding wall is required, and the connectivity of the wall and the cavity could be a major source of errors – a gap between the sliding wall and the cavity may act as a slot antenna to introduce additional loss. In addition, the resonant mode at the desired frequency should be removed which could be tricky in some cases since the power loss could be resulted from the resonance as well as the antenna.

UWB Wheeler Cap Method Broadband antennas are normally comparable or even larger than a wavelength. There has been a great deal of effort to make the Wheeler cap method suitable for broadband antenna measurement. Due to the enormous interest in ultra-wideband (UWB) antennas, the UWB Wheeler cap method was introduced in 2002 by Schantz (2002). The basic idea behind the method is that the cap should be much larger than $\lambda/2\pi$, and the radiated wave from the AUT can be treated as a ray (far field) as shown in Fig. 7. Thus, the total power reflection coefficient of the AUT inside the cap is

DOI 10.1007/978-981-4560-75-7_71-1

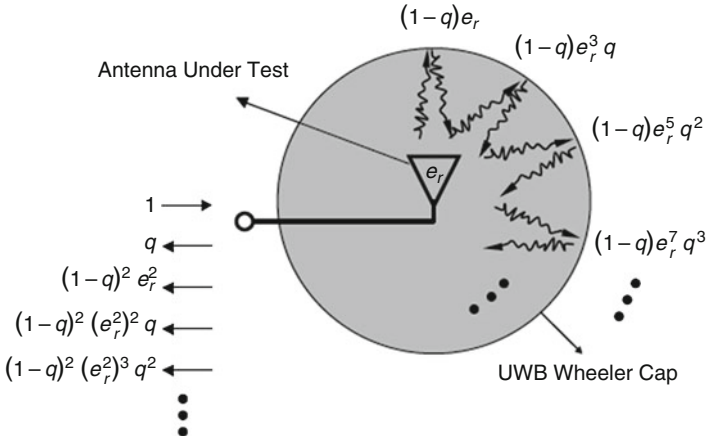


Fig. 7 The AUT inside a UWB Wheeler cap ($e_r = \eta$ is the antenna radiation efficiency) (Huynh 2004)

$$\left|S_{11\,\mathrm{cap}}\right|^2 = q + (1-q)^2\eta^2 + (1-q)^2\left(\eta^2\right)^2 q + \ldots = q + (1-q)^2\eta^2 \sum_{n=0}^{\infty}\left(\eta^2 q\right)^n \tag{14}$$

where $q = |S_{11\,\mathrm{free}}|^2$ is the antenna power reflection coefficient in free space (outside the cap) and the power reflection coefficient of the cap is assumed to be 1. $S_{11\,\mathrm{cap}}$ and $S_{11\,\mathrm{free}}$ are the antenna (voltage) reflection coefficients when it is inside the cap and free space, respectively. The solution of this equation for the radiation efficiency is found to be

$$\eta = \sqrt{\frac{\left(\left|S_{11\,\mathrm{cap}}\right|^2 - \left|S_{11\,\mathrm{free}}\right|^2\right)}{\left(1 - 2\times\left|S_{11\,\mathrm{free}}\right|^2 + \left|S_{11\,\mathrm{cap}}\right|^2 \times \left|S_{11\,\mathrm{free}}\right|^2\right)}} \tag{15}$$

This equation is the same as that in Eq. 9 but different from Schantz's original one which has an error.

Again, one needs to measure the antenna reflection coefficients inside and outside the cap in order to obtain the radiation efficiency. The measurement is straightforward and quick. The method shares many advantages of the conventional Wheeler cap method. Compared with the sliding wall cavity case, more resonant modes are excited in the UWB Wheeler cap at the frequencies of interest. The resonance is reflected on the S_{11} curve as sharp narrowband dips as shown in Fig. 5 which needs to be removed in the data processing. The common approach is to smooth the S_{11} by taking the largest value of the raw data over a frequency window (such as 100 MHz) and then use the processed S_{11} and Eq. 15 to compute the radiation efficiency. An example of the measured and simulated total antenna efficiencies of a monopole antenna is given in Fig. 8. More examples can be found from Schantz (2002) and Huynh (2004).

However, in reality, the cap cannot be made too large; thus, the ray approach may not be accurate. Also the loss of the cap has not been taken into account which could result in a considerable error (Lu et al. 2008). Thus, Eq. 15 is just an approximation, and the result tends to be more optimistic. But at higher frequencies, the mode density is very high, and the smooth technique may overestimate the loss of the antenna which results in smaller radiation efficiency.

Source-Stirred Cap/Chamber Method The concept of the "source-stirred chamber" was originally proposed in 1992 for electromagnetic compatibility (EMC) measurements (Huang and Edwards 1992), and the operational principle is similar to the reverberation chamber (to be discussed in the next section).

DOI 10.1007/978-981-4560-75-7_71-1

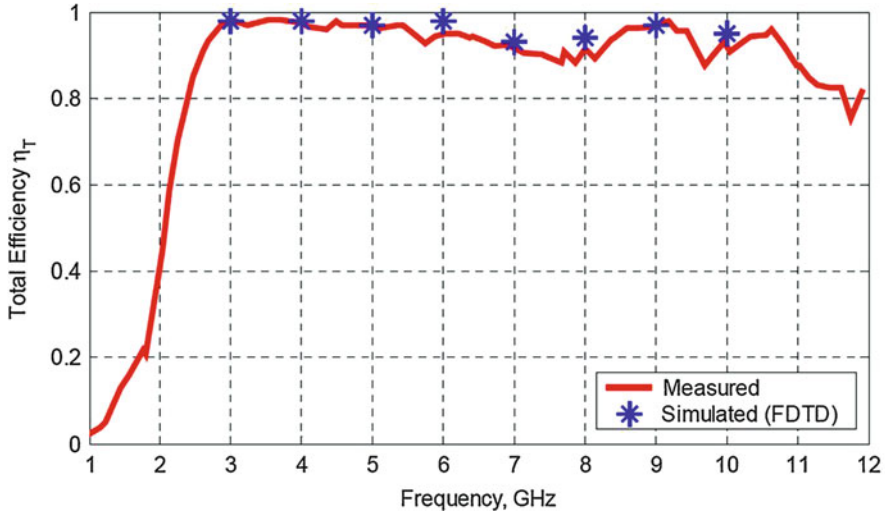


Fig. 8 The measured and simulated total efficiency of a UWB monopole antenna (Huynh 2004)

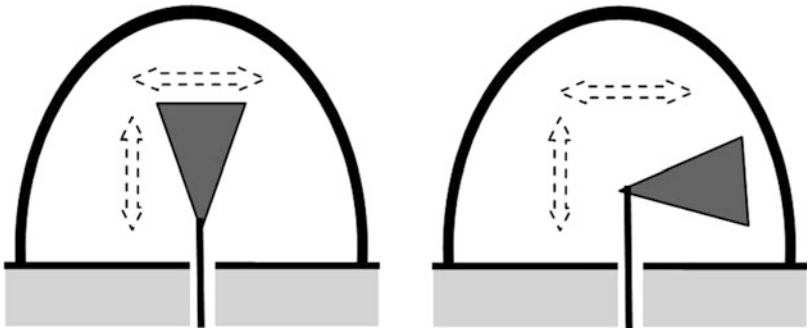


Fig. 9 Illustrative examples of the AUT placed inside the cap at different locations and orientations

Using this idea for antenna efficiency measurement, the antenna under test (AUT) should be placed in different locations inside the source-stirred cap/chamber (SSC) with different (better orthogonal) orientations (Huang et al. 2010, 2011) as illustrated in Fig. 9. The conducting cap/chamber should be large enough to accommodate significant field and mode changes. Having measured the reflection coefficients S_{11} of the antenna in free space and inside the SSC, one can use either Eq. 15 or Eq. 11 (they are about the same at over-mode cases) to calculate the radiation efficiency and plot the results obtained from each position on the same figure, as shown in Fig. 10 where three measurement results (with three orthogonal orientations represented by three colors) are displayed. One can link the peaks using a smoothed line which is the final antenna radiation efficiency – the peaks are not significantly affected by the resonant modes since they are considered to have the least effect by the SSC and mainly determined by the antenna (loss resistance). At nonresonant frequencies, the cap is considered as a short circuit in parallel with the radiation resistance as in the conventional Wheeler cap method.

It is apparent that this method is relatively simple to implement and requires very limited data processing (like the traditional Wheeler cap method): one just needs to stir the source (AUT) to different positions inside the cap and measure its impedance. Compared with the sliding wall cap/cavity method, this technique utilizes a fixed cap, and the resonant modes are tuned away by moving the AUT rather than sliding the wall. Compared with the UWB Wheeler cap method, this method has a similar data processing scheme, but there is no need to apply a window to smooth the data. Another important feature is that this method can better remove the resonant mode effects by changing both the antenna location and

DOI 10.1007/978-981-4560-75-7_71-1

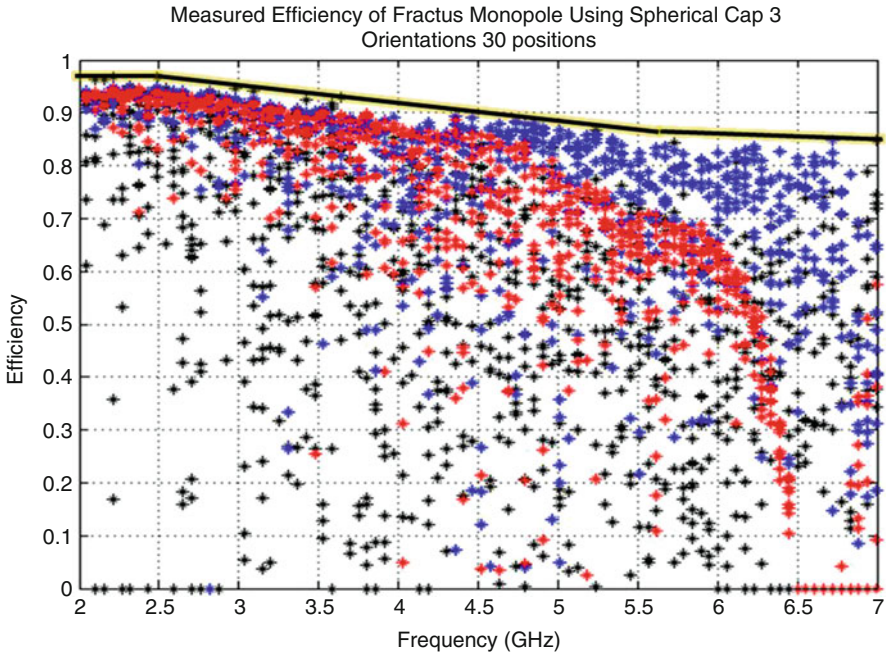


Fig. 10 An example: measured efficiency using the raw S_{11} data for three different AUT orientations (indicated by colors) inside the cap (Huang et al. 2010, 2011)

orientation inside the cap. Take Fig. 10 as an example; using the UWB Wheeler cap method, the antenna efficiency is obtained by linking the peaks of the red dots which are very small at higher frequencies (about 0 % between 6.5 and 7 GHz). This results in a considerable error since this particular orientation of the antenna inside the cap has a large loss over the higher frequency band. However, if we use the data from three orientations, the combined results can give much more accurate results since some resonant modes were not generated.

Since the measurement is conducted when the AUT is placed at different locations and orientations inside the cap/cavity, there is a section of cable inside the cap/cavity which could be a source of errors by introducing losses. The selection of the feeding cable (the type and length) is an important consideration. The cable effects inside the cavity must be minimized.

Reverberation Chamber Methods

Background

The reverberation chamber (RC) is an electrically large conducting cavity/chamber with one or two conducting mode stirrers to stir the field modes inside the chamber. Unlike an anechoic chamber (actually just the opposite), the boundary of the chamber is ideally a perfect reflector, and the cost of such a facility is not expensive. The field in the equipment under test (EUT) area is random, but the averaged field is relatively uniform. There are quite a few important parameters for a reverberation chamber. The lowest usable frequency (LUF) is one of them and is mainly determined by the chamber size and stirrer design. For a given chamber and a fixed EUT area, the lower the LUF is, the better the chamber is. If the operational frequency is below the LUF, the field in the EUT area is not random enough to generate the desired (averaged) uniform field; thus, all measurements should be conducted above the LUF inside the reverberation chamber – this is a limitation of the reverberation chamber. More detailed information can be found from IEC Standard 61000-4-21.

DOI 10.1007/978-981-4560-75-7_71-1

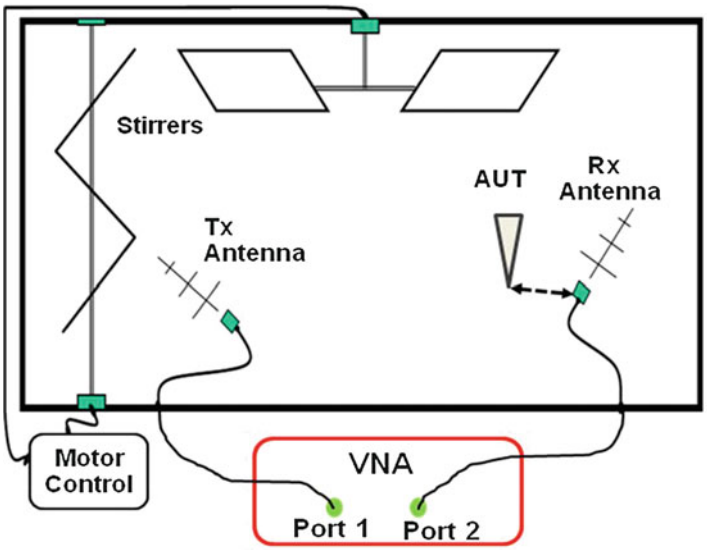


Fig. 11 A typical setup for antenna efficiency measurement inside a reverberation chamber

The reverberation chamber was originally introduced for electromagnetic compatibility (EMC) tests and measurements and is now emerging as a new facility for antenna measurements and characterization. It is particularly useful for the antenna efficiency measurement and MIMO (multiple-input and multiple-output) performance characterization using its multipath environment features (Rosengren and Kildal 2005; Fur et al. 2009; Holloway et al. 2012). For antenna efficiency measurement in a RC, it was first introduced using either a reference antenna with known efficiency or two identical antennas with the same efficiency – this could be a problem in some cases although the measurement is straightforward. More recently, the techniques for determining the antenna efficiency without the need for a reference antenna or two identical antennas have been introduced, and good measurement results have been obtained (Holloway et al. 2012). Thus, the methods with and without a reference antenna will be discussed.

Reverberation Chamber Method with a Reference Antenna The process of using a RC for antenna efficiency measurement is relatively straightforward. In addition to the RC, the equipment required is a transmitter and a receiver/spectrum analyzer or a network analyzer. As shown in Fig. 11, the power transmitted from the source Port 1 (say P_{Source}) will first encounter the reflection by the fixed transmitting antenna (due to impedance mismatch) which has a voltage reflection coefficient S_{11} and radiation efficiency η_F. Thus, the power radiated into the chamber is

$$P_{\text{Chamber}} = P_{\text{Source}}\left(1 - |S_{11}|^2\right) \cdot \eta_F \tag{16}$$

Part of this power arrives at the receiving antenna (which is either the reference antenna or the AUT), and it can be expressed as

$$P_{\text{in}} = P_{\text{Chamber}} - P_{\text{Loss}} \tag{17}$$

where P_{Loss} is the power loss due to the chamber and stirrer ohmic loss and power spreading.

The average power at the receiving antenna is assumed to be the same for both REF (reference) antenna and the AUT (a reasonable assumption for the averaged power in the EUT area), and this power is partially

DOI 10.1007/978-981-4560-75-7_71-1

consumed by the antenna (determined by the antenna efficiency) and partially reflected (due to mismatch between the antenna and cable connector); thus, the average power at the output of the receiving antenna is

$$\langle P_{\mathrm{Rec}} \rangle = P_{\mathrm{in}} \cdot \eta_{\mathrm{Rec}} \cdot \left(1 - |\langle S_{22\mathrm{Rec}} \rangle|^2\right) \tag{18}$$

The symbol $< >$ represents an ensemble average over a rotation cycle of the stirrer (typically 360°). $|\langle S_{22\mathrm{Rec}} \rangle|$ in a well-stirred chamber should be about the same as $|S_{22\mathrm{Rec}}|$ in free space (but smaller than $\langle |S_{22\mathrm{Rec}}| \rangle$). Thus, for the two different receiving antennas, the average power transmission coefficients are

$$\left\langle |S_{21\mathrm{REF}}|^2 \right\rangle = \frac{\langle P_{\mathrm{REF}} \rangle}{P_{\mathrm{Source}}} = \frac{P_{\mathrm{in}} \cdot \eta_{\mathrm{REF}} \cdot \left(1 - |\langle S_{22\mathrm{REF}} \rangle|^2\right)}{P_{\mathrm{Source}}} \tag{19}$$

and

$$\left\langle |S_{21\mathrm{AUT}}|^2 \right\rangle = \frac{\langle P_{\mathrm{AUT}} \rangle}{P_{\mathrm{Source}}} = \frac{P_{\mathrm{in}} \cdot \eta_{\mathrm{AUT}} \cdot \left(1 - |\langle S_{22\mathrm{AUT}} \rangle|^2\right)}{P_{\mathrm{Source}}} \tag{20}$$

Equations 19 or 20 to give

$$\eta_{\mathrm{AUT}} = \frac{\left\langle |S_{21\mathrm{AUT}}|^2 \right\rangle}{\left\langle |S_{21\mathrm{REF}}|^2 \right\rangle} \cdot \frac{\left(1 - |\langle S_{22\mathrm{REF}} \rangle|^2\right)}{\left(1 - |\langle S_{22\mathrm{AUT}} \rangle|^2\right)} \cdot \eta_{\mathrm{REF}} \tag{21}$$

This means that, in order to calculate the radiation efficiency of the AUT, one needs to measure/obtain:

- The S-parameters ($S_{21\mathrm{REF}}$, $S_{22\mathrm{REF}}$, $S_{21\mathrm{AUT}}$, and $S_{22\mathrm{AUT}}$) when the receiving port is connected to the reference antenna and the AUT, respectively
- The radiation efficiency of the reference antenna η_{REF}

From Eq. 21, it can be seen that the measured radiation efficiency has nothing to do with the transmitting antenna performance as long as enough power is radiated into the chamber. A calibration should be conducted to remove the cable loss. It should be pointed out that the reflection coefficient S_{22} is normally not sensitive to the stirrer position when the antenna is placed in the EUT area, but S_{21} is; thus, the S_{21} value in Eq. 21 should be the averaged S_{21} over a cycle of the stirrer rotation. It is very important to ensure that the received power is well above the noise level (Boyes et al. 2013).

Furthermore, the measurement accuracy is linked to the effectiveness of the stirrer and the resolution of the stirrer rotation. The higher the resolution is (or the smaller the step is), the better the accuracy is. But a smaller step means a longer measurement time. Thus, there is a trade-off. Figure 12 is an example of the simulated and measured radiation efficiency of a copper-based textile antenna in free space and on a body chest position. It can be seen that the difference between the measured and simulated results is typically less than 10 %, and the repeatability from one measurement to anther is very high. Another important feature of the method is that it can accommodate large antennas and supporting objects such as the human body, cars, and tanks – this is something that cap-based methods cannot offer since the caps are

DOI 10.1007/978-981-4560-75-7_71-1

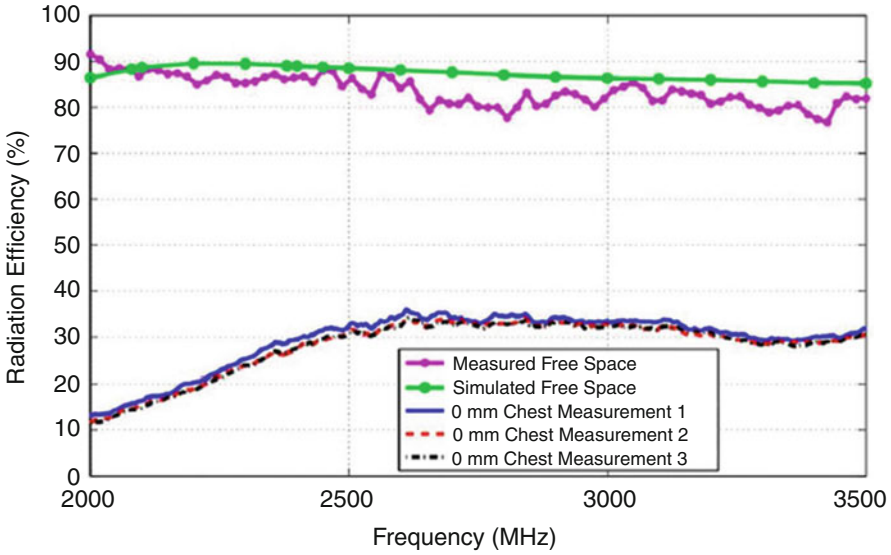


Fig. 12 The radiation efficiency of a copper-based textile antenna in free space and on-body (Boyes et al. 2013)

normally too small to host a human being, while a reverberation chamber is normally a large over-mode environment.

Reverberation Chamber Method Without a Reference Antenna The reference antenna efficiency required for Eq. 21 is sometimes not available in practice; thus, the RC method with a reference antenna cannot be used for determining the antenna efficiency. Recently, the RC method without a reference antenna method was proposed: three different approaches (one-antenna, two-antenna, and three-antenna approaches) were presented for determining the radiation and total efficiency of antennas (Holloway et al. 2012).

For the one-antenna approach, the measurement setup is similar to Fig. 11, but we only need to use the AUT and one port of the VNA. Other antennas are not needed; thus, it is simpler. It was shown (Holloway et al. 2012) that the total efficiency of the AUT can be computed using

$$\eta_{\text{Total}} = \sqrt{\frac{C_{RC}}{2\omega}\frac{\left\langle |S_{11s}|^2 \right\rangle_{RC}}{\tau_{RC}}} \tag{22}$$

where

ω is the angular frequency

$C_{RC} = \frac{16\pi^2 V}{\lambda^3}$ is a chamber constant and V is the volume of the chamber

τ_{RC} is the chamber time constant ($= Q/\omega$, Q is the chamber quality factor)

$\langle |S_{11s}|^2 \rangle$ represents the stirred energy contribution of the AUT S_{11} inside the RC

Thus, a measurement of $\langle |S_{11s}|^2 \rangle$ and τ_{RC} is all that is required to obtain the total radiation efficiency of a single antenna. If radiation efficiency is required, it can be obtained using

$$\eta = \frac{\eta_{\text{Total}}}{1 - |S_{11}|^2} = \sqrt{\frac{C_{RC}}{2\omega}\frac{\left\langle |S_{11s}|^2 \right\rangle}{\tau_{RC}}} \Big/ \left(1 - |S_{11}|^2\right) \tag{23}$$

DOI 10.1007/978-981-4560-75-7_71-1

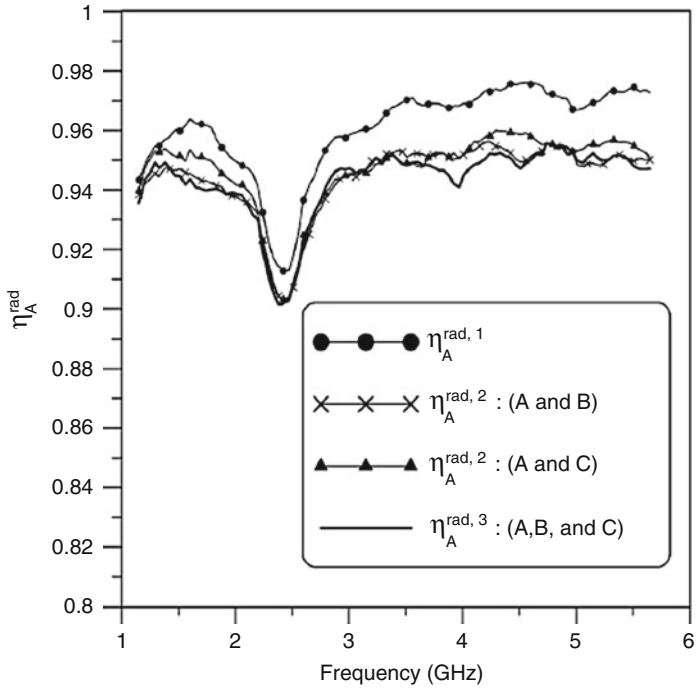


Fig. 13 Radiation efficiency: a comparison of three different approaches for antenna A (Holloway et al. 2012)

The measurement procedures for two-antenna approach and three-antenna approach are very similar. They require the measurements of $\langle |S_{11s}|^2 \rangle$ and τ_{RC} as well. The main differences are in the assumptions about the relationship between the power reflected into the antenna and the power received (Holloway et al. 2012). The three-antenna approach is more general and does not require such assumptions. The measurement results are very similar for all three approaches as we can see from the example in Fig. 13. The two-antenna approach seems to give the smallest uncertainty.

Compared with the RC method with a reference antenna, this method does not require a reference antenna with known radiation efficiency. But one has to calibrate the chamber carefully in the time domain to obtain the time constant – this is a time-consuming exercise. The good thing is that one just needs to calibrate the chamber once if the chamber performance/structure/loading is not changed, and the data can be used repeatedly.

Other Efficiency Measurement Methods for Small Antennas

There are many other methods and techniques developed over the years, but they are less popular or convenient for daily applications.

Radiometric Method

This is a classical measurement method (Ashkenazy et al. 1985); a low-noise receiver like a radiometer is required to measure the output noise power of the AUT in two different environments, typically a "warm target" (such as an anechoic chamber) and a "cold target" (such as the clear sky). A third measurement is required using a lossless standard antenna with a well-known antenna radiation pattern or a calibrated noise source. The antenna radiation efficiency can then be obtained using the results from these three measurements.

The problems of this method are the special equipment needed and the errors that may be caused by the inaccuracy of the noise sources, the mismatch of the AUT, and the drift of the receiver gain (Kraus and Marhefka 2002).

DOI 10.1007/978-981-4560-75-7_71-1

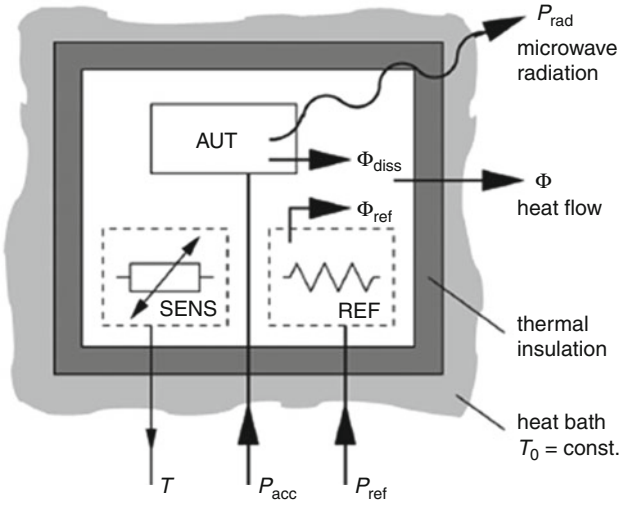


Fig. 14 The calorimetric method (Schroeder and Gapski 2006)

The Q-Factor Method

This technique is based on a comparison of the measured Q to ideal Q (Newman et al. 1975). The Q-factor is determined by energy stored, power radiated, and power dissipated. This method is also just suitable for electrically small antennas, and it is not as accurate and convenient as the Wheeler cap method; thus, it is not widely used.

Colorimetric Method

This interesting approach was proposed in 2006 (Schroeder and Gapski 2006). The operational principle is shown in Fig. 14, and it is based on measurement of antenna dissipation in a heat flow comparator. The most interesting feature is that neither a reference antenna nor site calibration is required. The measurement accuracy is high and suitable for wideband and small antennas. A main drawback seems to be the construction of the heat flow comparator. The measurement equipment and setup are not the standard in a typical antenna laboratory which may be an obstacle for becoming a main antenna efficiency measurement method.

Discussion and Comparison

The radiation efficiency is one of the most important parameters of an antenna, and many measurement methods have been developed. But so far, the "best method" has not been found: they all have their strengths and drawbacks. The search for the best antenna radiation efficiency measurement is still a task to be accomplished. Different methods are used in practice. The comparison of the measurement methods, especially about the accuracy and uncertainty based on published papers, is given in Table 1. The major features and drawbacks of each method are also provided. The provided accuracy limits and uncertainties are for a typical high-efficiency antenna case (for a low-efficiency case, 10 % would be quite significant). In practice, there are many other parameters, and considerations have to be taken into account, but not included here. For example, the Wheeler cap method seems to offer the best accuracy for electrically small antenna measurement, provided that the antenna impedance can be accurately measured – as it was mentioned earlier, this could be a problem for some antennas. If the antenna impedance was not measured accurately, the radiation efficiency error would be >3 %.

DOI 10.1007/978-981-4560-75-7_71-1

Table 1 Comparison of measurement methods

Measurement method	Accuracy (%)	Note
Pattern integration	<20	Suitable for most antennas. 3D radiation measurement required, time-consuming, and poor accuracy
Directivity/gain	<20	Suitable for a wide range of antennas, in particular directional antennas
Wheeler cap	<3	Suitable for electrically small antennas, cost-effective, quick, accurate, and little data processing
Sliding wall cavity	<5	Suitable for many small antennas, cost-effective. A special cavity required
UWB Wheeler cap	<10	Suitable for wideband and small antennas, cost-effective, quick, and simple data processing
Source-stirred cap/chamber	<5	Suitable for a wide range of antennas, cost-effective, quick, and little data processing
RC with a reference antenna	<10	Suitable for both large and small antennas and systems. A reference antenna with known radiation efficiency required, time-consuming, and the frequency >LUF
RC without a reference antenna	<10	Suitable for both large and small antennas and no reference antenna required. A lengthy chamber calibration required and the frequency >LUF
Radiometric	<15	Suitable for many antennas. Special equipment required
Q-factor	<10	Suitable for electrically small antennas, not as simple as the Wheeler cap method
Calorimetric	<5	Suitable for small antennas and systems (e.g., a mobile phone). A special heat flow comparator required

Conclusion

In this chapter, various methods used for measuring antenna radiation efficiency have been discussed. The classical pattern integration and directivity/gain methods are general, but the measurement could be time-consuming, and the measurement accuracy is normally poor. The Wheeler cap method has got many excellent features but is only suitable for electrically small antennas. A number of modified Wheeler cap methods have been discussed in details. The UWB Wheeler cap method and the source-stirred cavity/chamber method are gaining popularity for small antenna measurements. The latter may offer more accurate results for some antennas since it can better control the modes inside the cavity/chamber. The RC methods have received a lot of attention recently, and significant developments have been made. They are suitable for a wide range of antennas, large or small, even installed antennas, which is a very attractive feature of the RC. The application of the RC for diversity and MIMO antenna measurements is another interesting development of this useful facility.

Acknowledgment

The help from Mr Qian Xu, Dr Stephen Boyes, and Dr Yang Lu is gratefully acknowledged.

Cross-References

► Antenna Measurement/Setups-Introduction
► EMI/EMC Chamber Design, Measurement and Instrument
► Evaluation of Wearable and Implantable Antennas with Human Phantoms

DOI 10.1007/978-981-4560-75-7_71-1

► Near-Field Antenna Measurement Techniques
► Small Antennas (PIFA/PILA/Loading Antenna/etc)
► Ultra-wideband Antennas

References

Ashkenazy J, Levine E, Treves D (1985) Radiometric measurement of antenna efficiency. Electron Lett 21:111–112

Boyes SJ, Soh PJ, Huang Y, Vandenbosch GAE, Khiabani N (2013) Measurement and performance of textile antenna efficiency on a human body in a reverberation chamber. IEEE Trans Antennas Propag 61(2):871–881

Cho C, Kang J-S, Choo H (2014) Improved Wheeler cap method based on an equivalent high-order circuit model. IEEE Trans Antennas Propag 62:274–281

Fur GL, Lemoine C, Besnier P (2009) Performances of UWB Wheeler cap and reverberation chamber to carry out efficiency measurements of narrow band antennas. IEEE Antennas Wirel Propag Lett 8:332–335

Holloway C, Shah H et al (2012) Reverberation chamber techniques for determining the radiation and total efficiency of antennas. IEEE Trans Antennas Propag 60(4):1758–1770

Huang Y, Boyle K (2008) Antennas: from theory to practice. Wiley, Chichester

Huang Y, Edwards DJ (1992) A novel reverberating chamber: source-stirred chamber. In: Proceedings of IEEE 8th international conference on electromagnetic compatibility, Edinburgh, Sept 1992, pp 120–124

Huang Y, Narayanan R, Kadambi G (2003) Electromagnetic coupling effects on the cavity measurement of antenna efficiency. IEEE Trans Antennas Propag 51:3064–3071

Huang Y, Lu Y, Chattha H (2010) Antenna efficiency measurements. In: IEEE iWAT 2010, Lisbon, Mar 2010

Huang Y, Lu Y, Boyes S (2011) Source-stirred method for antenna efficiency measurements. In: EuCAP 2011, Rome

Huynh M (2004) Wideband compact antennas for wireless communication applications. PhD thesis, Virginia Polytecnic Institute and State University

Johnson RH, McRory JG (1998) An improved small antenna radiation efficiency measurement method. IEEE Antenna Propag Mag 40:40–48

Kraus JD, Marhefka R (2002) Antennas: for all applications, 3rd edn. McGraw-Hill, New York

Lu Y, Huang Y, Chattha H (2008) Effects of Wheeler cap loss on wideband antenna efficiency measurement results. In: IEEE iWAT 2008, Chiba, Japan

McKinzie III W (1997) A modified Wheeler cap method for measuring antenna efficiency. In: IEEE antenna and propagation symposium, July 1997, pp 542–545

Newman EH, Bohley P, Walter CH (1975) Two methods for the measurement of antenna efficiency. IEEE Trans Antennas Propag 23:457–461

Pozar DM, Kaufman B (1988) Comparison of three methods for the measurement of printed antenna efficiency. IEEE Trans Antennas Propag 36:136–139

DOI 10.1007/978-981-4560-75-7_71-1

Rosengren K, Kildal P-S (2005) Radiation efficiency, correlation, diversity gain and capacity of a six-monopole antenna array for MIMO system: theory, simulation and measurement in reverberation chamber. IEEE Proc Microwave Antennas Propag 152:7–16

Schantz HG (2002) Radiation efficiency of UWB antennas. In: Proceedings of the IEEE UWBST conference, Baltimore, USA

Schroeder W, Gapski D (2006) Direct measurement of small antenna radiation efficiency by calorimetric method. IEEE Trans Antennas Propag 54:2646–2656

Wheeler HA (1959) The radiansphere around a small antenna. Proc IRE 47:1325–1331

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_76-2

Mm-Wave Sub-mm-Wave Antenna Measurement

Heiko Gulan[a]*, Cyril Luxey[b] and Diane Titz[c,b]
[a]Institut für Hochfrequenztechnik und Elektronik (IHE), Karlsruher Institut für Technologie, Karlsruhe, Germany
[b]Laboratoire Electronique pour Objets Connectés, Université Nice Sophia Antipolis, Sophia Antipolis, France
[c]Lycée Jules Ferry, Cannes, France

Abstract

This chapter deals with mm-wave and sub-mm-wave probe-fed antenna measurements. First, worldwide state-of-the-art measurement solutions are presented. Then, the most interesting solutions to achieve the RF and mechanical parts of an mm-wave probe-fed measurement setup are described. In a second subpart, the focus lies on possible measurements and associated techniques to perform accurate characterization of probe-fed mm-wave antennas. Special attention is given on reflection coefficient measurements, gain measurements and associated calibration techniques, novel phaseless techniques developed for the characterization of circularly polarized antennas, hybrid total efficiency computations from simulation and measurements, near-field and phase measurements, and possible source of errors taking place in those mm-wave probe-fed measurement setups. A last subpart deals with measurements above 110 GHz as today no coaxial connection exists, and therefore rigid waveguide connections have to be employed at those frequencies. Lastly, open problems are discussed, and future directions for the coming years are proposed.

Keywords

Mm-wave measurement system; Probe-fed antenna; Microelectronic probe; Gain measurement; Calibration technique; Circularly polarized mm-wave antenna; Total efficiency; Near-field measurement; Phase measurement

Introduction

CMOS integrated circuits have already proven to be efficient at mm-wave frequencies. In parallel, there is today a strong interest for several license-free frequency bands in the millimeter-wave spectrum. For example, at 60 GHz, WiGig and Wi-Fi Alliance united in 2013. By 2015, it is expected that products operating in the 60-GHz frequency band will deliver multi-gigabit speeds, low latency, and security-protected connectivity between nearby devices (Wi-Fi). Other frequency bands that are strongly target are 77 GHz for automotive radar system and 94 GHz for imaging radar sensors.

Consequently, there is a strong need of mmW and sub-mmW measurement setups for antenna characterization. It is obvious that as the frequency increases, the size of the antenna decreases, and new problems are encountered with their characterization. In parallel, as the frequency increases, the interconnection losses also increase. Therefore, the antenna needs to be as close as possible to the circuitry. Recent developed solutions have proved that antenna-in-package (AiP) might be an efficient solution (Titz et al. 2012d; Pilard et al. 2012). Measurements in anechoic chambers using connector-fed

*Email: heiko.gulan@kit.edu

DOI 10.1007/978-981-4560-75-7_76-2

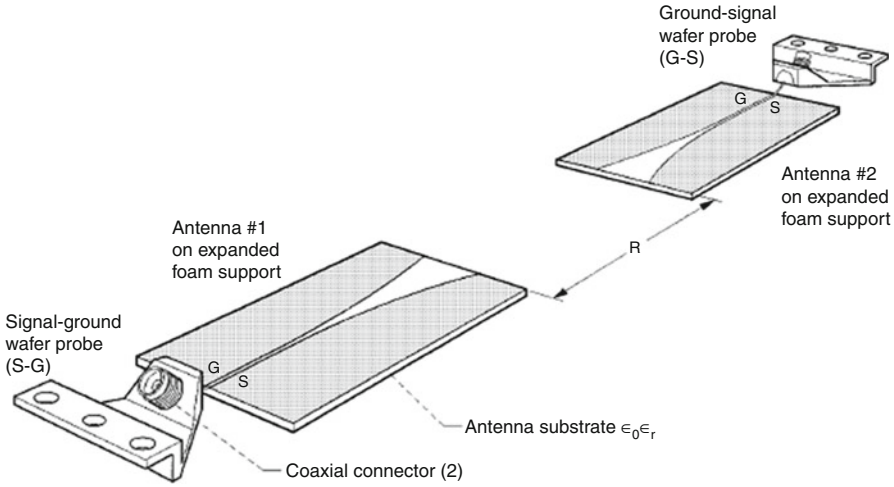


Fig. 1 Gain measurement of two identical antennas on a probe station proposed in Simons and Lee (1999)

antennas are still possible up to 110 GHz. But, above 30 GHz, the wavelength decreases below 10 mm, and the connector begins to be as big as or bigger than the antenna itself. Therefore, it can radiate more than the antenna itself and strongly disturb the radiation pattern from the antenna that wants to be characterized (Ranvier et al. 2009). Moreover, these connectors are difficult to solder to such small antennas. The other solution that can be used even above 110 GHz is to feed the antenna with a waveguide (Lamminen et al. 2008). However, this last solution is not possible or adequate for all topologies of antennas. The last possible solution already used for microelectronic circuit measurement is to feed the antenna under test with a microelectronic probe. This is particularly important for integrated antennas or even AiP as they will be fed exactly where the IC will be connected in the final application. It has been shown that even if it will slightly disturb the radiation pattern, its influence is less noticeable than a connector-fed solution at 60 GHz (Ranvier et al. 2009).

Therefore, in this chapter, the focus will be on probe-fed antenna measurements at mmW and sub-mmW frequencies. After a rapid overview of the worldwide existing setups, all the RF and mechanical parts needed in a probe-fed measurement setup are described. Finally, several measurement procedures are explained and referred to existing setups. A last subpart deals with measurements above 110 GHz, and lastly, open problems are discussed, and future directions for the coming years are proposed.

Building a Probe-Fed Measurement Setup

Overview of Worldwide Measurement Setups

The first on-wafer antenna measurements are reported in 1997 and 1999 by Simons and Lee (1997, 1999). In this setup, the metallic chuck has been replaced by a block of foam to avoid unwanted metallic reflections, and all the surrounding unavoidable metallic parts have been covered with microwave absorbers. Two identical antennas are placed at a distance R from each other, and the gain is computed from the measured S_{21} using the Friis equation. The antenna-to-antenna distance is chosen to be larger than the far-field distance. Using a calibration substrate, the feeding probe of the antenna under test (AUT) is calibrated to the tip of the probe. Measurements are done with an HP 8510C automatic network analyzer (Fig. 1).

DOI 10.1007/978-981-4560-75-7_76-2

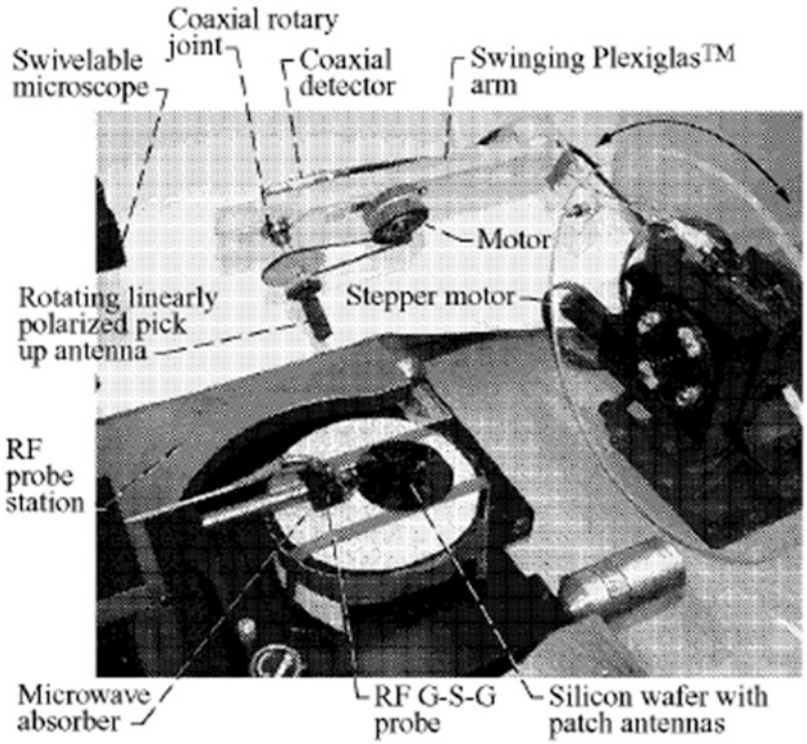


Fig. 2 Radiation pattern measurement of an antenna using a customized probe station proposed in Simons (2002)

However, with this first attempt, the position of the two antennas is fixed, and the gain in the end-fire direction can be measured only. In Simons et al. (2001) and Simons (2002), the classical RF probe station was modified to enhance the measurement possibilities. A small horn antenna or an open waveguide is attached to a Plexiglas arm which can be rotated (using a motor) to acquire the field intensity along an arc. The antenna is still fed using a microelectronic probe, and all the metallic parts remain covered by absorbers. Especially, the wafer where the antenna lies is separated from the metallic chuck by an absorber. Also, the acquisition waveguide can be rotated by 90° to measure co- and cross-polarization gains (Fig. 2).

This setup has still limitations as only the upper part of the sphere around the AUT can be measured because of the blockage of the chuck. The first entirely customized measurement setup for mmW antennas was proposed in Zwick et al. (2004). To avoid unwanted reflections from all the metallic parts of a probe station, the probe positioner was moved out of the probe station and placed in a small anechoic chamber on an isolation table. The AUT is held by a custom holder, and a standard gain horn can be rotated all around it, thanks to a motorized arm. The axis of the motor can be mounted in three different ways so that the radiation patterns can be measured in three different planes. The horn is set at 38 cm from the AUT to ensure a suitable condition for far-field operation. With this setup, S_{11} gain in one direction and in three different planes can be measured using a vector network analyzer (VNA). Special calibration procedures are detailed later in this chapter. For the first time, the radiation of the probe itself is investigated. A complete overview of this setup can be found in Zwick and Pfeiffer (2009; Figs. 3 and 4).

The first quasi-3-D radiation pattern measurement setup was proposed in Ranvier et al. (2009). Using two rotating arms, the gain of the AUT can be measured in all directions with the exception of the region which is blocked by the probe positioner. A very similar setup was developed in parallel and presented in Beer and Zwick (2010; Figs. 5 and 6).

DOI 10.1007/978-981-4560-75-7_76-2

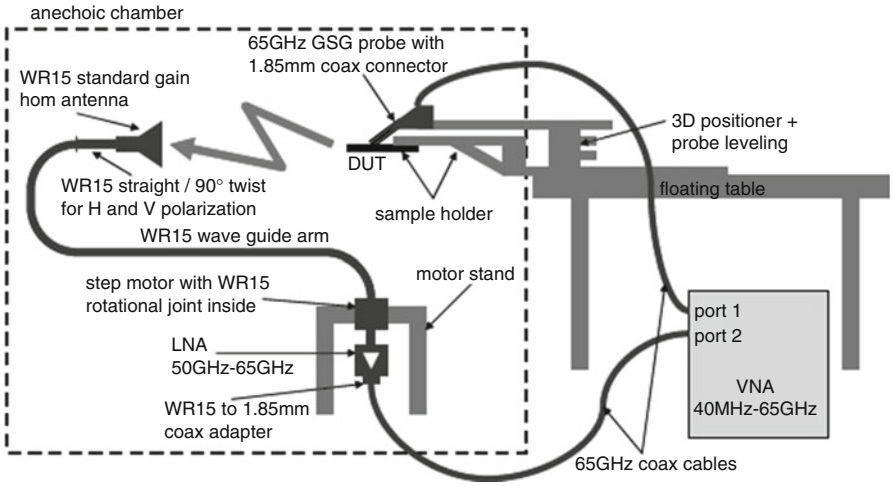


Fig. 3 Probe-fed measurement setup diagram proposed in Zwick et al. (2004)

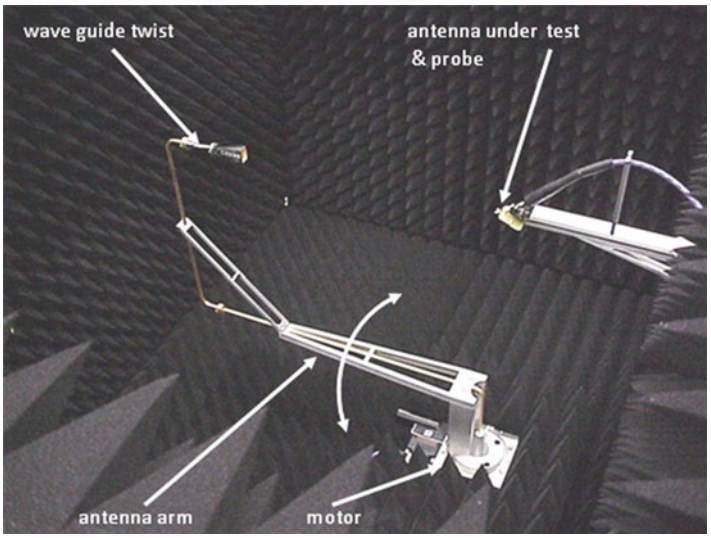


Fig. 4 Probe-fed measurement setup picture proposed in Zwick et al. (2004)

Several other measurement setups have been recently developed worldwide. Most of them have been listed and compared in Table 1. Most of them adopted a probe feeding technique for the AUT. Depending on the implemented RF parts, they are able to measure both S_{11} and gain radiation pattern or, at least, one of them. Depending on the mechanical parts, they are able to measure almost 3-D pattern or only plane cuts.

RF Parts

The RF equipment needed for mmW antenna measurements depends on the measurement which wants to be carried out. Basically, the measurements can be divided in two categories: reflection coefficient (S_{11}) and gain/radiation pattern measurements. As mentioned, the best way to feed an mmW antenna without disturbing it, especially if it is integrated either in a package or a chip, is to use a microelectronic probe. Of course, if the antenna is deported and is connected by a cable or waveguide in the final application, the

DOI 10.1007/978-981-4560-75-7_76-2

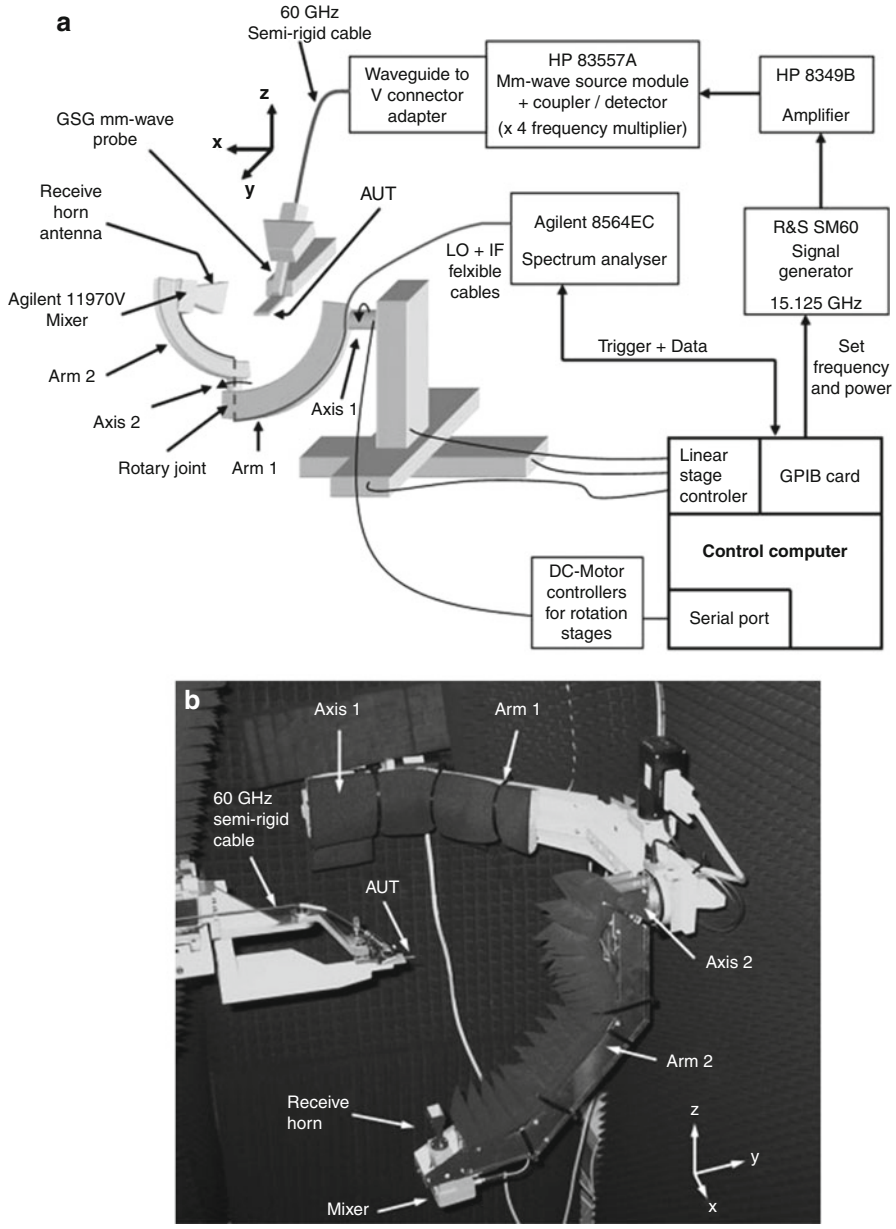


Fig. 5 3-D probe-fed measurement setup diagram (**a**) and picture (**b**) proposed in Ranvier et al. (2009)

probe is obviously not the best option. However, this tends to be less and less true as the operating frequency increases. So, this chapter mainly focuses on probe-fed measurements.

The selection of the microelectronic probe to be used depends on the frequency range and the pitch of the feeding pad of the AUT. Table 2 lists the available microelectronic probes with various pitches and associated losses. Table 3 lists some technical characteristics of the main commercial probes. As the size of the tip of the probe is very small and needs to be aligned on the antenna pads with precision, a microscope and a probe positioner are necessary. Indeed, microelectronic probes are very fragile and expensive, and they especially need to be treated with precaution.

For reflection coefficient measurement, a vector network analyzer (VNA) or PNA is needed. Using a reference calibration substrate (supplied by the probe manufacturer) and a proper calibration method (Fakharzadeh 2014), the reference plane of the measurement can be set at the tip of the probe. The most

DOI 10.1007/978-981-4560-75-7_76-2

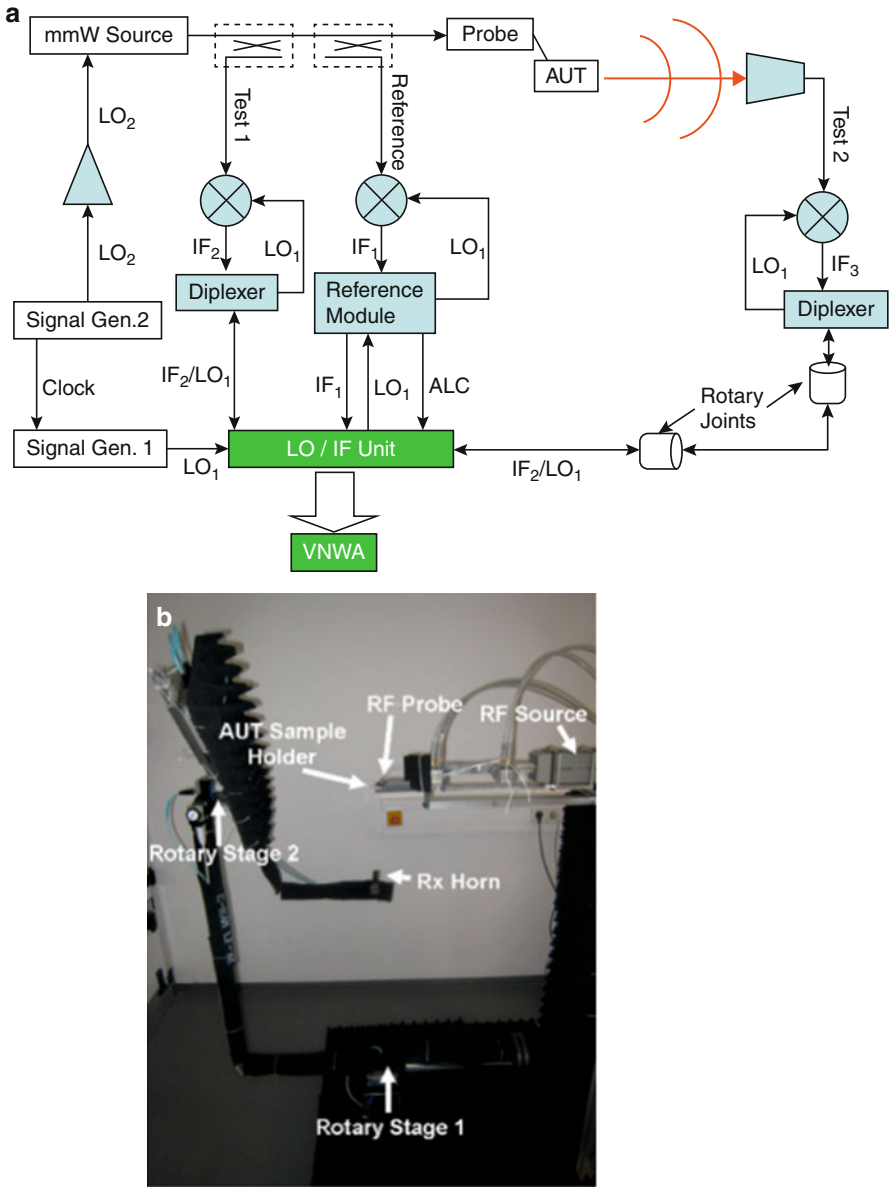


Fig. 6 3-D probe-fed measurement setup diagram (**a**) and picture (**b**) proposed in Beer and Zwick (2010)

commonly used method is a two-port method called the short-open-load-thru (SOLT) calibration. However, this method requires two probes and two positioners and thus is really difficult to achieve apart from a typical probe station. For antenna measurement, a one-port calibration is sufficient. Moreover, it can be done easily by using a short-open-load (SOL) method (Beer and Zwick 2010).

For gain/radiation pattern measurement, the RF equipment can be more varied. Gain and radiation patterns can be indeed measured separately. Gain measurement needs accurate RF equipment and calibration procedure, while radiation pattern can be normalized to the maximum received power. In each case, a signal needs to be generated at the right frequency and received adequately. It has been considered here that the AUT is the transmitting antenna, and a standard horn antenna is used at the receive side.

The signal can be generated by a VNA (Titz et al. 2012b). In this case, both S_{11} and gain measurements can be done with the same RF equipment. The visualization of the S_{11} before any gain measurement is also

Table 1 Selection and comparison of worldwide measurement setups for millimeter-wave antennas

University or company	Frequency range	Probe-fed	Antenna holder	AUT distance	S_{11} measurement	Gain measurement	RF equipment	Reference/ publication
NASA Lewis Research Center	27–29 GHz	Yes	Foam block	Adjustable	No	End-fire direction	ANA	Simons and Lee 1999
NASA Glenn Research Center	18–40 GHz	Yes	Chuck + microwave absorber	NA	Yes	Two planes −90° to +90°	ANA	Simons 2002
IBM T.J. Watson Research Center	50–65 GHz	Yes	Custom holder in FR4	38 cm	Yes	Three planes −90° to +90°	VNA	Zwick et al. 2004
Eindhoven University of Technology	40–67 GHz	Yes	Perspex table + polystyrene layer	~030 cm	Yes	Upper hemisphere	VNA	Akkermans et al. 2007
Carleton University, Ottawa	24 GHz	Yes	Probe station	40 cm	No	Broadside direction	VNA	Shamim et al. 2008
VTT	50–65 GHz	No (waveguide)	NA	77.2 cm	No	Full 3-D	Signal generator + amplifier + multiplier + mixer + spectrum analyzer	Lamminen et al. 2008
University of Michigan, Ann Arbor	2–40 GHz	Yes	Cavity filled of resonance absorber	Max 1.5 m	Yes	Upper hemisphere	VNA	Van Caekenberghe et al. 2008
Georgia Institute of Technology	55–65 GHz	Yes	Styrofoam shelf	57 cm	Yes	Quasi-3-D	PNA	Pan et al. 2008
National Cheng Kung University	55–65 GHz	Yes	Chuck + plastic box	Adjustable	No	End-fire direction	PNA	Hsu et al. 2008
Université Nice Sophia Antipolis	Up to 140 GHz	Yes	Foam holder	20 cm	Yes	Quasi-3-D	VNA	Ranvier et al. 2009
Purdue University	40 GHz	Yes	Aluminum + unclad FR4	NA	Yes	Upper hemisphere	VNA	Willmot et al. 2009
Uppsala University	70–90 GHz	Yes	Rigid foam	45 cm	Yes	Two planes −90° to +90°/54°	PNA + extenders	Cheng et al. 2009

(*continued*)

Table 1 (continued)

University or company	Frequency range	Probe-fed	Antenna holder	AUT distance	S_{11} measurement	Gain measurement	RF equipment	Reference/ publication
STMicroelectronics	40–67 GHz	Yes	Rexolite chuck	NA	Yes	Upper hemisphere	PNA	PIlard et al. 2009
Toshiba Corporate Research & Development Center	60 GHz	Yes	Teflon table	>23 cm	No	Three planes −90° to +90°	Signal generator + mixer + spectrum analyzer	Ito et al. 2009
National Cheng Kung University	55–65 GHz	Yes	Chuck	NA	No	One plane −90° to +90°	Signal generator + mixer + spectrum analyzer	Tsai et al. 2010
Karlsruhe Institute of Technology	Up to 170 GHz	Yes	Foam	60 cm	Yes	Quasi-3-D	VNA + extenders	Beer and Zwick 2010
CEA LETI	60–65 GHz	Yes	Chuck	40.3 cm	Yes	Upper hemisphere	VNA + extenders	Lanteri et al. 2010
IMEC	50–67 GHz	Yes	Hole in a metal table	50 cm	Yes	Lower hemisphere	VNA	Mohammadpour-Aghdam et al. 2011
Lab-STICC	50v67 GHz	Yes	Foam	5 cm	Yes	Upper hemisphere	VNA	Calvez et al. 2011
National University of Singapore	50–67 GHz	Yes	Absorbers	15 cm	Yes	Upper hemisphere	VNA	Sun et al. 2011
Institute for Infocomm Research	57–67 GHz	Yes	Low-permittivity plastic holder	80 cm	Yes	Quasi-3-D	VNA	Yeap et al. 2011

Digital Media and Communications Research Center, Samsung Electronics	57–66 GHz	Yes	Hollowed chuck	NA	Yes	Upper hemisphere	VNA	Hong et al. 2011
Eindhoven University of Technology	62.5 GHz	Yes	NA	NA	Yes	Upper hemisphere	VNA	Johannsen et al. 2012
City University of Hong Kong	60–75 GHz	No (connector)	NA	50 cm	Yes	Two planes −72° to + 72°	Signal generator + mmW source module + mixer + spectrum analyzer	Ng et al. 2012
Institute of Microelectronics, A*STAR	110–170 GHz	Yes	Chuck	10 cm	No	Broadside direction	Signal generator + extender + mixer + spectrum analyzer	Hou et al. 2012
National Taiwan University	66–82 GHz	Yes	Chuck	2 cm min	Yes	2-D planar scan	VNA	Tsai et al. 2013
Institute for Infocomm Research	110–170 GHz	Yes or (waveguide)	Chuck with absorber	80 cm	Yes	Upper hemisphere	VNA + extenders	Xu et al. 2013
Institute for Infocomm Research	220–325 GHz	No (waveguide)	Low-permittivity plastic holder	80 cm	Yes	Upper hemisphere	VNA + extenders	Qing and Chen 2014
Chang Gung University	50–70 GHz	Yes	Special holder	60 cm	Yes	Quasi-3-D	VNA	Chin et al. 2014

DOI 10.1007/978-981-4560-75-7_76-2

Table 2 Available microelectronic probes and feeding techniques associated with insertion loss

Operating frequency	Connector type	Waveguide type	Max. insertion loss	Typical pitch
DC-40 GHz	2.92 mm	–	2 dB	25−2540 μm
DC-50 GHz	2.4 mm	–	2 dB	25−1250 μm
DC-67 GHz	1.85 mm	–	2 dB	25−1250 μm
50–75 GHz	–	WR15	2 dB	50−1250 μm
60–90 GHz	–	WR12	2 dB	50−450 μm
75–110 GHz	–	WR10	2 dB	50−450 μm
DC-110 GHz	1 mm	–	1.4 dB	50−150 μm
90–140 GHz	–	WR8	3 dB	50−100 μm
110–170 GHz	–	WR6	4 dB	50−100 μm
140–220 GHz	–	WR5	4 dB	50−100 μm
220–325 GHz	–	WR3	4 dB	50–100 μm

Referneces: (Cascade 2015; GGB 2015)

Table 3 Commercially available microelectronic probes and technical characteristics

Frequency band	W-band (75−110 GHz)	D-band (110−170 GHz)	H-band (220−325 GHz)	H-band (220−325 GHz)
Probe name	110H-GSG-150-P	170-GSG-100-BT-M	325-GSG-80-BT	I325-T-GSG-100-BT
Probe type	Air coplanar	Air coplanar	Air coplanar	Infinity
Manufacturer	GGB Industries	GGB Industries	GGB Industries	Cascade Microtech
Pitch	150 μm	100 μm	80 μm	100 μm
Connection type	1 mm Coax	WR-6 waveguide	WR-3 waveguide	WR-3 waveguide
Max. loss	1.2 dB	2.1 dB	3.5 dB	7.5 dB

quite important to verify that the probe correctly touches the antenna pads. Current VNAs can generate a signal up to 67 GHz. For frequencies above 67 GHz, mmW extenders or mmW transmit/receive (T/R) modules can be used (Bisognin et al. 2014). However, such VNAs are quite expensive, and mmW antennas can be measured as well using a low-frequency signal generator (eventually followed by an amplifier) which frequency is then multiplied by an mmW source module to obtain the desired one (Titz et al 2012b). In this case, S_{11} has to be measured separately (and differently), but the input power delivered to the AUT can be higher hence improving the dynamic range of the measurement setup. In both cases, this input power has to be known to be able to compute the gain of the AUT. Another low-cost solution using a local oscillator at the right frequency is presented in Boone et al. (2011).

At the receive side, the horn antenna can be followed by several equipment. The VNA can be used also in reception. If the horn is fixed, a cable can directly link the horn antenna to the VNA. But at mmW frequencies, this cable should be fixed without moving for an accurate measurement (Hitzler et al. 2014). So, if cuts or parts of the 3-D radiation pattern have to be measured, a better solution is to use a mixer which is attached to the horn antenna to directly down-convert the signal to low frequency. An mmW extender can also be mounted close to the horn antenna, but those extenders are usually quite heavy, and the feasibility of this option depends on the mechanical rigidity of the assembly. The mixer has to be fed by a local oscillator (LO) signal, and it will transmit back an intermediate frequency (IF) signal. These two signals can be conveyed by flexible cables as their operating frequency is much lower than mmW. Some mixers are now integrated with a diplexer, so only one cable is necessary for proper operation.

DOI 10.1007/978-981-4560-75-7_76-2

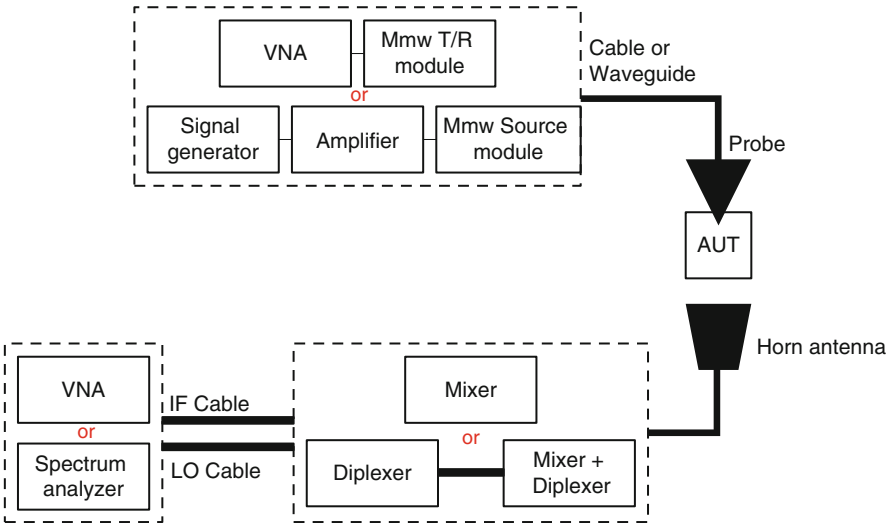


Fig. 7 RF equipment for gain measurement of an mmW antenna

The IF signal can then be received either by the VNA or by a spectrum analyzer. The LO signal needs to be generated by a signal generator. It can also be provided by a VNA or a spectrum analyzer (if they include such possibilities). A diplexer can be used before this equipment to separate the LO and IF signals if only one cable is used. If the reception is done by a VNA, another mixer has to be used for reference purpose. The dynamic range of a spectrum analyzer is usually higher than that of a VNA. So, if low-gain antennas are to be measured, a spectrum analyzer should be preferably used (Van Caekenberghe et al. 2008). Moreover, the noise floor of the spectrum analyzer can be lowered by modifying the bandwidth of the IF filter (Titz et al. 2012b). Figure 7 gives a quick sketch of the main options to measure the gain of an mmW antenna. These RF equipments are different from those used in anechoic chambers at lower frequencies (see chapter on ► Antenna Measurement/Setups-Introduction). An example of the needed RF equipment for measurements at 120 GHz is given in Fig. 8 (Bisognin et al. 2014).

Mechanical Parts

The mechanical parts are probably the most important in an mmW measurement setup. Due to the use of a microelectronic probe to feed the AUT, a fair mix and compromise have to be reached between classical anechoic chambers and microelectronic measurements.

AUT Holder

Indeed, the close environment of a probe tester is usually all metallic, composed of a chuck base, a microscope, the feeding probe(s), and the positioner. The first measurements made in this way proved that the radiation pattern was completely disturbed (Lanteri et al. 2010). Even if the authors from Murdock et al. (2011) proposed a way to remove the interference originating from a wafer probe, it is definitely safer to perform the characterization of the antenna in a very "clean" close environment. A first obvious step is to get rid of the metallic chuck as done in Zwick et al. (2004), Akkermans et al. (2007), Van Caekenberghe et al. (2008), and Pilard et al. (2009). In Zwick et al. (2004), no chuck was used, and the antenna was maintained by two substrate pieces of FR4 (flame resistant 4). In Pilard et al. (2009), the metallic chuck was replaced by a rexolite support. A better solution would be to use a material with a lower permittivity possibly close to one. Hence, the radiation pattern of the AUT would not be disturbed by its holding pieces: the chuck and the holder. In Akkermans et al. (2007), the chuck is made of polystyrene (permittivity close to one). In Van Caekenberghe et al. (2008), a cavity with absorbers is placed below

DOI 10.1007/978-981-4560-75-7_76-2

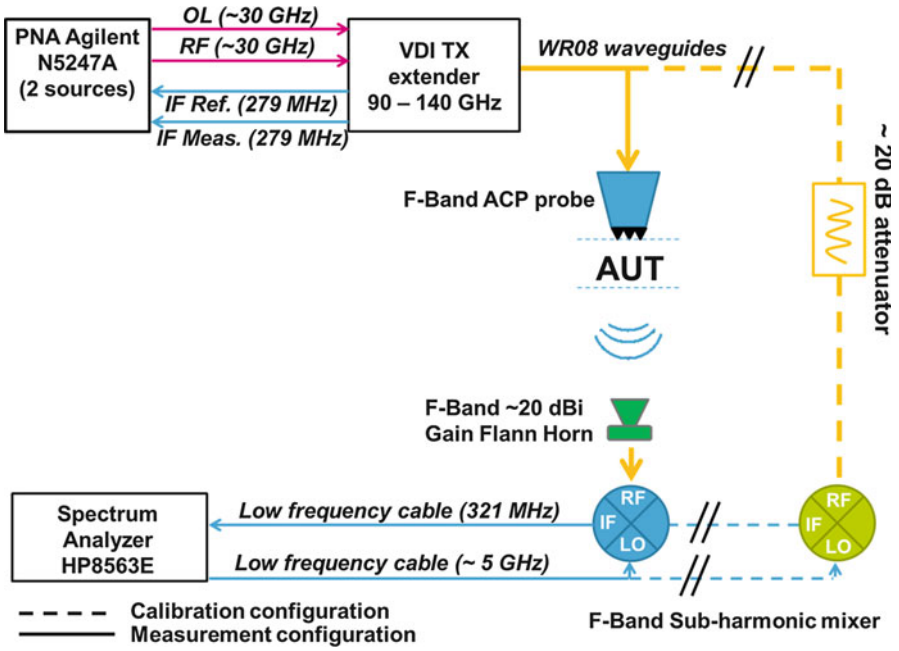


Fig. 8 90–140 GHz measurement setup – RF part from Bisognin et al. (2014)

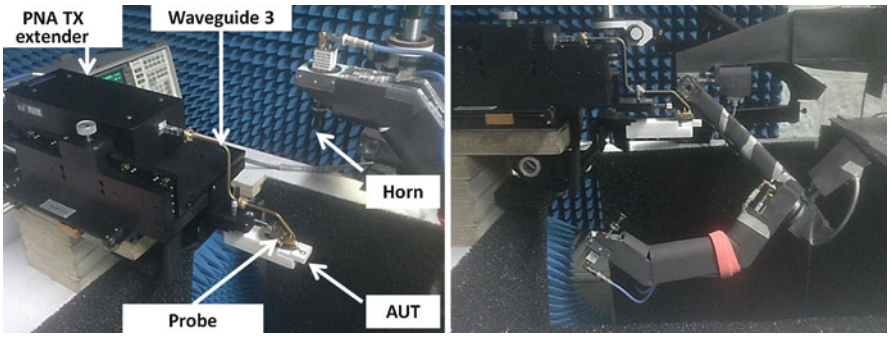


Fig. 9 Picture of the antenna holder in Bisognin et al. (2014)

the wafer. However, only the upper hemisphere of the radiation sphere is reachable by the receiving horn antenna. An even better solution would be to use a special holder made of foam. The radiation pattern below the antenna (if the antenna is fed by the probe on its top face) can thus be recorded, and the foam will not disturb the measurement at all as presented in Beer and Zwick (2010) and Titz et al. (2012a). In Bisognin et al. (2014), for each antenna, a specific foam holder is fabricated and screwed to another carrier foam. This last foam is 1 cm thick to be sufficiently rigid and horizontally planar. A cavity of the size of the AUT is created into the foam to properly maintain it. The special carrier is attached to a three-dimensional positioner, classically used in a probe station. This carrier greatly helps in leaving the space around the AUT free of any objects (Fig. 9). Of course, all remaining close metallic parts of the setup need to be covered with microwave absorbers (Reniers et al. 2014).

Horn Holder and Arms

The other important mechanical part of the setup is the receive area. A special fixture has to be realized to hold the standard horn antenna. To measure one or two cut planes, one arm “loaded” by a small horn antenna can be driven along an arc by a stepper motor as in Simons (2002) and Lanteri et al. (2010).

DOI 10.1007/978-981-4560-75-7_76-2

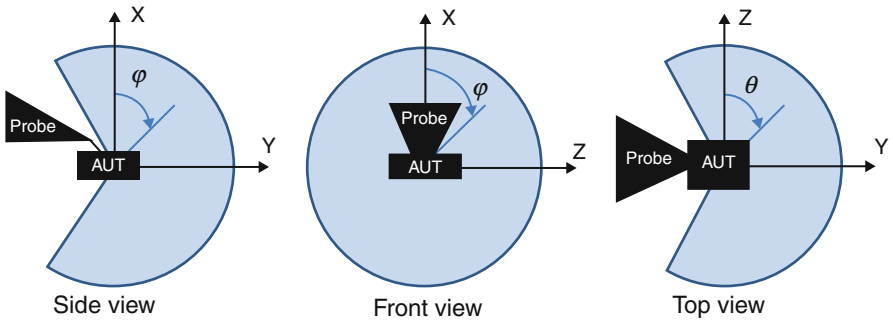


Fig. 10 Measurable parts of the sphere from the setup presented in Titz et al. (2012c)

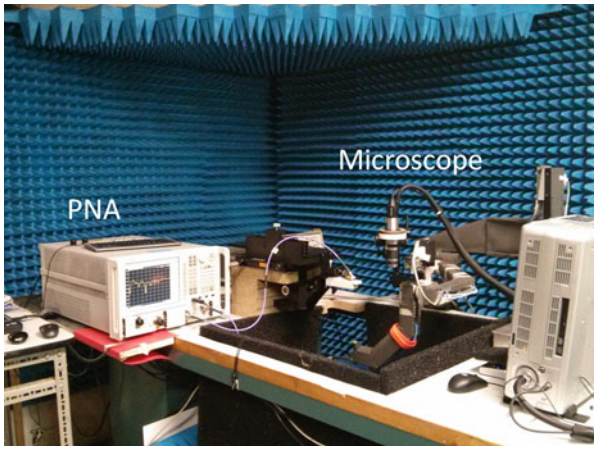


Fig. 11 Overall view of the measurement setup presented in Titz et al. (2012a)

With three different places for the motor, a wider range can be obtained as in Zwick et al. (2004) and Ito et al. (2009). To measure the upper hemisphere, an arm can be driven along a table as in Akkermans et al. (2007), or an arch can be made around the AUT and the horn can move along it as in Willmot et al. (2009) and Sun et al. (2011). The measurement of the lower hemisphere is proposed in Mohammadpour-Aghdam et al. (2011) using a hollowed table. However, to be able to measure almost 3-D pattern around the AUT, at least two arms are needed as in Ranvier et al. (2009), Beer and Zwick (2010), and Chin et al. (2014). Of course, the whole 3-D pattern cannot be measured as the probe positioned will always block a part of the sphere around the AUT (Fig. 10).

Depending on the distance between the receiving horn and the AUT, also the rigidity of the arms has to be taken into account. From Ranvier et al. (2009) to Titz et al. (2012a), the arms were modified from polyurethane to lightweight aluminum because the horn was slightly shaking during the rotation of the arm. Moreover, the weight of the horn and mixer (or the mmW extender if it exists) is not negligible and must be counter balanced if possible as demonstrated in Van Caekenberghe et al. (2008), Beer and Zwick (2010), and Mohammadpour-Aghdam et al. (2011). With a two-arm setup, the cable assembly has also to be treated with care. To avoid twisting the cables from the mixer, one or several rotating joints are necessary (if no diplexer is used).

Finally, the setup has to be positioned on a heavy and anti-vibration table to avoid any vibration during the measurement (which could damage the microelectronic probe). In Titz et al. (2012a), a table was purposely cut to let the rotating arms turning around the AUT (Fig. 11). The table was additionally

DOI 10.1007/978-981-4560-75-7_76-2

covered with absorbers to avoid any electromagnetic reflections. All the RF and mechanical parts are controlled from a laptop via a general purpose interface bus (GPIB) and a custom-made LabView code.

Measuring with a Probe-Fed Measurement Setup

A big task in antenna measurement setups is the calibration of the system in order to get the most accurate measurement results. The calibration is normally divided into three parts: a mechanical calibration to align the AUT exactly within the receiving part of the measurement setup, a gain calibration, and a probe calibration to get an accurate reflection coefficient. The latter two calibrations will be described in the following two sections.

S_{11} Measurement

For reflection coefficient measurements, a vector network analyzer (VNA) or PNA is needed. Using a reference calibration substrate usually supplied by the probe manufacturer and a proper calibration method, the reference plane of the measurement can be set at the tip of the probe. For antenna measurement, a one-port calibration is sufficient: a short-open-load (SOL) method is sufficient to achieve accurate measurements. The short and load standard calibrations are always done by placing the microelectronic probe over a new calibration area. The open standard is achieved by leaving the probe in air, at least 200 μm away from the calibration substrate. It should be noted that sufficient time is needed before starting a new calibration once the mm source or PNA are turned on (few hours). Once this calibration is done, a checking procedure is recommended to verify the accuracy of the achieved current calibration. Therefore, the microelectronic probe is specifically landed over another calibration site and connected to short, open, and load standards. Those measurements give always "worse" values than the values just recorded with the achieved calibration, but the calibration is considered to be sufficiently accurate if short and open S_{11} measurements stay within ± 0.05 dB and if the S_{11} load measurement is below -40 dB (see Fig. 12 for a realistic example achieved with the University of Nice measurement setup).

Gain Measurement and Calibration

If not only the radiation pattern of an AUT but also its gain is needed, a kind of gain calibration has to be taken out. From theory, there exist three kinds of gain calibration methods: the two-antenna method, the three-antenna method, and the gain-comparison method. However, in practice, almost only the gain-comparison method is used. In this method, the receiving power of a well-known standard antenna is measured in the same environment as the AUT will be. Then, the standard antenna is replaced by the AUT and the received power is recorded. With the power ratio of the signal received by those two antennas, the gain of the AUT can then be retrieved.

Figure 13 shows the main block diagram of an mmW antenna measurement setup like the one at Karlsruhe Institute of Technology (Beer et al. 2011). The first step before taking out a measurement is now the calibration of the waveguide output of the VNA extension module. This is done with a "short-open-load" calibration (Fig. 14).

After this is done, a standard horn antenna is placed at the exact same position at which later on the AUT will be mounted. Thus, the free-space losses as well as the transmission losses of the system can be taken into account. Figure 15 depicts the standard antenna measurement. Its gain G_{cal} is obtained from the gain of the horn G_{Horn} and the (negative) gain of the waveguide with which the horn is attached.

$$G_{\mathrm{cal}} = G_{\mathrm{Horn}} \cdot G_{\mathrm{Waveguide}} \tag{1}$$

DOI 10.1007/978-981-4560-75-7_76-2

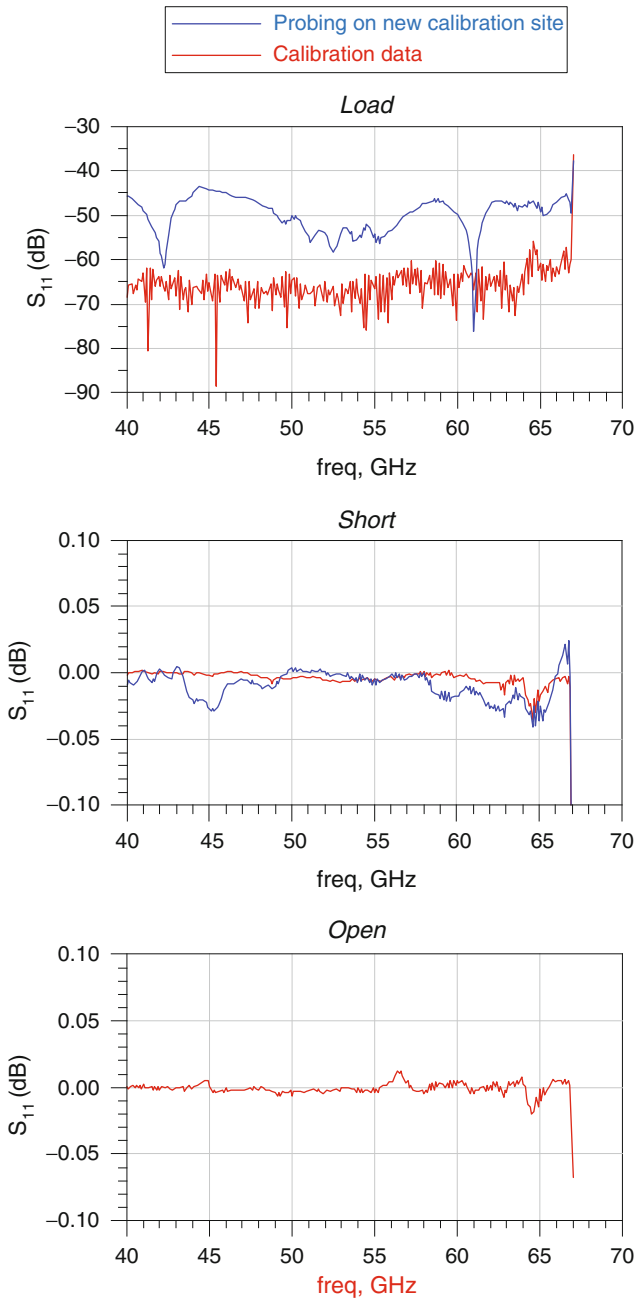


Fig. 12 Checking the calibration accuracy of an ACP 65 AW GSG 200 μm with an ISS: 101-190 substrate by landing the probe over a different site for short, open, and load standards (University of Nice measurement setup)

With the resulting $S_{21,\mathrm{cal}}$ of the transmission measurement incorporating the horn antenna, the system losses G_{System} which represent the system losses, the free-space loss, and the gain of the horn antenna itself can be computed.

$$G_{\mathrm{System}} = \frac{\left|S_{21,\mathrm{cal}}\right|^2}{G_{\mathrm{cal}}} \tag{2}$$

DOI 10.1007/978-981-4560-75-7_76-2

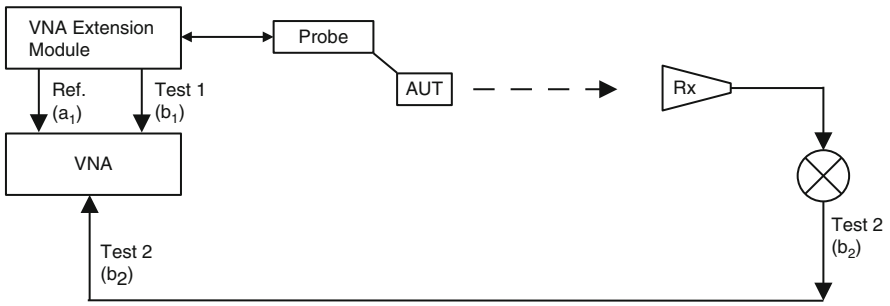


Fig. 13 Block diagram of an mmW antenna measurement setup

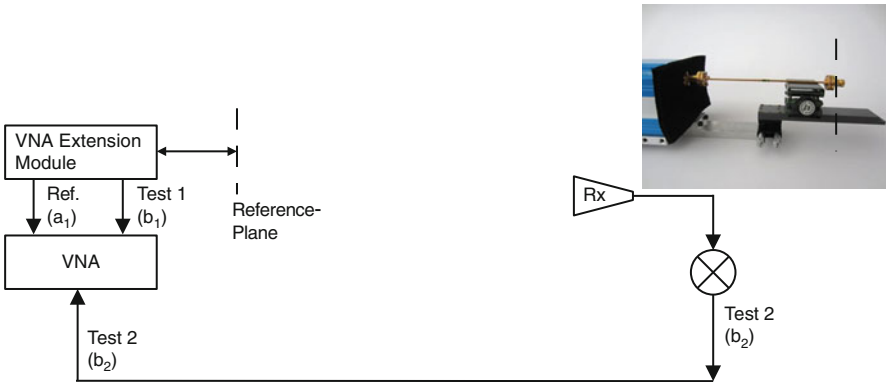


Fig. 14 One-Port SOL calibration at the output of the VNA extension module

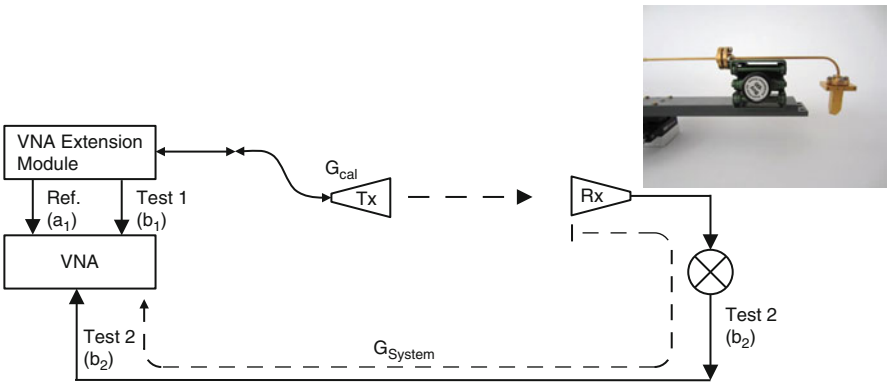


Fig. 15 Measurement of the standard gain antenna

To get now the correct gain of an AUT which is contacted with a probe, the losses of the probe itself G_{Probe} have to be known. Then, the gain of the AUT will be

$$G_{\text{AUT}} = \frac{\left|S_{21,\text{m}}\right|^2}{G_{\text{System}} \cdot G_{\text{Probe}}} \tag{3}$$

where $S_{21,\text{m}}$ is the measured transmission of the AUT. How G_{Probe} is calculated with a SOL calibration is described thereafter.

DOI 10.1007/978-981-4560-75-7_76-2

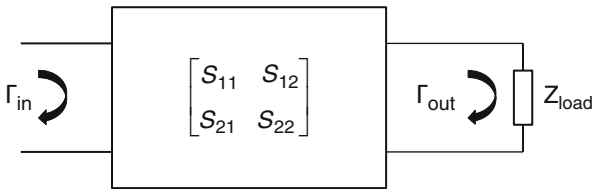


Fig. 16 Schematic of a quadripole loaded by Z_{load}

It should be noted that an alternative method from Purcell (1943) is presented in Brebels et al. (2013). This method is really appealing when applied at mm-wave frequencies, but so far, the initial conference publication was not followed by in-depth journal publication with more analysis and measurement data.

G_{Probe} Extraction Using the Three-Gamma Method

The losses from the probe can be taken from the manufacturer data sheet, but such data do not always exist and sometimes are just given as rough averaged numbers even not as a function of frequency. These probes are indeed different in terms of size, pitch, and design (see Tables 2 and 3). Consequently, they do not always behave as a 50 Ω series device and indeed induce different insertion losses. These losses are non-negligible in the link budget computed during the calibration procedure necessary for a proper extraction of the antenna gain. Therefore, knowing these losses could indeed further improve the overall accuracy of the measurement setup.

In the following, a way is described to extract the S-parameter of any probe using the three-gamma method (Titz et al. 2014).

A representation of a quadripole characterized by S_{ij} parameters and loaded by impedance Z_{load} is given in Fig. 16.

It has an input and output reflection coefficient named Γ_{in} and Γ_{out}. The input reflection coefficient depends on the output reflection coefficient:

$$\Gamma_{in} = S_{11} + \frac{S_{12}S_{21}\Gamma_{out}}{1 - S_{22}\Gamma_{out}} \tag{4}$$

which itself depends on the load Z_{load}. Then, if the quadripole (i.e., the probe in this case) is loaded by three different reference impedances (open, short, and load), three equations with three unknowns may be obtained:

$$\begin{cases} \Gamma_{in\_open} = S_{11\_probe} + \dfrac{S_{12\_probe}S_{21\_probe}\Gamma_{out\_open}}{1 - S_{22\_probe}\Gamma_{out\_open}} \\ \Gamma_{in\_short} = S_{11\_probe} + \dfrac{S_{12\_probe}S_{21\_probe}\Gamma_{out\_short}}{1 - S_{22\_probe}\Gamma_{out\_short}} \\ \Gamma_{in\_load} = S_{11\_probe} + \dfrac{S_{12\_probe}S_{21\_probe}\Gamma_{out\_load}}{1 - S_{22\_probe}\Gamma_{out\_load}} \end{cases} \tag{5}$$

By writing it as a matrix equation

$$\begin{bmatrix} \Gamma_{in\_open} \\ \Gamma_{in\_short} \\ \Gamma_{in\_load} \end{bmatrix} = \begin{bmatrix} 1 & \Gamma_{in\_open}\Gamma_{out\_open} & \Gamma_{out\_open} \\ 1 & \Gamma_{in\_short}\Gamma_{out\_short} & \Gamma_{out\_short} \\ 1 & \Gamma_{in\_load}\Gamma_{out\_load} & \Gamma_{out\_load} \end{bmatrix} \begin{bmatrix} S_{11\_probe} \\ S_{22\_probe} \\ -\Delta_s \end{bmatrix} \tag{6}$$

$$[\Gamma_{in}] = [\Gamma_{3\times3}][S_{probe}]$$

DOI 10.1007/978-981-4560-75-7_76-2

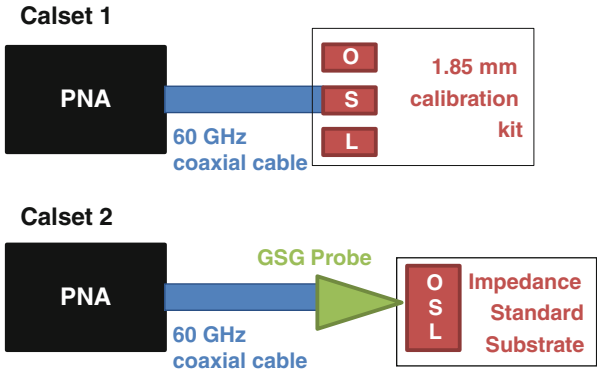


Fig. 17 Procedure for the three-gamma method

The three unknowns are now S_{11_probe}, S_{22_probe}, and Δ_s, defined in

$$\Delta_s = S_{11\_probe}S_{22\_probe} - S_{12\_probe}S_{21\_probe} \tag{7}$$

A passive probe is a reciprocal quadripole, and then S_{12_probe} is equal to S_{21_probe}. Thus, the S_{ij} parameters can be extracted from

$$\left[S_{probe}\right] = \left[\Gamma_{3\times3}\right]^{-1}\left[\Gamma_{in}\right] \tag{8}$$

and further the insertion loss of the loaded probe from the computed S_{21_probe}.

From the developed equations, two sets of data need to be measured: Γ_{in} and Γ_{out}. To do that, two different "one-port" calibrations must be done. The first calibration is done in a physical plane set before the probe, at the end of the coaxial cable (or the waveguide). With this calibration, it will be able to measure Γ_{in} when the probe is loaded with the three reference standards. The second calibration is set in a physical plane taken after the probe, with the help of a calibration substrate. It will give Γ_{out}. After these two calibrations, the probe can be loaded by the reference impedances (open, short, load), and S_{11} can be measured with the two calibrations, calsets 1 and 2. The procedure for the three-gamma method is summarized in Fig. 17.

The global losses of the probe depend on the impedance value of the AUT. Indeed, from the two-port power gain theory, it is known that the transducer power gain (G_T), defined as the ratio of the power (P_{ant}) delivered to the load (here the AUT) to the power (P_{source}) available from the source (here at the end of the cable or waveguide), is given by

$$G_T = \frac{P_{ant}}{P_{source}} = \frac{\left|S_{21\_probe}\right|^2\left(1 - \left|\Gamma_{ant}\right|^2\right)}{\left|1 - S_{22\_probe}\Gamma_{ant}\right|^2} \tag{9}$$

or expressed in decibels:

$$10\log_{10}(G_T) = 10\log_{10}\left(\left|S_{21}\right|^2\right) + 10\log_{10}\left(1 - \left|\Gamma_{ant}\right|^2\right) - 10\log_{10}\left(\left|1 - S_{22}\Gamma_{ant}\right|^2\right) \tag{10}$$

where there is a first term corresponding to the losses of a perfectly matched probe, a second term corresponding to the mismatch loss of the antenna, and a third additional term. G_T clearly depends on the parameters of the probe and the reflection coefficient of the AUT (Γ_{aut}). In some particular cases, this

DOI 10.1007/978-981-4560-75-7_76-2

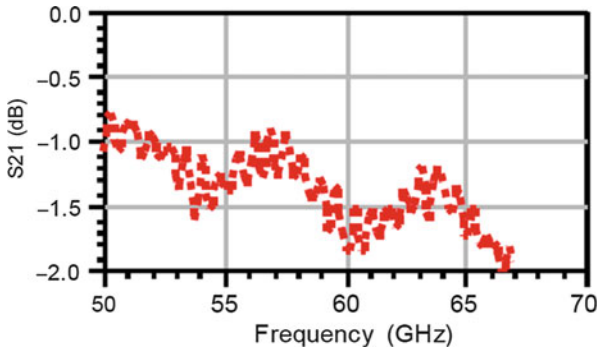


Fig. 18 Procedure for the three-gamma method

equation can be simplified. First, if the impedance of the antenna is purely 50 Ω (reflection coefficient is equal to zero), it becomes

$$G_{\mathrm{T}} = \left|S_{21\_\mathrm{probe}}\right|^2 \tag{11}$$

The global loss is then entirely due to insertion loss of the probe. If the probe is matched to 50 Ω, then S_{22} is equal to zero and it becomes

$$G_{\mathrm{T}} = \left|S_{21\_\mathrm{probe}}\right|^2 \left(1 - \left|\Gamma_{\mathrm{ant}}\right|^2\right) \tag{12}$$

The global loss is then indeed equal to the sum of the insertion loss of the probe and the mismatch loss of the AUT. In the previous part about gain measurement, the antenna is matched to 50 Ω, which means that, in fact, the realized gain is indeed computed. If now the IEEE gain from the measurement shall be computed to extract the mismatch loss, the insertion loss of the probe has to be substituted by the global losses. Therefore, the gain (G) can be extracted from the measured realized gain (G_{real}) using

$$\mathrm{G}(\varphi, \theta) = G_{\mathrm{real}}(\varphi, \theta) \frac{\left|1 - S_{22\_\mathrm{probe}}\Gamma_{\mathrm{ant}}\right|^2}{\left(1 - \left|\Gamma_{\mathrm{ant}}\right|^2\right)} \tag{13}$$

Further examples of such computations can be found in Titz et al. (2014). Figure 18 shows an example of the extracted losses of a 400-µm pitch probe given be 1 dB at 60 GHz by the manufacturer.

Circularly Polarized Antenna Measurements

It was demonstrated in Maltzev et al. (2010) that using circularly polarized (CP) antennas could be interesting when placed each at one side of a radio link, especially when the orientation of the two communicating devices is not a priori known. For 60-GHz WiGig applications, CP antennas could be of great interest to mitigate the effects of polarization mismatch, especially in multipath environments. Several methods have been used in the literature to compute the axial ratio (AR) of a CP antenna from specific measurements. In Liu et al. (2012) and Li et al. (2012), the AR was obtained by comparing the magnitude only of the received power levels of two orthogonal polarizations of the AUT: this is not always an accurate method especially when the polarization ellipse is tilted. In Zhou et al. (2009), another AR determination is presented with moderate accuracy versus simulation, but the computation method is not detailed and the AUT is waveguide fed. In Zhang et al. (2013), the problems encountered for the AR

DOI 10.1007/978-981-4560-75-7_76-2

computation are briefly discussed, namely, the difficulty to accurately record the transmitted phase by the AUT when rotating the receiving horn.

The method presented in the following has been developed to accurately compute the AR of probe-fed CP antennas. Doing this from phaseless gain measurements only, considering the phase acquisition at 60 GHz of any signal, is extremely challenging and difficult to achieve with strong accuracy (Titz et al. 2012a).

The AR is defined as the ratio of the major to the minor axis of the polarization ellipse. Thus, by plotting the ellipse, this AR can be computed. The tilt angle is the angle by which the ellipse is tilted from the horizontal polarization. Knowing the sense of polarization, the right-hand circularly polarized (RHCP) or left-hand circularly polarized (LHCP) components of the field (E_{RHCP}, E_{LHCP}) can be measured. They are related to the axial ratio through

$$AR = \left| \frac{|E_{RHCP}| + |E_{LHCP}|}{|E_{RHCP}| - |E_{LHCP}|} \right| \tag{14}$$

From the IEEE Standard Test Procedures for Antennas, four different methods to compute the AR of CP antennas from measurements can be counted: polarization-pattern method, rotating-source method, multiple-amplitude-component method, and phase-amplitude method. In the "polarization-pattern method" and the "rotating-source method," the horn which receives the linearly polarized magnitude (only) of the E-field of the CP AUT has to be rotated by 360° which is heavily time-consuming. These two methods cannot retrieve the sense of the field rotation which is not an important issue as the simulation data can be trusted for this purpose. The "multiple-amplitude-component method" consists in using six different receive antennas having each a different polarization. This is definitely not easy to design and fabricate at 60 GHz. Then, by post-processing the recorded data, the whole CP characteristics can be extracted. The "phase-amplitude method" allows a fast and complete characterization but needs a measurement setup with phase capability, which is quite difficult to implement with satisfactory accuracy at 60 GHz mainly because a 60-GHz moving cable is involved at the receive side. Another possibility is to attach an mmW source module to the horn, but this solution can be too heavy for the arm of the setup. A solution with less capability is to set the arm in a specific position (broadside, end-fire, or any other position where the AUT produces maximum gain) and measure gain and phase in this direction. In that case, the AR is computed for the parametric equation of an ellipse:

$$\vec{E} = \begin{cases} 0 \\ E_y = E_{0y} \cos\left(\omega t + \varphi_y\right) \\ E_z = E_{0z} \cos\left(\omega t + \varphi_z\right) \end{cases} \tag{15}$$

The method proposed at the University Nice Sophia Antipolis is a modified version of the multiple-amplitude-component technique. The idea is to compute the AR and tilt angle by rotating the receiving horn antenna at 0° (horizontal), 45°, 90° (vertical), and 135° (−45°) and achieve a fast four-point measurement operation. This is possible using two different holders for the receiving horn. The first holder is made for the 0° and 90° positions and the second holder for 45° and −45°. The four positions of the horn are shown in Table 4 for better understanding. When measuring linearly polarized antennas, only the 0° and 90° positions are required, but at least a third position (45° or −45°) is mandatory to be able to compute the polarization ellipse as shown in Fig. 19, when dealing with circularly polarized antennas.

The polarization pattern can be determined from the following equation with G_i being the realized gain of the AUT in the direction θ_i alternatively equal to 0°, 45°, 90°, or −45°:

DOI 10.1007/978-981-4560-75-7_76-2

Table 4 Positions of the horn antenna aperture for the measurement of CP antennas with the setup from University Nice Sophia Antipolis

Position	0°	90°	45°	−45°
Aperture view of the horn antenna				

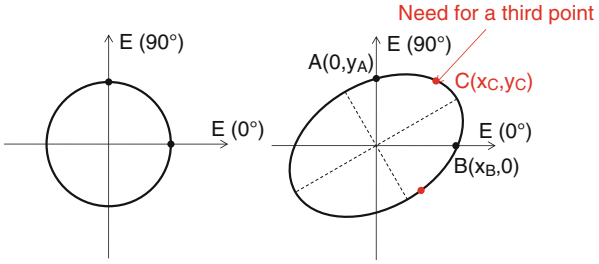


Fig. 19 Polarization circle or ellipse of a the electromagnetic field from a CP antenna

$$G_i = r_i^2 = \frac{b^2}{1 - e^2 \cos^2(\theta_i + \tau)} \quad \text{where} \quad i = 0,\ 45,\ 90,\ -45 \tag{16}$$

This equation corresponds to the ellipse equation in polar coordinates where b is the semiminor axis, e is the eccentricity, and τ is the tilt angle of the ellipse. b, e, and τ can then be computed from only three (out of four) measurements (e.g., G_0, G_{45}, G_{90}) from the following equations:

$$\left[\frac{1}{G_\varphi^2} + \frac{1}{G_\theta^2} + \frac{4}{G_{45}^2} - \frac{2}{G_\varphi G_\theta} - \frac{4}{G_\varphi G_{45}} - \frac{4}{G_\theta G_{45}}\right] b^4 + \left[\frac{4}{G_\varphi} + \frac{4}{G_\theta}\right] b^2 - 4 = 0 \tag{17}$$

$$e^2 = 2 - b^2\left(\frac{1}{G_\varphi} + \frac{1}{G_\theta}\right) \tag{18}$$

$$\begin{cases} \sin(2\tau) = 1 - \dfrac{2}{e^2}\left(1 - \dfrac{b^2}{G_{45}}\right) \\ \cos(2\tau) = \dfrac{b^2}{e^2}\left(\dfrac{1}{G_\theta} - \dfrac{1}{G_\varphi}\right) \end{cases} \tag{19}$$

The fourth measured point (G_{-45} in this example) can be used as a verification check. Indeed, from the four gain measurement points, four different ellipses can be computed, giving four different values for the AR and the tilt angle. When the axial ratio is low (below 3 dB), the four ellipses are nearly the same. When it increases and the polarization of the antenna becomes linear, the four ellipses can be quite different. In fact, the measured cross polarization is usually always higher than the simulated cross polarization, and thus the computed axial ratio is always lower than expected. The accuracy of this method is then worse when antennas with a low AR are measured. Considering having now four AR values and four tilt angle values, two methods are proposed to merge them into a single value.

The ellipse method consists in choosing the ellipse that is the closest to the fourth measured point. Another method that was also implemented plots a unique closer ellipse to the four measured points using

DOI 10.1007/978-981-4560-75-7_76-2

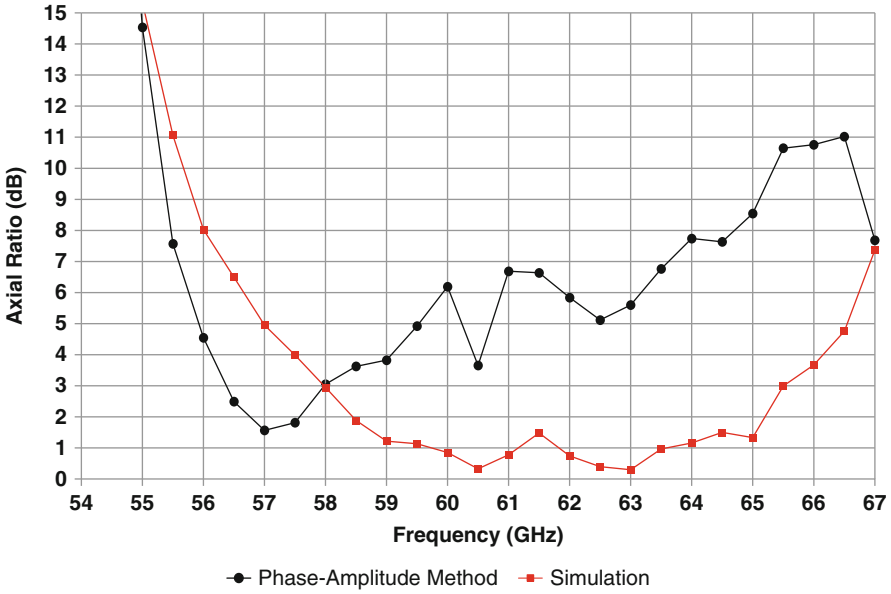


Fig. 20 Simulated and computed AR of a CP LTCC antenna array with the phase-amplitude method

a mean square (MS) optimization algorithm. This method is called the mean square method. Those methods are largely illustrated in Titz et al. (2012a) and have been successfully implemented in Rousstia and Herben (2013).

According to its definition, the AR is then determined using:

$$\mathrm{AR} = \frac{\mathrm{a}}{\mathrm{b}} \quad \text{with} \quad \mathrm{a} = \frac{\mathrm{b}}{\sqrt{1 - \mathrm{e}^2}} \tag{20}$$

The polarization sense cannot be extracted with the proposed methods, but the maximum and minimum values of the LHCP and RHCP realized gains (G_{RHCP}, G_{LHCP}) can be computed from:

$$\begin{aligned} \max(G_{\mathrm{RHCP}}, G_{\mathrm{LHCP}}) &= \frac{G_{\mathrm{tot}}}{2}\frac{(\mathrm{AR}+1)^2}{(\mathrm{AR}^2+1)} \\ \min(G_{\mathrm{RHCP}}, G_{\mathrm{LHCP}}) &= \frac{G_{\mathrm{tot}}}{2}\frac{(\mathrm{AR}-1)^2}{(\mathrm{AR}^2+1)} \end{aligned} \tag{21}$$

Depending on the setup, the presented method can be used to extract the axial ratio in 2-D cuts, 3-D, or only one direction (Titz et al. 2012a).

To demonstrate the accuracy of this method, Fig. 20 shows the simulated and computed AR of a CP LTCC antenna array with the phase-amplitude method, while Fig. 21 shows the same AR computed with the "ellipse method" and the "mean square method."

Efficiency Measurements

Several methods to determine the efficiencies of connector-fed antennas have been developed at low frequencies. For directive antennas, a first approach consists in computing the total efficiency from the ratio of the measured realized peak gain to the peak directivity. However, a very high-quality anechoic chamber is required as this method does not give accurate results if ripples are recorded when measuring the radiation patterns. Other measurement solutions include the Wheeler cap method (McKinzie 1997)

DOI 10.1007/978-981-4560-75-7_76-2

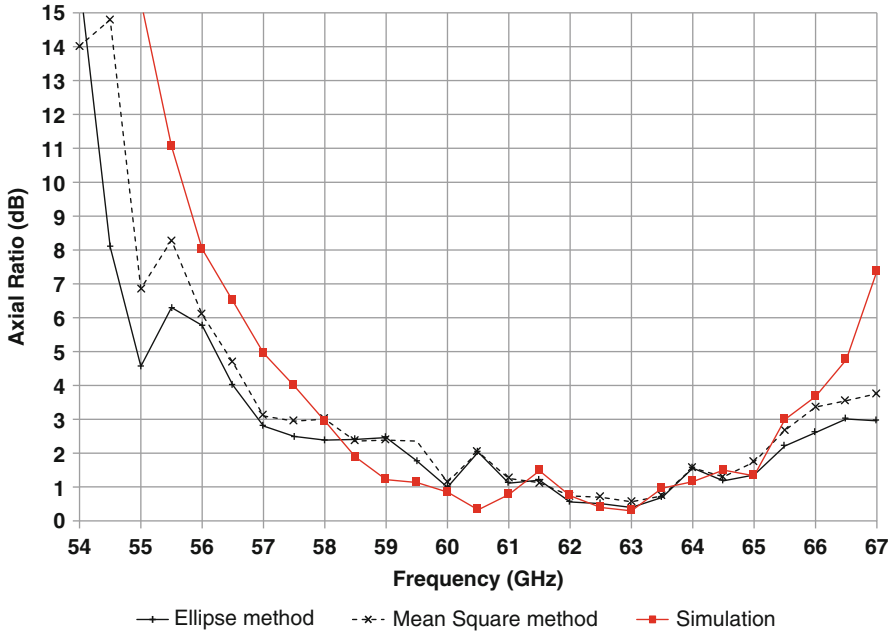


Fig. 21 Simulated and computed AR of a CP LTCC antenna array with the "ellipse method" and the "mean square method"

and the reverberation chamber (Kildal and Rosengren 2004). A Wheeler cap measurement of a probe-fed antenna at 60 GHz has already been presented in Fakharzadeh et al. (2011), but more results are needed to validate the setup. The main problem of this measurement system is the small hole etched in the cap to let passing the microelectronic probe to feed the antenna. Therefore, the expected Faraday cage behavior is not perfect, and the accuracy of the measures is degraded. The reverberation chamber is also a very good candidate to measure the efficiencies of low-frequency antennas, but only few attempts of reverberation chamber setups for probe-fed antennas at 60 GHz have been reported. In Asano et al. (2013), an interesting reverberation chamber operating at 60 GHz is presented, but so far, a 1.85 mm coaxial is used to feed the AUT. The last method consists in the integration of the gain radiation pattern of the antenna under test (AUT) over the full 3-D sphere. This is the method that will be presented in this part (Titz et al. 2012c). It has been developed at the University Nice Sophia Antipolis.

The total efficiency η_{tot} is the ratio of the total power radiated by an antenna to the incident power at the antenna port (IEEE Standard Test Procedures for Antennas 1979). It can be also defined using η_R and S_{11}:

$$\eta_{tot} = \frac{P_{rad}}{P_{inc}} = \eta_R\left(1 - |S_{11}|^2\right) \tag{22}$$

S_{11} is the reflection coefficient of the AUT, and η_R its radiation efficiency. If the realized gain G_{real} of the probe-fed AUT can be measured in any particular direction defined by θ and φ, the incident power at the antenna port can be deduced from

$$P_{inc} = \frac{4\pi U(\theta, \varphi)}{G_{real}(\theta, \varphi)} \tag{23}$$

where U(θ, φ) is the radiation intensity in one particular direction defined by θ and φ. Then P_{rad} is needed to compute the total efficiency. The radiated power P_{rad} is developed as follows:

DOI 10.1007/978-981-4560-75-7_76-2

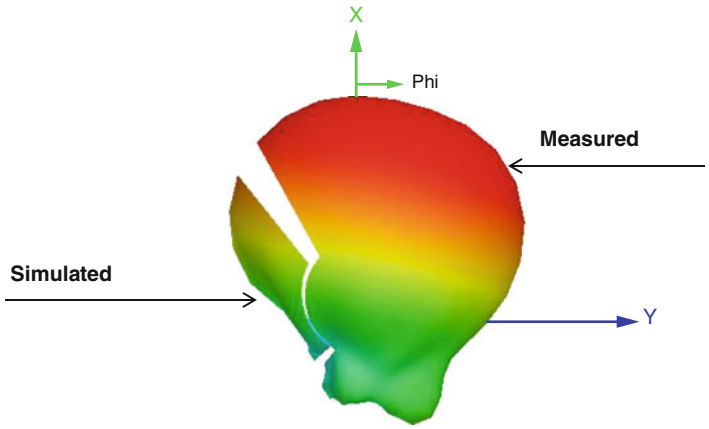


Fig. 22 Illustration of the hybrid method to compute the total efficiency of any probe-fed AUT from the measured 3-D patterns obtained with the setup in Titz et al. (2012c)

$$P_{\text{rad}} = \int_0^{2\pi} \int_0^{\pi} \text{U}(\theta, \varphi) \sin(\theta) \text{d}\theta \text{d}\varphi \tag{24}$$

η_{tot} can then be expressed as:

$$\eta_{\text{tot}} = \frac{P_{\text{rad}}}{P_{\text{inc}}} = \frac{\int_0^{2\pi} \int_0^{\pi} \text{U}(\theta, \varphi) \sin(\theta) \text{d}\theta \text{d}\varphi}{P_{\text{inc}}} = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} G_{\text{real}}(\theta, \varphi) \sin(\theta) \text{d}\theta \text{d}\varphi \tag{25}$$

As the realized gain is measured in θ and φ steps, the equation can be rearranged using discrete points M and N. For that, a gain value is associated to a certain portion of the sphere (surface); the smaller the angle steps are, the smaller the numerical error will be (Balanis 1997):

$$\eta_{\text{tot}} = \frac{1}{4\pi} \Delta\theta \Delta\varphi \sum_{\text{j}=1}^{\text{M}} \sum_{\text{i}=1}^{\text{N}} G_{\text{real}}\left(\theta_{\text{i}}, \varphi_{\text{j}}\right) \sin(\theta_{\text{i}}) \tag{26}$$

θ and φ should respectively vary from 0 to π and 0 to 2π to scan the whole 3-D sphere. However, this is not possible in most of the measurement setups as the probe positioned is blocking a part of the sphere. In Titz et al. (2012c), only 73.5 % of the total sphere's surface can be scanned because θ and φ might respectively vary from −30° to 220° in 10° steps and −30° to 210° in 2° steps. So, the total efficiency is only provided on these 73.5 %. This computed value is surely a lower boundary for the efficiency of the AUT, but it is still not satisfactory for accurate characterization. If the simulated and measured gain radiation patterns are in good agreement over 73.5 % of the sphere's surface, the simulated radiation pattern of the AUT can be trusted over the remaining portion of the sphere's surface (26.5 %) and therefore used to calculate the efficiency over this surface. Therefore, it is just a matter of adding the efficiencies of the measurement (73.5 % of the sphere's surface) and the simulation (26.5 % of the sphere's surface) to obtain the total efficiency of the AUT over the 3-D total sphere. Figure 22 illustrates the proposed hybrid method to compute the total efficiency of one particular AUT. As can be seen, one part of the radiation pattern is measured and the remaining portion is simulated. The radiation efficiency η_{R} can be computed with the measured reflection coefficient of the AUT.

DOI 10.1007/978-981-4560-75-7_76-2

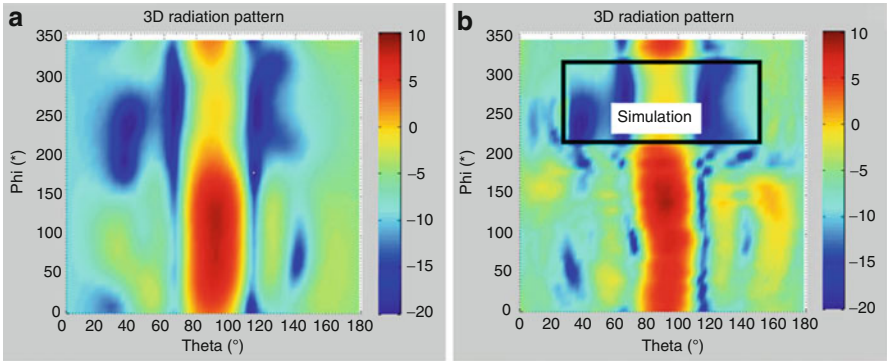


Fig. 23 Illustration of the hybrid method to compute the total efficiency of an end-fire LTCC antenna array at 60 GHz from the measured 3-D patterns obtained with the setup in Titz et al. (2012c). (**a**) Simulated 3-D field, 69.9 % of total efficiency. (**b**) Hybrid measured and simulated 3-D field, 69 % of total efficiency

This method has already proved to be accurate for directive and less directive antennas (Titz et al. 2012c). One example of measurements is shown in Fig. 23.

Near-Field and Phase Measurements

Accurate phase measurements at mm-wave frequencies are really tricky to achieve. So far, several planar scanners exist in many academic labs and companies (Janse Van Rensburg and Hindman 2008), but they rather cope with waveguide-fed AUT. From the magnitude and phase acquisition of the received signal by the open waveguide (also called the probe) over a planar surface, a near-field to far-field transformation enables to retrieve the far-field performance of the measured AUT. In Shamim et al. (2008), the near-field scanning probe is a printed circuit dipole at 24 GHz. After near-field to far-field transformation (for planar scanning), the patterns are compared with simulation. But, the agreement is not really good due to the insufficient scanning range around the antenna. Other planar near-field measurements have also been realized at 60 GHz in Hong et al. (2011) and Tsai et al. (2013) with an open waveguide or a horn antenna. As the measurement range is wider, the measurements are quite close to simulation, but still, ripples and shadow exist in the computed radiation patterns. With those planar near-field measurements, the mmW extender is placed close to the receiving waveguide parts. One setup also able to measure both waveguide-fed and probe-fed AUT seems to exist at NASA, but so far, not so much information is published and it is difficult to know exactly its capabilities and performance (Slater 1994). Recording the phase of the signal transmitted by the AUT allows to measure electrically large antennas at very short distances (in their near-field region). For such antennas (for primarily space applications for instance), the far-field region is too far away from the AUT, and therefore, there are no possibilities to physically place the received open waveguide at very long distance. The free-space loss added to the oxygen absorption of electromagnetic waves at certain mm-wave frequency bands is also an issue: sufficient signal is needed to stay above the sensitivity level of the receiving parts of the setup. Those electrically large antennas for space applications are mostly fed by waveguide, but recently, backhaul and fronthaul deployments with directive antennas in V- and E-band have been widely developed, and in this case, a near-field measurement setup with probe-fed capabilities might be needed for characterization purpose.

Near-field acquisition over a spherical surface is known to produce better results than planar scanning (in terms of near-field to far-field transformation). The main issue is to scan the spherical surface over the AUT with very high accuracy: this is extremely challenging and much more difficult than a planar scan due to mechanical challenges in spherical rotation. So far, rotating the open waveguide dedicated to acquisition over a slightly non-perfect sphere will produce small errors in magnitude but large phase error at mm-wave frequencies. Moreover, this phase error drastically increases when frequency increases. In

DOI 10.1007/978-981-4560-75-7_76-2

this case, correction techniques can reduce those errors. The main technique is to achieve a prior rotation before any measurement, with a laser tracker, and then to record the non-perfect spherical scan to be able, later on, to compensate mathematically the position of the arms for every measured point. This is a time-consuming procedure which considers that the rotating arms of the setup will scan space in exactly the same manner for the calibration step and the following measurement. The choice of the position of the mm-wave head at the receiver side is also of primary importance as one can choose to make it rotate during the measurements or make it stable and faraway from the receive horn, but in this case, the mm-wave head is connected to a 60-GHz cable which will move and pass through a rotary joint during the measurements. At higher frequencies, when cables are no more available and only waveguide connections exist, this option is very difficult to handle. Each solution has pro and cons but will anyway influence the accuracy of the acquisition in different manners.

Recently, two spherical near-field acquisition setups for probe-fed mm-wave antennas have been commercially released by Orbit (μ-Lab Station) and NSI (NSI-700S-360). Those setups have been sold to very few companies and academic labs, but so far, there are not so much available published data yet by the users' only suppliers' advertisement (Orbit Presentation; Janse Van Rensburg and Gregson 2014). Right now, those setups seem to be limited to 110 GHz for an accurate phase acquisition. As at low frequencies, having a commercial measurement system ready to plug is an appealing option, but main drawbacks usually come from low flexibility. If the software flexibility and usability tends to increase from year to year, the main drawback for mm-wave frequencies probe-fed antennas stays in the fact that every AUT to be measured is differently probed, and therefore, a custom foam holder is always necessary for testing. Solutions exist with generic foam support, but they probably will not fit any testing configuration.

In Arboleya et al. (2013), an interesting solution is presented. A planar near-field measurement setup is presented. As the phase measurement is the main difficulty with moving cables, a phaseless measurement technique is proposed using two sampling surfaces. This is probably the best approach for probe-fed measurements at very high mm-wave frequencies.

Measurements Above 110 GHz

As frequencies increase, it becomes more and more challenging to realize accurate magnitude measurement results. This is mainly due to four reasons which will be explained later on in more details:

- Size of antennas decreases which complicates a proper probe contact with the AUT.
- Calibrations usually become less accurate at higher frequencies.
- The dynamic range of the RF equipment (and therefore the setup) gets worse.
- Mechanical vibrations become more critical at higher frequencies.

Since the size of the antenna depends on the operating wavelength, the antenna becomes smaller, as the frequency increases. Thus, it has to be aligned with the receiving horn more precisely. Also, the pitch of the probe that contacts with the pads of the AUT normally decreases with higher frequencies to minimize the parasitic effects of the contact needles. For example, at V-band (50−75 GHz), a pitch width of 300 μm is quite common, whereas at H-band (225−300 GHz), a pitch width of 60−80 μm has to be used. A good microscope and a very fine positioner are therefore needed.

Another issue is the accuracy of the calibration at higher frequencies. At regular probe stations with two ports at frequencies above 110 GHz, normally, a "trough reflect line" (TRL) calibration is taken out. In this calibration, a short standard is used as "reflect," and two pieces of transmission line are used as "trough" and "line." Since the antenna measurement setup needs only one feeding port, a "short-open-load" (SOL) calibration only can be used (as explained in paragraph 1). The drawback of the SOL calibration is that it is

DOI 10.1007/978-981-4560-75-7_76-2

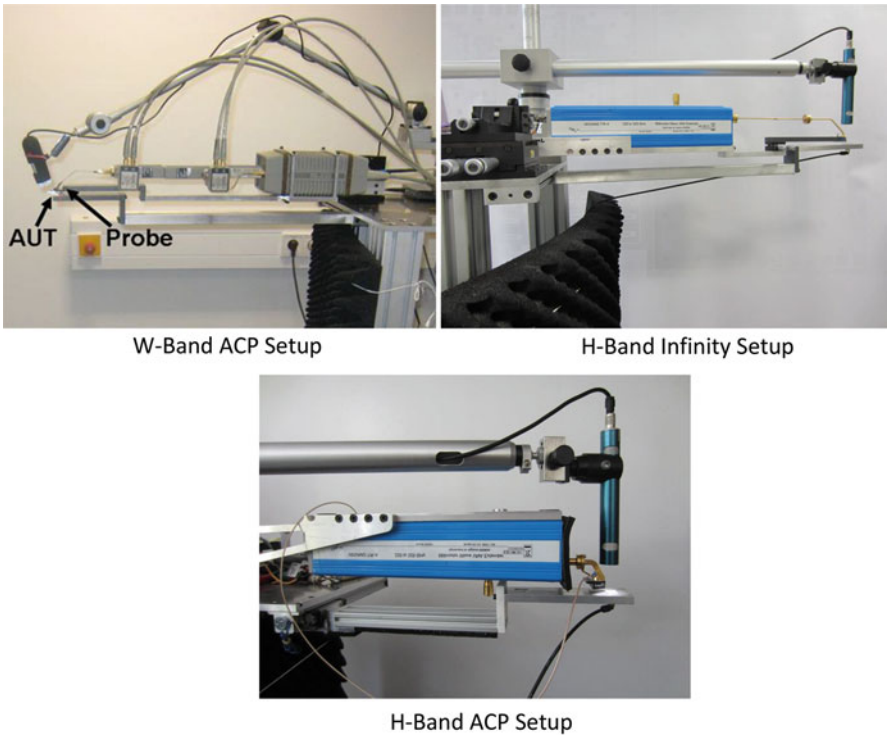


Fig. 24 The setups for W-/D- and H-band from Gulan et al. (2013)

difficult to realize an exact 50 Ω "load" fully constant over the complete frequency band. Thus, a calibration with an antenna measurement setup at frequencies over 110 GHz will usually be worse compared to a two-port calibration at a probe station. As an example, calibrations at different frequency ranges shall be compared in the following. The antenna measurement setup which is used is the one introduced in Beer and Zwick (2010) and Gulan et al. (2013), and the calibration has been done in the W-band (75−110 GHz), the D-band (110−170 GHz), and the H-band (220−325 GHz). The probes which have been used are indicated in Table 3, and the measurement setups are shown in Fig. 24.

If a SOL calibration is performed now, the reflection coefficient of the open probe in air (open-in-air) can be used to indicate the quality of the calibration. Even though this is not absolutely correct (because the probe tips will always act as an antenna), it gives a bit a feeling of the quality of the calibration. In Fig. 25, the reflection coefficient of the open-in-air for the different setups is shown. Figure 25a, b shows the reflection coefficient in W- and D-band. These calibrations have been taken out with so-called air coplanar probes (ACP probes) from GGB (Picoprobe). These probe types have three needles for ground-signal-ground connection. These three needles will always cause parasitic effect since they act like antennas and radiate. As it can be seen, the reflection coefficient of the probes which would be expected to be exactly 0 dB is between −0.2 dB and +0.2 dB in W-band and between −0.4 dB and +0.7 dB in D-band. These are quite reasonable values for SOL calibrations but worse than the ±0.05 dB expected in V-band as explained previously. At H-band, the calibration becomes even worse (Fig. 25c). The reflection coefficient stays between −0.5 dB and +0.5 dB in the 200−280 GHz range but begins to increase up to +2.8 dB at 325 GHz. This kind of probe with this kind of calibration is not suitable for this frequency range. An improvement consists in using Infinity probes from Cascade. These probes use a substrate with a printed coplanar waveguide line (CPW line) instead of needles as the contact part. Thus, the parasitic effects due to radiation are minimized. The drawback is that these probes are way more sensitive than the ACP probes; they have to be aligned very accurately with the AUT whereas slight unevenness can be compensated with the ACP probes. The performance of an SOL-calibrated Infinity probe can be seen in

DOI 10.1007/978-981-4560-75-7_76-2

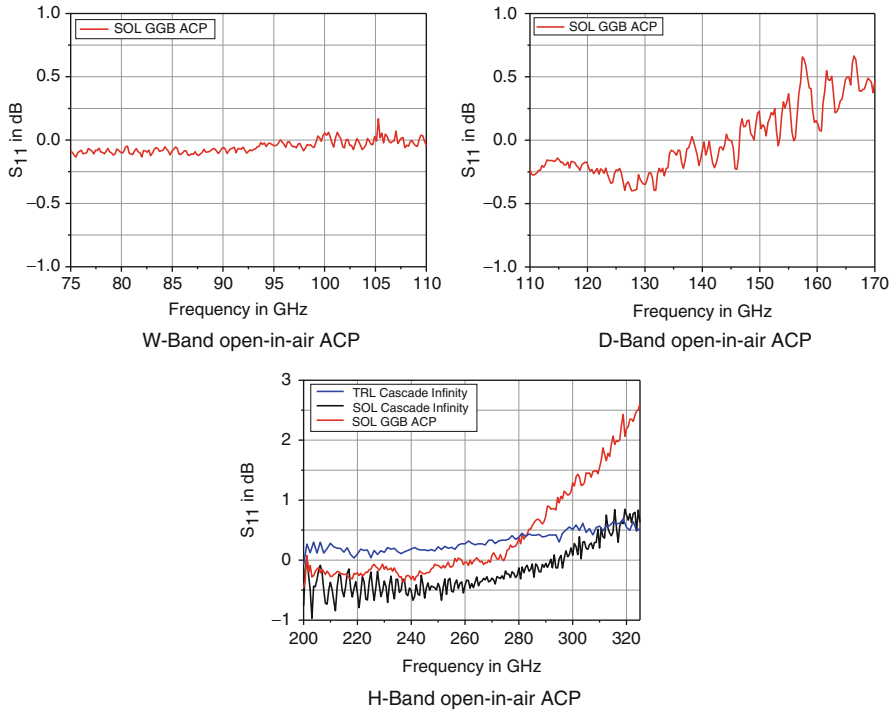


Fig. 25 Reflection coefficient of an open-in-air probe for different frequency ranges and setups

Fig. 25c. Compared to the calibration with ACP probes, the latter one remains in a +/−1 dB area even at higher frequencies. As a reference, a two-port TRL calibration which has been made at a probe station at KIT is also shown in Fig. 25c. Obviously, this calibration is more stable than the two SOL calibrations.

The next point which has to be taken into account when frequency increases is the dynamic range of the setup. A measurement above 110 GHz normally means a lower output power to feed the antenna, higher free-space losses, and higher conversion losses of the mixer which down-convert the received signals to an IF. As an example, the link budget of the antenna measurement setup from KIT is compared for different frequencies. The link budgets are shown in Table 5. In all cases, an AUT gain of 10 dBi is assumed.

While the input power of the VNA at 77 GHz and 120 GHz is quite high (−77 dBm and −81.2 dBm), it is strongly degraded at 220 GHz (−114.7 dBm) and 325 GHz (−130.5 dBm). The main components which have an influence on the link budget are the output source of the VNA, the free-space loss, and the conversion loss of the down-converter/mixer attached to the receiving antenna. Whereas the output power of the W-band source at 77 GHz is 0 dBm (Agilent 83558A mm-wave source module), it is only −9 dBm at 120 GHz (OML V06VNA2 millimeter-wave VNA extender) and decreases to −26 dBm at 325 GHz (OML V03VNA2 millimeter-wave VNA extender). In the case of this setup at H-band, Virginia Diodes offers a VNA extender with an output power of −9 dBm at 325 GHz (VDI Model: WR3.4-VNAX) which would give additional 18 dBm in the link budget. The free-space loss is determined by the distance between AUT and receiving antenna. In the case of the setup from KIT, the distance is set constant to 75 cm. At 325 GHz, the setup allows a far-field measurement of an AUT with about 35 dBi of gain. The down-converters used are passive diode mixers. The conversion loss normally increases with frequency. Especially, the down-converter at 325 GHz has a really high conversion loss of 45 dB. A good way to further increase the link budget would be the use of a low-noise amplifier (LNA) between the receiving antenna and the down-converter. However, broadband LNAs for H-band are not commercially available yet.

DOI 10.1007/978-981-4560-75-7_76-2

Table 5 Link budget of the antenna measurement setup introduced in Beer and Zwick (2010) at different frequencies

	77 GHz		120 GHz		220 GHz		325 GHz	
	Gain	Absolute power	Gain	Absolute power	Gain	Absolute power	Gain	Absolute power
LO output VNA		21 dBm		10 dBm		10 dBm		10 dBm
Output VNA extender (RF)		0 dBm		−9 dBm		−27 dBm		−26 dBm
Waveguide loss	−0.8 dB	−0.8 dBm	−3.5 dB	−12.5 dBm	−6.5 dB	−33.5 dBm	−5 dB	−31 dBm
Probe loss	−1.2 dB	−2 dBm	−1.7 dB	−13.2 dBm	−4.1 dB	−37.6 dBm	−7.5 dB	−38.5 dBm
AUT	+ 10 dBi	8 dBm	+10 dBi	−3.2 dBm	+10 dBi	−27.6 dBm	+10 dBi	−28.5 dBm
Free-space loss (75 cm)	−67.7 dB	−59.7 dBm	−72 dB	−75.2 dBm	−76.8 dB	−104.4 dBm	−80.2 dB	−108.7 dBm
Receiving antenna	+23 dB	−36.7 dBm	+23 dBi	−52.2 dBm	+22.7 dBi	−81.7 dBm	+25.2 dBi	−83.5 dBm
Input VNA extender (RF)		−36.7 dBm		−52.2 dBm		−81.7 dBm		−83.5 dBm
Conversion loss mixer	−39 dB	−75.7 dBm	−27 dB	−79.2 dBm	−31 dB	−112.7 dBm	−45 dB	−128.5 dBm
Cable and rotary joint losses	−2 dB	−77.7 dBm	−2 dB	−81.2 dBm	−2 dB	−114.7 dBm	−2 dB	−130.5 dBm
Input VNA (IF)		−77.7 dBm		−81.2 dBm		−114.7 dBm		−130.5 dBm

DOI 10.1007/978-981-4560-75-7_76-2

The following section will give a short overview of some known measurement setups above 110 GHz. Since the high millimeter-wave region is an area of ongoing research (and especially companies do not necessarily publish their available setups), it will be concentrated on three published setups from universities. These are the setups of:

- The Karlsruhe Institute of Technology, Germany
- The Institute for Infocomm Research, Singapore
- University Nice Sophia Antipolis, EpOC Lab, France

Even if some performance of the setup of the Karlsruhe Institute of Technology (KIT) was described in the previous section, it is compared to other setups in this section. It was first published in 2009 as a measurement setup for V- and W-band (50−75 GHz and 75−110 GHz) (Beer and Zwick 2010) and later extended to D- and H-band (Beer et al. 2011; Gulan et al. 2013). It consists of an Agilent PNA-X with frequency extenders for the different bands. Thus, it is able to measure from 110 to 325 GHz but with a gap between 170 and 225 GHz. However, the lower bound of the H-band can be reduced to 200 GHz by tuning the internal mixer accordingly. The AUT (or the standard antenna for calibration purposes) is either directly connected to a VNA extender from OML (if it is waveguide fed) or it is fed by a microelectronic probe which is in turn attached to the VNA extender. The receiving antenna is mounted on two arms with two rotary stages as shown in Fig. 6b. Thus, it is possible to measure nearly a complete sphere (except the part, which is blocked from the table) around the AUT. Since a second VNA extender module would be too bulky and heavy to be mounted behind the receiving antenna, a kind of self-developed extension module is used. As it can be seen in Fig. 26, the LO output signal of the port 2 of the VNA is amplified and transmitted to a mixer which is directly attached to the receiving antenna. The back path IF signal is transmitted back to the VNA so that the extension module behaves like a commercially available module in receiving mode. For a more convenient transmission of the LO and IF signal along the rotating arms, two diplexers are used to transmit the signals trough the same cable. Thus, the use of only one cable is sufficient. The distance between the AUT and the receiving antenna is fixed to 75 cm for D- and H-bands.

The setup of the Institute of Infocomm Research (I^2R) was introduced in 2014 (Qing and Chen 2014) and actually consists of two different setups for D-band and H-band, respectively. The D-band setup is shown in Fig. 27a, b. Like the KIT setup, it consists of an Agilent PNA-X with OML extension modules. However, in this setup, OML modules are used for both ports of the VNA. The setup is capable to measure two vertical cutting planes. The $\varphi=0°$ plane can be measured from $-90°$ to $+90°$, and the $\varphi=0°$ can be measured from $-90°$ to about $+35°$. The setup can measure probe-fed as well as waveguide-fed AUTs. The setup for H-band is shown in Fig. 27c, d. The electrical components are the same as the ones used in the D-band setup but with adapted OML extension modules. The mechanical setup, however, is different. It can only measure one horizontal plane from $-90°$ to $+90°$ and up to now only waveguide-fed antennas.

The setup of the University Nice Sophia Antipolis (UNICE), EpOC Lab, has been introduced in 2011. It was developed at V-band setup and has been extended to F-band (90−140 GHz) in the meantime. It consists of an Agilent PNA-X with a frequency extension module from Virginia Diodes as source and an HP spectrum analyzer as receiver, which uses an external mixer from OML. To move the receiving antenna around the AUT, it uses two rotating arms. Thus, it is able to measure also three-dimensional radiating patterns. The distance between AUT and receiving antenna is fixed to 20 cm. In contrast to the two previous described setups, radiation efficiency and circular polarization performance measurement methods have been specifically developed. The setup is depicted in Figs. 8 and 9. Table 6 recapitulates the properties of those three setups.

DOI 10.1007/978-981-4560-75-7_76-2

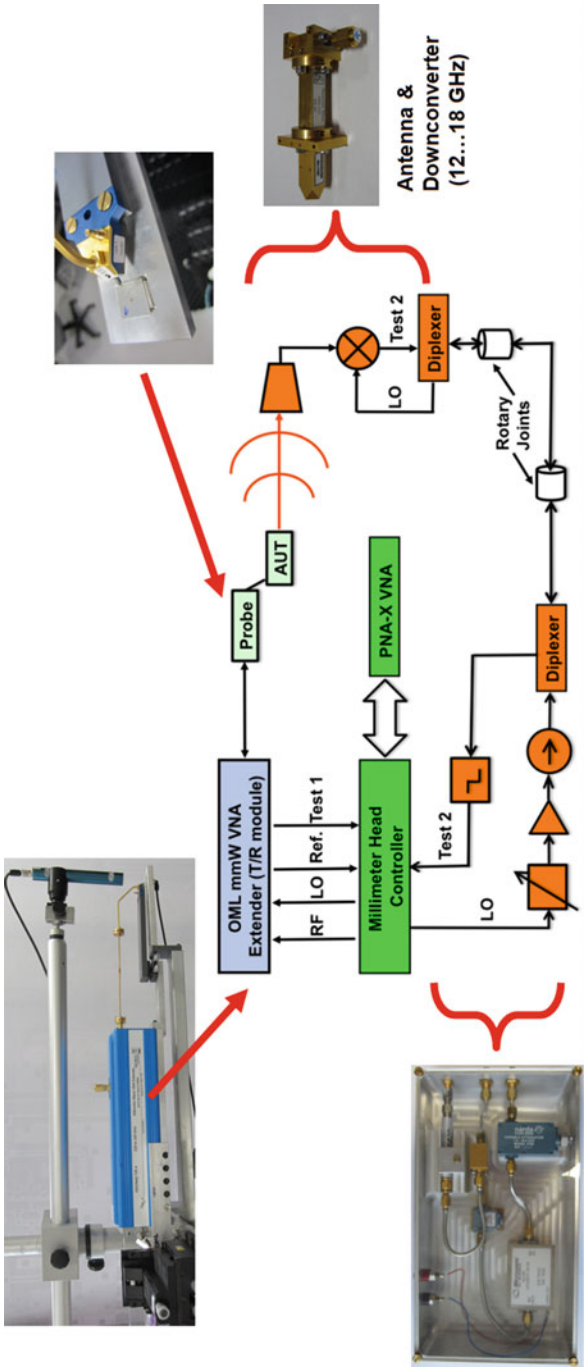


Fig. 26 Block diagram and selected components of the KIT measurement setup (Gulan et al. 2013)

DOI 10.1007/978-981-4560-75-7_76-2

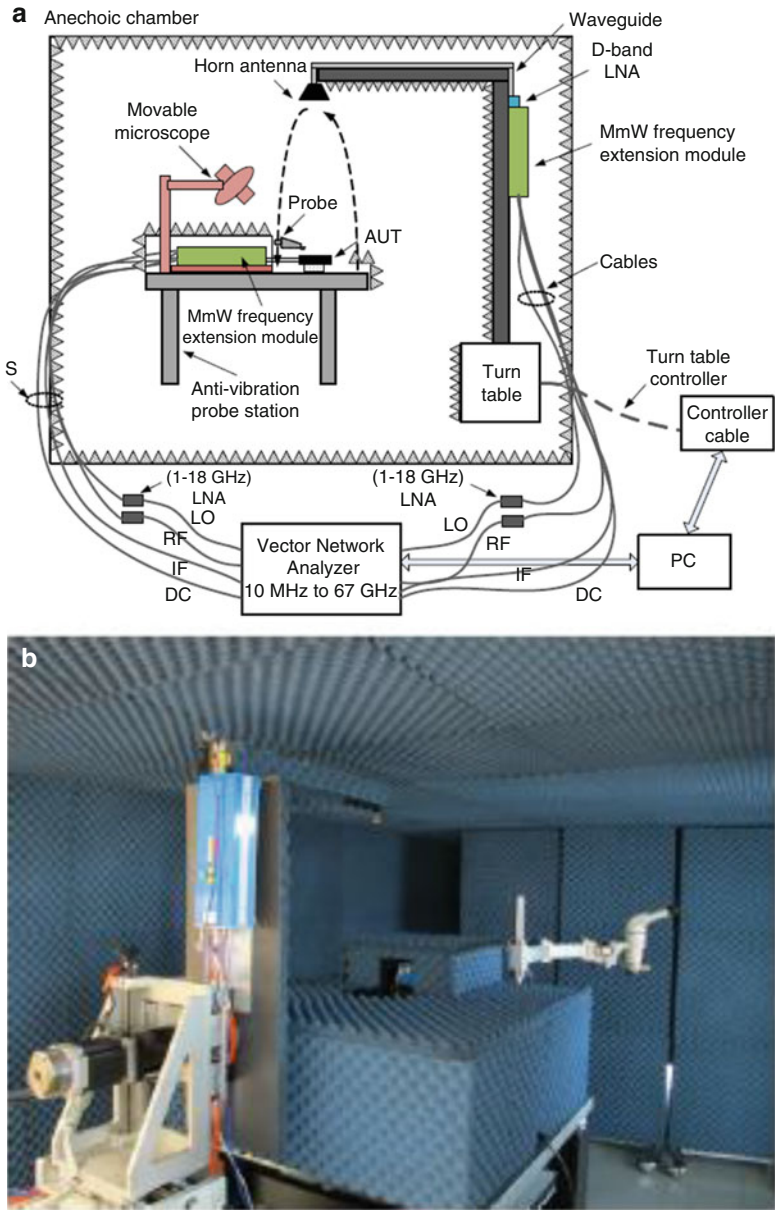


Fig. 27 (continued)

DOI 10.1007/978-981-4560-75-7_76-2

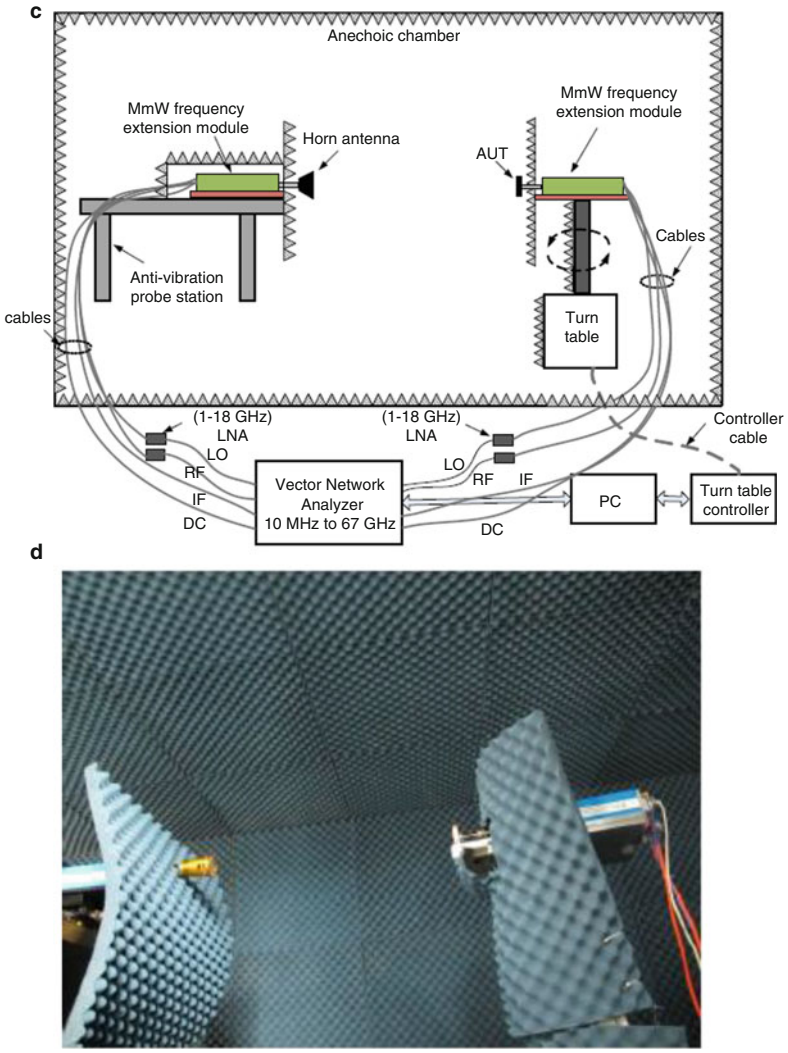


Fig. 27 Block diagram and pictures of the I^2R measurement setup (From Qing and Chen 2014)

Sources of Errors

In the previous sections, several measurement setups (and associated calibrations) using microelectronic probes to feed the AUT have been presented. The main sources of errors usually come from the calibration accuracy. For example, in the V-band, the setup of University Nice Sophia Antipolis is given within ± 0.8 dB accuracy. In the D-band, this setup gives ± 1.2 dB accuracy. Those numbers have been obtained from the uncertainty of every block in the mm-value chain (transmit, receive, and free-space loss). However, there are also other contributors to the global error produced during a measurement. The microelectronic probe is still a huge metallic part which is closely placed to the AUT. Depending on the way the AUT radiates, some power will definitely hit the microelectronic probe to be scattered, appearing in the opposite direction of the AUT and generating unwanted ripples in the radiation pattern. In Zwick and Pfeiffer (2009), the radiation of an unloaded probe is measured, and a maximum of -15.5 dBi radiation is found. This value is expected to be lower when the probe is loaded by the AUT that is mostly by a radiator having almost 50 Ω input impedance. The 3-D pattern of a loaded probe was measured in Ranvier et al. (2009). The average gain was computed to be around -28 dBi and the maximum gain around -22 dBi. The proposed solution was to cover the microelectronic probe with absorbers, improving the radiation by 2 dB, but so far, no global solution was proposed to this problem.

DOI 10.1007/978-981-4560-75-7_76-2

Table 6 Capabilities of three selected measurement setups above 110 GHz

Setup	Frequency range	Source	Extension modules	Measurement possibilities	Antenna polarization
KIT	D-/H-band	Agilent PNA-X	OML	Gain/3-D radiation pattern	Linear
IIR	D-/H-band	Agilent PNA-X	OML	2-D radiation pattern	Linear
UNICE	F-band	Agilent PNA-X	VDI	Gain/3-D radiation pattern/ efficiency	Linear/circular

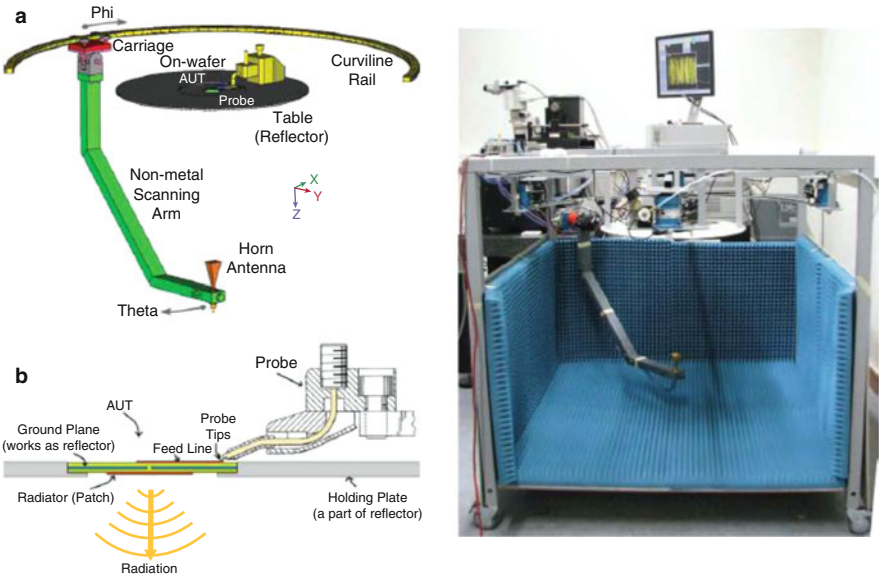


Fig. 28 Schematic and picture of the measurement setup proposed in Mohammadpour-Aghdam et al. (2011)

This phenomenon can be treated in different ways: block the radiation of the probe or including it into the calibration procedure. In Mohammadpour-Aghdam et al. (2011), the RF probe was modeled using a commercial software, and it was demonstrated to have a dipole-like omnidirectional radiation pattern. The authors proposed to shield this radiation (from the region that has to be scanned) by placing a metallic reflector plate behind the microelectronic probe. Following this idea, a robust mechanical setup was proposed by the authors (Fig. 28). The sensitivity was improved by 20 dB. However, only the lower hemisphere can be measured, with an AUT radiating only in the opposite direction of the probe.

In Fu et al. (2011), Tsai et al. (2013), and Reniers et al. (2014), several custom probes have been used. The probe tip feeder is extended to increase the distance between the AUT and the body of the microelectronic probe. This solution can reduce the back scattering but also the blocking by the fixture. The probe body can also be reversed as in Fu et al. (2011). However, those custom probes are more expensive than regular probes and provide higher insertion loss.

In Murdock et al. (2011), a probe radiation de-embedding technique is proposed, but two measurements are needed instead of only one.

DOI 10.1007/978-981-4560-75-7_76-2

Future Directions and Open Problems

Today, lots of academics working at mm-wave frequencies have implemented their own measurement system and testing techniques for probe-fed antennas at 60 GHz (in fact, until 67 GHz due to the available PNAs and mm-wave head). In this characterization domain, the main difficulties come from the fact that two engineering worlds are colliding: the antenna world and the microelectronic world. For example, antenna researchers or engineers are usually not aware of probing on silicon, and microelectronic researchers or engineers do not know how to probe on organic substrates usually needed for antenna solutions. The main consequence is the following: every measurement system is custom made and does not look like the one of the neighbor as every academic is gaining knowledge by trial and errors with his own experiments. In parallel, two commercial measurement systems have been released on the market by antenna companies. The lack of microelectronic competencies from those companies makes their measurement systems perfectible, but they are brand new, which implies that the feedback from their users will obviously allow those commercial companies to improve their proposed measurement systems.

In the future, this current situation (the distribution of the measurement setups) will definitely evolve if some applications strongly emerge around 60 GHz. Two markets are currently hot topics around 60 GHz: WiGig applications and backhaul/fronthaul mm-wave links. It will probably be possible to avoid probe-fed feeding to characterize backhaul/fronthaul antenna solutions: SMPM connectors offer today appealing performance, and they are even used at the initial stage of an antenna concept when no functional electronic hardware is yet available. However, the measurements of WiGig probe-fed antenna modules will be of paramount importance especially at the prototyping and mock-up stages of the product. In fact, if the WiGig standard really emerges, the certification of every product to be released on the market will be mandatory. In this case, the required measurements (with functional hardware), performance, and safety threshold will be detailed. Manufacturers of commercial measurement setups will for sure develop the exact needed facility to be certified. Indeed, we are not speaking about probe-fed measurements as functional hardware will be available. As the mobile phone industry at low frequencies, the product to be characterized will be powered up with no physical connection, but it will be wirelessly linked with a "small-cell" emulator to retrieve its characteristics. However, the whole antenna community will benefit from those newest ups because the involved companies will definitely propose a probe-fed option available with their setup, as the probe-fed characterization of the antenna module during the prototyping and mock-up stages will be necessary to avoid any deviation from expected performance. If no standardized application emerges, it is not sure if commercial measurement setups will rise up and dump the market. The specific absorption rate (SAR) testing and certification area is a good example of what could happen or not.

Building a functional probe-fed measurement setup above 110 GHz is today a hot and challenging topic for many academics. As explained in the previous section, only few players entered this game as it is expensive and not straightforward. It should be noted that only researchers with competencies at 60 GHz made some attempts. There is an interest for the available 120−140 GHz and the 200−280-GHz frequency bands, but so far, commercial companies are a bit shy for the moment as they do not yet see any immediate applications. In spherical setup, mechanical stability is the key issue at those frequencies because it will directly translate into magnitude and phase errors. It should be also noted that, today, only waveguide connections exist (microelectronic probes are only waveguide fed) which makes the RF parts of the setup less flexible than at 60 GHz. Therefore, probing an antenna on a foam support is more far difficult than at 60 GHz as the rigidity of the waveguide makes the probing contact difficult to properly ensure. However, the situation could quickly change with the recent introduction of novel connectors (Anritsu 2015). If magnitude and phase acquisition are highly challenging above 110 GHz, planar scanners seem to be the best mechanical solutions. Some results have been even reported by Smith

DOI 10.1007/978-981-4560-75-7_76-2

et al. (2012) and at NASA (Slater 1994), but so far, no probe-fed AUT is described in those papers. If one would like to keep spherical scanning around the AUT when measuring above 110 GHz, phaseless acquisition techniques and appropriate processing for near-field to far-field transformation are probably the best options (Arboleya et al. 2013). More efforts should focus in this direction. The question of available probes above 300 GHz is also of strong interest: interesting innovative solutions have been already designed and measured (Reck et al. 2011a, b). Those probes are fully custom made. Finally, quasi-optical measurement solutions could be an alternative but they need to gain maturity (Neshat et al. 2012; Hirata et al. 2001). When increasing frequency, one main question remains the signal generation being solid state or from photonic solution (Alouini et al. 2014); in this last case, the radiation characterization could be completely different.

Acknowledgment

The authors would like to deeply acknowledge Aimeric Bisognin, Frederic Devillers, Frederic Gianesello, and Romain Pilard.

Cross-References

- ► Anechoic Chamber Design
- ► Antenna Measurement/Setups-Introduction
- ► Millimeter-wave Antennas and Arrays
- ► On-Chip Antennas

References

Akkermans JAG, van Dijk R, Herben MHA (2007) Millimeter-wave antenna measurement. In European microwave conference, Munich, Germany, pp 83–86

Alouini M, Rolland A, Pouget L, Brunel M, Vallet M, Frein L, Loas G, Lampin JF, Ducournau G (2014) Bridging the gap between THz and microwave photonics through optoelectronic generation of interleaved combs. In: International topical meeting on microwave photonics (MWP), Valencia, Spain

Anritsu (2015) Coax connector design above 110 GHz. Microw J 3:28–32

Arboleya A, Alvarez Y, Las-Heras F (2013) Millimeter and submillimeter planar measurement setup. In: IEEE antennas and propagation society international symposium, Orlando, Florida

Asano J, Hirokawa J, Nakano H, Hirachi Y, Isono H, Ishii A, Ando M (2013) Manufacture and performance of a 60GHz-band high-efficiency antenna with a thick resin layer and the feed through a hole in a silicon chip. IEICE Trans Commun 96(12):3108–3115

Balanis C (1997) Fundamental parameters of antennas. In: Balanis C (ed) Antenna theory: analysis and design / Constantine A. Balanis, 2nd edn. Hoboken, NJ: Wiley-Interscience, pp 28–115

Beer S, Zwick T (2010) Probe based radiation pattern measurements for highly integrated millimeter-wave antennas. In: European conference on antennas and propagation, Barcelona, Spain

Beer S, Gulan H, Rusch C, Adamiuk G, Zwick T (2011) A double-dipole antenna with parasitic elements for 122 GHz system-in-package radar sensors. In: European conference on antennas and propagation, Rome, Italy, pp 1903–1906

Bisognin A, Titz D, Ferrero F, Jacquemod G, Rilard R, Gianesello F, Gloria D, Laporte C, Ezzeddine H, Lugara D, Luxey C (2014) Probe-fed measurement system for F-band antennas. In: European conference on antennas and propagation, The Hague, Netherlands

Boone J, Krishnan S, Stefanakos E, Goswami Y, Bhansali S (2011) Coplanar-waveguide-fed folded dipole slot antenna for wireless local area network applications and V-band frequency operations. IET Microw Antennas Propag 6(5):583–587

Brebels S, Bravo N, Libois M, Soens C (2013) Absolute gain measurement of integrated millimeter wave antennas. In: 35th ESA Antenna Workshop. on. Antenna and Free Space RF Measurements. ESTEC, Noordwijk, The Netherlands

Calvez C, Pilard R, Person C, Coupez J, Gallee F, Gianesello F, Ezzeddine H, Gloria D (2011) Millimeter-wave antenna designs for 60 GHz applications: SoC and SiP approaches. Int J Microw Wirel Technol 3(2):179–188

Cascade (2015) Webpage http://www.cmicro.com/files/Probe-Selection-Guide.pdf. Retrieved 10 Apr 2015

Cheng S, Yousef H, Kratz H (2009) 79 GHz slot antennas based on substrate integrated waveguides (SIW) in a flexible printed circuit board. IEEE Trans Antennas Propag 57(1):64–71

Chin K, Jiang W, Che W, Chang C, Jin H (2014) Wideband LTCC 60-GHz antenna array with a dual-resonant slot and patch structure. IEEE Trans Antennas Propag 62(1):174–182

Fakharzadeh M (2014) Antenna measurement in probe station. In: 2014 IEEE International Symposium on Antennas and Propagation, Memphis, Tennessee, U.S.A.

Fakharzadeh M, Nezhad-Ahmadi M, Biglarbegian B, Safavi-Naeini S (2011) Efficiency measurement of millimeter wave on-chip antennas. In: 2011 IEEE International Symposium on Antennas and Propagation, Spokane, Washington, U.S.A, pp 3141–3143

Fu Y, Vuong T, Dussopt L, Ndagijimana F (2011) Characterization of integrated antennas at millimeter-wave frequencies. In: European conference on antennas and propagation, Rome, Italy, pp 160–163

GGB (2015) Webpage http://www.ggb.com. Retrieved 10 Apr 2015

Gulan H, Beer S, Diebold S, Rusch C, Leuther A, Kallfass I, Zwick T (2013) Probe based antenna measurements up to 325 GHz for upcoming millimeter-wave applications. In: 2013 IEEE International Symposium on Antennas and Propagation, Orlando, Florida, U.S.A.

Hirata R, Ishii H, Nagatsuma T (2001) Design and characterization of a 120 GHz millimeter-wave antenna for integrated photonic transmitters. IEEE Trans Microwave Theory Tech 49(11):2157–2162

Hitzler M, Bader S, Waldschmidt C (2014) Key aspects of robot based antenna measurements at millimeter wave frequencies. In: European conference on antennas and propagation, The Hague, Netherlands, pp 392–396

Hong W, Goudelev A, Baek K, Arkhipenkov V, Lee J (2011) 24-element antenna-in-package for stationary 60-GHz communication scenarios. IEEE Antennas Wirel Propag Lett 10:738–741

Hou D, Xiong Y, Goh W, Hu S, Hong W, Madihian M (2012) 130-GHz on-chip meander slot antennas with stacked dielectric resonators in standard CMOS technology. IEEE Trans Antennas Propag 60(9):4102–4109

Hsu S, Wei K, Hsu C, Ru-Chuang H (2008) A 60-GHz millimeter-wave CPW-Fed Yagi antenna fabricated by using 0.18-μm CMOS technology. IEEE Electron Device Lett 29(6):625–627

IEEE Standard Test Procedures for Antennas (1979), ANSI/IEEE 149

Ito T, Tsutsumi Y, Obayashi S Shoki H, Morooka T (2009) Radiation pattern measurement system for millimeter-wave antenna fed by contact probe. In: European microwave conference, Rome, Italy, pp 1543–1546

Janse Van Rensburg D, Gregson S (2014) Parametric study of probe positioning errors in articulated spherical near-field test systems for mm-wave applications. In: IEEE conference on antenna measurements & applications focus on antenna systems Antibes, France

Janse Van Rensburg D, Hindman G (2008) An overview of near-field sub-millimeter wave antenna test applications. In: COMITE conference, Prague, Czech Republic

Johannsen U, Smolders A, Remiers A, Dommele A, Huang M (2012) Integrated antenna concept for millimeter-wave. In: European conference on antennas and propagation, The Hague, Netherlands pp 2560–2563

Kildal P, Rosengren K (2004) Correlation and capacity of MIMO systems and mutual coupling, radiation efficiency, and diversity gain of their antennas: simulations and measurements in a reverberation chamber. IEEE Commun Mag 42(12):104–112

Lamminen A, Säily J, Vimpari A (2008) 60-GHz patch antennas and arrays on LTCC with embedded-cavity substrates. IEEE Trans Antennas Propag 56(9):2865–2873

Lanteri J, Dussopt L, Pilard R, Gloria D, Yamamoto S, Cathelin A, Hezzeddine H (2010) 60 GHz antennas in HTCC and glass technology. In: European conference on antennas and propagation, Barcelona, Spain

Li Y, Chen Z, Qing X, Zhang Z, Xu J, Fend Z (2012) Axial ratio bandwidth enhancement of 60-GHz substrate integrated waveguide-fed circularly polarized LTCC antenna array. IEEE Trans Antennas Propag 60(10):4619–4626

Liu C, Guo Y, Bao X, Xiao S (2012) 60-GHz LTCC integrated circularly polarized helical antenna array. IEEE Trans Antennas Propag 60(3):1329–1335

Maltzev A, Perahia E, Maslennikon R, Sevastyanov A, Lomayev A, Khoryaev A (2010) Impact of polarization characteristics on 60-GHz indoor radio communication systems. IEEE Antennas Wirel Propag Lett 9:413–416

McKinzie WE III (1997) A modified wheeler cap method for measuring antenna efficiency. In: 2007 IEEE International Symposium on Antennas and Propagation, Honolulu, Hawai'i, USA, pp 542–545

Mohammadpour-Aghdam K, Brebels S, Enayati A, Faraji-Dana R, Vandenbosch G, DeRaedt W (2011) RF probe influence study in millimeter-wave antenna pattern measurements. Int J RF Microwave Comput Aided Eng 21(4):413–420

Murdock J, Ben-Dor E, Gutierrez F, Rappaport T (2011) Challenges and approaches to on-chip millimeter wave antenna pattern measurements. In: IEEE MTT-S international microwave symposium, Baltimore, MD, USA

Neshat M, Hailu D, Nezhad-Ahmadi M, Rafi G, Safavi-Naeini S (2012) Gain measurement of embedded on-chip antennas in mmW/THz range. IEEE Tran Antennas Propag 60(5):2544–2549

Ng K, Wong H, So K, Chan C, Luk K (2012) 60 GHz plated through hole printed magneto-electric dipole antenna. IEEE Trans Antennas Propag 60(7):3129–3136

NSI-700S-360 http://ww2.nearfield.com/Sales/datasheets/NSI-700S-360.htm. Retrieved 10 Apr 2015

Orbit Presentation http://www.keysight.com/upload/cmc_upload/All/19March15WebcastSlides.pdf?&cc=CH&lc=ger. Retrieved 10 Apr 2015

Pan B, Li Y, Ponchak G, Tentzeris M, Papapolymerou J (2008) A low-loss substrate-independent approach for 60-GHz transceiver front-end integration using micromachining technologies. IEEE Trans Microwave Theory Tech 56(12):2779–2788

Pilard R, Montusclat S, Gloria D, Le Pennec F, Person C (2009) Dedicated measurement setup for millimetre-wave silicon integrated antennas: BiCMOS and CMOS high resistivity SOI process characterization. In: European conference on antennas and propagation, Berlin, Germnay, pp 2447–2451

Pilard R, Titz D, Gianesello F, Calascibetta P, Riviere JM, Lopez J, Coffy R, Saugier E, Poulain A, Ferrero F, Luxey C, Brachat P, Jacquemod G, Gloria D (2012) HDI organic technology integrating

built-in antennas dedicated to 60 GHz SiP solution. In: IEEE international symposium on antennas and propagation, July 8-14, 2012, in Chicago, IL, USA

Purcell EM (1943) A method for measuring the absolute gain of microwave antennas (RL-41-9). Technical reports of Radiation Laboratory, Massachusetts Institute of Technology

Qing X, Chen ZN (2014) Measurement setups for millimeter-wave antennas at 60/140/270 GHz bands. In: International workshop on antenna technology, Sydney, Australia

Ranvier S, Kyro M, Icheln C, Luxey C, Staraj R, Vainikainen P (2009) Compact 3-D on-wafer radiation pattern measurement system for 60 GHz antennas. Microwave Opt Technol Lett 51(2):319–324

Reck TJ, Chen L, Zhang C, Arsenovic A, Groppi C, Lichtenberger AW, Weikle RM II, Barker NS (2011a) Micromachined probes for submillimeter-wave on-wafer measurements-part I: mechanical design and characterization. IEEE Trans Terahertz Sci Technol 1(2):349–356

Reck TJ, Chen L, Zhang C, Arsenovic A, Groppi C, Lichtenberger AW, Weikle RM II, Barker NS (2011b) Micromachined probes for submillimeter-wave on-wafer measurements-part II: RF design and characterization. IEEE Trans Terahertz Sci Technol 1(2):357–363

Reniers A, Van Dommele A, Huang M, Herben M (2014) Disturbing effects of microwave probe on mm-wave antenna pattern measurements. In: European conference on antennas and propagation, The Hague, Netherlands, pp 247–250

Rousstia M, Herben M (2013) High performance 60-GHz dielectric rod antenna with dual circular polarization. In: European radar conference, Nuremberg, Germany, pp 359–362

Shamim A, Roy L, Fong N, Tarr N (2008) 24 GHz on-chip antennas and balun on bulk Si for air transmission. IEEE Trans Antennas Propag 56(2):303–311

Simons R (2002) Novel on-wafer radiation pattern measurement technique for MEMS actuator based reconfigurable patch antennas. In: 24th Annual antenna measurement techniques association meeting and symposium, Cleveland, 3–8 Nov 2002

Simons R, Lee R (1997) Characterization of miniature millimeter-wave vivaldi antenna for local multipoint distribution service. In: ARFTG conference, Portland, Oregon, USA, pp 95–100

Simons R, Lee R (1999) On-wafer characterization of millimeter-wave antennas for wireless applications. IEEE Trans Microwave Theory Tech 47(1):92–96

Simons R, Chun D, Katehi L (2001) Reconfigurable array antenna using microelectromechanical systems (MEMS) actuators. In: IEEE antennas and propagation society international symposium, Boston, Massachusetts, pp 674–677

Slater D (1994) A 550 GHz near-field antenna measurement system for the NASA submillimeter wave astronomy satellite. In: Antenna Measurement Techniques Association Conference. Norwood, Massachusetts

Smith S, Archer J, Timms G, Smart K, Barker S, Hay S, Granet C (2012) A millimeter-wave antenna amplitude and phase measurement system. IEEE Trans Antennas Propag 60(4):1744–1757

Sun M, Zhang Y, Guo Y, Karim M, CHuen O, Leong M (2011) Integration of circular polarized array and LNA in LTCC as a 60-GHz active receiving antenna. IEEE Trans Antennas Propag 59(8):3083–3089

Titz D, Ferrero F, Luxey C (2012a) Development of a millimeter-wave measurement setup and dedicated techniques to characterize the matching and radiation performance of probe-fed antennas. IEEE Antennas Propag Mag 54(4):188–203

Titz D, Kyrö M, Ferrero F, Ranvier S, Luxey C, Brachat P, Jacquemod G, Vainikainen P (2012b) Calibration methodology for a 3D radiation pattern set-up of probe-fed millimeter-wave antennas. Microwave Opt Technol Lett 54(5):1183–1189

Titz D, Ferrero F, Brachat P, Jacquemod G, Luxey C (2012c) Efficiency measurement of probe-fed antennas operating at millimeter-wave frequencies. IEEE Antennas Wirel Propag Lett 11:1194–1197

Titz D, Pilard R, Gianesello F, Ferrero F, Luxey C, Brachat P, Jacquemod G, Gloria D (2012d) Industrial HTCC SiP solution for 60 GHz applications. IEEE Antennas Wirel Propag Lett 11:576–579

Titz D, Pilard R, Ferrero F, Gianesello F, Luxey C, Jacquemod G (2014) Improved measurement accuracy of probe-fed Mm-wave antennas using the three Γ method. IEEE Antennas Wirel Propag Lett 13:103–105

Tsai K, Yeh L, Kuo P, Chuan H (2010) Design of 60-GHz CPW-fed CMOS on-chip integrated antenna-filter. In: European conference on antennas and propagation, Barcelona, Spain

Tsai Z, Wu Y, Chen S, Lee T, Wang H (2013) A V-band on-wafer near-field antenna measurement system using an IC probe station. IEEE Trans Antennas Propag 61(4):2058–2067

μ-Lab Station http://www.orbitfr.com/content/products/micro-lab. Accessed 19 Sep 2015

Van Caekenberghe K, Brakora K, Hong W, Jumani K, Liao D, Rangwala M, Wee Y, Zhu X, Sarabandi K (2008) A 2–40 GHz probe station based setup for on-wafer antenna measurements. IEEE Trans Antennas Propag 56(10):3241–3247

Wi-Fi http://www.wi-fi.org/discover-wi-fi/wigig-certified. Retrieved 10 Apr 2015

Willmot R, Kim D, Peroulis D (2009) A Yagi–Uda array of high-efficiency wire-bond antennas for on-chip radio applications. IEEE Trans Microwave Theory Tech 57(12):3315–3321

Xu J, Chen Z, Qing X, Hong W (2013) 140-GHz TE20-mode dielectric-loaded SIW slot antenna array in LTCC. IEEE Trans Microwave Theory Tech 61(4):1784–1793

Yeap S, Chen Z, Qing X (2011) Gain-enhanced 60-GHz LTCC antenna array with open air cavities. IEEE Trans Antennas Propag 59(9):3470–3473

Zhang B, Zhang Y, Titz D, Ferrero F, Luxey C (2013) A circularly-polarized array antenna using linearly-polarized sub grid arrays for highly-integrated 60-GHz radio. IEEE Trans Antennas Propag 61(1):436–439

Zhou R, Liu D, Xin H (2009) Design of circularly polarized antenna for 60 GHz wireless communications. In: European conference on antennas and propagation, Berlin, Germany, pp 3787–3789

Zwick T, Baks C, Pfeiffer U, Liu D, Gaucher B (2004) Probe based MMW antenna measurement setup. In: IEEE antennas and propagation society international symposium, Monterey, California, pp 747–750

Zwick T, Pfeiffer U (2009) Special millimeter-wave measurement techniques. In: Liu D, Pfeiffer U, Grzyb J, Gaucher B (eds) Advanced millimeter-wave technologies: antennas, packaging and circuits, 1st edn. Wiley, Chichester, pp 729–769

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_81-1

EMI/EMC Chamber Design, Measurement, and Instrument

Boon Kuan Chung*
Faculty of Engineering and Sciences, Universiti Tunku Abdul Rahman, Kuala Lumpur, Malaysia

Abstract

An EMI/EMC anechoic chamber represents a substantial investment. A number of careful considerations must be weighted in an anechoic chamber project. The screened room must be designed to provide an environment free of extraneous signals. The suitable type of RF absorber must be chosen to line the entire inner surface of the shielded room in order to simulate a free-space environment with no reflection from the walls, ceiling, and floor. The test site must meet the required performance specified in EMC standards. For frequency range below 1 GHz, the normalized site attenuation (NSA) method is specified by CISPR 16 and ANSI C63.4 for validation of semi-anechoic chamber where the floor is not covered with RF absorber. The reference site method is recently added to the standards as a better option to improve the site validation accuracy. For frequency range above 1 GHz, EMC radiated emission measurements require the use of full anechoic chamber where the floor is also covered with RF absorber. The site voltage-standing-wave ratio (SVSWR) method is specified by CISPR 16-1-4 for test site validation. A time-domain reflectivity (TDR) method proposed by ANSI C63.4 offers numerous benefits compared to the CISPR method. It can produce equivalent values of SVSWR without physically moving the antenna, potentially more accurate in validating the chamber quietness performance, and the measurement process is much less time consuming.

Keywords

Electromagnetic; EMI; EMC; Interference; Compatibility; Radiated emission; Immunity; Compliance; Product certification

Introduction

Electromagnetic radiation occurs whenever there is an acceleration of electric charges. Hence all electrical devices may radiate RF wave either intentionally or unintentionally. If the radiation is strong enough, it may have some effects on the operation of another electrical device located nearby. When this occurs unintentionally, it is known as electromagnetic interference (EMI).

EMI noise may be coupled from the source to the victim through electromagnetic fields in the space by mean of radiated coupling. In addition, if two appliances share the same power source (such as the main outlets within a building), EMI noise generated by an electrical appliance may appear on the power leads of the other by means of conducted coupling.

Electromagnetic compatibility (EMC) refers to the capability of two or more electrical devices to operate simultaneously without mutual interference. An electrical device must not emit too much EMI noise into the environment. It must also be able to tolerate a reasonable amount of inflicted noise and

*Email: chungbk@utar.edu.my
*Email: bkchung1429@gmail.com

DOI 10.1007/978-981-4560-75-7_81-1

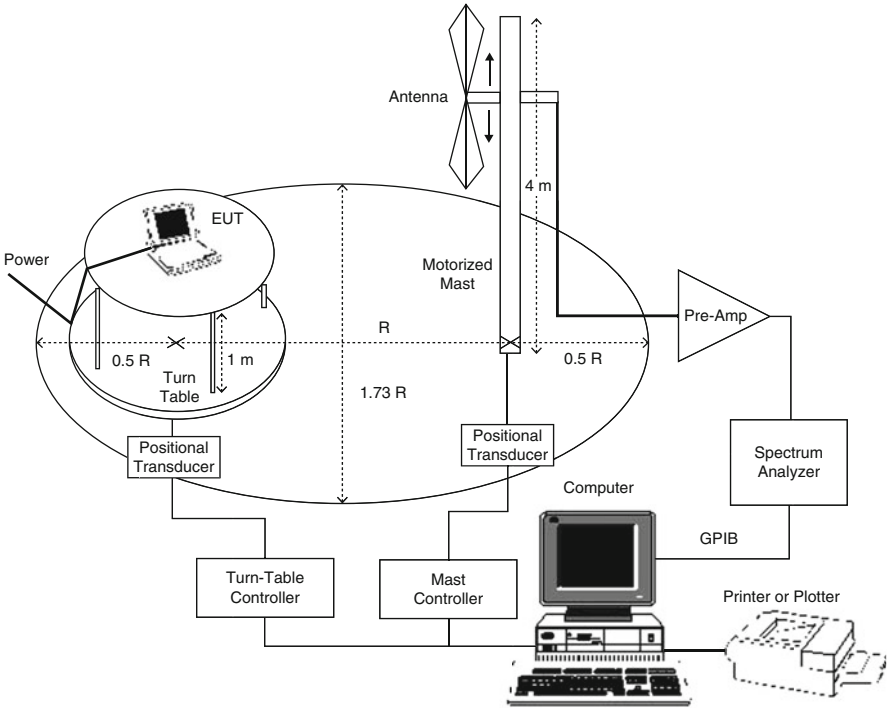


Fig. 1 Typical instrumentation for EMC radiated emission test

function properly in its intended electromagnetic environment. Due to the increasing EMI noise pollution to the environment, EMC standards and regulations have been established worldwide to regulate the absolute limits for both radiated and conducted emissions from electrical devices and the immunity of the devices to EMI emissions.

The International Electrotechnical Commission (IEC) has formed two technical committees devoted to drafting of EMC standards. The TC77 is responsible for basic standards on electromagnetic compatibility between equipment including networks. The CISPR (International Special Committee on Radio Interference) is responsible for setting limits and measurement methods of radio interference characteristics of potential interference sources.

CENELEC (European Committee for Electrotechnical Standardization) is responsible to produce EMC standards for European countries. For telecommunications equipment, ETSI (European Telecommunications Standards Institute) is responsible to produce the standards. Generally, CENELEC and ETSI use IEC/CISPR publications as the basis for preparation of European EMC standards. For example, the EN55022 (Limits and methods of measurement of radio disturbance characteristics of information technology equipment) is equivalent to CISPR Publication 22.

EMC requirements are regulated by the FCC (Federal Communications Commission) in the USA. The technical requirements are detailed in Code of Federal Regulations CFR 47, Part 15 and Part 18. The American standard for EMC measuring instrumentation and test methods is ANSI C63.4. It is roughly equivalent to CISPR Publication 16 in most areas.

Open area test site (OATS) is specified in all the EMC standards for conducting the radiated emission test, in the frequency range of 30–1000 MHz. The OATS is a large outdoor area with a perfectly conducting ground plane and free of obstacles and interfering ambient fields. The test site may be enclosed with nonconductive materials such as canvas or air-supported plastic for protection against weather. The instrumentation is as shown in Fig. 1. The large conductive floor is required so that radiation directed toward the floor can be reflected to the receiving antenna. The sum of the reflected wave and the

DOI 10.1007/978-981-4560-75-7_81-1

wave traveling directly from the equipment under test (EUT) to the receiving antenna will be measured. As the motorized mast carries the receiving antenna to move from 1-m to 4-m height, the conductive floor may cause the electric field amplitude to double at certain angles, due to constructive interference of the direct wave and ground-reflected wave. The radiated emissions must be measured at all directions around the EUT. This is achieved by using a nonconductive rotating turntable as the pedestal for the EUT and a fixed location of receiving antenna. For an automated measurement system, a broadband antenna (such as biconical antenna and log periodic antenna) should be used so that a high-speed scanning over a wide frequency range can be accomplished while the turntable is rotating continuously. This is to search for the EUT's angle and antenna height in which the radiated emission is the highest. The antenna polarization is arranged in the vertical orientation in a sequence of measurements. It is then arranged for horizontal polarization in another sequence of measurements. The EUT must be in its complete system configuration, with all interconnected peripheral equipment connected such that the emission is maximized. The RF field amplitude is then measured by the receiver which is basically a spectrum analyzer. The computer is programmed to automate the measurement process.

The measuring distance, R, between EUT and receiving antenna can be 3 m, 10 m, or 30 m depending on the device class of the EUT. The distance is defined from the nearest perimeter of the EUT to the reference point of the antenna. The ellipse of Fig. 1 defines the area that must be flat and free of reflecting obstacles. The control room where the test instrumentation is housed must be some distance away from the site or constructed directly below the ground plane. The 10-m range is by far the best test environment, whereas the 3-m range has been shown to cause immense problems of measurement uncertainty (Garn et al. 1993).

The characteristics of a minimum standard OATS are defined in CISPR 16 and ANSI C63.4. The required site shall be validated through normalized site attenuation (NSA) measurements at frequencies from 30 to 1000 MHz for both horizontal and vertical polarizations. The instrumentation configuration is the same as the setup for the EUT compliance tests except the EUT is replaced by a known transmitting source. The measured NSA data shall be compared to that calculated for an ideal site. A test site is considered acceptable for radiated emission measurements if the measured results are within ± 4 dB of the theoretical NSA for an ideal site. The ± 4 dB tolerance includes instrumentation calibration errors, measurement technique errors, and errors due to site anomalies. If the measuring distance is 3 m, the mutual impedance correction factors for tuned dipole antennas must be applied to the measurement results.

Absorber-lined semi-anechoic chamber can be used as an alternative site for radiated emission measurements provided it fulfills the NSA requirements. A semi-anechoic chamber is a shielded enclosure with the walls and ceiling covered with RF wave absorbers, except the floor. Hence, there is no RF reflection from the walls and ceiling. It is designed this way in order to emulate the OATS where RF reflection from the conductive floor is required. However, a single-point NSA measurement is insufficient to pick up possible reflections from the walls and ceiling because the RF absorbing materials are not ideal. For semi-anechoic chamber, the test volume is defined as the volume to accommodate the largest equipment to be tested. In validating the site, transmit antenna shall be placed at various points within the test volume with both horizontal and vertical polarizations. The measurements are carried out using broadband antenna. In the European standard, the characteristics of semi-anechoic chamber to be used as an alternative test site are defined in CISPR 16-1-4:2010 (previously defined by EN50147-2:(CENELEC 1997) which was withdrawn in 2008). Semi-anechoic chamber provides a controlled environment with all-weather capability and minimum electromagnetic interference.

On the other hand, the anechoic chamber is often used for radiated immunity measurements according to the EN61000-4-3. In this measurement, a strong RF field is radiated from a transmitting antenna toward the EUT. The standard requires the field at the EUT position to have uniform amplitude distribution from

DOI 10.1007/978-981-4560-75-7_81-1

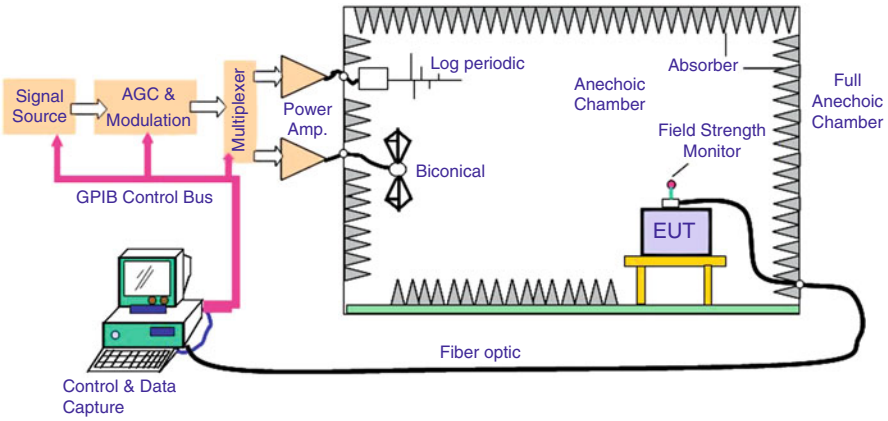


Fig. 2 Radiated RF immunity test setup

80 to 2700 MHz (EN61000-4-3:2006+A2:2010). The field strength at 12 of the 16 points over a grid of plane area (1.5 × 1.5 m with the lowest point at 0.8 m above the floor) must be within the tolerance of −0 dB/ +6 dB. Reflection and resonance effects may cause standing wave in the chamber. In order to meet the field uniformity requirements, the semi-anechoic chamber is converted to a full anechoic chamber by placement of additional RF absorbers on the conductive floor.

A typical radiated immunity test setup is shown in Fig. 2. An electromagnetic field is radiated toward the EUT. Amplitude modulation with 1-kHz signal at 80 % modulation depth is applied and the carrier frequency is swept from 80 to 2700 MHz. Additional test at 900 MHz with 200-Hz pulse modulation may be required by certain product-specific standards. Illumination of the electromagnetic field is applied at four faces of the EUT with both vertical and horizontal polarizations. The applicable electric field strength depends on the expected operating electromagnetic environment. The standard field strengths for severity levels 1, 2, and 3 are 1 V/m, 3 V/m, and 10 V/m, respectively. The most susceptible configuration and operating mode of the EUT are to be set up. A set of parameters that represent the operation of the EUT are monitored from the control room using a CCTV system or other remote monitoring means. For accurate control of the field strength, an isotropic field probe is used to monitor the electric field and adjust the transmitter power level accordingly. The probe is battery powered, and fiber-optic link (instead of electric cable) is used for transferring the measurement data to the control room so that disturbance to the field and EMI leakage to the outdoor space can be minimized.

Changing a semi-anechoic chamber to a full anechoic chamber each time is labor intensive and time consuming. Hence, the study on the use of full anechoic chamber as an alternative site to semi-anechoic chamber for radiated emission measurements was a hot topic of research for some times. The prEN50147-3 (CENELEC 2000) draft standard for emission measurements in fully anechoic chambers was developed by CENELEC. If this test method is accepted as a published standard, a fully anechoic chamber can then be used to perform both radiated emission and immunity tests. However, the draft standard has since been withdrawn in 2012.

EN55022:2010 requires information technology equipment (ITE) shipped into Europe to comply with the new requirement of EMC radiated emission above 1 GHz since October 2011. Full anechoic chamber is specified for the measurement up to 400 GHz. Validation of the test site is to be conducted using the new site voltage-standing-wave ratio (SVSWR) method specified in CISPR 16-1-4 published in February 2007. For the US market, measurement of radiated emission above 1 GHz up to 40 GHz is also required by the FCC. The American National Standards Institute (ANSI) had developed a draft proposal for C63.4 called the time-domain reflectivity (TDR) measurement for validation of test site.

DOI 10.1007/978-981-4560-75-7_81-1

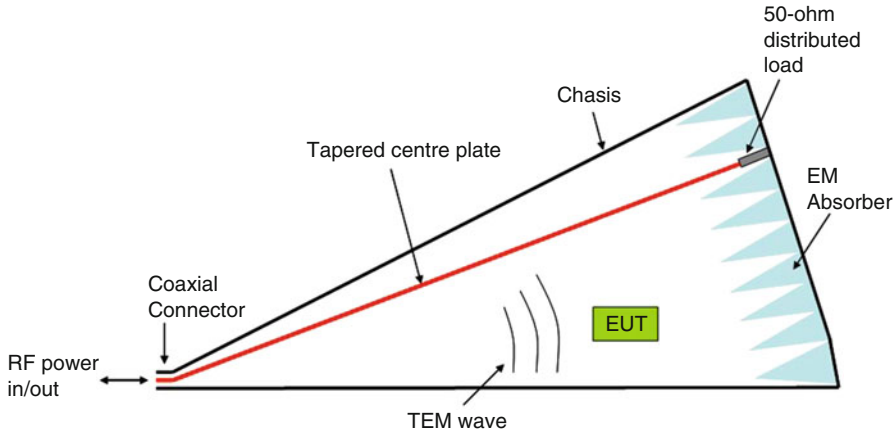


Fig. 3 Geometry of a GTEM cell

EMC Test Facilities

An OATS must be built on a flat terrain free of reflective obstacles such as buildings, fences, etc. The OATS is exposed to the local electromagnetic ambient environment as well as the weather. EUT emissions may be masked off by the background interference if the frequencies coincide. The background interference (from navigational beacons, telemetry, vehicle ignition, machinery noise, etc.) is often unstable or intermittent, and it makes EUT emissions difficult to be identified. Strong background signals such as FM broadcast transmissions may cause compression and false indications in the spectrum analyzer. A countermeasure to reduce the ambient interference, although not very effective, is to build the OATS in a remote area far away from urbanization.

A screened room can eliminate the problem of ambient interference. With suitable RF absorbers covering the walls and ceiling, undesired reflections can be minimized so that the semi-anechoic chamber (SAC) emulates an OATS. Radiated emission measurement results from the SAC are recognized as having a direct correlation to measurement at an OATS.

Radiated immunity testing is required by the European EMC standards. The OATS is not suitable for radiated immunity test because the transmission of high-power broadband signal for such test in an outdoor environment is prohibited by the radio regulations. For a SAC, RF absorbers may be placed on the floor so that the required field uniformity with maximum variation less than 6 dB over the test volume can be achieved. All these circumstances make SAC much more attractive than OATS as EMC test facility.

The process of achieving product compliance in an electronic product development cycle typically involves a pre-compliance scan to identify the radiated emission spectral lines that are potentially noncompliant to EMC regulations. The pre-compliance test is done using lower-cost test facility and equipment that may not fully meet the standard requirements. Based on the pre-compliance test results, corrective action can be undertaken on the electronic product to eliminate any undesirable emissions that might exceed the regulatory limits, and the outcome can be quickly evaluated in a pre-compliance test facility that is more convenient and accessible. The total test time can be minimized and the overall cost can be reduced. The worst-case EUT configuration can be identified for the final test in a full-compliance test laboratory.

With the absence of background interference, a screened room without RF absorber lining can be used to scan for frequencies of any emissions from an electronic device. However, standing wave will be reinforced by reflections between the opposing walls and between the ceiling and the floor, causing cavity resonance to be established in the screened room. Due to the resonant modes, a large change in the field strength may be encountered with a small change in the antenna position. The coupling between the

DOI 10.1007/978-981-4560-75-7_81-1

source and the receiver antenna at slightly different frequencies can be different by up to 60 dB. The antenna impedance and antenna factor will be much different from the calibrated values. Hence, the emission measurements will be highly unrepeatable, and the results cannot be compared with those obtained in a previous measurement or in other environment. The screened room is therefore not suitable for pre-compliance test.

The gigahertz transverse electromagnetic (GTEM) cell has been widely used for radiated emission/immunity test. It is a tapered coaxial structure. One example is shown in Fig. 3. The taper apex is equipped with a coaxial connector for measurement of EUT emission voltage, or launching of RF wave toward the EUT for immunity test. The other end of the structure is a wide bandwidth termination made of resistors and RF absorbers. There are various sizes of GTEM cells. For desktop equipment testing, a GTEM cell can be 25-ft long with a cross section of 11 ft by 13 ft at the flared end to provide the required test zone volume.

In emission test, the EUT is modeled as a sum of electric and magnetic dipole moments (Brunaugh and Osburn 1991) located at the test zone. The dipole moments are assumed to be in phase. Three apex voltages corresponding to three orthogonal rotations of the EUT are measured. The vector sum of these voltages is equal to the sum of the dipole moments. The total radiated power emitted by the EUT can be determined from the propagation constant and the sum of the dipole moments. The analysis assumes that the transverse electromagnetic (TEM) mode is the dominant propagation mode. Higher-order TE and TM modes are weakly coupled to the GTEM cell, and the broadband match provided by the termination suppresses the formation of the higher-order modes.

To predict the radiated emission of the EUT that will be measured at an OATS, the EUT is modeled as a resonant dipole fed with the total radiated power as calculated from the apex voltages. The prediction method takes into account ground reflection and the effect of receiving antenna scanning in height from 1 m to 4 m. The FCC will accept the test data generated from this technique as equivalent to OATS provided the correlation can be established. Emissions are to be measured using both GTEM and OATS methods for the specific product type as a test case. The correlation will allow for the GTEM results to be accepted in the future.

The dominant radiation source of many products is the cables. The ability to orient the cables for maximum emission is required for the EMC test. The major drawback of the GTEM cell is in measuring cable emissions. Orthogonal rotations of a cable are not very practical within the limited space of the GTEM cell. Products that have test setup or cabling dimensions larger than about 1.5 m cannot be accurately tested in a GTEM cell.

Mode-stirred chamber (MSC) or reverberation chamber is another technique for performing radiated emission/immunity tests (Corona et al. 1976). It is a screened room without RF absorbing material so that the MSC acts as a high-Q resonant cavity. The size of the MSC is typically many wavelengths long. The field distribution in the chamber is very complex and rapidly varying with spatial displacement of the probe. At five times the first cavity resonant frequency, the mode spacing in frequency space is very dense. A motorized conducting paddle wheel or vane is located inside the chamber to provide mode-stirring action. A field maximum occurs at each point in the chamber during one rotation of the paddle. The resonant modes cause the field density within the chamber to become nearly constant over the chamber volume. Effectively the gain of any antenna is forced to become unity.

The MSC can be used to measure differences in radiated emission by an EUT when changes to the EUT are made. The property of unity gain results in repeatable measurements since the effects of cable movement are minimized. The total radiated power measured with the MSC is different from the total radiated power that would be generated by the same EUT at an OATS. Hence, the MSC is only good for statistical averaging of data for improvement/degradation comparisons and for radiated immunity tests. Emission measurements made in a MSC cannot be used for certification purposes.

DOI 10.1007/978-981-4560-75-7_81-1

Large semi-anechoic chambers (SAC) with a measurement distance of 10 m between the EUT and the receiving antenna are by far the best test environments for EMC radiated emission test. The required characteristics of the SAC are defined in ANSI C63.4 and CISPR 16-1-4:2010. Most of the European product standards require the emission measurement to be performed at 10-m range. The FCC also requires emission measurement to be performed at 10-m range for Class A devices, while the use of 3-m range is acceptable for Class B devices. Large measurement uncertainty is inherent in the measurement at 3-m range (Garn et al. 1993). In addition to mutual coupling between the source and the receiver, the antenna factors supplied by the antenna manufacturer may be affected by up to 5 dB in the low-frequency range due to interactions with the nearby walls (Hansen et al. 2000). At higher frequencies, reflection from the wooden turntable may cause a dip of as much as 20 dB in the site attenuation frequency response. Despite the problems, the high cost of 10-m range forces most manufacturers of electronic products to invest in 3-m range SAC for pre-compliance tests. Full-compliance 10-m range SACs are only available at a few test laboratories.

The study on the use of full anechoic chamber (FAC) for radiated emission measurements has been rather intensive. It has long been recognized that the presence of ground plane in OATS and SAC introduces a considerable measurement error especially for horizontal polarization measurements below 70 MHz. Another factor that motivated the development of FAC is the requirement to perform radiated immunity test in such facility. Movement of absorber in and out when converting a SAC to a FAC each time is very labor intensive and time consuming, and it reduces the efficiency of the test laboratory.

In a FAC, the height scan of the receiving antenna in radiated emission test is not necessary because there is no reflection path that could introduce phase cancelation of the field at certain antenna height and frequency. The test time can be significantly reduced. The quasi-free-space environment eliminates the need to consider height-dependent antenna factors for the measurements. A method for correlating the FAC test result to OATS measurement has been developed by Ristau D. and Hansen D. (2000). Under far-field conditions, computing free-space field strength for different measurement distance is straightforward. The levels measured at 10-m distance will be 10.46 dB lower than the levels at 3-m distance since the field strength is inversely proportional to distance (i.e., decays at a rate of 1/r). Over a conductive ground plane, this relationship is not valid due to additional reflections from the floor.

EM Shielded Room Design

The ideal environment for making RF wave measurements should be free from electromagnetic wave not directly associated with the measurements. The isolation prevents extraneous energy from masking measurements of RF wave of interest and protects nearby systems from interference caused by electromagnetic energy generated by radiation source used in the measurement instrumentation.

For various types of RF measurements, shielded rooms are widely used to provide the required isolation from the increasingly crowded electromagnetic environment. A room space completely enclosed by conductive plates will be shielded against electromagnetic radiation. However, leakage of electromagnetic waves is unavoidable due to ventilation holes and access door to the shielded enclosure, as well as cables penetrating the shielded enclosure for signal/power interconnection purposes. The shielding effectiveness is also affected by different construction methods which give rise to varying degrees of imperfection. In the design of a shielded room, the considerations should include the purpose of the shield, the type of equipment involved, the measurement of distance, and the sensitivity of the information being processed.

A conductive shield absorbs and reflects the energy of an approaching wave by allowing induced current to flow on its surface. Smooth flow of the induced current is required to provide efficient shielding

DOI 10.1007/978-981-4560-75-7_81-1

effect. The presence of holes or slots diverts the current and reduces the shielding effectiveness. In all practical cases, openings in a shield are necessary for ventilation, wiring, and access door into a screened enclosure. The effects of the holes can be analyzed as slot antennas, and the length of slot perpendicular to the direction of current flow determines the amount of radiation leakage. A good shielding practice must ensure all slots are shorter than a quarter wavelength at the highest operating frequency. Overlapping conductive sheets without electrical bonding or continuity will not eliminate the effects of the slot. High conductivity metallic gaskets (such as wire knit mesh or beryllium copper "finger stock") are commonly used to close the gaps. Some conductive gaskets designed solely for electrostatic protection are not suitable for EMI shielding. Any increase in local impedance will allow EM fields to penetrate and degrade the shielding effectiveness. Conductive paint or caulk is sometimes used to bond the joints.

Specialized "honeycomb" vents are commonly used to facilitate airflow into and out of the shielded enclosure. The openings are formed using a large number of small hollow waveguide tubes welded together in a "honeycomb" fashion. A waveguide has a high-pass response. The cutoff frequency depends on the cross-sectional dimensions. The waveguide cross section must be small enough so that the cutoff frequency is much higher than the highest operating frequency. Below the cutoff frequency, an EM field entering the waveguide will be attenuated exponentially with the waveguide length.

No conductor should pass through the opening hole or waveguide (except nonconductive fiber-optic cable). A wire passing through a hole can cause much more harms than the hole itself. The wire guides the RF field through the hole and then reradiates it into the air space. Feedthrough capacitors, filters, and common-mode choke are commonly mounted directly in the hole to bypass or impede the flow of the interference current in wires intended for low-frequency signal or power. For cables intended to carry high-frequency signals, the outer conductor of the coaxial cable must be bonded all around the hole. Ferrite ring can be fitted on the coaxial cable to suppress common-mode interference current.

The metals commonly used in screened room fabrication are sheet copper, aluminum, or steel. The 24-gauge galvanized/zinc-plated steel sheet is widely used to laminate onto wooden boards. Theoretically, the sheet metals can offer more than 120-dB shielding effectiveness. However, in all practical cases, shielding effectiveness of screened enclosure in excess of 120 dB is rarely to be achieved (Hemming 1992). It is due to the presence of holes/seams, high-impedance discontinuities, and wires/cables passing through holes in the shield. Shielding integrity can only be maintained when doors, vents, filters, and piping are carefully designed and constructed. On the other hand, resonance effects may cause certain areas within a screened room to exhibit poorer performance.

Smaller screened enclosure can be constructed as modular clamp-up structure using prefabricated wooden panels which are laminated with sheet metals. The panels are tied together to form the six-sided free standing room. Metal-to-metal seal between the panels is provided by the clamping unit which comprises two metal straps brought together by a number of screws. The shielding performance can be easily degraded by discontinuity between the panels and the presence of cracks or breaks between the panels. Attentions to meticulous details are critical in the installation work. Electrical continuity requires cleanliness around the edges of the panels. Painting of the shielding panels is not recommended since the paint may be capillary under the metal clamping straps and degrades the shielding properties of the system.

Modular clamp-up construction has the disadvantage that people moving about the enclosure tend to cause the floor panels to become loose as a result of varying loads and constant stresses. Corrosion of the clamped seams will also degrade the shielding effectiveness. It may incur expensive maintenance and recertification. Water may cause the wooden core to swell and damage the shielding integrity. Hence, modular screened room is not suitable for climate that has continuous high humidity.

When the required room size exceeds 240 ft^2, welded construction must be considered. Large modular system may not be practical because expensive steel support structure will be required to carry the weight

DOI 10.1007/978-981-4560-75-7_81-1

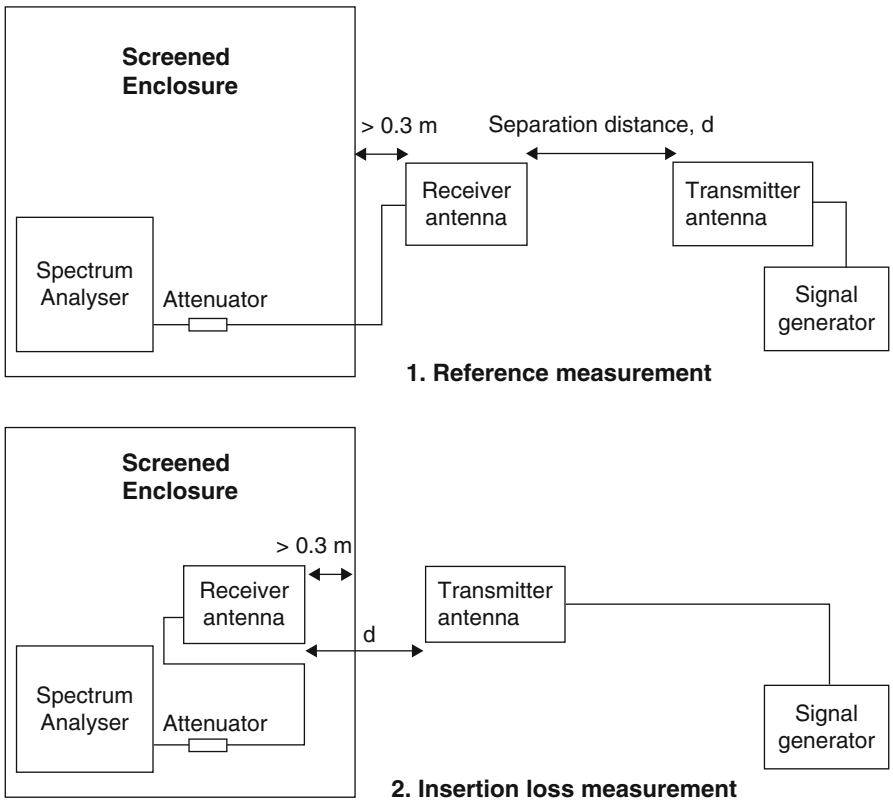


Fig. 4 Method of measuring shielding effectiveness

of the ceiling panels of the screened enclosure. Welded enclosure provides reliable structural and long-term high-performance shielding. Low-carbon steel may be used if reduction of material cost is desirable. However, it has the tendency to rust when exposed to moisture unless it is protected with paints containing a large amount of zinc. Galvanized steel is not recommended because it may produce toxic gas in the welding process.

Many welded shielding enclosures could not meet the expected performance due to lack of attention to proper shielding design and poor field practices. The constructability of the design must be considered for all seams to be properly welded. Sufficient access must be available for welding of corner seams. Well-qualified welders must be engaged to do the work and good quality control must be in place. Metal inert gas (MIG) or shielded metal arc welding process must be used so that high-quality electromagnetic properties can be obtained. The welds must be free of slag, inclusions, gas pockets, worm holes, cracks, or incomplete fusion.

Buckling of the shielding sheet due to heat applied during the seam welding process is difficult to control. It can be reduced using a pan welding technique. The required pan-shaped sheet metal is formed by means of cutting and bending the metal sheet edges. Adjacent pans can be temporarily clamped together for welding at the edge of the pans. The main disadvantages are the amount and cost of shop fabrication and reduced density for transportation.

Butt welding requires no shop fabrication, resulting in lower material cost. Rectangular metal sheets in the largest available size are laid out side by side to be welded together onto a backing structure. For the shield floor, the backing structure can be formed with a network of metal straps or square tubing embedded in concrete slab. Buckling is most troublesome in this welding technique. Tack welding at some intervals along the seam prior to running a continuous weld may ease the buckling effect.

DOI 10.1007/978-981-4560-75-7_81-1

Architectural shielding system where metal foil or sheet metal is built into the walls, floor, and ceiling of existing or new construction is used for application requiring moderate shielding performance. Electrical continuity is maintained by the use of soldering and shielding tapes to join together the metal shields.

The required shielding effectiveness is primarily determined by the operational purpose of the facility. The appearance, HVAC, physical layout, and environmental requirements also depend on the operational purpose. For the construction of 10-m range anechoic chamber for microwave research and EMC tests, a welded shielding enclosure would be the most preferred type.

The methods of measuring the shielding effectiveness (SE) of screened enclosure are specified in MIL-STD-285, IEEE Std 299, EN50147-1, and other government standards (such as NSA 65-6, NSA 73-2A, etc.). The basic principle of the measurements is a method of substitution. A reference reading is made outside the screened enclosure at a prescribed edge-to-edge separation between the transmitter and the receiver antennas. Next, the measurement is performed with the shield barrier placed in between the antennas at the same separation. SE is measured from the decrease of attenuation setting on the receiver which produces the same reference reading. An illustration is given in Fig. 4.

The IEEE Std 299 divides the measurement frequencies into low range (0.009–20 MHz), resonant range (20–300 MHz), and high range (0.3–18 GHz). For the low range, low-impedance magnetic field is measured using 0.3-m-diameter loops placed 0.3 m from the wall. Spot frequency points for shielding measurements are chosen by the owner, and the required levels of SE are specified based on the functions of the screened enclosure. For the resonant range, vertical and horizontal polarization measurements are performed using biconical or dipole antennas placed at separation distance of 2 m. The geometric shape and physical size of the screened enclosure can significantly affect the measurements. Testing at or very near the enclosure resonant frequency shall be avoided. Dipole, horn, biconical, Yagi, log periodic, and other linear antennas are recommended for the high range. Plane-wave test is the most important requirement of a screened enclosure since shielding is primarily needed to provide isolation of ambient electromagnetic interference wave or radiated emissions from the EUT placed in the screened enclosure.

The RF power into the transmitter antenna shall be adequate to obtain the required measurement dynamic range. In all measurements, the receiver shall be located inside the screened enclosure with the transmitter external. This procedure allows the sensitive receiver to be protected from the field of the transmitter antenna and other ambient fields. Ferrite loaded cable shall be used to suppressed the parasitic reflections. Effects from nearby objects or personnel shall be minimized.

The measured insertion loss by method of substitution contains the properties of the shield barrier intermingled with the properties of the measuring circuits (antennas, cabling, attenuator, signal source, and receiver). In general, the properties of the shield barrier cannot be separated from those of the measuring circuits. Hence, the insertion loss cannot be used to estimate the voltage or current induced in an arbitrary internal circuit by an arbitrary external source. One exception is the special case where the source of interference is like the transmitter antenna system used in the SE measurement and the victim circuit to be protected is like the receiver antenna system. The same ambiguity is not experienced by RF circuit designer in characterizing filters, attenuators, connectors, and other RF devices because the devices are measured with 50-Ω terminations and used in 50-Ω circuits. Parasitic coupling of the electromagnetic field with the surrounding devices is usually not an issue of concern especially in the characterization of coaxial devices. In contrast, resonance and other reflective condition effects in shielding measurement may cause a large variation of SE measured at different frequencies.

DOI 10.1007/978-981-4560-75-7_81-1

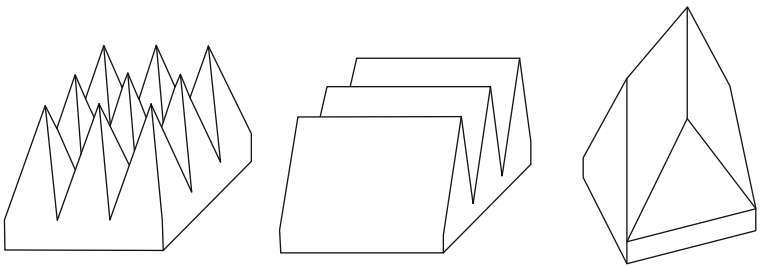


Fig. 5 Absorbers of the shape of (**a**) pyramid, (**b**) wedge, and (**c**) twisted pyramid

RF Absorbers

For antenna, radar, and EMC radiated emission measurements, the shielded enclosure must be sufficiently large so that the receiver can be placed at a far-field distance away from the radiation source. Minimum reflection should take place from the inside of the enclosure walls. Hence, RF absorbing material which has low reflection coefficient and sufficient attenuation is used over the entire inner surfaces of the enclosure to create an anechoic environment. In this case, the strength of the shielding structure must be strong enough to support the weight of the RF absorber. The shielded enclosure must also be sufficiently wide so that the distance between the radiation source and the RF absorber is beyond the reactive-radiative coupling boundary for the absorber to work efficiently.

The choice of RF absorber to be used is a function of the purpose of the anechoic chamber. The required electrical performances as well as the absorber shape, thickness, weight, mounting method, cost, and logistic must be considered. For antenna and radar measurements above 1 GHz, a 24-in.-thick pyramid absorber may be suitable. However, it would not be able to provide a sufficiently low reflectivity at lower frequencies. Standing wave will be reinforced by reflections between the opposing walls and between the ceiling and the floor, causing cavity resonance to be established in the shielded enclosure. The absorber may not give sufficient damping of Q factors of the resonant modes.

Absorbers made of pyramidal shape carbon impregnated urethane foam can provide an extremely broadband performance. The wave impedance gradually tapers from that of free space at the tip of the pyramid to those of a lossy medium at the back surface. However, in order for it to function efficiently, the height of the pyramid must be greater than a quarter wavelength. If the lowest operating frequency is 1 GHz, a 24-in. pyramidal foam absorber may offer a reflectivity level below −40 dB over the frequency range from 1 to 40 GHz at normal incidence. The percentage carbon loading of the foam material can be adjusted to provide optimum performance. For lower frequency of 30 MHz, 6-ft-thick absorbers have been used for construction of EMC anechoic chambers. Although the taper length is several wavelengths for frequencies above 1000 MHz, the long absorbers restricted the application of these chambers to approximately 1000 MHz (German 1982). To reduce the taper length to less than quarter wavelength, a different carbon loading for the foam pyramids must be used to achieve low reflectivity. The impedance taper will be too abrupt if the carbon loading is too high. Consequently, the incident wave will be reflected from the region near the pyramid tips instead of penetrating into the absorber medium. On the other hand, less reflection and more penetration of the incident wave into the absorber medium may result if the carbon loading is too low, but the wave may not be sufficiently absorbed. With a carbon loading to give reflectivity of −20 dB for the 200–1000-MHz range, the pyramids may not provide adequate performance at higher frequencies (Holloway et al. 1997) where the absorber reflectivity must be better than −40 dB for the typical applications in designing antenna and radar cross-section measurement chamber.

Twisted pyramid structure has been used in the fabrication of thick absorbers due to its improved mechanical strength. The pyramid is “twisted” by 45° with respect to the base as shown in Fig. 5. The tips

DOI 10.1007/978-981-4560-75-7_81-1

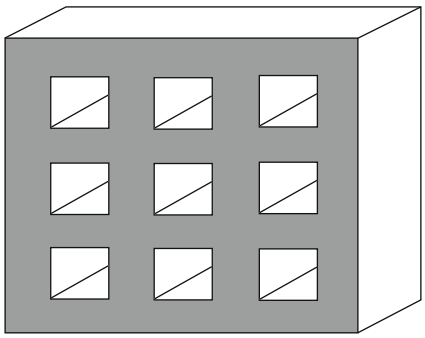


Fig. 6 Ferrite grid RF absorber

of the pyramid will droop less with age than the tips of the standard pyramid. However, some trade-off in its electrical performance will be incurred. Scattering from wedge-shaped absorbers has been reported by DeWitt and Burnside (1988). Better absorbing characteristic compared to pyramidal shape can be obtained for wave directions nearly parallel to the ridge of the wedge.

Resistive material can be made by dispersing microscopic particles of metal in a nonconducting matrix. Hollow pyramid-shaped substrate with a height of quarter wavelength can be coated with the resistive material to make a lightweight wideband absorber. This kind of long pyramid-shaped absorber with adequate mechanical strength has been used for construction of EMC test chamber.

Broadband absorption can also be achieved by increasing the material loss tangent with depth from a flat outer surface. A practical approach is to use multiple layers of carbon impregnated foam. The layers have increasing impedance from the front layer toward the bottom layer, hence achieving an electrical taper that yields a material with good reflectivity level over a wide frequency range. This kind of absorber can be installed at the right-angled corners between two walls or between the wall and the ceiling to absorb electromagnetic wave where long pyramid-shaped absorber is hard to install.

Recently, many EMC test chambers have utilized ferrite tile absorber to obtain low-frequency wideband reflectivity performance. The ferrite tile (NiZn) with a thin profile of 0.25 in. may offer a reflectivity level below −15 dB over the frequency range of 30 to 600 MHz. The reflectivity typically deteriorates rapidly at higher frequencies (Holloway et al. 1997). A ferrite grid is shown in Fig. 6. Combination of the square air section and the ferrite material can be considered as a composite substance having an equivalent complex permittivity and permeability. The filling factor or fraction of space occupied by ferrite can be selected in order to engineer a reflectivity null at a desired frequency. Ferrite grid with a thickness of 0.75 in. may cover the frequency range of 30–1000 MHz (Mayer et al. 1998). Exponentially tapered ferrite grid has been shown to widen the bandwidth to 2400 MHz (Kim et al. 1996).

The ability of ferrite material to absorb electromagnetic wave over a very wide bandwidth is attributed to the availability of ferrite sintering process to precisely control the dielectric constant and complex permeability characteristics of the material. The required relationships are (Mayer and Chaumat 1992)

$$\mu'' \geq 3\mu' \tag{1}$$

$$\frac{d}{\lambda} = \frac{1}{2\pi\mu''} \tag{2}$$

and

$$\varepsilon' = 3\mu' \tag{3}$$

DOI 10.1007/978-981-4560-75-7_81-1

where λ is the wavelength, $\mu = \mu' - j\mu''$ is the complex permeability, $\varepsilon = \varepsilon' - j\varepsilon''$ is the complex permittivity, and d is the thickness of the material. The thickness is much less than a quarter wavelength. As λ decreases with frequency, the material is made to have μ'' decrease at the same rate for relations Eq. 2 to be fulfilled over the desired bandwidth. The wave impedance is close to the impedance of free space, hence allowing the incident wave to penetrate the ferrite tile without significant reflection at the air-to-tile interface. Because the ferrite is magnetically lossy, the penetrating wave is absorbed as it travels through the tile. The wave is subsequently reflected by the metal backing and travels toward the tile/air interface.

The reflectivity of ferrite tiles and grids can be improved by simply adding a dielectric layer between the ferrite and the metal wall (Holloway et al. 1997) of the screened room. Commercial plywood with dielectric constant of about 2 can be used as the ferrite backing. Optimum performance over a wider bandwidth can be achieved by a proper selection of the dielectric thickness.

Ferrite tiles that are effective from 30 to 600 MHz have been combined with thin urethane foam pyramids and wedges that are optimized to operate from 200 to 1000 MHz. This configuration is called a hybrid absorber. It combines the advantages of ferrite tiles and pyramid foam absorbers to cover the EMC frequency range of 30 to 1000 MHz. The carbon loading of the pyramids must be suitably selected such that the low-frequency performance of the ferrite tiles is not degraded.

A general guideline on the selection of RF absorbers has been given by Holloway et al. (1997). For a 10-m SAC, the absorber reflectivity shall be lower than -20 dB at normal incidence throughout the 30–1000-MHz frequency range. Additionally, the reflectivity shall be lower than -15 dB at 45° incident angle. These characteristics may be provided by a 2.5-m pyramidal absorber. Hybrid absorber consisting of a 24-in. pyramid and 0.25-in. ferrite tile may be used to reduce the required screened room size. Ferrite grid (with a thickness of 0.75 in.) and multilayered ferrite tiles are commonly used for construction of 3-m SAC. The hybrid absorber and ferrite grid/tile are very expensive. In addition, their weight (>30 kg/m^2) is generally heavier than foam absorbers. The structural strength of the shielded room must be strong enough to support the weight of the ferrite material. The special room structure adds to the cost of constructing the anechoic chamber.

Small- and medium-size foam absorbers are conventionally glued onto the chamber wall using rubber-based contact adhesive. Pyramid longer than 3 ft is usually supplied with metal base plate to be installed onto a rail-and-clip mounting system on the chamber wall in order to ensure the installation reliability. Ferrite tiles can be installed using glue. However, due to the weight and sharp edges of the tile, overhead net must be installed as safety measure to prevent fallen tile from injuring or killing the personnel who enters the chamber room. A more secure method is to install a plywood panel onto the wall and then fasten the ferrite tile through a center hole onto the plywood panel using self-tapping screws. The gap between adjacent tiles can then be precisely controlled to give better reflectivity performance in contrast to using adhesive in which some glue may get in between the tiles. The plywood panel also acts as the dielectric backing to improve the performance of the ferrite tile.

A good quality absorber alone does not guarantee a good performance for the anechoic chamber in terms of "quietness" of the quiet zone, field uniformity, and site attenuation. The lining of the absorbers must be done carefully to minimize the discontinuity between the neighboring absorber panels. Gaps and misalignment of the pyramids may give a much poorer reflectivity than the manufacturer specification.

In addition to electrical and mechanical specifications of the absorber, the fire retardant property of the absorber must also be considered in order to comply with building safety codes. Lifespan and maintenance needs of the absorber may also be an aspect to consider in the absorber selection criteria.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_81-1

Test Site Validation Method (Below 1 GHz)

Radiated emission measurement results may be affected by any imperfections in the test site and the characteristics of the receiving antenna. Therefore, the test site must be validated before it can be used for EMC compliance tests. The validation method involves a series of normalized site attenuation (NSA) measurements to be carried out for the given transmitter-receiver separation and heights above a flat conductive surface for both horizontal and vertical polarizations. Antenna spacing used for making site attenuation measurements shall be the same as the spacing used for the EUT radiated emission measurements at frequencies from 30 to 1000 MHz. The measured NSA results shall be compared to that calculated for an ideal OATS. A test site is considered acceptable for EMC compliance test if the measured NSA results are within ±4 dB of the theoretical NSA for an ideal site.

Site attenuation is defined as the ratio of the voltage input to a matched and balanced lossless tuned dipole radiator, V_T, to that at the output of a similarly matched balanced lossless tuned dipole receiver antenna, V_R. From the Friis transmission formula, the site attenuation, SA, is given by

$$SA = \frac{V_T}{V_R} = \sqrt{\frac{P_T}{P_R}} = \sqrt{\frac{d^2\lambda^2}{(A_{\text{eff}})_T (A_{\text{eff}})_R}} \tag{4}$$

where d is the antenna spacing, λ is the wavelength, $(A_{\text{eff}})_T$ is the effective area of the transmitter antenna, and $(A_{\text{eff}})_R$ is the effective area of the receiver antenna.

Antenna factor, AF_R, is defined as the ratio of the electric field strength, E_R, to the receiver antenna output voltage, V_R. The relationship between $(A_{\text{eff}})_R$ and AF_R is given by

$$(A_{\text{eff}})_R = \frac{\text{Received Power}}{\text{Power Density}} = \frac{\eta}{R_o} \cdot \frac{V_R^2}{E_R^2} = \frac{\eta}{R_o AF_R^2} \tag{5}$$

where η is the intrinsic wave impedance (377 Ω) and R_o is the reference impedance (50 Ω). From the reciprocity theorem, the same relationship is also true for the transmitter antenna, given by

$$(A_{\text{eff}})_T = \frac{\eta}{R_o AF_T^2} \tag{6}$$

Substituting Eqs. 5 and 6 into Eq. 4, SA can be rewritten as

$$SA = \frac{V_T}{V_R} = \frac{d \cdot \lambda \cdot R_o}{\eta} AF_T \cdot AF_R \tag{7}$$

Normalized site attenuation (NSA) is defined as

$$NSA = \frac{SA}{AF_T \cdot AF_R} = \frac{V_T}{V_R} \cdot \frac{1}{AF_T \cdot AF_R} \tag{8}$$

In the NSA measurement, V_T is measured with the two coaxial cables disconnected from the two antennas and connected to each other via an adapter. V_R is measured with the coaxial cables reconnected to the two antennas and the maximum signal measured with the receive antenna scanned in height from 1 m to 4 m. With all terms in dB, the measured NSA is given by:

DOI 10.1007/978-981-4560-75-7_81-1

Fig. 7 NSA test setup in an EMC test chamber

$$NSA = V_T - V_R - AF_T - AF_R \tag{9}$$

The antenna factors may be affected by mutual coupling if the transmitter and the receiver are in the near field. For 3-m antenna spacing, mutual impedance correction factors, ΔAF_{total}, must be applied. The measured NSA is then given by

$$NSA = V_T - V_R - AF_T - AF_R - \Delta AF_{\text{total}} \tag{10}$$

V_T and V_R can be measured using a spectrum analyzer. Alternatively, SA or ratio of V_T to V_R can be measured using a vector network analyzer. With a full 2-port calibration performed at the cable ends, the measured S_{21} with the cables connected to the respective antennas gives the value of SA with vector error correction for source mismatch, load mismatch, and transmission tracking errors. Additional 6-dB attenuators can be fitted to the transmitter and receiver antenna terminals to improve the impedance matching.

A discrete frequency measurement method can be used with tuned dipole antennas for OATS validation. Swept frequency method must be used with linearly polarized broadband antennas for semi-anechoic chamber (SAC) validation so that any resonance or undesired reflections that cause large peaks and nulls in the NSA frequency response can be spotted. The frequency sweep speed shall be much faster than the receiver antenna height scan speed. Figure 7 shows the photograph of an NSA test setup in an EMC test chamber. A bilog-periodic antenna is mounted on a nonconductive antenna mast for height scan. Alternatively, swept frequency measurements can be performed at discrete antenna heights, so long as the height step is small enough to detect the maximum signal.

In an EMC compliance test scenario, the EUT size is considerably larger than a point source. When the EUT rotates with the turntable, the source position could be moved. For a SAC, a single-point NSA measurement is insufficient to pick up possible reflections from the chamber structure and RF absorbers covering the walls and ceiling of the chamber. The test volume to accommodate the largest equipment to be tested shall be defined. Volumetric NSA measurements shall be performed with the transmitter antenna placed at various points within the test volume for both horizontal and vertical polarizations.

The positions of the transmitter and receiver antennas for the volumetric NSA measurements are illustrated in Fig. 8. The separation distances are measured with respect to the center of the antenna. The position of the receiver antenna shall be relocated to maintain constant distance. For vertical polarization, the lower tip of the antenna shall be more than 25 cm from the floor. For horizontal polarization measurements in the left and right positions, the center of the antenna is moved toward the center of the test volume so that the extreme tip of the antenna is at the periphery of the test volume diameter. With

DOI 10.1007/978-981-4560-75-7_81-1

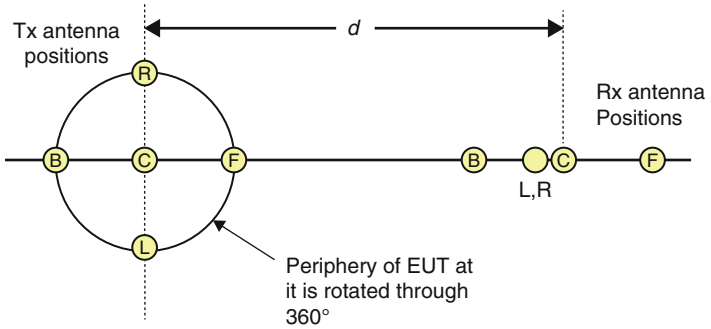


Fig. 8 Antenna positions for volumetric NSA measurements

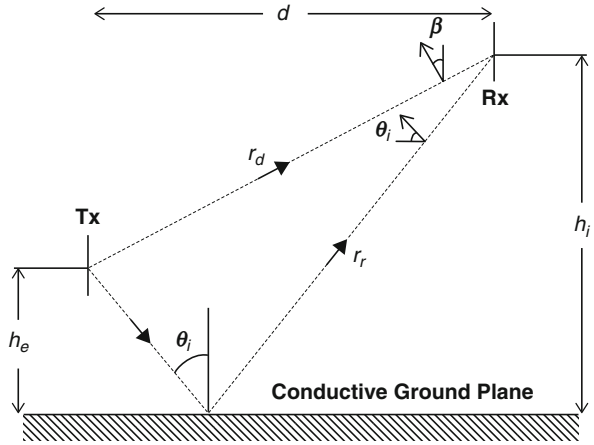


Fig. 9 Geometry of the transmitter and receiver antennas for NSA measurements

5 positions (center, left, right, front, and back), for 2 polarizations and for 2 heights (1 m and 1.5 m for vertical polarization, 1 m and 2 m for horizontal polarization), a maximum of 20 separate NSA measurements are required.

The NSA measurements for the back position would not be required if the closest point of the RF absorber is at a distance greater than 1 m from the rear boundary of the test volume. Further, if the test volume is less than 1-m depth, by 1.5 m in width, and by 1.5 m in height including the turntable, horizontal polarization NSA measurements need to be made at only the center and front positions for 1-m and 2-m heights. The vertical polarization NSA measurements need to be made for only 1-m height at the center, front, left, and right positions. The number of required measurements can be reduced to 8 separate NSA measurements.

The theoretical NSA model has been developed by Smith et al. (1982). For a transmitter power of 1 picoWatt, the theoretical NSA is given by

$$NSA_{TH} = -20\log(f) + 48.9 - 20\log|E_D|_{\max} \tag{11}$$

where f is the frequency in MHz and $|E_D|_{\max}$ is the maximum field strength detected by the receive antenna scanned in height from 1 m to 4 m. For the measurement geometry in Fig. 9, the field strengths $|E_{DH}|$ and $|E_{DV}|$ for horizontal polarization and vertical polarization, respectively, are given by

DOI 10.1007/978-981-4560-75-7_81-1

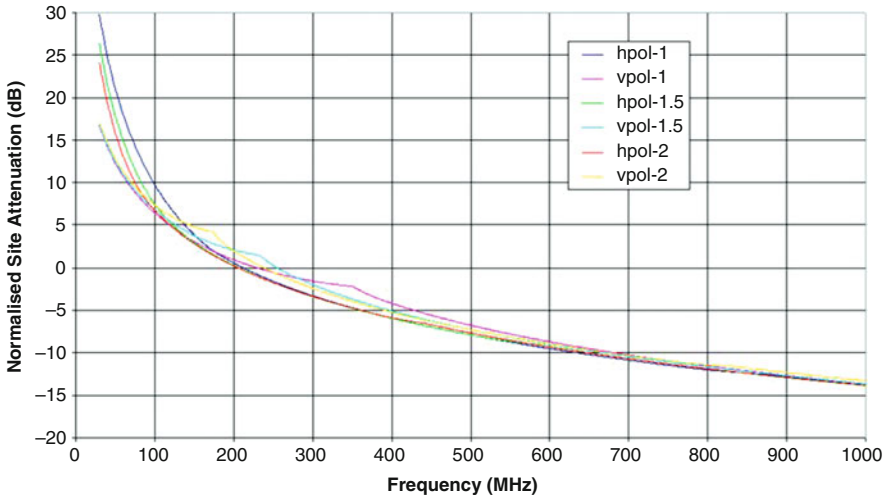


Fig. 10 Theoretical NSA for transmitter antenna height 1 m, 1.5 m, and 2 m ($d = 10$ m)

$$|E_{DH}| = \frac{\sqrt{49.2}\left\{{r_r}^2 + {r_d}^2 - 2r_d r_r \cos\left[\frac{2\pi}{\lambda}(r_r - r_d)\right]\right\}^{1/2}}{r_r r_d} \tag{12}$$

$$|E_{DV}| = \frac{\sqrt{49.2}d^2\left\{{r_r}^6 + {r_d}^6 + 2{r_d}^3{r_r}^3 \cos\left[\frac{2\pi}{\lambda}(r_r - r_d)\right]\right\}^{1/2}}{{r_r}^3{r_d}^3} \tag{13}$$

$$r_d = \sqrt{(h_i - h_e)^2 + d^2} \tag{14}$$

$$r_r = \sqrt{(h_i + h_e)^2 + d^2} \tag{15}$$

The antennas are theoretical half-wave dipoles with antenna gain of 1.64. However, an isotropic radiation pattern is assumed. Large peaks and nulls in received signal strength will be found in the height scan. The receiver antenna height h_i is selected to give maximum field strength. Figure 10 shows the theoretical NSA for transmitter antenna height 1 m, 1.5 m, and 2 m. Figure 11 shows the receiver antenna heights at which maximum field strength is expected to be detected. Detection of maximum signal eliminates the sensitivity of the measurements to nulls due to cancelation between the direct signal and the ground reflection. However, at frequencies below 190 MHz (for $h_e = 1$ m and $d = 10$ m), maximum field strength for horizontal polarization can only be detected at a receiver antenna height above 4 m. Below 4-m height, the direct signal will not be in phase with the ground reflection. Hence, larger measurement errors are anticipated at the low-frequency range.

Accurate antenna factors must be used in determining the measured NSA. Due to coupling to the ground plane, the antenna factors may vary with the height at which the antenna is positioned. The height-dependent corrections for antenna factors are available for tuned dipole antennas (Sugiura 1992). For broadband antennas, the free-space antenna factors are usually used. The height-dependent characteristics are not completely understood; therefore, they are arbitrarily assumed negligible in most cases or lumped into the error budget (±4 dB) of ANSI C63.4 standard. A study by Mann and Marvin (1994) has shown that the antenna factors of biconical antenna (30–300 MHz) may vary significantly (±1 dB) with antenna height above a ground plane.

DOI 10.1007/978-981-4560-75-7_81-1

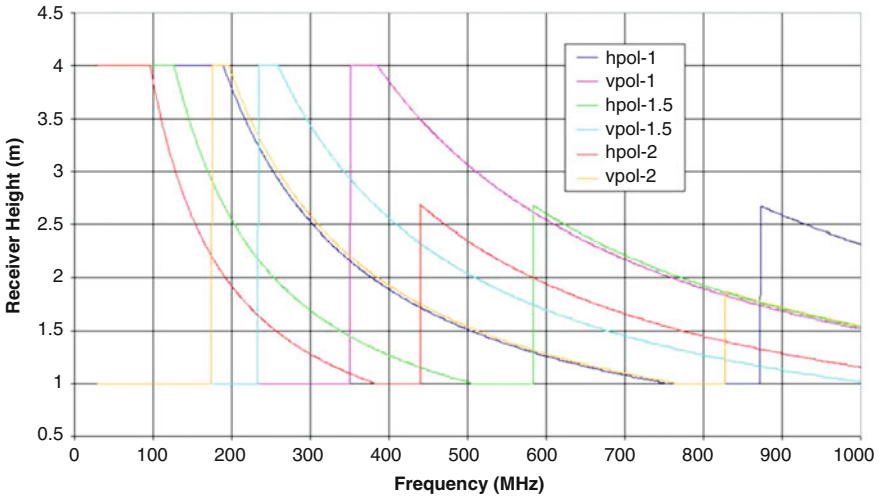


Fig. 11 Receiver antenna heights at which maximum field strength will be detected for transmitter antenna height 1 m, 1.5 m, and 2 m ($d = 10$ m)

A well-constructed OATS may have a measured NSA that deviates from theoretical NSA by typically ±2 dB. If the OATS is used as the reference site, the antenna factors for an antenna individually calibrated according to ANSI C63.5 can defer from the actual value by ±2 dB (Mann and Marvin 1994). The antenna is acceptable for measurement of EMI. For NSA measurement, the pair of antennas may give rise to a measurement uncertainty of ±4 dB. This is already the total error budget of the site validation requirements. Hence, antenna factors provided with the antenna are generally inadequate. The transmitter and receiver antennas that will be used for test site validation purpose must be calibrated in pair at a reference site to obtain the total antenna factor $AF_{\text{total}} = AF_T + AF_R$ so that the error is minimized (Garn et al. 1992). This method is known as dual antenna factor method. In a study by Müllner and Garn (2001), it was shown that AF_{total} of broadband antenna varied with polarizations and heights of transmitter antenna at the test volume (1 m, 1.5 m, and 2 m).

Antenna factors are usually calibrated for horizontal polarization only (ANSI C63.5-1998). The same set of antenna factors is used for vertical polarization. The measured NSA for vertical polarization may have a larger deviation from the theoretical NSA compared to horizontal polarization. In the practical situation, the antenna at vertical polarization is unbalanced (because half of the antenna elements are nearer to the conductive ground plane, hence larger coupling with the ground). A balance-to-unbalance transformer or balun is usually fitted to the antenna terminals to improve the balancing. Residual unbalanced effect will cause undesirable current to flow on the outer braid of the coaxial cable. Ferrite rings can be fitted over the cables to reduce the unbalanced effect.

In the horizontal polarization measurement, the coaxial cables dropping vertically to the ground are orthogonal to the antenna elements. In the vertical polarization, the antenna elements will be parallel to the cables. Hence, parasitic reflections from the cables and also the antenna mast will cause larger errors in vertical polarization NSA measurements than in horizontal polarization. Some errors are also contributed by the mutual coupling between the cables and the antenna elements. Fitting ferrite suppression rings over the cables can reduce the parasitic reflections.

The separation distance between the transmitter and receiver antennas is measured with respect to the center of the antennas. For a log periodic or bilog-periodic antenna, the size of the antenna is typically close to 1.5-m length. The phase center of the antennas varies with frequencies. Hence, the separation distance between the transmitter and receiver antennas effectively changes with frequencies. This may

DOI 10.1007/978-981-4560-75-7_81-1

contribute to some deviations from the theoretical NSA model that assumes point sources. In a 3-m range, larger deviation is expected.

Free-space antenna factors are calibrated in a uniform plane-wave field. In an OATS or SAC, the field distribution is nonhomogeneous due to ground reflections. The field amplitude and phase variation across the length and width of a broadband antenna may be very large. The voltage measured at the antenna terminals is therefore not strictly related to the field strength but the response of the antenna in integrating the nonhomogeneous field distribution. The geometry-dependent antenna factors will be different from the free-space antenna factors.

The theoretical NSA model assumes far-field radiation between the transmitter and receiver antennas that the field strength decays as a function of $1/d$ where d is the separation distance. At low frequencies where the wavelengths approach 10 m, the additional $1/d^2$ term which must be considered at the near-field may cause the measured NSA to deviate from the theoretical NSA. The mutual coupling between the antennas is another error factor that must be corrected. An exact analysis of the mutual impedance correction factors is complex and subject to several assumptions. The approximate values are available for tuned resonant dipoles for specific site geometry with 3-m separation distance. The effect of varying coupling path length due to height scanning of receiver antenna is neglected. For broadband antennas, mutual impedance correction factors are unknown, therefore assumed zero for any site geometry. If the separation distance is less than a wavelength, the change in antenna factor will become significant.

The transmitter and receiver antennas shall be aligned to point at each other, especially for the case of log periodic antenna and bilog-periodic antenna. Misalignment due to antenna tilting to the left, right, down, or up may cause the transmitter antenna beam to be slightly off the end-fire direction of the receiver antenna. Together with misalignment of transmitter-receiver polarization, some measurement errors may be incurred.

The theoretical NSA model assumes maximum power transfer between the signal source and the transmitter antenna and between the receiver antenna and the spectrum analyzer input. Practical broadband antennas will have varying impedance over the wide frequency range of 30–1000 MHz. Reflections within the cables due to impedance mismatch may cause the measurement errors to exceed the NSA tolerance. The use of attenuators (6–10 dB), one at the terminals of each transmitter and receiver antenna, will reduce the mismatch effect.

Other factors that may affect the accuracy of NSA measurement are antenna separation distance error, frequency error, antenna factor frequency interpolation, cable loss, and instrumentation calibration error. Out of the ±4 dB tolerance recommended by the standards, only 1 dB is allocated for site anomalies while the remaining 3 dB is to take care of instrumentation errors and measurement technique errors. As a result, considerable attentions on the setup details and measurement procedure are required in order to demonstrate compliance with the site validation requirements.

The CISPR 16-1-4 Amd.1 Ed. 3.1 (2012) was published in July 2012 to replace Ed. 2.0 (2007). The amendment introduces an alternative site validation method called reference site method (RSM) (Müllner and Garn 2001). In this method, site attenuation $SA_{\text{Ref\_site}}$ is measured at a reference test site with validated performance, using a pair of transmitting/receiving broadband antennas (biconical and log periodic to cover 30 MHz–1 GHz). The same pair of antennas is then used to perform site attenuation SA_{Site} measurement at the chamber that is being verified. The site attenuation deviation (DSA), defined as the difference between SA_{Site} and $SA_{\text{Ref\_site}}$, will give a direct comparison of the two test sites. Deviation not exceeding ±4 dB is considered acceptable. Since the same antennas have consistency between validations, uncertainties such as polarization and height dependency of antenna factors, antenna radiation pattern effect, and non-traceable mutual coupling play no role in site validation anymore. The RSM proves to be a more reliable, traceable, and time-saving site validation method.

DOI 10.1007/978-981-4560-75-7_81-1

Table 1 The upper frequency of EMC radiated emission measurement range

Highest frequency generated	FCC part 15 digital device	CISPR 22/EN55022
<108 MHz	1000 MHz	1000 MHz
108–500 MHz	2000 MHz	2000 MHz
500–1000MHz	5000 MHz	5000 MHz
>1 GHz	5× the frequency but no higher than 40 GHz	5× the frequency but no higher than 400 GHz

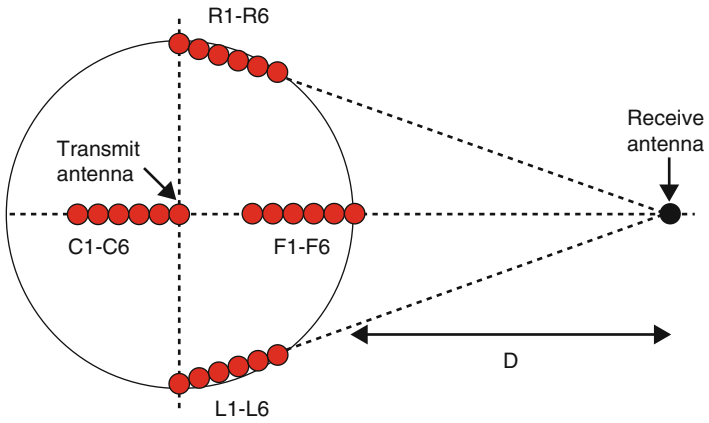


Fig. 12 SVSWR measurement locations and positions

In the design of a semi-anechoic chamber (SAC), the volumetric NSA measurement requirements must be considered. A modeling technique to predict the NSA performance prior to the SAC construction must include the effects of placing the transmitter antenna at various locations in the test volume, scanning the receiver antenna from 1- to 4-m height, reflections from the ground plane, and reflections from the absorber-lined walls and ceiling.

Test Site Validation Method (Above 1 GHz)

The CISPR 16-1-4 requires the use of fully anechoic chamber (FAC) for testing above 1 GHz. With the ground plane covered with RF absorbing material to suppress any reflection, FAC simulates a free-space environment. The upper frequency of radiated emission measurement for the CISPR 22/EN55022 and FCC rules is a function of the highest frequency generated within the product, as summarized in Table 1. If the measurement is performed at 3-m range, the emission level measured can be easily extrapolated for 10-m range by subtracting away 10.46 dB since the field strength is inversely proportional to distance.

The test site validation method specified in CISPR 16-1-4 is called site voltage-standing-wave ratio (SVSWR) method. As with any VSWR measurement, the objective is to measure the maximum and minimum values of a standing wave over a specified travel distance. The ratio of these values is the VSWR. For an ideal free-space environment where there is no reflected wave, VSWR will be unity, i.e., no standing wave. For a practical FAC, reflected waves from the nonideal RF absorbers and chamber structure will give rise to some variation of field strength with respect to distance due to constructive and destructive interferences. Larger value of this SVSWR indicates greater imperfection of the FAC.

The CISPR 16-1-4 proposes to measure the SVSWR by moving a transmitting antenna along a straight line in the chamber and measuring the received voltage with the receiver antenna in the normal location

DOI 10.1007/978-981-4560-75-7_81-1

used for product testing. The test locations and positions are shown in Fig. 12. The received voltages over the broad frequency range (with maximum step size of 50 MHz) are to be measured at only six physical positions of the transmitting antenna for each of the four volumetric locations: 0, 2, 10, 18, 30, 40 cm. It is a compromise to save testing time. Nonuniform spacing is chosen by CISPR 16-1-4 so that the chances of sampling the maximum and minimum values of a standing wave are not totally negated at some frequency points of the sweeping frequencies. Although the true maxima or minima of the standing wave will not be sampled at most frequency points, i.e., a result that is always lower than the true SVSWR will be obtained, the error will not have a direct contribution to measurement uncertainty of EMC product testing.

At each of the four test locations within the test volume, SVSWR measurement must be done for two polarizations and at least two heights. The sampled field strength should be corrected for distance for the six positions. Hence the positioning accuracy of the transmitting antenna directly impacts the repeatability of the SVSWR measurement. The antennas used must meet the antenna pattern requirement of CISPR 16-4-1. Misalignment of the transmitting antenna to the receiving antenna may result in a changed received voltage. The acceptable SVSWR is 6 dB. The site validation procedure is recommended to be repeated annually:

$$E_{\text{relative}} = \frac{E_R}{E_D} = \frac{1 - SVSWR}{1 + SVSWR} \tag{16}$$

The maxima and minima in SVSWR measurement are related to the direct path signal E_D and reflected signal E_R. The ratio of reflected signal to direct signal E_{relative} is related to SVSWR by Eq. 16. Larger value of E_{relative} above zero indicates poorer quietness performance. SVSWR of 6 dB is equivalent to E_{relative} of -10.66 dB. Modern vector network analyzer (VNA) can be used to measure E_D and E_R without the need to physically move the antennas by applying a time-domain reflectivity (TDR) technique. The transmitting and receiving antennas are connected to two separate ports of the VNA, and the ratio of received voltage vector to transmitted voltage vector is determined. Inverse Fourier transform of the frequency-domain data to time-domain data will show the direct signal and reflected signal as impulses arriving at the receiving antenna at different times. In time domain, the direct signal will arrive at the receive antenna first, and the reflected signal will arrive later. By applying time gating, the effect of the direct signal can be separated from the reflected ones. Fourier transform of the time-gated impulses will produce the frequency-domain data of E_D and E_R. Subsequently, E_{relative} and SVSWR can be determined over the measurement frequency range. This TDR method is proposed in ANSI C63.4 draft standard.

From the explanation above, it seems that the SVSWR by spatial movement (CISPR 16-1-4) and SVSWR by time-domain reflectivity (TDR) may produce equivalent data. However, some differences may result due to antenna pattern effect, direction of reflected wave (related to location of chamber imperfection), effect of multiple reflected waves, and the under-sampling pitfall of the SVSWR by spatial movement discussed earlier.

The benefits of the TDR method are numerous. The TDR method does not require physically moving the antenna to a few discrete positions, and the SVSWR from time domain represents the true value of the site (higher than CISPR method). It is much less time consuming. Exact distance between transmit and receive antennas is not required, ditto the field strength values corrected for distance. In addition, TDR method is an excellent tool for troubleshooting site imperfection. Directive horn antenna can be used as the transmit antenna, and its direction can be rotated to identify the azimuth angle that produces the largest peak of E_R. The distance displayed on the VNA screen is the round-trip distance that the impulse traveled from the transmit antenna to the chamber imperfection and back to the receive antenna. The location of chamber imperfection can be confirmed if the peak increases when a large piece of metal is held at the suspected location. Corrective action can then be taken to improve the chamber quietness performance.

DOI 10.1007/978-981-4560-75-7_81-1

Conclusion

An anechoic chamber with a floor area of 20 × 10 m and 8-m height requires a substantial investment both financially and in building space. A 10-m range EMC chamber typically cost more than US$ 2 million. The important considerations in an anechoic chamber project may include the specifications, procurement process for various components and instrumentation, logistical arrangements (contract, shipping, insurance, haulage, time to delivery, storage space), construction process, and performance evaluation. The design must include the necessary features such as access doors, ventilation, lighting, electrical power supply, filters, connection panels, cables, pneumatic air supply, turntable, motorized antenna mast, etc.

The geometry of the chamber structure must not be too complex so that the screened room can easily be constructed, the shielding effectiveness can be under controlled, and the RF absorbers can properly be lined on the chamber walls, ceiling, and floor. Efficient use of available room space is important so that additional screened room can be constructed next to the anechoic chamber for use as the control room to house the test and measurement equipment.

The suitable type of RF absorber must be chosen to achieve the required chamber performance in terms of the purpose of the chamber (antenna, radar, emission, or immunity), measurement frequency range, available room space, required test volume, performance level (NSA, SVSWR, etc.), financial budget, screened room structural strength, and building safety.

Radiated emission measurement results may be affected by any imperfections in the test site. Therefore, the test site must be validated before it can be used for EMC compliance tests. For frequency range below 1 GHz, the NSA method is specified by CISPR 16 and ANSI C63.4 for validation of semi-anechoic chamber where the floor is not covered with RF absorber. Dual antenna factor method is recommended to reduce the measurement error. A better alternative is to use the reference site method to eliminate uncertainties such as polarization and height dependency of antenna factors, antenna radiation pattern effect, and non-traceable mutual coupling effect on the site validation accuracy. For frequency range above 1 GHz, full anechoic chamber to simulate free-space environment must be used. The SVSWR site validation method which involves measuring received voltages at a few discrete positions is specified by CISPR 16-1-4. A TDR method proposed by ANSI C63.4 offers numerous benefits compared to the CISPR method. It can produce equivalent values of SVSWR without physically moving the antenna, potentially more accurate in validating the chamber quietness performance, and the measurement process is much less time consuming.

Cross-References

- ▶ Anechoic Chamber Design
- ▶ Antenna Measurement Setups-Introduction
- ▶ Mm-Wave Sub-mm-Wave Antenna Measurement
- ▶ Radio Wave Propagation Models

References

Brunaugh EL, Osburn JDM (1991) Radiated emissions test performance of the GHz TEM cell. IEEE 1991 Int Symp EMC 1–7. Doi:10.1109/ISEMC.1991.148168

Corona P, Latmiral G, Paolini E (1976) Use of reverberating chamber for measurements of radiated power in the microwave frequency range. IEEE Trans EMC 18:54–59

DeWitt BT, Burnside WD (1988) Electromagnetic scattering by pyramidal and wedge absorbers. IEEE Trans Antennas Propag 36:971–984

European Committee for Electrotechnical Standardisation (1997) Anechoic chamber – part 2: alternative test site suitability with respect to site attenuation. CENELEC, EN50147-2

European Committee for Electrotechnical Standardisation (2000) Anechoic chambers, part 3: emission measurements in fully anechoic chambers. CENELEC prEN50147-3

Garn H, Müllner W, Kremser H (1992) A critical evaluation of uncertainties associated with the ANSI C63.5 antenna calibration method and a proposal for improvements. IEEE Int Symp EMC 485–490. Doi:10.1109/ISEMC.1992.626151

Garn H, Zink E, Kremser R (1993) Problems with radiated emission testing at 3m distance according to CISPR 11 and CISPR 22. In: Proceedings of the 1993 I.E. EMCS symposium, Dallas, TX, pp 216–221. Doi:10.1109/ISEMC.1993.473744

German RF (1982) Comparison of semi-anechoic chambers and open-field test site attenuation measurements. In: Proceedings of 1982 I.E. international symposium on electromagnetic compatibility, Santa Clara, 8–10 Sept, pp 260–265

Hansen D, Ristau D, Lilienkamp P (2000) Correcting OATS antenna factors for small fully anechoic chambers. IEEE Int Symp EMC 219–223. Doi:10.1109/ISEMC.2000.875566

Hemming LH (1992) Architectural electromagnetic shielding handbook: a designer and specification guide. IEEE Press, Piscataway

Holloway CL, DeLyser RR, German RF, McKenna P, Kanda M (1997) Comparison of electromagnetic absorber used in anechoic and semi-anechoic chambers for emission and immunity testing of digital devices. IEEE Trans EMC 39(1):33–47

Kim DI, Takahashi M, Anzai H, Jun SY (1996) Electromagnetic wave absorber with wide-band frequency characteristics using exponentially tapered ferrite. IEEE Trans EMC 38(2):173–177

Mann SM, Marvin AC (1994) Characteristics of the skeletal biconical antenna as used for EMC applications. IEEE Trans EMC 36(4):322–330

Mayer F, Chaumat JP (1992) Broadband absorbers and applications in absorptive ground-mats (AOATS's) and wall coverings of absorber lined chambers (ALC's). Reg Symp EMC 6.2.3/1–6.2.3/8. Doi:10.1109/ISEMC.1992.257561

Mayer F, Ellam T, Cohn Z (1998) High frequency broadband absorption structures. IEEE Int Symp EMC 2:894–899

Müllner W, Garn H (2001) From NSA to site-reference method for EMC test site validation. IEEE Int Symp EMC 2:948–953

Ristau D, Hansen D (2000) Correlating fully anechoic to OATS measurements. IEEE Int Symp EMC 1:219–223

Smith AA Jr, German RF, Pate JB (1982) Calculation of site attenuation from antenna factors. IEEE Trans EMC 24(3):301–316

Sugiura A (1992) Correction factors for normalised site attenuation. IEEE Trans EMC 34(4):461–470

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_83-1

Evaluation of Wearable and Implantable Antennas with Human Phantoms

Koichi Ito[a]*, Chia-Heisn Lin[a] and Ho-Yu Lin[b]
[a]Center for Frontier Medical Engineering, Chiba University, Chiba, Japan
[b]Graduate School of Engineering, Chiba University, Chiba, Japan

Abstract

Recently, body-centric wireless communications (BCWCs) have become a very active area of research because of their numerous applications such as security, electric money, smart homes, personal entertainment, and identification systems. Most BCWCs systems have been focused on the development of wearable antennas for on-body and off-body communications and implantable antennas for in-body communications. These antennas are evaluated with different kinds of human-body phantoms. Particularly, it is almost impossible to use a real human body to evaluate implantable antennas experimentally. Instead, human-body phantoms are inevitable for experiments.

This chapter describes evaluation of wearable and implantable antennas with human phantoms. As for tested antennas, two compact wearable dual-mode (on-body mode at 10 MHz and off-body mode at 2.45 GHz) antennas and two UHF band implantable antennas are exemplified in this chapter.

Keywords

Body-centric wireless communications (BCWCs); On-body; Off-body; In-body; Dual-mode; Phantom

Introduction

Recently, body-centric wireless communications (BCWCs) have become a very active area of research because of their numerous applications such as security, electric money, smart homes, personal entertainment, and identification systems (Hall and Hao 2006; Hall et al. 2009; Zimmerman 1996). In particular, many researchers consider personal healthcare as the biggest potential application of BCWCs, where many kinds of wireless medical devices are attached to the human body, as shown in Fig. 1 (Fujitsu website). However, the environment within and around the human body is often conducive to the transmission of wireless signals. Therefore, as an interface between the transceiver and the propagation environment, antennas should be carefully designed in personal healthcare systems.

BCWCs provide the wireless and maintenance-free communication link between the human body and the surroundings through wearable and implantable antennas. Most BCWCs systems have been focused on the development of wearable antennas for on-body and off-body communications and implantable antennas for in-body communications. These antennas are evaluated with different kinds of human phantoms that are close to the human tissues.

In this chapter, section "Body-Centric Wireless Communications" overviews BCWCs including the IEEE standard and some potential applications of both wearable and implantable antennas/devices.

*Email: ito.koichi@faculty.chiba-u.jp

DOI 10.1007/978-981-4560-75-7_83-1

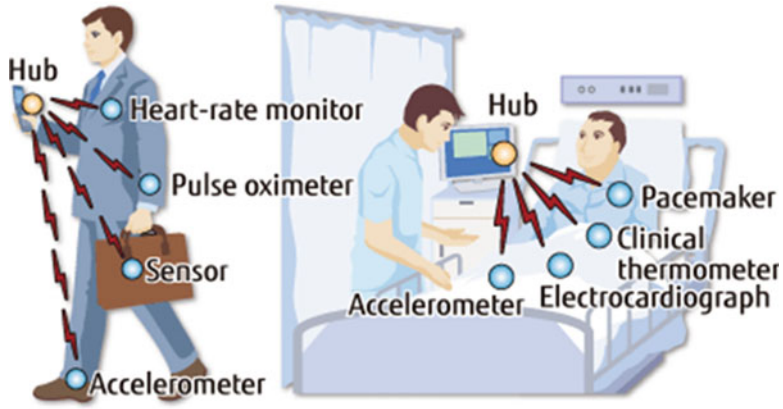


Fig. 1 Medical application of BCWCs (Fujitsu website)

Section "Human-Body Models" summarizes the basics of human-body phantom models to evaluate wearable and implantable antennas. Human-body phantoms are categorized into two groups, namely, numerical phantoms and experimental phantoms.

Then, section "Evaluation of Wearable Dual-Mode Antennas with Human Phantoms" describes evaluation of wearable dual-mode antennas with human phantoms. Here, a dual-mode antenna and a switchable antenna for on-body (10 MHz) and off-body communications (2.45 GHz) are discussed. For on-body communications, *E*-field distributions and received voltages of the receivers on the human body are analyzed, while for off-body communications, reflection coefficients and radiation patterns are discussed.

It is almost impossible to use a real human body to evaluate implantable antennas experimentally. Instead, human-body phantoms are inevitable for experiments. So far, there are few literatures discussing the measured radiation performance of the antennas embedded in the human phantoms. Section "Evaluation of Implantable Antennas with Human Phantoms" describes evaluation of implantable antennas with human phantoms. In this section, dipole antennas are adopted for in-body communications. An upper-body human phantom consisting of a human head, shoulder, and trunk is fabricated to evaluate the performances of the antennas embedded in the human-body phantom.

Body-Centric Wireless Communications

BCWCs can be classified in three communication scenarios (Hall and Hao 2006), as shown in Fig. 2:

- Communications between on-body devices, called *on-body* communications
- Communications with implanted devices, called *in-body* communications
- Communications between on-body and off-body devices, called *off-body* communications

Short-range, low-power, high data rate and reliable wireless communication protocol for use in close proximity to, or inside, a human body are the merits of IEEE 802.15.6-2012 (IEEE standard website_b). Data rates, up to 10 Mbps, are offered to serve a wide and evolutionary set of personal entertainment and healthcare services. This standard supports the combination of security, reliability, quality of service, low power, data rate, and interference protection needed to address the breadth of unique body-centric wireless communication applications. The frequency bands allocated for the IEEE 802.15.6 are summarized in Table 1. Physical channels are classified into three types in accordance with the operating frequency, namely, narrowband (NB), ultrawide band (UWB), and human-body communications (HBC).

DOI 10.1007/978-981-4560-75-7_83-1

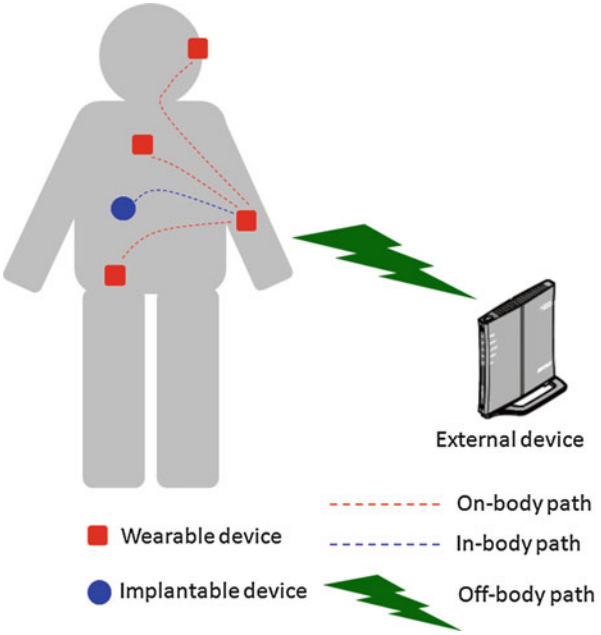


Fig. 2 A schematic of body-centric wireless communications

Table 1 Frequency bands allocated for IEEE 802.15.6 (IEEE standard website_a)

PHY	NB-PHY		UWB-PHY	HBC-PHY
Frequency	402–405 MHz	Implanted device	3.1 − 10.6 GHz	21 MHz (BW:5.25 MHz) 32 MHz (BW: 8 MHz)
	420–450 MHz	Medical telemeter		
	863–870 MHz	ISM band (Europe)		
	902–928 MHz	ISM band (USA)		
	2.36–2.40 GHz	Medical device (USA)		
	2.4 GHz	ISM band (All the world)		
Distance	2–10 m			
Modulation	BPSK, QPSK, GMSK			
Speed	1–10 Mbps			

When wireless devices are used on the human body, they are equipped with what are known as wearable antennas. They are applied for use in on-body as well as off-body communications, two important scenarios involving BCWCs. In on-body communications, the communication range is limited to the human body itself. Both the wearable transmitter and the receiver are mounted on the surface of the body, and biophysical signals can be transmitted along the human body. On the other hand, off-body communications facilitate communication from the surface of the human body to a nearby external device. At least one antenna is located on the human body, and biophysical signals can then be transmitted wirelessly from the antenna to the external device. A number of biomedical applications can be achieved using both on-body and off-body communications in one system.

In-body communication is one of the body area network (BAN) protocols that may be among the most important future communication applications (Hall and Hao 2006), offering advantages such as unattended data transfer, digital data telemetry over mobile phone networks, instant access to relevant medical records, and tracking of patients implanted with critical devices. With the ongoing development of modern technologies, in order to reduce the pressure on the already-overburdened healthcare system and improve the patient's comfort and care, implantable technologies integrating with wireless functions

DOI 10.1007/978-981-4560-75-7_83-1

have been continually anticipated as a future tool, because of their potential to replace cable connectivity with biological telemetry monitoring to provide contributions to diagnosis, therapy, and disease prevention; such integrating technologies can also control the functioning of the system and report the patient's status (Chow et al. 2013; Kiourti and Nikita 2012a, b; Merli et al. 2008; Karacolak et al. 2009). Today the in-body technology is being used in medical application, where a cardiac pacemaker is embedded into the human body to cure cardiac arrhythmia. Besides, such revolutionary advances of technologies have had major impacts in sensor prosthesis (cochlear implant, retinal prosthesis) and treatment of diseases or impairments (Parkinson's diabetes). Some implantable devices (IMDs) of insulin pumps, deep brain stimulation, glucose monitoring, and endoscopy are a few examples that could be considered to take advantage of remote monitoring and control of an implant unit. Here, animal microchip implants use identifying integrated circuits for tracking and tagging, and it has been hypothesized that in-body technology can be applied in the healthcare field to transfer human-body information such as temperature, blood pressure, and cardiac rate. On the other hand, it is even more attractive because of implants inside the human body providing high privacy and personal security to protect patient's information such as a patient's name, stored diagnosis and history of treatments, and some medical characteristics of allergies and medications (Halperin et al. 2008). Moreover, IMDs can be used particularly for the important institutions of a state.

Human-Body Models

Due to a large amount of communication devices that are used close to the human body, such as cell phones, Bluetooth earphones, and tablet computers, human-body phantoms have become essential tools for not only designing and evaluating performances but also testing safety of such devices. For simulation and measurement of wearable and implantable antennas, various types of human-body models or phantoms have been developed and presented. In order to save computation resource and carry out experiment easily, *simple phantom models* are widely used. However, in some particular cases, for example, to evaluate SAR (specific absorption rate) distribution in the human body, realistic *high-resolution human phantom models* are necessary.

Numerical Phantoms

Simple Phantom Models

Homogeneous (square and cylinder) and layered phantom models are widely utilized to study antenna characterization. Lin et al. reported three kinds of simple two-thirds arm phantom models and discussed the characteristics of a wearable antenna on the three phantoms, as shown in Fig. 3. These phantoms are sometimes used to evaluate implantable antennas which are to be embedded in the human arm. Figure 4 shows a whole-body block model with an antenna (Ito et al. 2009) which has homogenous dielectric property equal to the two-thirds muscle at each frequency and consists of nine rectangular parts that represent head, torso, shoulders, arms, and legs.

High-Resolution Phantom Models

To evaluate the safety of radio waves, it is necessary to make precise evaluations of SAR within the human body. High-resolution whole-body voxel human models with anatomical structures, and models that simulate the body tissues in a simple fashion, are employed to estimate the behaviors of radio waves inside the human body and SAR distribution. National Institute of Information and Communications Technology (NICT) and a few universities in Japan developed various numerical human-body models with the

DOI 10.1007/978-981-4560-75-7_83-1

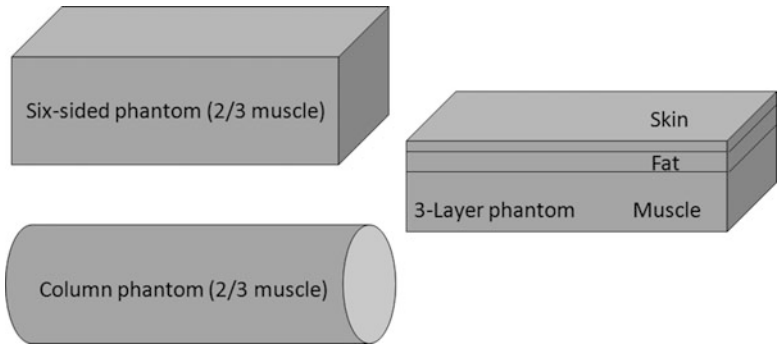


Fig. 3 Three different types of human arm phantom models

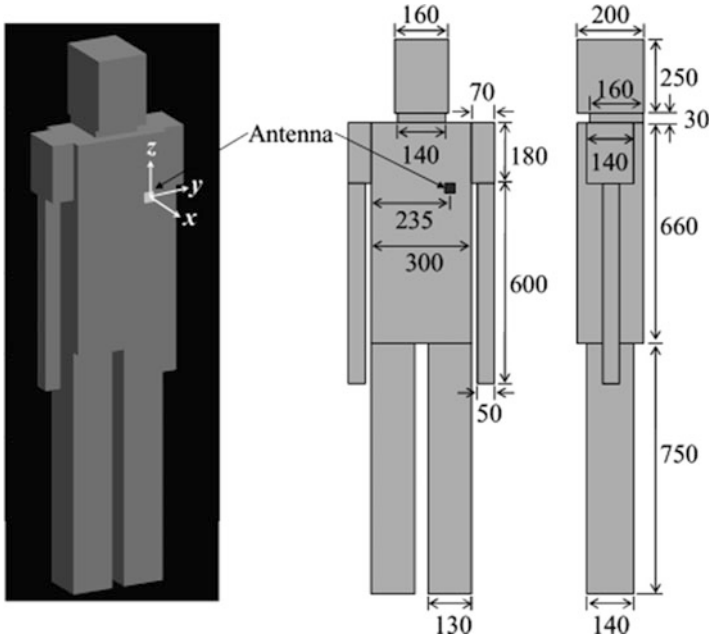


Fig. 4 Whole-body block human model (Ito et al. 2009. © 2009 IEEE)

aim of evaluating the safety of radio waves with respect to the human body and are making them available to the public, as shown in Fig. 5 (NICT website_human model).

Experimental Phantoms

In general, experimental human-body phantoms can be classified into three different types, namely, *liquid* phantoms, *semisolid* phantoms, and *solid* phantoms.

Liquid Phantoms

Liquid phantoms have the average tissue-equivalent materials that have the same electrical characteristics as the tissues in the human body, in the considered frequency range. Liquid phantoms are widely used in SAR measurement and studies, using direct measurement of the electric fields inside the phantom with a small probe. As specified in IEEE Std 1528–2003 (IEEE standard website_c), the permittivity of the phantom shell should be less than 5, the loss tangent less than 0.05, and the thickness of the shell should be 2.0 ± 0.2 mm for the frequencies in the range from 0.3 to 3 GHz. From the documents of MCL-T (MCL-T, London, UK), liquid phantoms are only available up to 6 GHz. As an example, Fig. 6 shows a liquid head phantom in the shell (NICT website_SAR measurement).

DOI 10.1007/978-981-4560-75-7_83-1

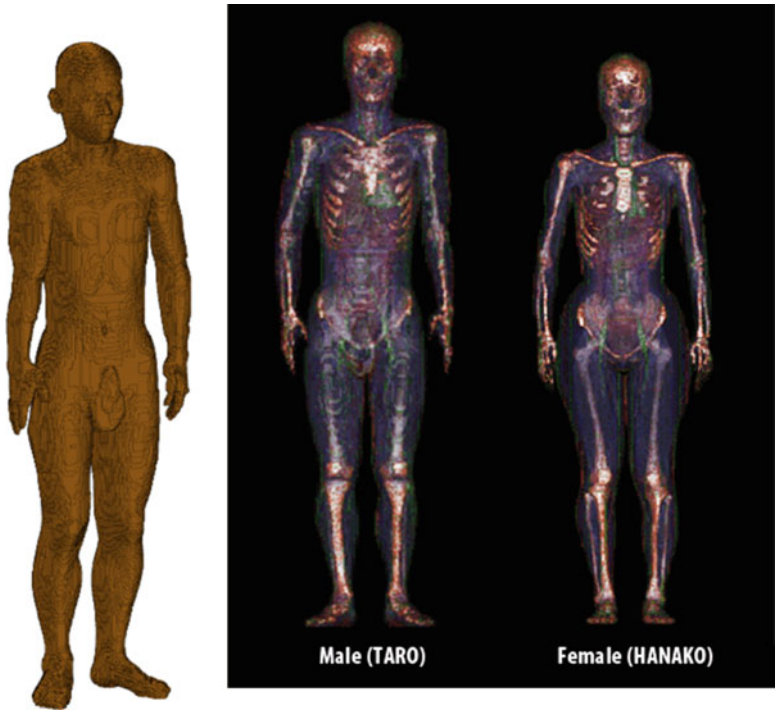


Fig. 5 Realistic numerical human-body models (NICT website_human model)

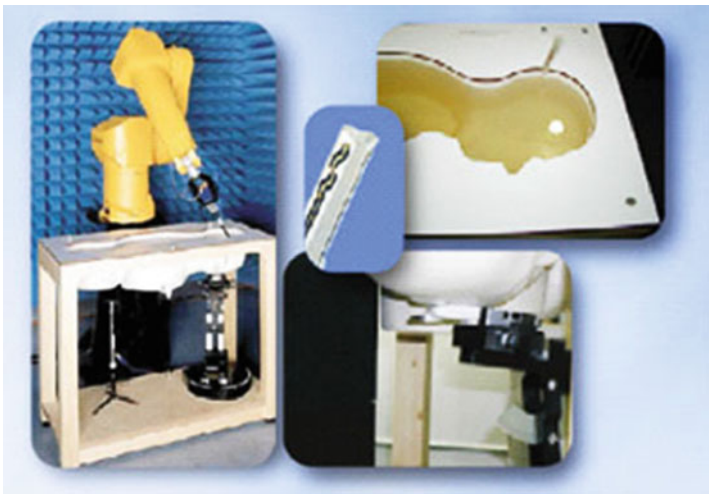


Fig. 6 A head phantom with a predetermined standard size and shape for measurement (NICT website_SAR measurement)

Semisolid Phantoms

By using coagulant agent, it is easy to develop semisolid phantoms capable of self-shaping. Semisolid phantoms are low cost, easy to fabricate, and easy to manufacture into arbitrary shape and multiple layers. They are available at frequencies up to 11 GHz and can be applied for evaluation of antenna performance and SAR measurement. Electrical properties of semisolid phantoms can be adjusted over a wide frequency range. In a single composition ratio, the electrical properties are almost equal to biological tissue over the frequency range of $0.3 - 2.5$ GHz or $3.1 - 10.6$ GHz. Figure 7 presents two types of semisolid phantoms proposed by Chiba University in Japan.

DOI 10.1007/978-981-4560-75-7_83-1

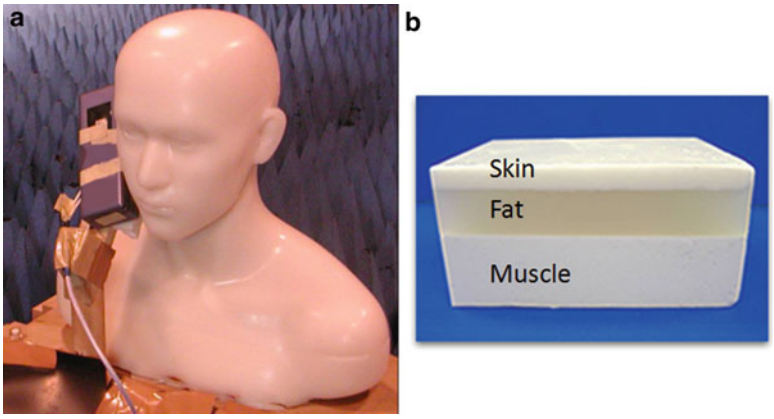


Fig. 7 Tissue-equivalent human phantom. (**a**) Upper body phantom. (**b**) Layered phantom

Fig. 8 Carbon-silicon phantoms (MCL-T, UK)

Solid Phantoms

For measuring the SAR on the surface of the human body, a solid phantom is a good candidate, as shown in Fig. 8 (MCL-T, UK). Compare with liquid and semisolid ones, solid phantoms are capable of keeping their shape for a period of time and designed to fit a wide range of complex permittivity. However, solid phantoms are expensive and require special equipment and complex manufacturing procedures.

Evaluation of Wearable Dual-Mode Antennas with Human Phantoms

Recently, the advance of electronic devices is leading to the creation of body-centric wireless communications (BCWCs), in which wireless devices are attached to the human body. With these devices, as an important application of BCWCs, remote health monitoring will become real. In remote health monitoring, at least two modes of communications are required: on-body communications (collecting medical data) and off-body communications (exchanging data with outside networks). For on-body communications, relatively low frequency bands (tens of MHz) are more suitable (Ito et al. 2012; Fujii et al. 2005; Ito et al. 2009), while for the off-body communications, 2.45 GHz ISM band (2.40 − 2.48 GHz) is a good candidate (Lin et al. 2012a). Thus, a dual-mode antenna is a key component for remote health monitoring. Therefore, two wearable antennas with on-body and off-body modes will be introduced in this section.

DOI 10.1007/978-981-4560-75-7_83-1

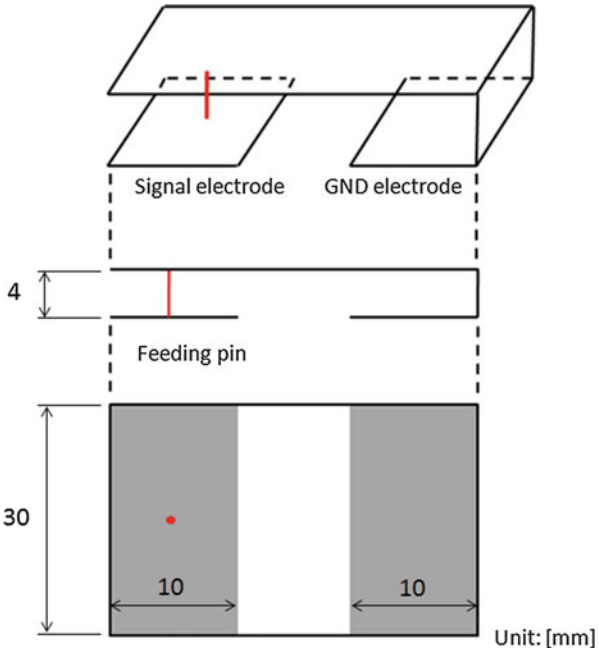


Fig. 9 Structure of the dual-mode antenna

Antenna Design and Human Body

Figure 9 shows a dual-mode wearable antenna (Lin et al. 2012b). It consists of a feeding pin, a signal electrode, and a GND electrode. The size of the antenna is $30 \times 33 \times 4$ mm^3. The height of the signal electrode is 4 mm, and the feeding pin is located in the center of the signal electrode. For the off-body communications, the operating frequency is 2.45 GHz. In this antenna design, in order to concern the equipment packaging, a plastic ($\varepsilon_r = 3$) with 2 mm in thickness is placed between the antenna and the human body. Moreover, in order to confirm the performance of the antenna on the human body and simplify measurement process, a part of chest phantom (electrical constant $\varepsilon_r = 35.2$, conductivity $\sigma = 1.16$ S/m at 2.45 GHz) (Lin et al. 2012b) is adopted to replace the whole human body. However, radiation characteristics of the antenna on the human body should be realized; therefore, the proposed dual-mode antenna on a high-resolution human model in 1.75 m provided by the National Institute of Information and Communication Technology (NICT) in Japan is also concerned in this work, as shown in Fig. 10. Table 2 summarized the computational parameters of the antenna design.

On-Body Mode

As shown in Fig. 11a, one transmitter is attached to the right chest, while four receivers are attached to the left chest, right and left wrists, and abdomen, respectively. Figure 11b presents the variation in received voltage for the four different on-body links on the human body. From the simulated result, the received voltages are around -110 to -160 dB depending on the distance between the transmitter and the receiver. Figure 12 illustrates the *E*-field distributions inside and outside the whole human body at 10 MHz. The *E*-field distribution is discussed according to the received voltage that is calculated from the electric field. The observation plane includes the feeding point of the transmitter. From the result, it can be seen that the level of the electric field distribution are smooth around the human body at 10 MHz. As a result, it can be concluded that the lower frequency such as 10 MHz is suitable for on-body communications.

DOI 10.1007/978-981-4560-75-7_83-1

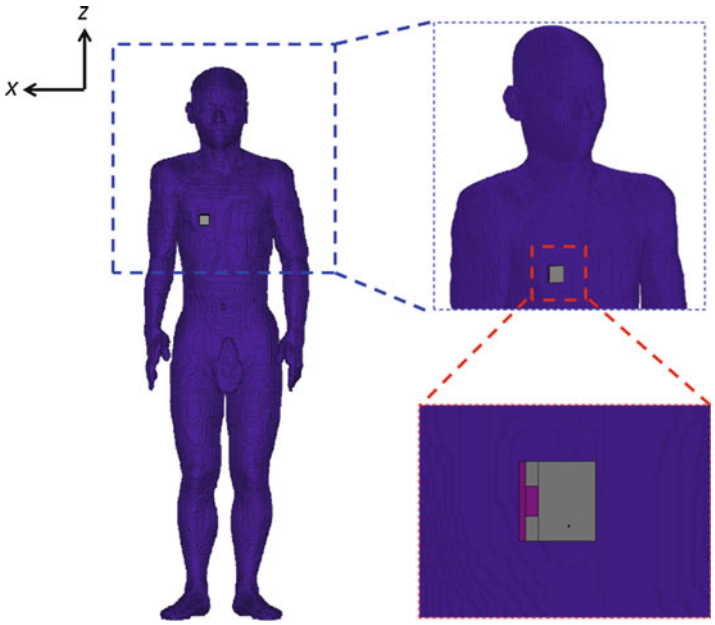


Fig. 10 Antenna on the high-resolution human model

Table 2 Computational parameters for two modes

Mode	On-body mode	Off-body mode
Frequency	10 MHz	2.45 GHz
Cell size	1–2 mm	0.5–2 mm
Computation domain	$1.5 \times 1.5 \times 2\ m^3$	$1 \times 1 \times 2\ m^3$
Number of PML layers	16	8

Off-Body Mode

Off-body mode can be applied to send the received body signals (10 MHz on-body mode) to an external device. The performance such as reflection coefficient and radiation pattern of the antenna on a human body is discussed here. Figure 13 shows simulated and measured reflection coefficients when the antenna is located on a part of chest phantom with the dimension of $200 \times 200 \times 70\ mm^3$ and then attached onto a high-resolution human model. It was found from the result that this antenna works in the ISM band (2.45 GHz). Also, simulated and measured results are in good agreement. The simulated radiation patterns on the chest phantom and the high-resolution human model at 2.45 GHz in *xy*- and *yz*-planes are shown in Fig. 14. From the results, the radiation patterns in both planes are relatively omnidirectional with no deep nulls in the front half-sphere of the human body. Therefore, the antenna is a good candidate for off-body communications.

Switchable Dual-Mode Antenna

When simultaneous operation in both on-body and off-body communications is not necessary, switchable scheme between the on-body and off-body modes is useful in practice. A compact switchable antenna for BCWCs was reported by Lin et al. in which a simple switching circuit is integrated into the shorting plate of PIFA (Lin et al. 2012). As a result, enough gain is obtained since it is equivalent to a PIFA antenna at 2.45 GHz when the switching is turned on. In addition, when the switch is turned off, the antenna is similar to a metal electrode operating as on-body mode at HF band (10 MHz).

DOI 10.1007/978-981-4560-75-7_83-1

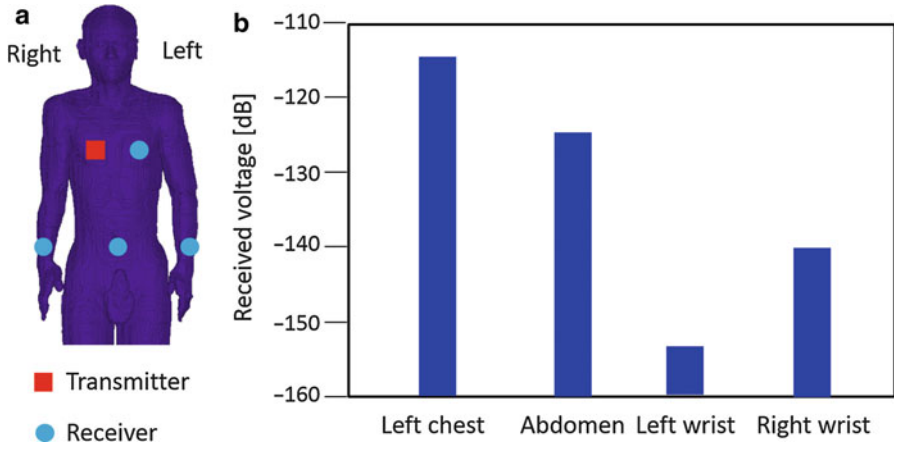


Fig. 11 Simulation. (**a**) Location of the devices. (**b**) Received voltages

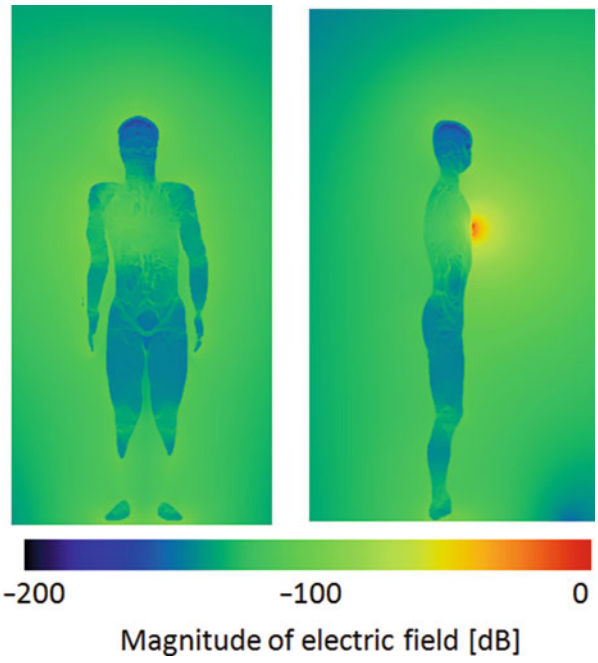


Fig. 12 Electric field distribution at 10 MHz on high-resolution human body

Figure 15 shows the structure of the switchable planar inverted-F antenna. The antenna includes a ground plane, radiator, feeding pin, and a switching circuit board to replace the conventional shorting plate, in which a pin diode was used as the switch component. The radiating top patch and the ground plane have the same dimension of $30 \times 12\ \text{mm}^2$. The height of the antenna is 8 mm. The feeding pin is on the center line of the antenna. In the lower frequency band, the characteristics of the antenna are equivalent to an electrode operating at 10 MHz, while the pin diode located on the circuit board is in the OFF state. The excitation signal for lower frequency is a 10 MHz sinusoidal wave. In a higher frequency band (2.45 GHz ISM band), the antenna operates as a PIFA, while the pin diode is in the ON state. As shown in Fig. 15, the area of the switching circuit board is $10 \times 8\ \text{mm}^2$. It is fabricated on PTFE fluorocarbon polymers (dielectric constant 2.17, 0.8 mm in thickness) with no bottom layer of copper used. In this antenna design, the pin diode HVU131 from Renesas Tech. was selected; thus, a resistor of 200 Ω is used

DOI 10.1007/978-981-4560-75-7_83-1

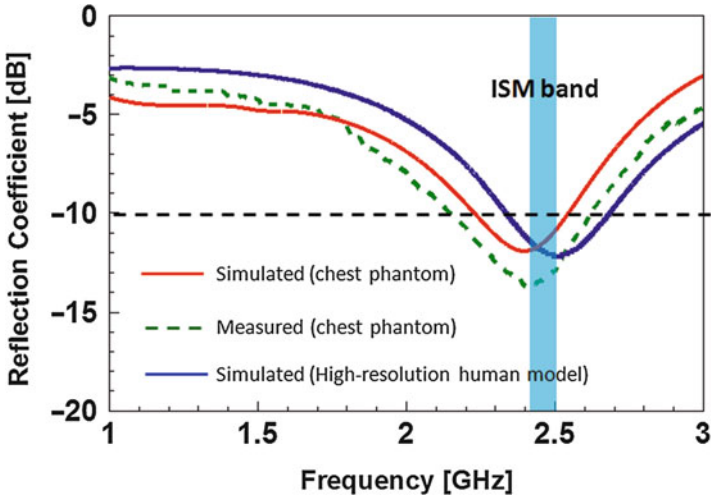


Fig. 13 Simulated and measured reflection coefficients

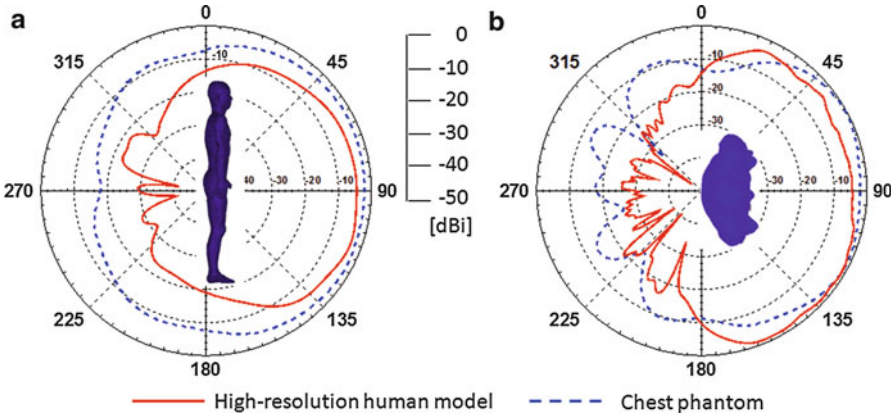


Fig. 14 Radiation patterns. (**a**) *yz*-plane. (**b**) *xy*-plane

to provide the bias for the diode. In addition, an inductor of 33 nH plays the RF choke and a capacitor of 100 pF is used to block the DC voltage, which is similar to the design in Li et al. (2009). The efficiency of the antenna was 46 % at 2.45 GHz as a PIFA. In this study, in all the cases, the radiator is positioned near the phantom and the ground plane is far side, and the space between the antenna and the arm phantom is set to be 2 mm.

On-Body Mode

As shown in Fig. 16a, one transmitter is located to the right wrist while seven receivers are placed to the head, left chest, right chest, abdomen, left wrist, right ankle, and left ankle, respectively. Figure 16b illustrates the simulated *E*-field distributions inside and around the whole human body at 10 MHz. The observation plane includes the feeding point of the transmitter. From the result, it can be seen that the level of the electric field distribution are smooth around the human body at 10 MHz.

During measurement, the transmitter was fed by the 10 MHz sinusoidal wave from a 10 MHz oscillator and the excitation voltage was 4.1 $V_{p\text{-}p}$, while the pin diode on the switching circuit was set in the OFF state, as shown in Fig. 17 a, b. The receiver was connected to a Spectrum Analyzer E4403B with a 50 Ω loading resistance. In order to reduce the effects of the cable and the earth ground on the received voltage, a ferrite transformer is connected between the receiver and the cable from the spectrum analyzer, as shown

DOI 10.1007/978-981-4560-75-7_83-1

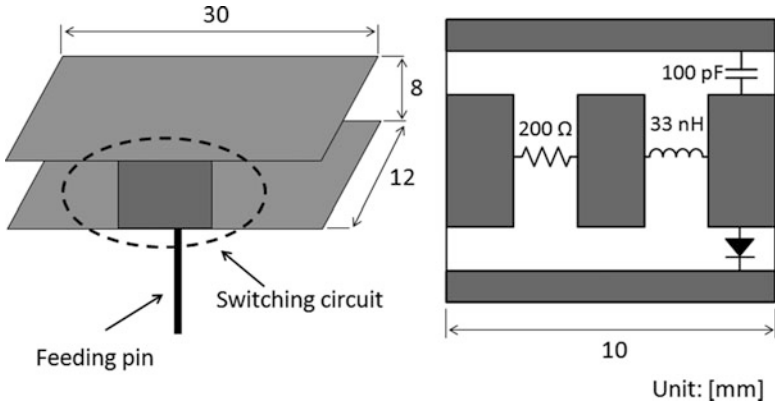


Fig. 15 Switchable planar inverted-F antenna (3D view and switching circuit)

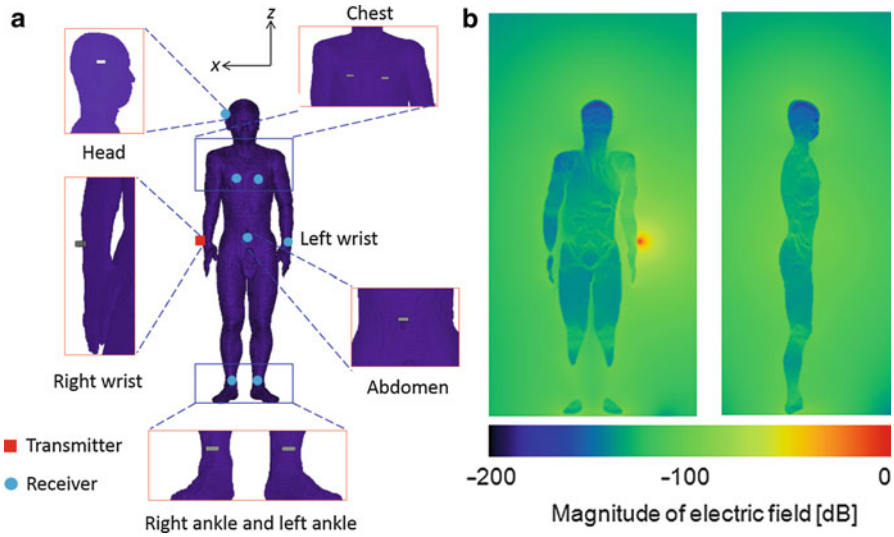


Fig. 16 Simulation. (**a**) Location of a transmitter and seven receivers. (**b**) Simulated *E*-field distribution at 10 MHz

in Fig. 17c. The measurement setup is shown in Fig. 18. The transmitter is excited by a 10 MHz oscillator and each receiver is connected to a spectrum analyzer. In this study, seven sensors are located on the human body which includes the left chest, right chest, abdomen, head, left wrist, left ankle, and right ankle. Measured variation in received voltage for the seven different on-body links on the human body is presented in Fig. 19. From the results, the received voltages are around −98 to −117 dB depending on the distance between the transmitter and the receiver.

Off-Body Mode

Figure 20 shows the simulated and measured reflection coefficients when the antenna is located on a simple arm phantom with the dimension of $50 \times 50 \times 300\ \text{mm}^3$ while the diode was in the ON state (3 V is supplied by battery box), and the results with a high-resolution human model are also given in the same figure. It was found from the result that this antenna works in ISM band (2.45 GHz). The measured bandwidth was 350 MHz (2.30–2.65 GHz) and the measured result was close to the simulated ones.

The simulated and measured radiation patterns on the arm phantom and a high-resolution human model at 2.45 GHz in *xy*- and *xz*-planes are shown in Fig. 21. From the results, the radiation patterns in both planes are relatively omnidirectional with no deep nulls in the front half-sphere of the human body.

DOI 10.1007/978-981-4560-75-7_83-1

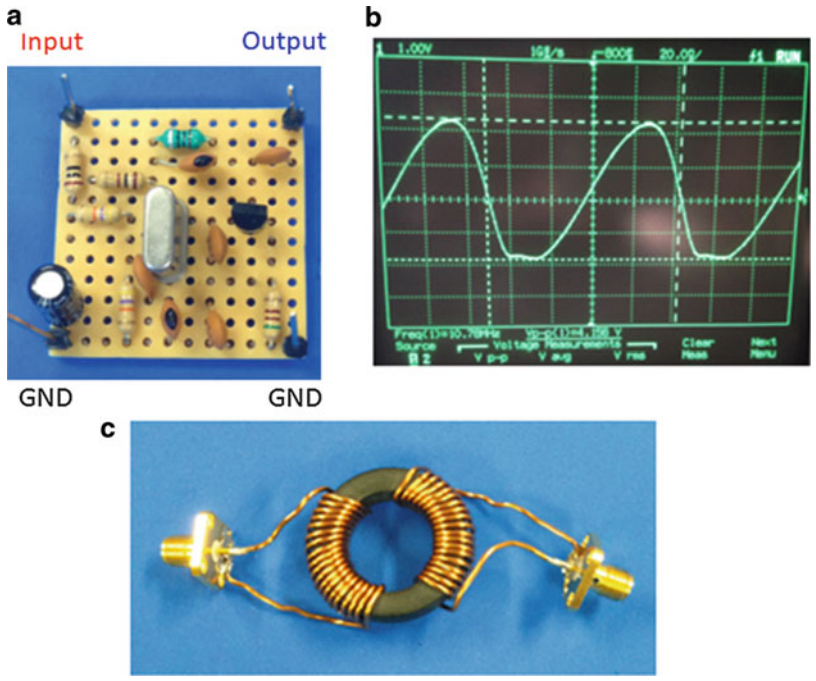


Fig. 17 Measurement. (**a**) 10 MHz oscillator. (**b**) Output signal. (**c**) Transformer

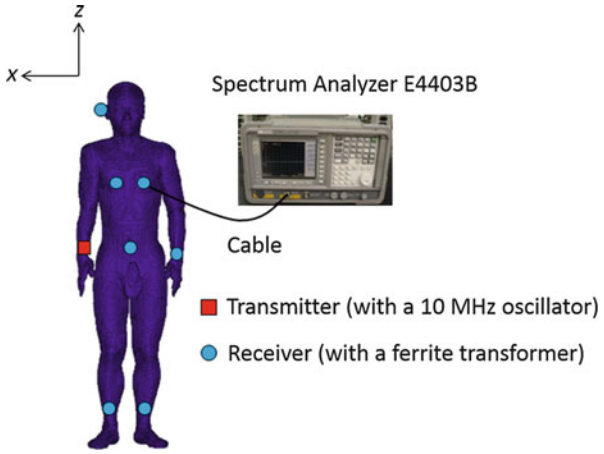


Fig. 18 Measurement setup for received voltage

Therefore, the antenna is a good candidate for off-body communications. The average gain in the two planes of the arm phantom is 5 dBi higher than in Lin et al. (2012), as shown in Fig. 21. It can also be noted that the performance of radiation patterns on the developed phantom is close to the high-resolution human model. Therefore, fabricated phantom can be applied to evaluate wearable antenna performance.

Evaluation of Implantable Antennas with Human Phantoms

It is almost impossible to use a real human body to evaluate implantable antennas experimentally. Instead, human-body phantoms are inevitable for experiments. In general, implantable antennas or devices should

DOI 10.1007/978-981-4560-75-7_83-1

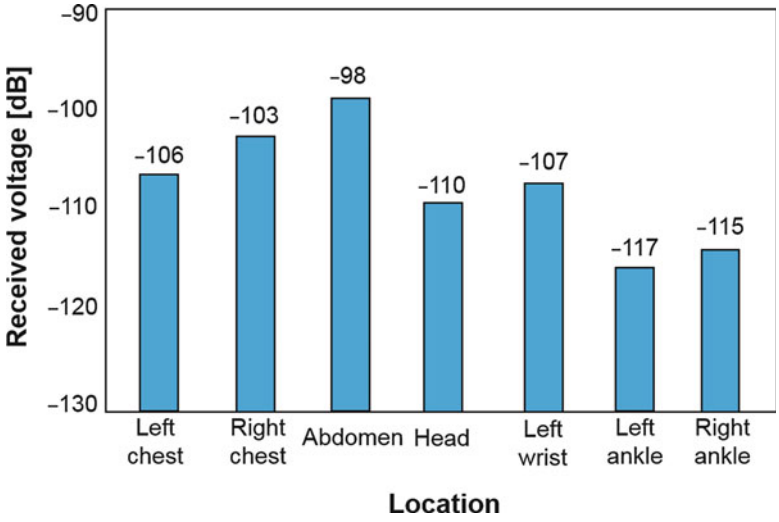


Fig. 19 Measured received voltages at seven locations

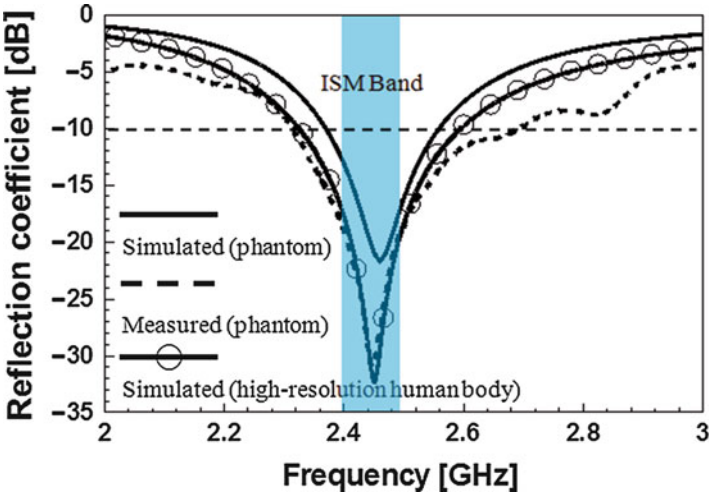


Fig. 20 Simulated and measured reflection coefficients

be small in size, and this will increase the complexity of measurements. This section describes a folded dipole and a simple dipole as typical implantable antennas and how to evaluate such small implantable antennas with human-body phantoms.

Antenna Geometry and Simulation Model

Figure 22a shows the geometry and dimensions of the UHF antenna (Lin et al. 2013). The antenna is fabricated from a narrow strip with a width of 0.1 mm and bent into a folded structure; the whole size is 20.3 mm (L) $\times$ 0.8 mm (w) $\times$ 0.8 mm (h) and the gap is 0.5 mm (g). The L of the antenna is about 0.065 wavelength in the human body at 0.953 GHz. The feed point is in the center of the antenna so as to make it a symmetric structure, and the input impedance is assumed to match a 50 Ω system. With such a slender structure, the radiating elements of the antenna are extremely close, and g, h, and w would be sensitive parameters for impedance matching. In addition to the antenna structure, a coating is generally introduced to decrease the effect of the human body. In this study, a glass coating ($\varepsilon_r = 5.0$) with a whole size of 20.5 $\times$ 1.0 $\times$ 1.0 mm^3 is used for covering the antenna as shown in Fig. 22b, but it can be removed in

DOI 10.1007/978-981-4560-75-7_83-1

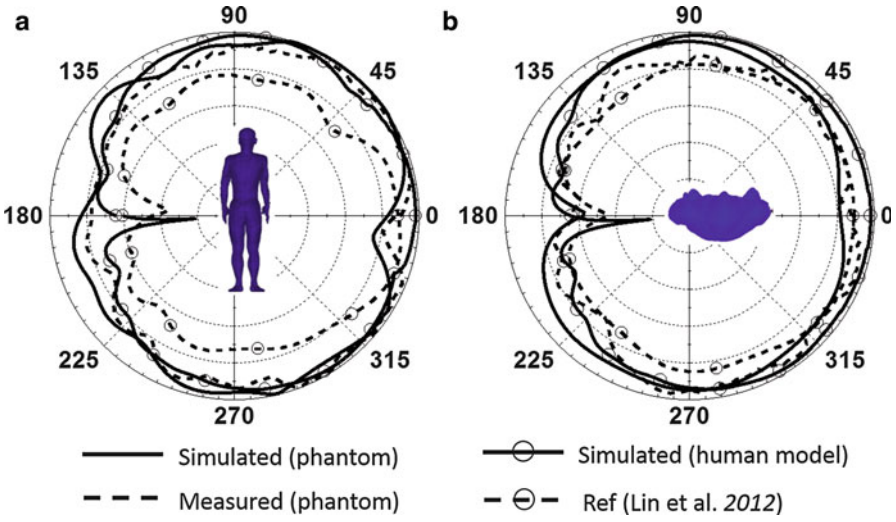


Fig. 21 Radiation patterns at 2.45 GHz. (**a**) *xz*-plane. (**b**) *xy*-plane

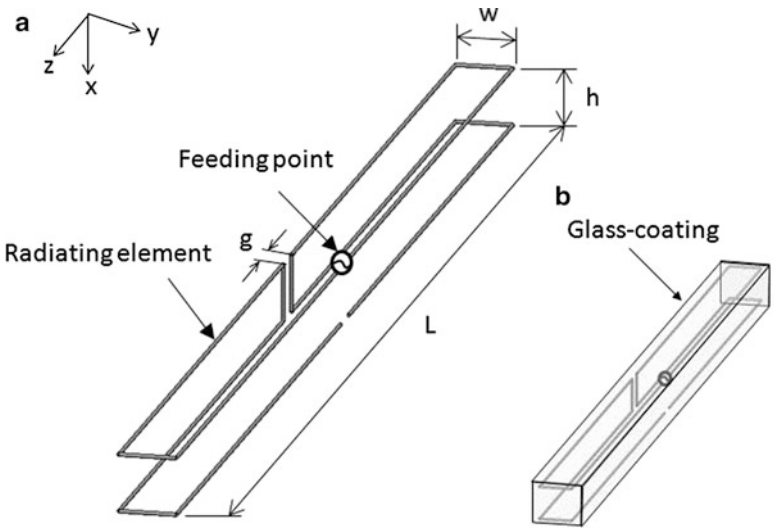


Fig. 22 A folded dipole antenna. (**a**) Antenna structure. (**b**) Glass coating (Lin et al. 2013. © 2012 IEEE)

experiment when the permittivity of the glass coating is close to that of human fat tissue ($\varepsilon_r = 5.45$), which implies that its influence on antenna performance is not strong.

Since the antenna is applied as an implant, it should be evaluated with a realistic human-body model. Figure 23a shows an image of the human body in which the antenna is embedded into the arm and the assumptive communication range is between 90° and 270°. According to Lin et al. (2013), a layered phantom that can be used to represent a part of human arm to confirm the antenna performances, therefore, in the simulation, a model of the human arm using layered phantom, is used as shown in Fig. 23b. The phantom consisted of skin ($\varepsilon_r = 41.14$, $\sigma = 0.88$ S/m), fat ($\varepsilon_r = 5.45$, $\sigma = 0.05$ S/m), muscle ($\varepsilon_r = 54.91$, $\sigma = 0.96$ S/m), and bone ($\varepsilon_r = 12.4$, $\sigma = 0.15$ S/m). These dielectric constants are the values at 0.953 GHz. Moreover, the antenna is generally embedded into the fat layer of the phantoms at a depth (d) of 2.0 mm from the surface of the skin layer, since the loss there is less than at the skin and muscle layers.

DOI 10.1007/978-981-4560-75-7_83-1

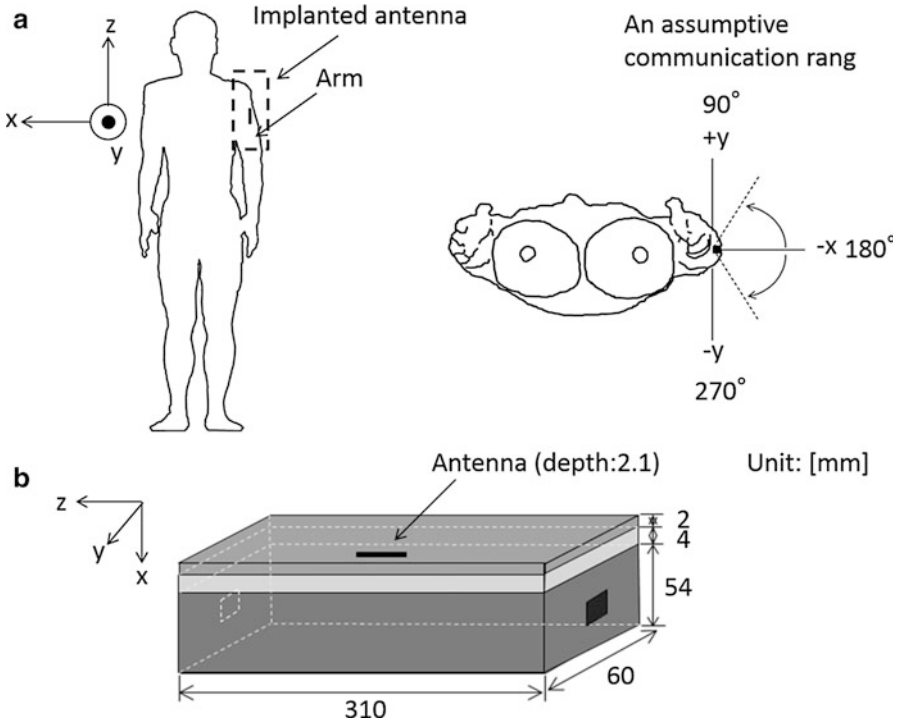


Fig. 23 Human-body models. (**a**) Realistic model. (**b**) Simple arm model

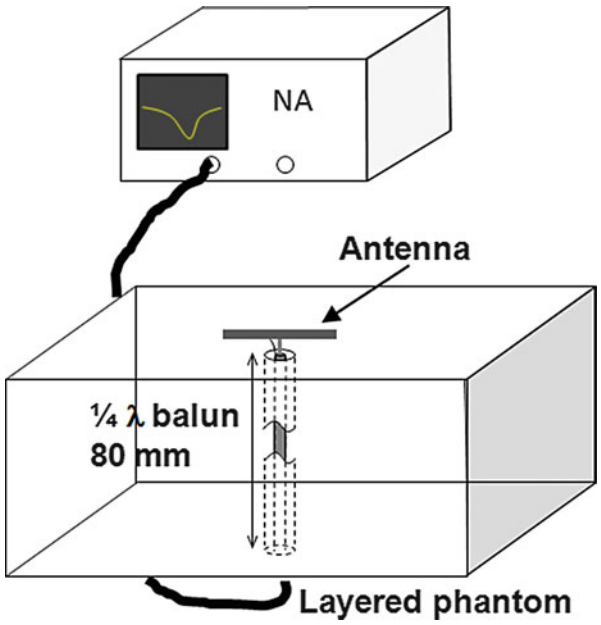


Fig. 24 An image of measurement setup

Measurement Setup and Experimental Result

Figure 24 shows the measurement setup of this study. In order to keep maximum transmission efficiency in the experiment, the antenna is connected to a quarter-wavelength balun that has a length of 79 mm (0.953 GHz) and then is embedded into the fat layer of the phantom. Photographs of the fabricated antenna, a quarter-wavelength balun, and the layered phantom are shown in Fig. 25a, b, respectively.

DOI 10.1007/978-981-4560-75-7_83-1

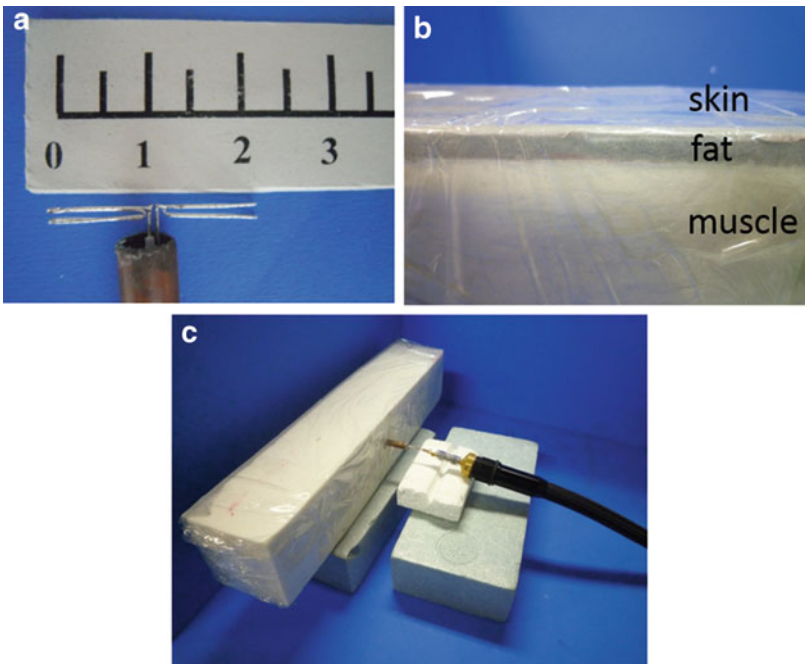


Fig. 25 Measurement. (**a**) Antenna. (**b**) Layered phantom. (**c**) Setup

Table 3 The permittivity and conductivity of the test phantom at 0.953 GHz (Lin et al. 2013. © 2012 IEEE)

Tissues	Permittivity			Conductivity [S/m]		
	Target value	Measured value	Error %	Target value	Measured value	Error %
Skin	41.14	42.06	2.2	0.88	0.91	3.4
Fat	5.45	5.5	0.9	0.05	0.046	8
Muscle	54.91	53.5	2.5	0.96	0.99	3.1

Figure 25c shows the measurement for testing the reflection coefficient. Tables 3 and 4 show the target and measured electric constant (ε_r and σ) of the human tissues at 0.953 GHz and the compositions of the skin, fat, and muscle. When compared to the target values, all the measured results are below or approaching the 5 % error mark according to Ito et al. (2012). Figure 26 shows the calculated and measured reflection coefficients of the antenna in the layered phantom. It can be confirmed that the measured result of the antenna can cover the desired frequency band (0.951–0.956 GHz) and has a good agreement with the calculated result. The calculated and measured radiation pattern results at 0.953 GHz are shown in Fig. 27, which includes the required antenna gain of −35 dBi and its cover region. The radiation performance of the antenna could also be affirmed, although a slight difference is presented in upper half planes of *xz*-plane and *xy*-plane due to experimental errors such as the position of the antenna and fabricated human arm phantom. In addition, the radiation performance was plotted in the lower half planes of *xz*-plane and *xy*-plane. Here the quarter-wavelength balun is used to ensure maximum transmission efficiency. The measured and calculated results are in agreement, in that even if a coaxial cable is used in the experiment, the feeding gap of the simulation differs accordingly. The maximum antenna gains are about −23.5 and −26.0 dBi in the *x*-*z* and *xy*-planes; at -*x* direction the antenna gains are −28.7 and −30.3 dBi, respectively. Therefore, wireless communication is available since the antenna gain exceeds the required antenna gain of −35 dBi, calculated by the link budget.

DOI 10.1007/978-981-4560-75-7_83-1

Table 4 The compositions of the layered phantom (Lin et al. 2013. © 2012 IEEE)

Materials	Human Tissues (%)		
	Skin	Fat	Muscle
Deionized water	32.3		85
Sodium dehydroacetate			0.05
TX-151			2.1
NaCl			0.5
Polyethylene powder			9.7
Agar			2.65
Silicon		62.5	
Glycerin		25	
Gelatinous	3.2		
Silicone emulsion	64.5		
Aluminum powder		12.5	

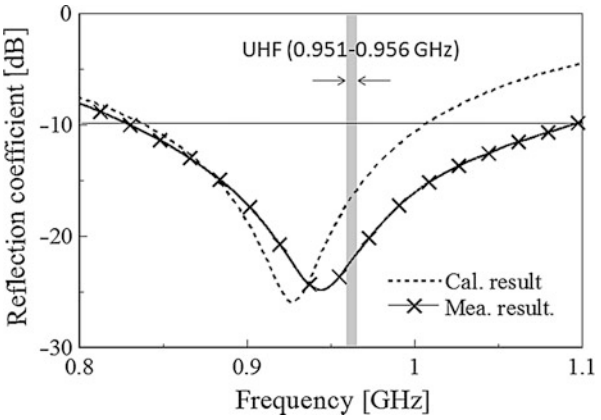


Fig. 26 Simulated and measured reflection coefficients (Lin et al. 2013. © 2012 IEEE)

Measurement of Implantable Antenna with Physical Human-Body Phantom

In this section, in order to confirm the far-field radiation pattern of implantable antennas by experiment within a human body, an upper-body human phantom consisting of a head, shoulder, trunk, and simple human arm is fabricated and measured. A dipole antenna designed to operate at 1.27 GHz is assumed to be embedded in the human arm. The antenna performance is simulated by the finite-difference time-domain (FDTD) method.

Human-Body Models for Simulation

Because the antenna is used for implants, the antenna performance in the 1.27-GHz band should be assessed with a simple human arm model and high-resolution human-body model as shown in Figs. 28a, b.

DOI 10.1007/978-981-4560-75-7_83-1

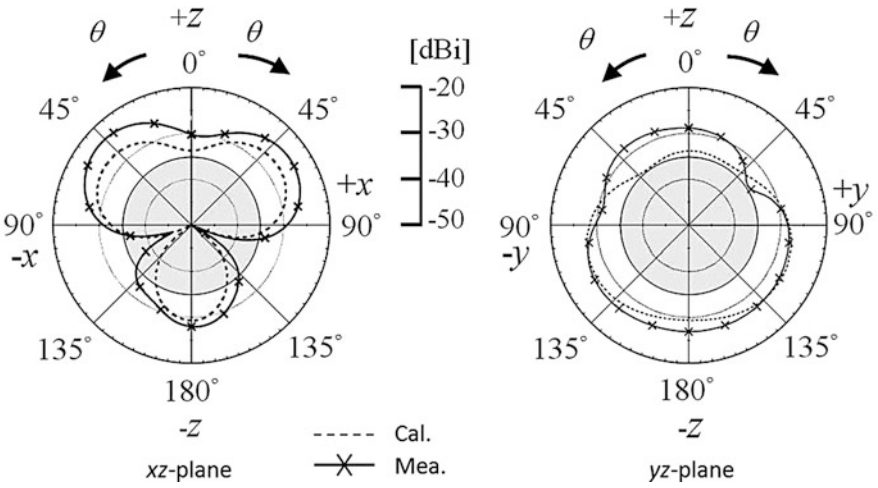


Fig. 27 Radiation patterns at 0.95 GHz in *xz*-plane and *yz*-plane (Lin et al. 2013. © 2012 IEEE)

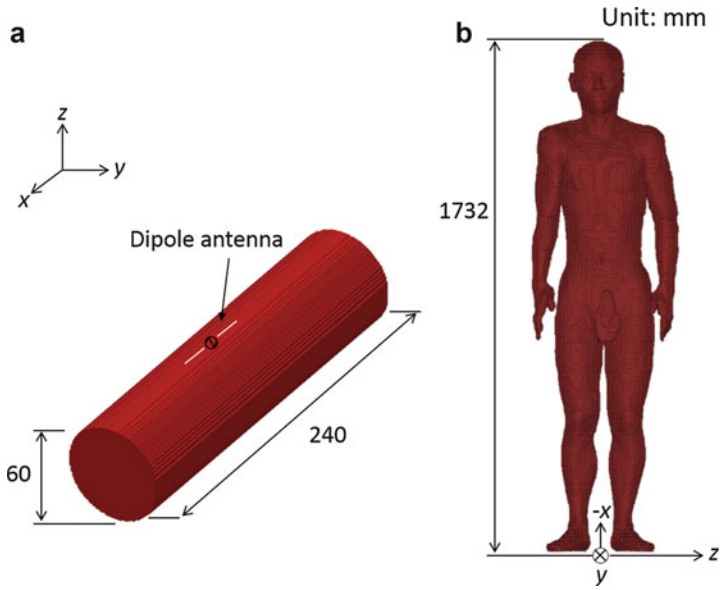


Fig. 28 Simulation models. (**a**) Simple human arm. (**b**) High-resolution human model

In order to maintain simulation consistency between these models, a half-wavelength dipole antenna is embedded into these models at a depth of 4 mm. In addition, the dipole antenna is designed to operate in the 1.27 GHz band; therefore, it has a length of 26 mm.

Figure 28a shows the simple human arm model consisting of a 2/3 muscle-equivalent phantom ($\varepsilon_r = 36.8$ and $\sigma = 0.75$ at 1.27 GHz). This model has a cylindrical structure with a length of 240 mm ($\approx \lambda$) and a diameter of 60 mm. In addition, a high-resolution human model based on the average adult Japanese volunteers (height, 173.2 cm; weight, 65 kg) is introduced in the simulation (Rajagopalan and Rahmat-Samii 2010), and its dielectric constant is the same as the simple human arm model. The cell size of the dipole antenna and that of the human model are set as 0.05 and 2.0 mm, respectively. Thus, the cells of the dipole antenna and human model are nonuniform. Perfectly matched layer boundary conditions (eight layers) are used to truncate the simulation region, and the voltage source is set as the model source. The computational parameters are listed in Table 5.

DOI 10.1007/978-981-4560-75-7_83-1

Table 5 Computational parameters

	Cylindrical model	Human-body model
Cell size	0.05–2 mm	
Computational domain ($\triangle x$, $\triangle y$, $\triangle z$)	0.4 m × 0.16 m × 0.27 m	2.0 m × 1.0 m × 1.0 m
PML layers	8 layers	

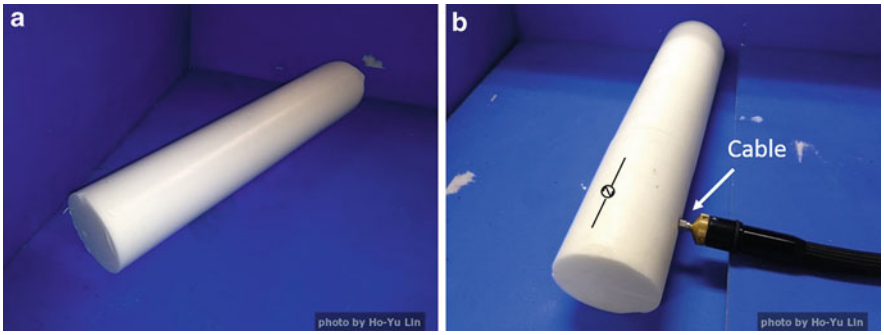


Fig. 29 Measurement. (**a**) Fabricated arm phantom. (**b**) Setup

Fabricated Human Phantom and Measurement Setup

Figure 29a, b depicts the fabricated human arm phantom and measurement setup for reflection coefficient, respectively. The composition of the human arm phantom and the target as well as measured dielectric constants of the 2/3 muscle-equivalent phantom at 1.27 GHz are described in Sani et al. (2009), respectively. All the measured results approach the 5 % error mark when compared to the target values according to Sani et al. (2010). As illustrated in Fig. 29b, the dipole antenna is embedded into the fabricated phantom at a depth of 4 mm. It should be noted that the cable is connected to the antenna, and it does not touch the human phantom, as current might flow through the phantom to cause a change in the input impedance of the antenna.

Figure 30 shows the fabricated upper-body human phantom consisting of a human head, shoulder, and trunk. This phantom has the same dielectric constant and composition as the simple human arm phantom. Figure 31 shows the measurement setup for radiation pattern of the dipole antenna. In the measurement setup in Fig. 31a, the fabricated upper-body human phantom is placed on the turntable, and the simple human arm phantom including the antenna is then connected to the human phantom to represent a human arm. Before measuring far-field radiation patterns, it should be confirmed that these phantoms are fixed properly onto the turntable when rotating because the phantoms are heavy. Photographs of the measurement setup for testing far-field radiation patterns of the dipole antenna are shown in Fig. 31b, c.

Reflection Coefficient and Far-Field Radiation Pattern

Figure 32 shows the simulated and measured reflection coefficients when the antenna is embedded into the simple human arm phantom. As a result, it is confirmed that the antenna can satisfy the desired frequency band of 1.27 GHz. Further, the measured results have reasonable agreement with the simulated results, even though there is a minor difference due to measurement errors such as the position of the dipole antenna and the dielectric constant of the human body. The far-field radiation pattern of the dipole antenna embedded in the human body is also studied. Figure 33 plots the simulated and measured far-field radiation patterns at 1.27 GHz. From the results in *yz*-plane, it is clearly observed that both radiation patterns are mainly in the $+z$ direction (0°); however, the radiation patterns in the $-z$ direction are attenuated owing to the presence of the human body. Furthermore, it is found that the magnitude of the radiation pattern in the $-z$ direction in measured results is higher than in the simulated results. The main reason for this is because the fabricated upper-body human phantom has only one arm that is slightly

DOI 10.1007/978-981-4560-75-7_83-1

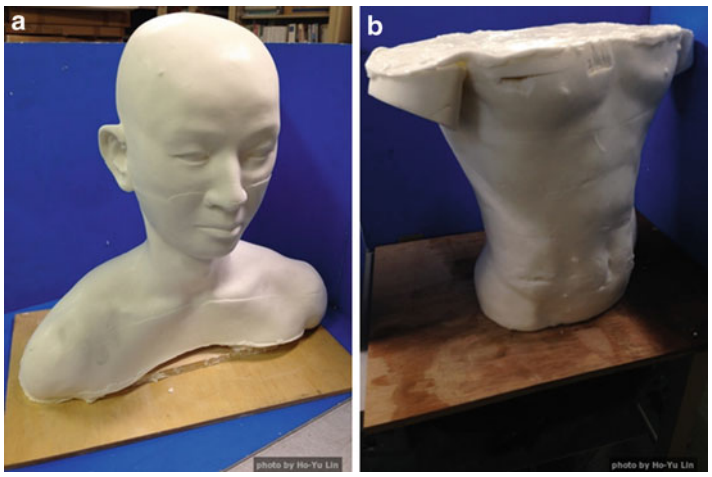


Fig. 30 Fabricated upper-body human phantom. (**a**) Head and shoulder. (**b**) Trunk

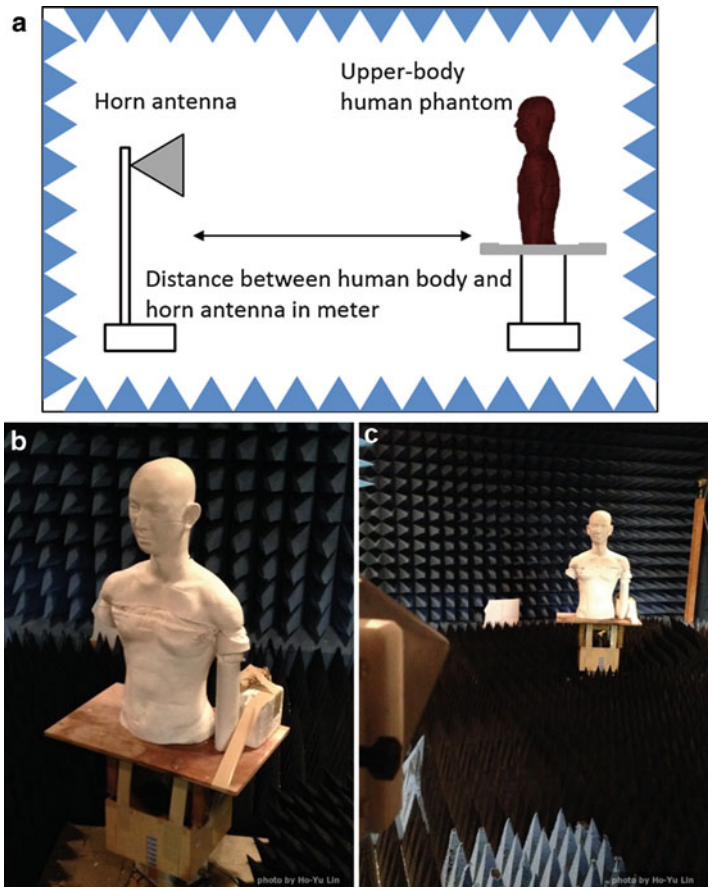


Fig. 31 Measurement setup for radiation pattern of the dipole antenna

different from the simulated model. Moreover, the surface current on the cable line may cause undesired radiation to influence the radiation pattern. In addition, the maximum antenna gains in the simulation and measurement are −19.3 and −17.4 dBi, respectively. In the upper half plane, the simulated and measured results have good agreement. Therefore, the far-field radiation pattern of the antenna in the human body can be confirmed using this measurement setup.

DOI 10.1007/978-981-4560-75-7_83-1

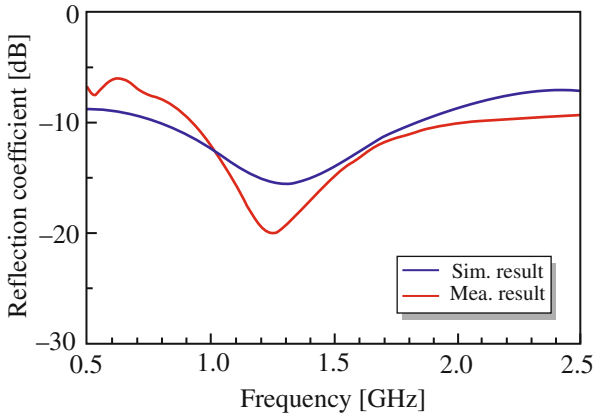


Fig. 32 Reflection coefficient of the implantable antenna

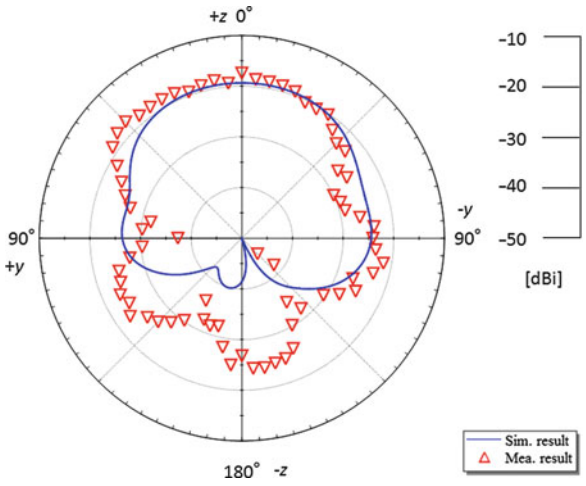


Fig. 33 Simulated and measured radiation patterns with upper-body human phantom

Conclusion

Body-centric wireless communications (BCWCs) systems have an important role in the fields of sports training, biomedicine, and personal healthcare. BCWCs provide wireless and maintenance-free communication link between the human body and the surroundings through wearable and implantable antennas. In order to evaluate such wearable and implantable antennas experientially, human-body phantoms are inevitable. In this chapter, firstly, some typical numerical and experimental phantoms have been summarized. Then, detailed evaluation techniques and results of wearable and implantable antennas with specific human-body phantoms have been described. In the future, so-called "advanced human phantoms" will be needed to evaluate more complicated and sophisticated antennas as well as scenarios. For example, "mm-wave phantoms" to evaluate mm-wave antennas attached to the human body and "dynamic phantoms" to simulate movement of the human body.

DOI 10.1007/978-981-4560-75-7_83-1

Cross-References

- ▶ Antenna Sensors in Passive Wireless Sensing Systems
- ▶ Antennas and EM Issues in Medical Diagnosis and Treatment Systems
- ▶ Antennas in Body-centric Sensor Network Devices
- ▶ Antennas in Handheld Devices
- ▶ Implanted Antennas in Biomedical Telemetry
- ▶ Small Antennas (PIFA/PILA/Loading Antenna/etc.)

References

Chow EY, Morris MM, Irazoqui PP (2013) Implantable RF medical devices: the benefits of high-speed communication and much greater communication distances in biomedical applications. IEEE Microwave Mag 14(4):64–73

Fujii K, Takahashi M, Ito K, Hachisuka K, Terauchi Y, Kishi Y, Sasaki K, Itao K (2005) Study on the transmission mechanism for wearable device using the human body as a transmission channel. IEICE Trans Commun E88-B(6):2401–2410

Fujitsu website. http://www.fujitsu.com/global/vision/2014/chapter1/02/. Accessed 24 July 2015

Hall PS, Hao Y (2006) Antennas and propagation for body-centric wireless communications. Artech House, Norwood

Hall PS, Hao Y, Ito K (2009) Guest editorial for the special issue on antennas and propagation on body-centric wireless communications. IEEE Trans Antennas Propag 57(4):834–836

Halperin D, Kohno T, Heydt-Benjamin TS, Fu K, Maisel WH (2008) Security and privacy for implantable medical devices. IEEE Pervasive Comput 7(1):30–39

IEEE standard website_a. http://ieee802.org/15/pub/TG6.html. Accessed 23 July 2015

IEEE standard website_b. http://standards.ieee.org/findstds/standard/802.15.6-2012.html. Accessed 23 July 2015

IEEE standard website_c. https://standards.ieee.org/findstds/standard/1528-2003.html. Accessed 30 September 2015

Ito K, Haga N, Takahashi M, Saito K (2009) Electric field distributions around a human body with a small antenna in the frequency range of 2.5 MHz to 2.5 GHz. In: Proceeding of the international workshop on antenna technology 2009, Santa Monica

Ito K, Haga N, Takahashi M, Saito K (2012) Evaluation of body-centric wireless communication channels in a range from 3 MHz to 3 GHz. Proc IEEE 100(7):2356–2363

Karacolak T, Cooper R, Topsakal E (2009) Electrical properties of rat skin and design of implantable antennas for medical wireless telemetry link budget analysis. IEEE Trans Antennas Propag 57(9):2806–2812

Kiourti A, Nikita KS (2012a) A review of implantable patch antennas for biomedical telemetry: challenges and solutions. IEEE Antennas Propag Mag 54(3):210–228

Kiourti A, Nikita KS (2012b) Miniature scalp-implantable antennas for telemetry in the MICS and ISM bands: design, safety considerations and link budget analysis. IEEE Trans Antennas Propag 60(8):2568–3575

Li Z, Du Z, Gong K (2009) Compact Reconfigurable Antenna Array for Adaptive MIMO Systems, IEEE Antennas and Wireless Propagation Letters, 8, 1317–1320

Lin CH, Saito K, Takahashi M, Ito K (2012a) A compact planar inverted-F antenna for 2.45 GHz on-body communications. IEEE Trans Antennas Propag 60(9):4422–4426

Lin CH, Li Z, Ito K, Takahashi M, Saito K (2012b) Dual-mode antenna for on-/off-body communications (10 MHz/2.45 GHz). Electron Lett 48(12):1383–1385

Lin HY, Takahashi M, Saito K, Ito K (2013) Performance of implantable folded dipole antenna for in-body wireless communication. IEEE Trans Antennas Propag 61(3):1363–1370

MCL-T broadband tissue equivalent liquid: 30 MHz to 6 GHz. MCL-T, London

Merli F, Bolomey L, Meurville E, Skrivervik AK (2008) Implanted antenna for biomedical applications. In: IEEE Antennas and propagation society international symposium (AP-S), San Diego, pp 1–4

NICT website_human model. http://emc.nict.go.jp/bio/model/model01_1_e.html. Accessed 23 July 2015

NICT website_SAR measurement. http://emc.nict.go.jp/bio/phantom/index_e.html. Accessed 23 July 2015

Rajagopalan H, Rahmat-Samii Y (2010) Novel ingestible capsule antenna designs for medical monitoring and diagnostics. In: Antennas and Propagation (EuCAP), Barcelona, pp 1–5

Sani A, Alomainy A, Hao Y (2009) Numerical characterization and link budget evaluation of wireless implants considering different digital human phantoms. IEEE Trans Microwave Theory Tech 57(10):2605–2613

Sani A, Rajab M, Foster R, Hao Y (2010) Antennas and propagation of implanted RFIDs for pervasive healthcare applications. Proc IEEE 98(9):1648–1655

Zimmerman TG (1996) Personal area networks: near-field intrabody communication. IBM Syst J 35(3/4):609–917

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_86-1

Antenna Sensors in Passive Wireless Sensing Systems

Haiying Huang*
University of Texas Arlington, Arlington, TX, USA

Abstract

Antenna sensors have received considerable interests in recent years due to their passive wireless operation, simple configuration, compact size, multiplexing capability, and multimodality sensitivity. Based on the principle of antenna backscattering, an antenna sensor can be wirelessly interrogated at middle range distances without an onboard battery. Since the antenna serves the dual function of sensing and communicating, an antenna sensor can be implemented with minimum number of components. As narrowband resonators, antenna sensors can be easily multiplexed based on the principle of frequency division multiplexing. In addition, different types of antenna sensors that are sensitive to a variety of physical measurands have been demonstrated. Due to these unique features, antenna sensor technology could play a vital role in our drive toward ubiquitous sensing. This chapter provides a comprehensive review of this exciting technology with detailed descriptions on the operational principle and wireless interrogation of batteryless antenna sensors. Four application examples of using antenna sensors for moisture, dynamic strain, temperature, and crack sensing are discussed. Future research directions and open problems are suggested.

Keywords

Antenna sensor; Batteryless wireless sensor; Dielectric sensor; Strain sensor; Crack sensor; Bio-chemical sensor; Moisture sensor; Temperature sensor; Dielectric substrate; Antenna backscattering; Amplitude modulation; Impedance switching; RFID

Introduction

Sensors, which convert physical measurands into electric signals, are becoming more and more prevalent in this Internet age. As our world becomes increasingly connected, sensors play the fundamental role of integrating the physical domain into the cyberspace. The advent of wireless sensors has fundamentally transformed sensor technologies in the past decades. By eliminating electric wiring from conventional sensors, wireless sensors are inexpensive and easier to install, which makes distributing sensors over a large area and with high density a reality. A typical wireless sensing system is shown in Fig. 1; it composes of two sub-systems, namely, the wireless sensor and the wireless reader. The wireless sensor is responsible for encoding the sensing information into a radio-frequency (RF) signal that can be wirelessly transmitted via an antenna, while the function of the wireless reader is to decode the received RF signal to extract the original sensing information. In most cases, an interrogation signal is transmitted from the wireless reader to the wireless sensor via the uplink. In response, the wireless reader sends the sensing information back to the wireless reader via the downlink. Traditionally, antennas only serve as a communication device for wireless sensors. Recent developments, however, have extended the role of the antennas for sensing.

*Email: huang@uta.edu

DOI 10.1007/978-981-4560-75-7_86-1

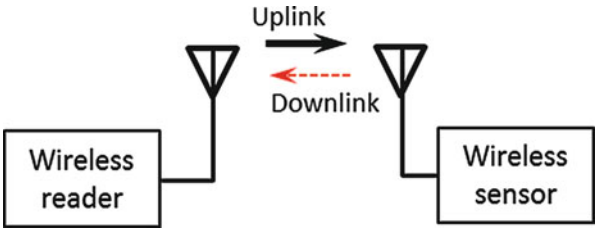


Fig. 1 A typical wireless sensing system consisting of a wireless sensor and a wireless reader

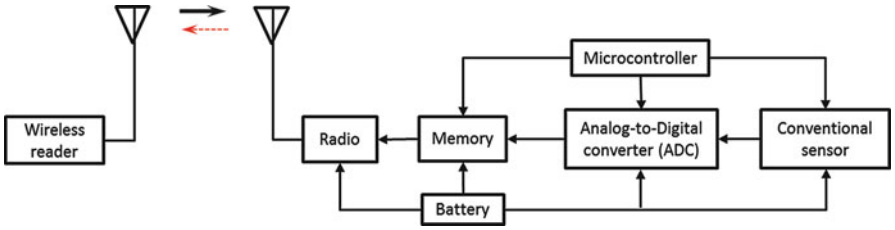


Fig. 2 Architecture of active wireless sensor nodes

This chapter aims to introduce the readers to this new exciting technology with an in-depth discussion of its operational principle, wireless interrogation schemes, application examples, and future directions. For the readers who are not familiar with wireless sensor technologies, a brief review of common wireless sensors is given first.

Active Wireless Sensors

A wireless sensor can be broadly classified as either being active or passive, depending on whether it needs onboard batteries or not. The most common active wireless sensors are based on the Berkeley mote platform, which is illustrated in Fig. 2. This kind of wireless sensor essentially functions as a microcomputer with the capability of wirelessly receiving data from and transmitting data to the wireless reader, also known as the base station. The "brain" of an active wireless sensor is the microprocessor, which controls and schedules the operations of other electronic components including the analog-to-digital converter (ADC), memory, radio, etc. The sensing signal is usually acquired in an analog format using a conventional sensor, e.g., a strain gauge, an accelerometer, a thermocouple, etc., which is then converted to digital signals by the ADC. The digital signals can be transmitted directly by the radio or processed by the microprocessor before transmission. Due to the digital nature of the wireless communication protocol and the onboard radio, wireless sensor nodes can provide reliable and interference-free data communication for a distance of more than 100 m. Active wireless sensor technology has been mature enough that several standards, such as the Bluetooth and ZigBee standards, have been developed, which facilitates the commercialization of such sensors. Currently, active wireless sensors can be purchased off-the-shelf at less than 50 dollars. One major limitation of the active wireless sensors is their requirement for onboard batteries since the electronic components as well as the sensors consume substantial amount of power. Replacing or recharging the battery introduces a significant maintenance challenge. Even though various energy harvesting techniques have been investigated to address this limitation, there is still a significant gap between the power consumption of the wireless sensors and the small energy level that can be harvested from ambient sources.

DOI 10.1007/978-981-4560-75-7_86-1

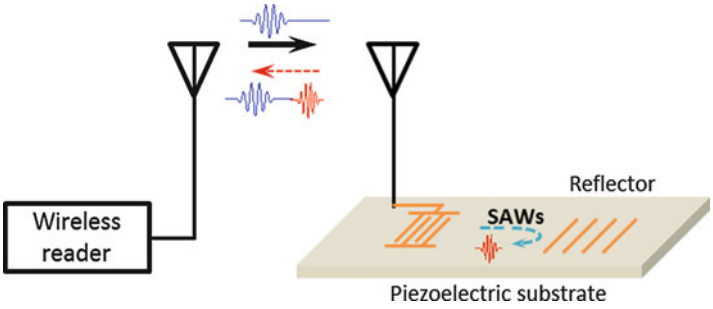


Fig. 3 SAW-based passive wireless sensing system

Passive Wireless Sensors

Passive wireless sensors are viable alternatives to the battery-powered wireless sensors, especially for applications that can tolerate shorter communication ranges. A major difference between an active wireless sensor and a passive wireless sensor is that a passive wireless sensor does not have a radio on board. As such, these sensors either consume no power at all or their power consumption is so low that energy harvested from other sources is sufficient to power them up. Passive wireless sensors can be broadly classified into four categories, namely, the surface acoustic wave (SAW)-based, inductive coupled, radio-frequency identification (RFID) enabled, and antenna sensors. The SAW-based and inductive coupled wireless sensors predate the invention of active wireless sensors. RFID systems, initially developed mainly as an electronic device to replace the barcode for product and inventory tracking, have been extended for low-cost, ubiquitous, passive wireless sensing in recent years. The antenna sensors are the latest arrival of the passive wireless sensor family. They share the same communication principles as the RFID sensors, but they do not need an external sensor and can thus achieve a much simpler configuration than most RFID-enabled sensors.

Passive Wireless SAW Sensors

The schematic illustration of a typical passive wireless SAW sensor is shown in Fig. 3. The wireless sensor consists of a piezoelectric substrate connected to an antenna. Reflectors are patterned on the piezoelectric substrate at short distances from the antenna connection. Taking advantage of the piezoelectric effect, the RF signal received by the antenna is directly applied to excite SAWs propagating along the substrate surface. Upon reaching the reflectors, the SAWs are reflected and converted back to the RF signals at the antenna connections, which is then re-transmitted by the antenna. Because the propagation speeds of the SAWs are much smaller than that of the RF waves, the two-way conversion between the RF waves and SAWs introduces a large time delay between the incident and re-transmitted RF signals. The delay between these two signals is dependent of the distances between the antenna and the reflectors as well as the wave speeds of the SAWs. A physical measurand (e.g., strain, temperature, moisture, etc.) that has an influence on the SAW wave speed can thus be detected from the time delay of the re-transmitted signal. In order to keep the SAW propagation losses reasonably low, the RF frequency used for exciting the SAWs is usually limited to tens of megahertz (MHz). Higher RF frequencies would introduce large SAW propagation losses and thus limit the interrogation distance of the wireless SAW sensors. This frequency limitation makes the miniaturization of the sensor antenna extremely difficult. Moreover, since the wave speeds of the SAWs can be influenced by multiple variants, decoupling their effects on the wave speeds of the SAWs is a challenging task.

DOI 10.1007/978-981-4560-75-7_86-1

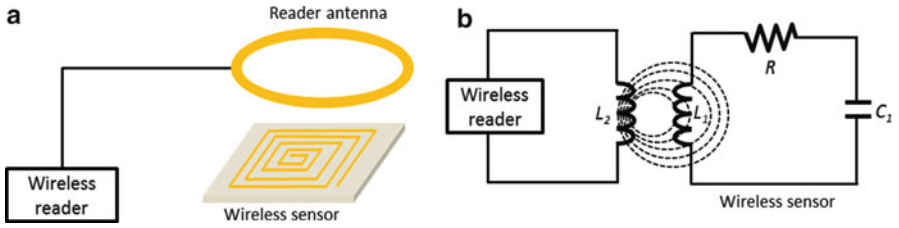


Fig. 4 Inductive coupled passive wireless sensing; (**a**) configuration; (**b**) lump element model

Passive Wireless Inductive Coupled Sensors

The schematic diagram of a simplest inductive coupled wireless sensing system is shown in Fig. 4a. The sensor node is a spiral resonant circuit in the form of a conductive trace patterned on a dielectric substrate. As such, it can be represented by an equivalent lump element circuit having an inductor and a capacitor in parallel (see Fig. 4b). The inductance of the resonant circuit is determined by the geometric parameters of the conductive trace, while its capacitance is related to the gap distance between the conductive traces and the dielectric properties of the substrate. Physical measurands that change these geometric and dielectric parameters therefore will change the inductance and/or capacitance values of the resonant circuit, which in turn cause the resonance frequency to shift. When a reader antenna is placed in the near field of the spiral, the spiral resonant circuit is inductively coupled to the electromagnetic (EM) field of the reader antenna. Therefore, the energy passing through the reader antenna is largely absorbed by the spiral resonant circuit when the interrogation frequency matches with the resonant frequency of the spiral circuit. Sweeping the supplied frequency and monitoring the power of the signal reflected by the reader antenna can thus determine the resonance frequency of the spiral resonance. Compared to SAW-based passive wireless sensors, inductively coupled wireless sensors have a much simpler configuration. A major limitation of such sensors is that the interrogation distance is typically limited to be a few centimeters since the spiral circuit has to be placed in the near field of the reader antenna.

Passive Wireless RFID Sensors

A passive wireless RFID system is shown in Fig. 5a. The RFID reader broadcasts an interrogation signal to the RFID tag, while the RFID tag responds by sending a signal encoded with the identification (ID) number. An RFID tag consists of an antenna and an RFID chip. A detailed diagram of the RFID chip is shown in Fig. 5b. One function of the RFID chip is to harvest energy from the interrogation signal. Once sufficient energy is harvested, the RFID chip is activated to retrieve the ID number from its memory and transmits the ID number via impedance switching of the antenna load. Once received by the RFID reader, the backscattered signal can be processed to retrieve the ID number.

Four common RFID sensor topologies that can incorporate sensing function into an RFID tag are shown in Fig. 6. The Wireless Identification and Sensing Platform (WISP), initially developed by Intel Research, probably is the most complicated and versatile passive wireless RFID sensor. Similar to an active wireless sensor, a WISP sensor is equipped with a microprocessor, an ADC, an external memory, and one or more conventional sensors, as shown in Fig. 6a. The differences between a WISP sensor and an active wireless sensor are as follows: (1) a WISP sensor is powered by energy harvested from the interrogation signal while an active wireless sensor is powered by a battery and (2) a WISP sensor transmits the data by modulating the backscattered signal and thus does not have an onboard radio while an active wireless sensor transmits the data through the radio. Even though a WISP sensor consumes substantially less power than an active wireless sensor, the energy provided by the RFID reader is still not sufficient enough to power the WISP sensor continuously in most cases. Similar to the active wireless sensors, the power management of the WISP sensors remained to be a challenge. An alternative for

DOI 10.1007/978-981-4560-75-7_86-1

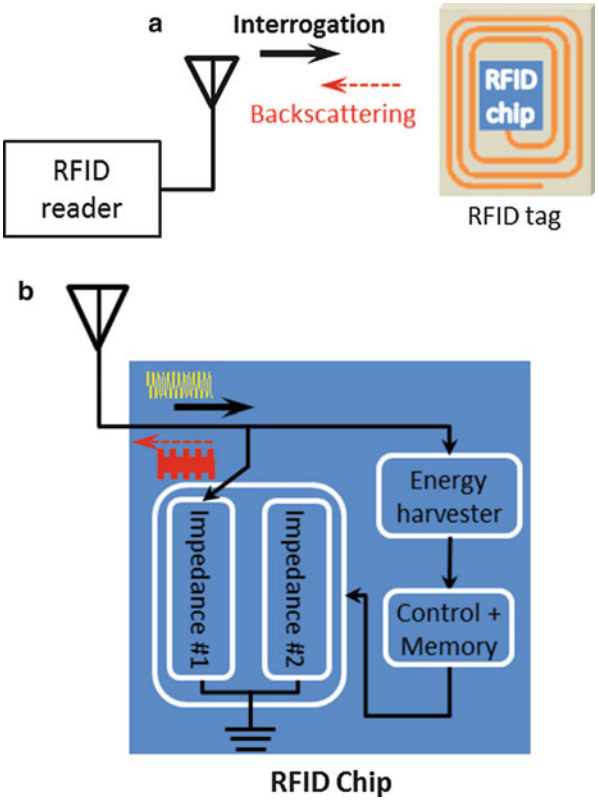


Fig. 5 Passive wireless RFID sensing system; (**a**) system configuration; (**b**) diagram of RFID tag

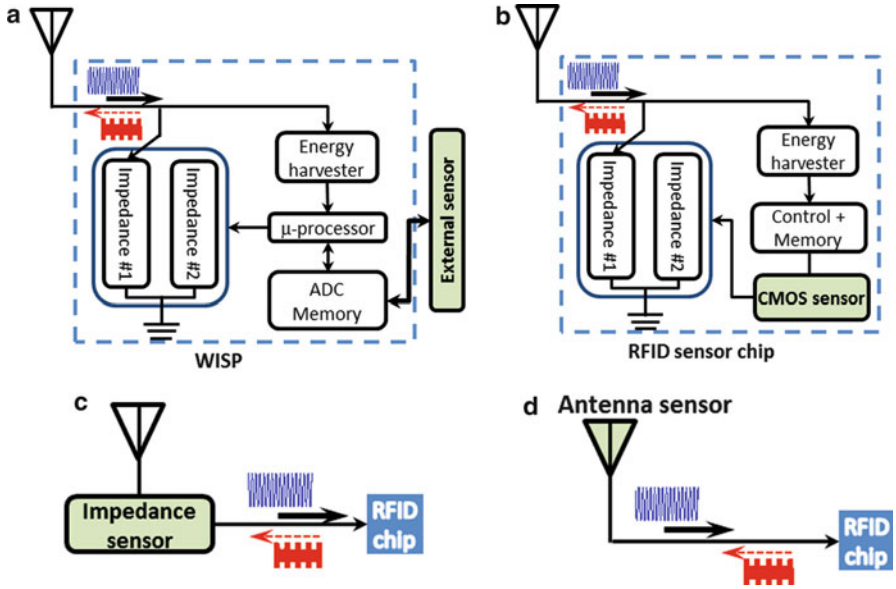


Fig. 6 Four common RFID sensor topologies: (**a**) WISP; (**b**) CMOS-based RFID sensor; (**c**) impedance load-based RFID sensor; (**d**) RFID-enabled antenna sensor

integrating sensing function into an RFID tag is to implement low-power on-chip CMOS sensors into the RFID chip (see Fig. 6b). This approach is compatible with massive integrated circuit (IC) fabrication techniques and can result in very compact, low-power, and sensitive RFID sensor chips. The development

DOI 10.1007/978-981-4560-75-7_86-1

of these chips, on the other hand, required customized chip design and fabrication. In addition, the types of sensor that can be implemented into a CMOS IC chip are also limited. The third type of RFID sensors uses an analog impedance sensor to change the load of the RFID tag antenna, as shown in Fig. 6c. The impedance sensor is designed to change its impedance with the desired physical measurands. Since the radar cross section (RCS) of a loaded antenna is depended of the load impedance, the impedance change of the sensor can thus be detected from the backscattered RFID signal, which in turn can be correlated to the physical parameter that causes the impedance change. The advantage of such sensors is that they do not need special RFID chips and thus can keep the cost of the sensor node low. The fourth type of RFID sensor eliminates the external sensor by employing the RFID tag antenna as the sensor. These RFID sensors can be considered as a subset of passive wireless antenna sensors and therefore will be discussed in detail in this chapter. Among these four types of RFID sensors, the first two encode the sensor data digitally while the other two encode the sensor data in an analog format. In general, digital encoding is more reliable but consumes higher power, while analog encoding is very power efficient but is more susceptible to EM interferences.

Passive Wireless Antenna Sensors

Passive wireless antenna sensors probably have the simplest configurations among all wireless sensors. As shown in Fig. 7, a passive wireless antenna sensor consists of only two components, i.e., an antenna and a modulator. The antennas can be a regular antenna fabricated on conventional dielectric materials or coated with functionalized materials. The main purpose of the modulator is to modulate the interrogation signal received by the antenna so that the signal backscattered by the antenna, i.e., the antenna backscattering, can be separated from the signals backscattered by the surrounding structures, i.e., the structural backscattering. Depending on the sensor design, the modulator can be completely passive or consumes very low power. The sensing function is performed by detecting the change in the radiation characteristics of the antenna sensor, such as the resonance frequency shift. Antenna sensors that are sensitive to moisture, gas, temperature, strain, displacement, crack, etc., have been demonstrated. The unique advantages of these passive wireless antenna sensors can be summarized as the following:

- **Simple configuration**: The antenna itself serves the dual function of communication and sensing. Therefore, no external sensor is needed.
- **Passive operation**: The sensing information is directly encoded in the antenna backscattering. As such, no onboard battery is needed at the wireless sensor node.
- **Multiplexing capability**: The antenna sensors can be multiplexed based on the principle of frequency division multiplexing, spatial division multiplexing, code division multiplexing, etc. This sensor multiplexing capability enables densely distributed sensor networks and parallel interrogation of multiple antenna sensors.
- **Multimodality sensing**: The antenna can be designed to be sensitive to various physical measurands.

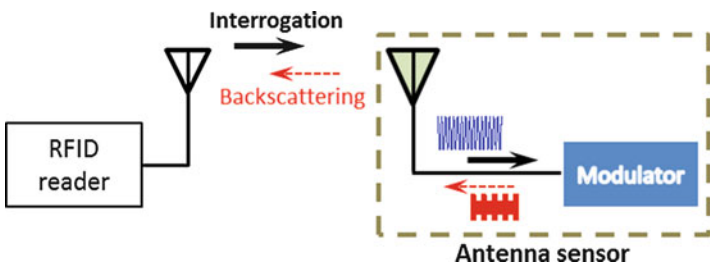


Fig. 7 Passive wireless sensing system based on antenna sensor

DOI 10.1007/978-981-4560-75-7_86-1

- **Low cost**: The antenna sensors can be fabricated on inexpensive materials using conventional printed circuit board fabrication techniques.
- **Planar and conformal**: The antenna sensors can be fabricated on low-profile, planar, flexible substrates that completely conform to the surface it attached to.

Antenna as Sensor

Historical Review

While the first antenna was demonstrated by Heinrich Rudolph Hertz in 1887, using antennas for sensing purposes is a more recent invention. The first antenna sensor found in literature was published by Gagnadre et al. (1995), in which they described a 1.5 GHz circular patch antenna for measuring the moisture content of sludge samples. The sludge sample was contained inside a plastic beaker and placed on top of the circular patch antenna. The moisture content of the sample was calculated from the measured effective dielectric constant of the antenna sensor using the particular model and the Bottcher model. They discovered that their method was able to differentiate about 0.3 % difference in the moisture content of the sludge sample when the moisture content was high. Various antenna sensors, designed for measuring the dielectric constant of snow and soils (Denoth 1997), gas (McGrath et al. 2004; Verma et al. 2013), relative humidity (Chang et al. 2006), soil moisture (You et al. 2010), and glucose (Wiwatwithaya et al. 2011), were demonstrated using wired connections. The first wireless antenna sensor was demonstrated by Sidén et al. (2007). They used a pair of regular RFID tags for relative humidity measurements. One of the RFID tags was covered with a water absorbent material while the other tag was open. Relative humidity was measured from the differences in the turn-on powers of these two RFID tags. Following a similar approach, Bhattacharyya et al. (2010) investigated RFID tags for displacement, filling level, and temperature threshold sensing.

Antenna temperature sensors can be developed by exploiting the temperature dependency of the dielectric material. The first antenna temperature sensor was demonstrated by Bhattacharyya et al. (2011) for temperature threshold detection. The sensor was constructed from a shape memory polymer (SMP) sheet sandwiched between an RFID tag and a metallic plate. The SMP changes its dielectric constant when the temperature crosses a threshold, which in turn can be detected from the turn-on power of the RFID tag. This sensor, however, can only detect two temperature states. Continuous temperature measurements were demonstrated by Yang et al. (2012) and Qiao et al. (2013). Both works used a slotted patch antenna; (Yang et al. 2012) integrated a water pocket in the substrate of the patch antenna, while Qiao et al. (2013) used a high-density polyethylene $Ba_{0.3}Sr_{0.7}TiO_3$ (HDPE-BST) material as the antenna substrate to achieve enhanced temperature sensitivity. Recently, Jiang et al. (2014) investigated a conventional microwave substrate (Rogers Corporation, RO3005) for temperature sensing. Temperature sensitivities of 133.1 ppm/°C and 142.6 ppm/°C, which are equivalent to 665 kHz/°C and 856 kHz/°C, were achieved for the two fundamental resonant frequencies of a rectangular patch antenna. All of the antenna sensors discussed above were based on the principle of measuring the dielectric constant of the antenna and thus can be classified as antenna dielectric sensors. One challenge of developing antenna dielectric sensors is that the relationship between the antenna radiation parameters and the physical measurand is determined by the material property, which in most cases cannot be modelled analytically. As a result, most of the works on the antenna dielectric sensors are experiment based.

The first mathematical derivation of the antenna sensing principle was presented by Tata et al. (2009). Based on the transmission line model, the relationship between the antenna frequency shift and the

DOI 10.1007/978-981-4560-75-7_86-1

applied strain relationship was rigorously derived. The analytical model predicted that the normalized frequency shift of a patch antenna is linearly proportional to the applied strain. This prediction was validated using a rectangular patch antenna installed on a cantilever beam. Following Tata's work, other antenna types, including circular patch (Daliri et al. 2010), folded dipole (Merilampi et al. 2010, 2011) antennas, meander-line dipole (Occhiuzzi et al. 2011), folded patch (Yi et al. 2011), planar dipole (Jang and Kim 2012), embroidered dipole (Hasani et al. 2013), and slotted patch (Salmani et al. 2011; Yi et al. 2014) antennas, were also demonstrated for strain sensing. All these antenna sensors are based on the same principle, i.e., the applied strain changes the dimension of the antenna radiation element and thus causes a shift of the antenna resonant frequency. A different principle was exploited by Bhattacharyya et al. (2009) and Mohammad and Huang (2012a) for displacement and pressure sensing. In Bhattacharyya et al. (2009), an RFID tag antenna was placed below a metallic plate. The capability of the antenna sensor to measure beam deflection was demonstrated by attaching the metallic plate on a simply supported beam; the deflection of the beam changes the distance between the metallic plate and the tag antenna and thus alters the efficiency of the tag antenna, which can be detected from the turn-on power as well as the backscattered signal power. In Mohammad and Huang (2012a), a pressure sensor was constructed by separating a loop antenna and a metallic film using a foam material. Pressure applied on the sensor reduces the distance between the loop antenna and the metallic plate, which shifts the resonant frequency of the loop antenna. A linear relationship between the applied pressure and the resonant frequency shift was demonstrated experimentally.

The first antenna crack sensor was developed by Matsuzaki et al. (2009) using a dipole antenna made of carbon fiber-reinforced plastic (CFRP). They demonstrated that the antenna resonant frequency increases correspondingly when a crack shortens the antenna element. Therefore, it is possible to detect the crack location from the measured antenna resonant frequency increase. Kalansuriya et al. (2013) demonstrated an RFID-enabled crack sensor based on the same principle. Since the antenna sensor can only detect the crack when it severs the dipole element, they also exploited using a tag array to detect the propagation of a crack. Mohammad and Huang (2010) demonstrated an antenna sensor that can detect not only the crack growth but also the crack opening. Later on, the same group also demonstrated that it is possible to use a rectangular patch antenna to detect the crack orientation (Mohammad et al. 2012) and the propagation of a hidden crack (Erdmann and Huang 2011). Similar to the antenna strain sensors, these antenna crack sensors operate on the principle that a crack changes the electric dimensions of the antenna and thus can be detected from the resulting frequency shift. Mohammad and Huang (2012b) also demonstrated an antenna shear sensor based on the same principle.

A list of antenna sensors is given in Tables 1, 2, 3, and 4, grouped based on their sensing measurands in chorological orders. This list includes all the published work on antenna sensors the author found in the literature. The references marked with an asterisk demonstrated wireless interrogation of the antenna sensor, while the references without the asterisk only demonstrated the operation of the antenna sensor using wired connections.

Principle of Operation

The operational principle of an antenna sensor is explained in this section using the microstrip rectangular patch antenna as an example. The operational principles of the other antenna types are similar and can be found in the references given above. A microstrip rectangular patch antenna, as shown in Fig. 8a, consists of four components, i.e., the ground plane, the dielectric substrate, the radiation patch, and the transmission feed line. Both the ground plane and the radiation patch are made of conductive materials. Separated by the dielectric substrate, they form an EM resonance that radiates at specific frequencies. The radiation patch is "fed" by the microstrip transmission line, through which an incident signal can be supplied. This incident signal is either transmitted by the radiation patch or reflected by it. Therefore, the radiation

DOI 10.1007/978-981-4560-75-7_86-1

Table 1 List of antenna dielectric sensors

References	Measurand	Antenna				Sensing parameter
		Type	Size	Frequency	Material	
(Gagnadre et al. 1995)	Moisture of sludge	Circular patch	36 mm radius	1.95 GHz		Frequency shift
(Denoth 1997)	Dielectric of snow or soils	Monopole	86 mm long	0.1–2 GHz		Drive point impedance
(McGrath et al. 2004)	Gas	Circular patch	~8.6 mm radius	5.5 GHz	CNT	Frequency shift
(Chang et al. 2006)	Relative humidity	Meander patch		~5.36 GHz	Polyimide	Frequency shift
(Sidén et al. 2007)*	Relative humidity	Two RFID tags	95 × 8 mm each	865–868 MHz	Paper superstrate	Threshold power
(Bhattacharyya et al. 2009)*	Displacement	RFID UHF tag	80 × 80 mm^2	915 MHz	Metallic backplane	Threshold power
(Marrocco et al. 2009)*	Filling level of plastic bottle	Planar dipole antenna				Turn-on and signal power
(You et al. 2010)	Soil moisture	Regular and slotted patch	31.8 × 31.8 mm^2	1.5 & 3 GHz	FR4	Frequency shift and return loss
(Wiwatwithaya et al. 2011)	Glucose	U-shaped patch		~1.9 GHz		Return loss
(Potyrailo et al. 2011)*	Conductivity	RFID tag		13.56 MHz		Feed point impedance
(Verma et al. 2013)	Gas	Patch antenna	41 × 41 mm^2	2.4 GHz	Duroid and CNT	Frequency shift

* wireless interrogation demonstrated

Table 2 List of antenna temperature sensors

References	Measurand	Antenna				Sensing parameter
		Type	Size (mm^2)	Frequency	Material	
(Bhattacharyya et al. 2010)*	Temperature threshold	RFID tag		902–928 MHz	Water and Al plate	Signal strength
(Bhattacharyya et al. 2011)*	Temperature threshold	RFID tag	130 × 20	902–928 MHz	SMP and metal plate	Turn-on time
(Potyrailo et al. 2011)*	Temperature			13.56 MHz		Feed point impedance
(Yang et al. 2012)*	Temperature	Slotted patch	38 × 38	900 MHz	Water pocket	Frequency shift
(Qiao et al. 2013)*	Temperature	Slotted patch	44.5 × 44.5	900 MHz	HDPE-BST	Frequency shift
(Jiang et al. 2014)	Temperature	Rectangular patch	11.8 × 9.8	4.85 and 5.95 GHz	Duroid RO3006	Frequency shift

efficiency of the patch antenna can be defined by the ratio between the reflected power and the incident power, also known as the return loss. Since the patch antenna functions as a resonator, its return loss is frequency dependent. The S_{11} parameter of a patch antenna, which represents the return loss of the patch antenna as a function of frequency, is shown in Fig. 8b. Three radiation parameters can be extracted from the S_{11} parameter, i.e., the resonant frequency, the resonant bandwidth, and the baseline losses. The

DOI 10.1007/978-981-4560-75-7_86-1

Table 3 List of antenna mechanical sensors

References	Measurand	Antenna				Sensing parameter
		Type	Size	Frequency	Material	
(Tata et al. 2009)	Strain	Rectangular patch	4 × 5.3 mm^2	17.2 and 20.5 GHz	Kapton	Frequency shift
(Bhattacharyya et al. 2009)*	Displacement	RFID tag	80 × 80 mm^2	915 MHz	Metal and tag	Turn-on and signal power
(Daliri et al. 2010)	Strain	Circular patch		1.5 GHz	FR4	Frequency shift
(Merilampi et al. 2010)*	Strain	Folded dipole	100 mm long	915 MHz	PVC and fabric	Signal strength
(Merilampi et al. 2011)*	Large deformation	Folded dipole	8 × 97 mm^2	866 MHz	PVC and fabric	Turn-on and signal power
(Occhiuzzi et al. 2011)*	Strain	Meander line dipole	36 × 36 mm^2	868–956 MHz		Signal power
(Salmani et al. 2011)	Strain	Slotted patch	19.8 × 17.9 mm^2	5 and 9.26 GHz	Duroid 5880	Frequency shift
(Yi et al. 2011)*	Strain	Folded patch	56.6 × 50 mm^2	921 MHz	Duroid 5880	Frequency shift
(Caizzone et al. 2012)*	Surface deformation	RFID grid	148 × 10 mm^2		Foam	Turn-on and signal power
(Jang and Kim 2012)	strain	Planar dipole	17 × 16 mm^2	8–12 GHz	Plastic	Frequency shift
(Mohammad and Huang 2012a)	Pressure	Loop antenna	6.5 mm radius	6.5 GHz	Rogers 4350B	Frequency shift
(Mohammad and Huang 2012b)	Shear deformation	Patch antenna	15 × 12.75 mm^2	4.6 and 5.6 GHz	Foam	Frequency shift
(Yi et al. 2013)*	Strain and crack	Folded patch	6 × 6.9 cm^2	915 MHz	Duroid 5880	Frequency shift
(Hasani et al. 2013)*	Strain	Embroidered dipole	150 mm long	915 MHz	Polyester fabric	Turn-on power
(Yi et al. 2014)*	Strain	Slotted antenna	44 × 48 mm^2	900 MHz	Duroid 5880	Turn-on power
(Yao et al. 2014)*	Dynamic strain	Rectangular patch	21.6 × 18 mm^2	5.38 GHz	Duroid 5880	Frequency shift

resonant frequency is identified as the frequency at which the return loss is a local minimum; at this frequency, most of the incident power is radiated by the antenna, and thus, little energy is reflected back. The resonant bandwidth is usually measured as the frequency span of the resonant valley at a given return loss, e.g., at −6 dB. The baseline loss represents the overall loss of the antenna that is independent of the resonance. In theory, all of these three radiation parameters can be used to transduce the physical measurand into a measurable radiation parameter. In practice, however, the resonant frequency shift is the most commonly used sensing parameter. One of the advantages of the resonant frequency is that it is reference-free, i.e., the measured resonant frequency is not influenced by other parameters such as the interrogation range, transmission losses, etc. Therefore, it can provide absolute, robust, and reliable measurements.

Analytical Model of Antenna Resonant Frequency and Its Sensitivity to Measurands

The resonant frequency of a rectangular patch antenna can be calculated as

DOI 10.1007/978-981-4560-75-7_86-1

Table 4 List of antenna crack sensors

References	Measurand	Antenna Type	Size	Frequency	Material	Sensing parameter
(Matsuzaki et al. 2009)	Crack	Dipole antenna	365 mm long	400 MHz	Carbon composite	Frequency shift
(Mohammad and Huang 2010)	Crack opening and growth	Rectangular patch	15 × 12.5 mm^2	6.1 and 8.6 GHz	Kapton	Frequency shift
(Deshmukh et al. 2011)*	Crack growth	Rectangular patch	15 × 12.8 mm^2	7.1 GHz	Kapton	Frequency shift
(Erdmann and Huang 2011)	Hidden crack	Rectangular patch	12.75 × 14 mm^2	6.6 and 7.8 GHz	Kapton superstrate	Frequency shift
(Mohammad et al. 2012)	Crack orientation	Rectangular patch	15 × 12.75 mm^2	5.75 and 6 GHz	Kapton	Frequency shift ratio
(Kalansuriya et al. 2013)*	Crack	Dipole	160 mm long	915 MHz		Signal power

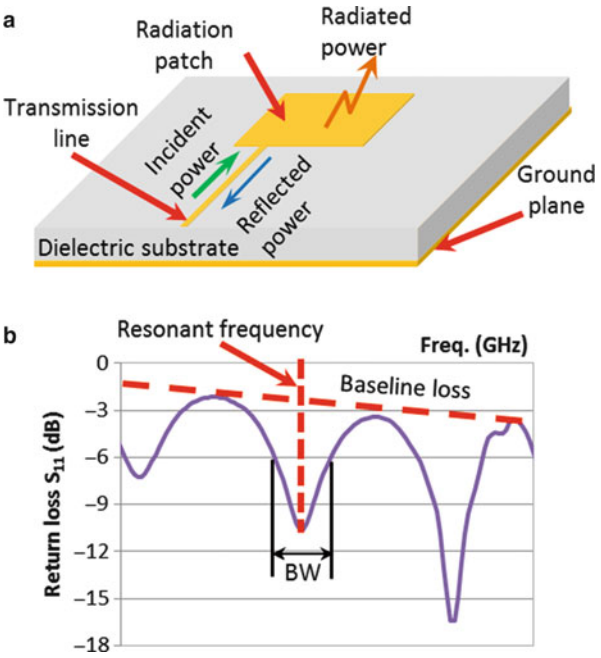


Fig. 8 (**a**) Configuration of a microstrip rectangular patch antenna; (**b**) S_{11} representation of antenna radiation characteristics

$$f_{mn} = \frac{C}{2\pi\sqrt{\varepsilon_{\text{reff}}}}\sqrt{\left(\frac{m\pi}{L_e}\right)^2 + \left(\frac{n\pi}{W_e}\right)^2}, \tag{1}$$

where c is the speed of light and $\varepsilon_{\text{reff}}$ is the effective dielectric constant. The electric length L_e and the electric width W_e can be calculated from the geometric dimensions of the radiation patch and the fringe extensions as

$$L_e = L + \Delta L \text{ and } W_e = W + \Delta W. \tag{2}$$

DOI 10.1007/978-981-4560-75-7_86-1

(m, n) represents the resonant orders; for an antenna with a rectangular radiation patch, the fundamental (1,0) mode, i.e., the TM_{10} mode, has the current flow along the length direction while the fundamental (0,1) mode, i.e., the TM_{01} mode, has the current flow along the width direction. Since Eq. 1 is almost identical for the TM_{10} and TM_{01} modes, the derivations given below will be focused on the TM_{10} mode. The resonant frequency of the TM_{10} mode can be simplified from Eq. 1 as

$$f_{10} = \frac{c}{2\sqrt{\varepsilon_{\text{reff}}}} \frac{1}{L_e}. \tag{3}$$

When the substrate height h is much smaller than the dimensions of the radiation patch, the line extension ΔL can be neglected. Therefore, Eq. 3 is reduced to

$$f_{10} = \frac{c}{2L\sqrt{\varepsilon_{\text{reff}}}}. \tag{4}$$

The relationship between the antenna frequency shift δf_{10} and the changes in $\varepsilon_{\text{reff}}$ and L can be derived by taking the full derivative of Eq. 4, i.e.,

$$\delta f_{10} = \frac{\partial f_{10}}{\partial \varepsilon_{\text{reff}}} \delta \varepsilon_{\text{reff}} + \frac{\partial f_{10}}{\partial L} \delta L, \tag{5}$$

in which

$$\frac{\partial f_{10}}{\partial \varepsilon_{\text{reff}}} = \left(-\frac{1}{2\varepsilon_{\text{reff}}}\right) \frac{c}{2L\sqrt{\varepsilon_{\text{reff}}}} = \left(-\frac{1}{2\varepsilon_{\text{reff}}}\right) f_{10} \quad \text{and} \quad \frac{\partial f_{10}}{\partial L} = \left(-\frac{1}{L}\right) \frac{c}{2L\sqrt{\varepsilon_{\text{reff}}}} = \left(-\frac{1}{L}\right) f_{10}. \tag{6}$$

Normalizing δf_{10} with respect to f_{10} results in

$$\frac{\delta f_{10}}{f_{10}} = -\frac{1}{2} \frac{\delta \varepsilon_{\text{reff}}}{\varepsilon_{\text{reff}}} - \frac{\delta L}{L}. \tag{7}$$

Equation 7 serves as the theoretical foundation for the antenna sensors. The first term represents the sensitivity of the antenna resonant frequency to the effective dielectric constant changes of the antenna, which can be exploited for environment (e.g., temperature, moisture, gas, etc.) and biochemical sensing. The second term represents the sensitivity of the antenna resonant frequency with respect to the dimensional change of the radiation patch, which can be exploited for strain, crack, and temperature sensing.

Dielectric Sensing

An antenna dielectric sensor can be a regular patch antenna with an exposed radiation patch as shown in Fig. 8a or a patch antenna with a superstrate, i.e., the radiation patch is covered with a dielectric material, as shown in Fig. 9. In the latter case, the effective dielectric constant of the antenna sensor is contributed by both the substrate and the superstrate. The effective dielectric constant of the patch antenna can then be calculated by introducing two filling factors q_1 and q_2 as (Bogosanovich 2000)

$$\varepsilon_{\text{reff}} = q_1\varepsilon_1 + q_2\varepsilon_2 + (1 - q_1 - q_1), \tag{8}$$

DOI 10.1007/978-981-4560-75-7_86-1

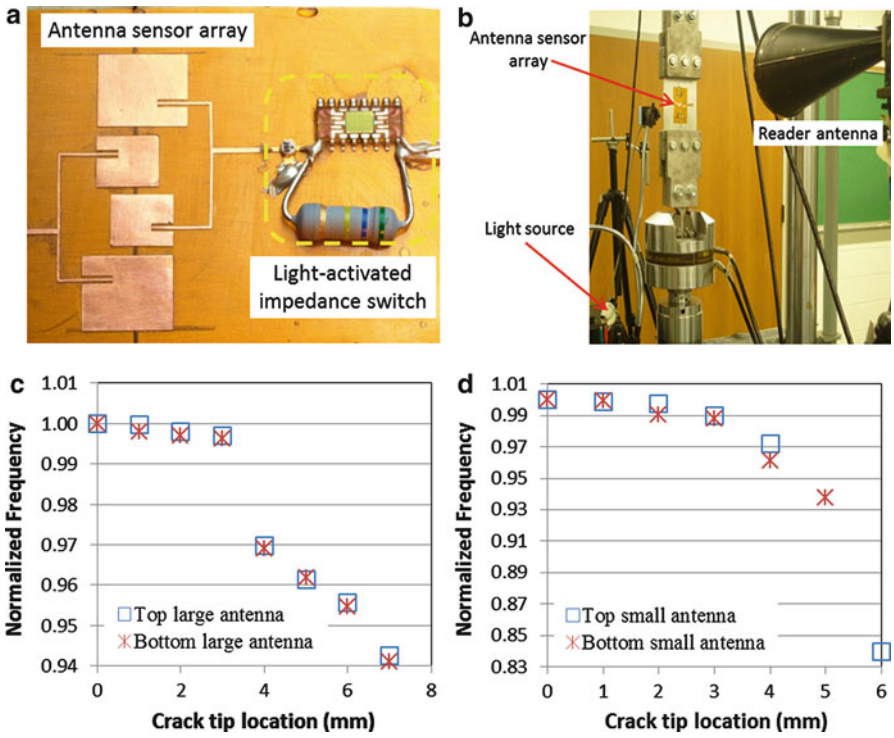


Fig. 9 An antenna dielectric sensor with its radiation patch covered by a dielectric superstrate. The superstrate can be functionalized material for selected sensing

where ε_1 and ε_2 are the dielectric constants of the substrate and superstrate, respectively. The third term accounts for the medium above the superstrate, which is assumed to be air. The superstrate material can be regular materials such as paper for moisture sensing or functionalized material such as carbon nanotubes (CNT) for gas sensing. The relationship between the effective dielectric constant and the physical measurand is usually calibrated by experiments. Representing the physical measurand as μ and assuming the effective dielectric constant can be represented as a function of μ, i.e., $\varepsilon_{\text{reff}} = f_\mu(\mu)$, the relationship between the normalized frequency shift and the physical measurand can then be derived from Eq. 7 as

$$\frac{\delta f_{10}}{f_{10}} = -\frac{1}{2\varepsilon_{\text{reff}}}\frac{\partial f_\mu}{\partial \mu}\delta\mu - \frac{\delta L}{L}, \tag{9}$$

which reduces to

$$\frac{\delta f_{10}}{f_{10}} = -\frac{1}{2f_\mu(\mu)}\frac{\partial f_\mu}{\partial \mu}\delta\mu \tag{10}$$

assuming the length of the patch antenna remains to be a constant.

Strain Sensing

As a direct measurement of the material deformation, strain is one of the most important mechanical properties to be sensed. Since strain can be easily correlated to the stress, external load, vibration, and boundary conditions of a load carrying structure, it is frequently used for evaluating the structural integrity of engineering components. There are two types of strains; the normal strain is defined as the relative change in the size of a structure with respect to its original size and the shear strain is defined as the change

DOI 10.1007/978-981-4560-75-7_86-1

in angle with respect to an original 90° angle. The normal strain along the length direction of the radiation patch is therefore defined as

$$\varepsilon_L = \frac{\delta L}{L}. \tag{11}$$

Assuming the effective dielectric constant is independent of the normal strain, i.e., $\delta\varepsilon_{\text{reff}}/\varepsilon_{\text{reff}} = 0$, and the temperature remains constant, Eq. 7 reduces to

$$\frac{\delta f_{10}}{f_{10}} = -\frac{\delta L}{L} = -\varepsilon_L, \tag{12}$$

which indicates that the normalized frequency shift is linearly proportional to the applied strain. In case the structure is subjected to bidirectional loadings, the strain along the length and width directions of the radiation patch can be calculated from the normalized resonant frequency shifts of the TM_{10} and TM_{01} modes, i.e.,

$$\begin{cases} \dfrac{\delta f_{10}}{f_{10}} = -\dfrac{\delta L}{L} = -(\varepsilon_L - \upsilon\varepsilon_W) \\ \dfrac{\delta f_{01}}{f_{01}} = -\dfrac{\delta W}{W} = -(\varepsilon_W - \upsilon\varepsilon_L) \end{cases} \Rightarrow \begin{cases} \varepsilon_L = -\dfrac{1}{1-\upsilon^2}\left(\dfrac{\delta f_{10}}{f_{10}} + \upsilon\dfrac{\delta f_{01}}{f_{01}}\right) \\ \varepsilon_W = -\dfrac{1}{1-\upsilon^2}\left(\upsilon\dfrac{\delta f_{10}}{f_{10}} + \dfrac{\delta f_{01}}{f_{01}}\right) \end{cases}, \tag{13}$$

where υ is the Poisson's ratio of the structural material. Based on the principle of strain transformation, the normal and shear strains along an arbitrary direction can be calculated from the normal strains along two orthogonal principle directions. Therefore, the strain state of a structure can be uniquely measured using one rectangular patch antenna by aligning the antenna radiation patch along the principle stress directions.

Based on Eqs. 12 and 13, the strain sensitivity of the antenna sensor, measured as the amount of frequency shift per unit strain (i.e., $\delta f_{mn}/\varepsilon$), is linearly proportional to the corresponding antenna resonance frequency f_{mn}. Therefore, an antenna with higher resonant frequencies is more sensitive to strain changes. In addition, it is desirable for a strain sensor to have small size so that it can offer fine spatial resolution. Based on these two considerations, the antenna sensor employed for strain sensing should have the highest resonant frequencies possible, under the constraints of manufacturing uncertainties, interrogation distance, and the costs of the modulation and reader circuits.

Temperature Sensing

Temperature is an important parameter to measure in many applications, including manufacturing process control, environmental monitoring, food production, human and animal health monitoring, etc. For the purpose of temperature sensing, the effective dielectric constant of an antenna temperature sensor should be linearly proportional to the temperature change ΔT, i.e.,

$$\frac{\delta\varepsilon_{\text{reff}}}{\varepsilon_{\text{reff}}} = \alpha_\varepsilon \Delta T. \tag{14}$$

α_ε is called the thermal coefficient of dielectric constant (TCD_k) of the substrate material (Carver and Mink 1981). Due to the thermal expansion effect, the dimensions of the radiation patch are also sensitive to the temperate changes as well. Denoting the coefficient of thermal expansion (CTE) of the antenna sensor as α_T, we can calculate the thermally induced length change as

DOI 10.1007/978-981-4560-75-7_86-1

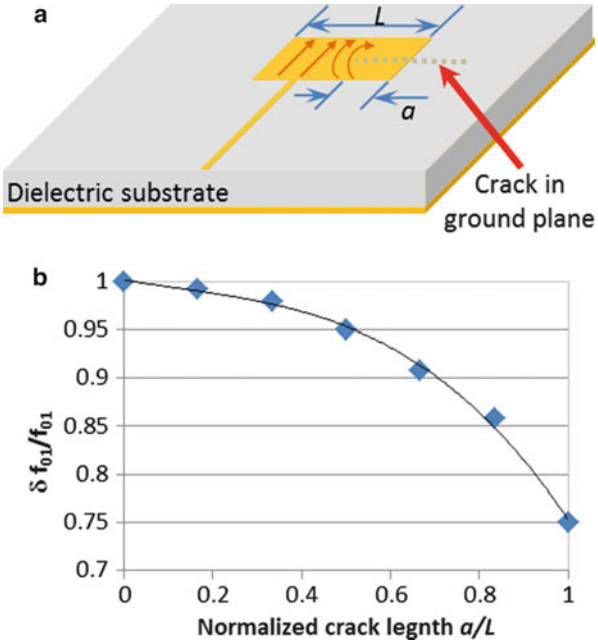


Fig. 10 (**a**) Effect of crack on current distribution of radiation patch; (**b**) reduction of antenna resonant frequency with crack growth

$$\frac{\delta L}{L} = \alpha_T \Delta T. \tag{15}$$

Substituting Eqs. 14 and 15 into 7, we have

$$\frac{\delta f_{10}}{f_{10}} = -\left(\frac{1}{2}\alpha_\varepsilon + \alpha_T\right)\Delta T = k_T \Delta T. \tag{16}$$

k_T, therefore, represents the temperature sensitivity of the antenna resonant frequency. Since both the TCD_k and the CTE are material properties, the frequency shift $\delta f_{01}/f_{01}$ of the TM_{01} mode should have the same temperature sensitivity value k_T, assuming the substrate material is isotropic.

The strain and temperature sensitivity of the antenna resonant frequency are derived separately up to this point. Similar to many other strain sensors, the effects of strain and temperature on the antenna resonant frequencies are coupled. In case the antenna sensor is subjected to uniaxial strain, these two effects can be decoupled by utilizing the two orthogonal antenna radiation modes. The normalized frequency shifts of the TM_{10} and TM_{01} modes under a length direction strain ε_L and a temperature change ΔT can be expressed as

$$\begin{cases} \dfrac{\delta f_{10}}{f_{10}} = k_T \Delta T - \varepsilon_L \\ \dfrac{\delta f_{01}}{f_{01}} = k_T \Delta T - \upsilon \varepsilon_L \end{cases} \Rightarrow \begin{cases} \varepsilon_L = -\dfrac{1}{\upsilon - 1}\left(\dfrac{\delta f_{10}}{f_{10}} - \dfrac{\delta f_{01}}{f_{01}}\right) \\ \Delta T = -\dfrac{1}{k_T}\left(\dfrac{\delta f_{10}}{f_{10}} + \varepsilon_L\right) \end{cases}. \tag{17}$$

If the antenna sensor is subjected to biaxial strains and a temperature change simultaneously, two antenna sensors placed in close vicinity have to be used to extract these three parameters.

DOI 10.1007/978-981-4560-75-7_86-1

Crack Sensing

The presence of cracks is a direct indication of structural damage. If left unattended, some cracks could cause catastrophic structural failures. From a fracture mechanics point of view, in order to accurately predict the effect of the crack on the structural integrity, the crack location, length, and orientation have to be known. The antenna sensor is unique in its capability to provide such detailed crack information. In the derivation of Eq. 7, the radiation patch and the ground plane of the patch antenna were assumed to be perfectly continuous, i.e., there is no slot, cut, or crack in them. If the radiation patch or the ground plane has a crack, however, the crack will disturb the current flow of the radiation patch. As shown in Fig. 10a, a crack parallel to the length direction of the radiation patch will force the current along the width direction to flow around it and thus elongates the electric width of the radiation patch. Similarly, a crack parallel to the width direction will elongate the electric length. In the case of a slanted crack that is parallel neither to the length nor to the width direction of the radiation patch, the crack increases both electric dimensions of the microstrip patch antenna, which causes the antenna resonant frequencies to shift lower. Unfortunately, there is no analytical model that can conveniently calculate the effects of crack on the antenna resonant frequencies. In order to establish the relationship between the crack length and the antenna resonant frequency shifts, either numerical EM simulation or experimental calibration has to be carried out. The resonance frequency – crack growth relationship for an antenna sensor with a crack parallel to its length direction is shown in Fig. 10b. The crack sensitivity of the antenna sensor is defined as the amount of frequency shift per unit crack growth, i.e., $\delta f_{mn}/da$, where da is the crack growth. When the crack just reaches the edge of the radiation patch, the crack sensitivity is relatively low. The crack sensitivity increases as the crack propagates deep into the radiation patch. After the crack tip reaches beyond the half length of the radiation patch, the frequency shift-crack growth relationship is almost linear.

Wireless Interrogation Schemes

Historic Review

Functioning as an EM resonator, a passive antenna can be wirelessly interrogated via near-field or far-field communications. When a reader antenna is placed in the near field of an antenna sensor, these two antennas are inductively coupled. Therefore, the impedance or scattering parameter (S-parameter) of the sensor antenna can be measured directly from the S_{11} parameter of the reader antenna (Potyrailo et al. 2011; Jang and Kim 2012; Daliri et al. 2012). The reading range of the near-field communication, however, is comparable with that of the inductively coupled sensors. The far-field interrogation of the antenna sensors that can offer a much longer range is therefore the focus of this section. Since a passive antenna sensor does not have a radio on board, wireless interrogation is carried out based on the signal reflected by the antenna sensor, i.e., the antenna backscattering. The theory of antenna backscattering can be traced back to papers published by King (1949) and Harrington (1963). The first wireless measurement of the antenna backscattering in an anechoic chamber was presented by Newman and Forrai (1987), which was used to validate the integral equation and method of moments (MOM) solution for calculating the plane wave scattered by a rectangular microstrip antenna. The effect of a load impedance on the backscattering of a microstrip patch antenna was simulated by Pozar (1987) using a spectral domain Green's function. Later on, Aberle et al. (1991) extended this method to evaluate the input impedance and the antenna backscattering of probe-fed microstrip patch antennas. Other numerical simulation methods, such as the finite-difference time-domain (FDTD) method, enable calculating the antenna backscattering over a large bandwidth (Reineix et al. 1993). These theoretical works laid the foundation for the development of RFID systems. Penttila et al. (2006) analyzed the RCS of passive RFID tags for the purpose of system optimization. They also performed scattering pattern measurements of the tag antennas

DOI 10.1007/978-981-4560-75-7_86-1

with four different loads. The differential RCS of the RFID tag was analyzed by Nikitin et al. (2007) and validated with experiment measurement. All these measurements, however, were performed in anechoic chambers.

The first demonstration on the wireless interrogation of an antenna sensor outside of an anechoic chamber was published by Sidén et al. (2007). In this work, a regular RFID tag was employed for relative humidity measurement, and the turn-on power of the RFID tag was selected as the sensing parameter. Since the turn-on power is not only related to the antenna efficiency but also influenced by the read-tag distance and the transmission losses, a reference RFID tag without the sensing material has to be placed at an equal distance from the RFID reader. The relative humidity was then measured from the difference in the turn-on powers of these two RFID tags. Even though the backscattered signal power was introduced by Bhattacharyya et al. (2009) and Marrocco et al. (2009) as an additional sensing parameter for the RFID-enabled antenna sensors, it can only offer a relative measurement and thus share the same limitation as the turn-on power. Wireless measurement of the antenna resonant frequency outside of an anechoic chamber was demonstrated by Tata et al. (2009) and Deshmukh and Huang (2010). Using delayed time-domain reflectometry (TDR) and normalized TDR, they validated the linear relationship between the antenna resonant frequency shifts and the applied strain levels experimentally. Wireless measurement of the antenna resonant frequency using RFID tags was demonstrated by Yi et al. (2011), Qiao et al. (2013), and Hasani et al. (2013). To measure the resonant frequency of an RFID tag antenna, the frequency and/or the power level of the interrogation signal has to be swept. Since it takes some time for the RFID chip to harvest sufficient power for each frequency/power setting, measuring the resonant frequency of the RFID tag antenna can take a long time. Recently, dynamic interrogation of the wireless patch antenna sensor was demonstrated by Yao et al. (2014) based on the principles of amplitude modulation (AM) and frequency-modulated continuous wave (FMCW) radar.

A list of published work on the wireless interrogation of antenna sensors is given in Table 5. Again, the list may not be exhaustive but included all the relevant references the author found in literature.

Theoretical Foundation

An antenna is a wireless transmission device that can receive and radiate EM signals. The power transmission between the reader and an antenna sensor is shown in Fig. 11. Assuming the power transmitted by the reader antenna is P_{tx} and the distance between the reader and the sensor is R, the power received by the sensor antenna can be calculated from the Friis equation as

$$P_{rx} = P_{tx} G_r G_s |\rho_r \rho_s|^2 \left(\frac{\lambda}{4\pi R}\right)^2, \tag{18}$$

where λ is the wavelength, G_r and G_s are the gains of the reader and sensor antenna, and ρ_r and ρ_s are the polarization loss factors of these two antennas, respectively. Since the antenna sensor is connected to a load through the microstrip feed line, the signal received by the antenna propagates along the microstrip feed line until it reaches the load. If the load impedance does not match with the antenna impedance perfectly, a portion of the received power will be reflected. Therefore, the reflected power can be calculated from the impedance mismatch between the antenna and the load as

$$P_\Gamma = |\Gamma|^2 P_{rx} \text{ and } \Gamma = \frac{Z_L - Z_a}{Z_L + Z_a}, \tag{19}$$

in which Z_L is the load impedance and Z_a is the antenna impedance. This reflected power traces its path

DOI 10.1007/978-981-4560-75-7_86-1

Table 5 A list of published wireless interrogation schemes for passive wireless antenna sensors

		Antenna sensor		Sensing	Wireless interrogation		
References	Measurand	Type	Frequency	parameter	Technique	Parameter	Distance (m)
(McGrath et al. 2004)	Gas	Circular patch	5.5 GHz	Frequency shift	Concept of direct wireless interrogation. No implementation		
(Sidén et al. 2007)	Relative humidity	RFID tags	865 MHz	Power difference	RFID	Turn-on power	1.47
(Tata et al. 2009)	Strain	Patch	20.5 GHz	Frequency shift	Delayed TDR	Signal power	0.8
(Bhattacharyya et al. 2009)	Displacement	RFID tag	915 MHz	Turn-on and signal power	RFID	Turn-on and signal power	
(Marrocco et al. 2009)	Filling level of bottle	RFID tag		Turn-on and signal power	RFID	Turn-on and signal power	0.5
(Deshmukh and Huang 2010)		Patch antenna	6.3 and 7.4 GHz	Resonant Frequency	Normalized TDR	Signal power	1.34
(Bhattacharyya et al. 2010)	Temperature threshold	RFID tag	915 MHz	Signal power	RFID		0.5–1.5
(Merilampi et al. 2010)	Strain	Folded dipole	915 MHz	Signal power	RFID	Signal power	0.5
(Merilampi et al. 2011)	Large displacement	Folded dipole	866 MHz	Turn-on and signal power	RFID	Turn-on and signal power	1.5
(Occhiuzzi et al. 2011)	Strain	Meander line	900 MHz	Signal power	RFID	Signal power	0.6
(Bhattacharyya et al. 2011)	Temperature threshold	RFID tag	915 MHz	Frequency contribution	RFID	Turn-on and signal power	
(Yi et al. 2011)	Strain	Folded patch	921 MHz	Frequency shift	RFID	Turn-on power	0.3–0.6
(Xu and Huang 2012a)	Strain	Patch	5.46 GHz	Frequency shift	Amplitude modulation	Signal power	0.84
(Caizzone et al. 2012)	Surface deformation	RFID grid	870 MHz	Turn-on and signal power	RFID	Turn-on and signal power	0.1
(Jang and Kim 2012)	Strain	Planar dipole	8–12 GHz	Frequency shift	Direct VNA S_{11}	S_{11}	0.15
(Daliri et al. 2012)	Strain	Circular patch	1.5 GHz	Frequency shift	Direct VNA S_{11}	S_{11}	0.05
(Yi et al. 2013)	Surface and crack	Folded patch	915 MHz	Frequency shift	RFID	Turn-on power	0.3
(Qiao et al. 2013)	Temperature	Slotted patch	915 MHz	Frequency shift	RFID	Turn-on power	0.5
(Hasani et al. 2013)	Strain	Embroider patch	915 MHz	Frequency shift	RFID and VNA	Turn-on power	0.4 and 0.2
(Yi et al. 2014)	Strain	Slotted patch	910 MHz	Frequency shift	RFID	Turn-on power	0.91
(Yao et al. 2014)	Dynamic strain	Patch	5.38 GHz	Frequency shift	Amplitude modulation	Signal power	0.9

DOI 10.1007/978-981-4560-75-7_86-1

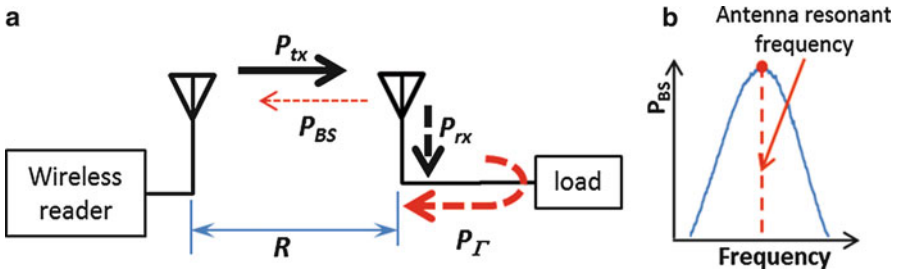


Fig. 11 Wireless transmission based on antenna backscattering; (**a**) power transmission; (**b**) power-frequency relationship of antenna backscattering

back and is subsequently reradiated by the antenna sensor. The power of the backscattered signal received by the wireless reader can again be calculated from the Friis equation as

$$P_{BS} = P_\Gamma G_r G_s |\rho_r \rho_s|^2 \left(\frac{\lambda}{4\pi R}\right)^2 = \left|\frac{Z_L - Z_a}{Z_L + Z_a}\right|^2 P_{TX} G_r^2 G_s^2 |\rho_r \rho_s| 4 \left(\frac{\lambda}{4\pi R}\right)^4. \tag{20}$$

The antenna impedance Z_a is related to the antenna S_{11} parameter and a characteristic impedance Z_0 as

$$Z_a = \left(\frac{1 + S_{11}}{1 - S_{11}}\right) Z_0. \tag{21}$$

Since the S_{11} parameter is a function of the interrogation frequency, the power of the antenna backscattering is frequency dependent as well. For an antenna sensor with a mismatched load, the antenna resonant frequency can then be identified as the frequency at which the antenna backscattering has the highest power, as shown in Fig. 11b.

In theory, the antenna resonant frequency can be wirelessly determined from the antenna backscattering. In practice, however, this is not an easy task since the antenna sensor is usually surrounded by other structures that also reflect EM waves. The signal received by the reader is therefore a combination of the antenna backscattering and structural backscattering (also known as the background clutter). In most cases, the structural backscattering is so strong that it could completely jam the antenna backscattering, creating the well-known "self-jamming" problem. In order to separate the antenna backscattering from the structural backscattering, a modulator has to be implemented at the antenna sensor. Depending on the modulation schemes used in the antenna sensor, the wireless reader has to be equipped with a corresponding demodulation circuit in order to extract the antenna backscattering from the received signal. A few common modulation schemes are described below.

Time-Gated TDR

Time-Gated TDR with Delay Line

The simplest way to interrogate a wireless antenna sensor is to connect a long transmission line to the antenna sensor, as shown in Fig. 12a. Since the received signal has to travel along the long transmission line before it gets reflected by the load, a time delay is introduced between the antenna backscattering and the structural backscattering. In order to provide the needed separation between these two backscatterings, a microstrip meander line is usually employed. The experiment setup for the wirelessly interrogation of such an antenna sensor is shown in Fig. 12b. Assuming the wireless reader is a vector network analyzer (VNA) with a single reader antenna, the S_{11} parameter measured by the VNA is converted to a time-

DOI 10.1007/978-981-4560-75-7_86-1

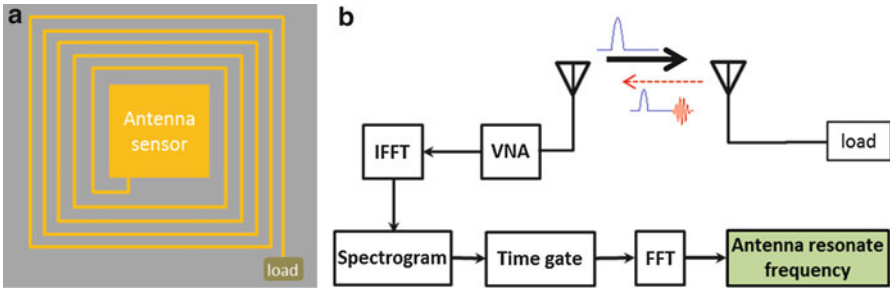


Fig. 12 Passive wireless antenna sensor with a delay line; (**a**) sensor configuration; (**b**) wireless interrogation and data processing scheme

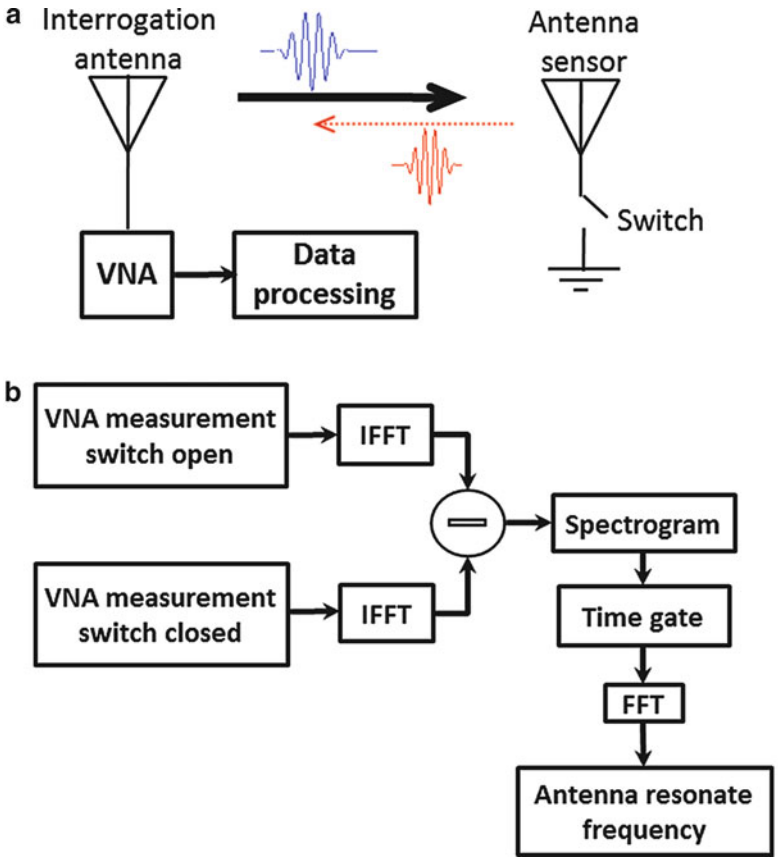


Fig. 13 Antenna sensor with an impedance switch; (**a**) wireless interrogation scheme; (**b**) data processing procedure

domain signal by the inverse fast Fourier transformation (IFFT) operation. Short-time fast Fourier transformation (STFFT) can then be applied to calculate the spectrogram of the backscattered signal. Once the starting and ending times of the antenna backscattering are identified from the spectrogram, the signal can be time gated to extract the antenna backscattering. Finally, the antenna resonant frequency can be determined from the frequency spectrum of the time-gated signal calculated by fast Fourier transformation (FFT).

DOI 10.1007/978-981-4560-75-7_86-1

Normalized TDR by Impedance Switching

Introducing a long transmission line to delay the antenna backscattering is simple, but the transmission line increases the overall size of the antenna sensor. A technique to remove the long transmission line is to terminate the antenna sensor with an impedance switch, as shown in Fig. 13a. The load of the antenna sensor is switched between two states, e.g., a short and an open load; a short load reflects the received signal with a 180° phase shift while an open load reflects the received signal without any phase shift. In comparison, the structural backscattering is identical regardless of the switching state of the antenna. Subtracting the received signal acquired at the two switching states therefore cancels out the structural backscattering while doubles the antenna backscattering. The data processing procedure for such a wireless interrogation scheme is shown in Fig. 13b. The S_{11} parameters of the reader antenna are acquired using a VNA at the two switching states. Both measurements are converted to time-domain signals using the IFFT operation. After subtracting these two time-domain signals to remove the structural backscattering, the time gating procedure described above can be applied to extract the antenna resonant frequency.

One limitation of the time-gated TDR is that the frequency resolution of the FFT operation is usually limited to be around 20 MHz for an antenna sensor with resonant frequencies of a few gigahertzes (GHz). In addition, the two methods described above use a VNA as the wireless reader. Even though it is convenient for lab tests, a VNA is too expensive and too bulky for practical sensor applications. Low-cost wireless readers based on the principles described above have been demonstrated. Please see sections "Gated TDR Based on Frequency Stepping and Delay Switches" and "Gated TDR Based on FMCW Radar" for more details.

RFID-Enabled Wireless Interrogation

When an antenna sensor is loaded with an RFID chip, the antenna backscattering can be measured using two parameters, i.e., the turn-on power that is required to activate the RFID chip and the differential backscattered signal power at the two states of the RFID chip. As discussed in the introduction section, there are two fundamental functions of an RFID chip; one is to harvest energy from the interrogation signal and the other is to switch the impedance of the tag antenna. Based on Eqs. 18 and 19, the power received by an RFID chip can be calculated as

$$P_{\text{chip}} = \left(1 - |\Gamma|^2\right) P_{tx} G_r G_s |\rho_r \rho_s|^2 \left(\frac{\lambda}{4\pi R}\right)^2, \tag{22}$$

which is also a function of the interrogation frequency. Since the RFID chip requires a minimum power level to be activated, the power received by the antenna sensor can then be measured by varying the transmitted power P_{tx} and tracking at what power level the RFID chip is turned on. By sweeping the interrogation frequency, the antenna resonant frequency can then be determined as the frequency at which the turn-on power is the minimum. This approach requires varying both the transmitted power level and the interrogation frequency. Therefore, the time needed to determine the antenna resonant frequency is quite substantial. The differential backscattered signal power approach measures the differences in the powers received by the RFID reader at the two ID digits, i.e., digit "1" or "0." Assuming digit "1" is generated when the antenna is connected to impedance #1 while digit "0" is generated when the antenna is connected to impedance #2, the power received by the RFID reader at these two states can be calculated as

$$P^{(1)} = |\Gamma_1|^2 P_{tx} G_r^2 G_s^2 |\rho_r \rho_s|^4 \left(\frac{\lambda}{4\pi R}\right)^4 + P_{\text{structure}} \tag{23}$$

DOI 10.1007/978-981-4560-75-7_86-1

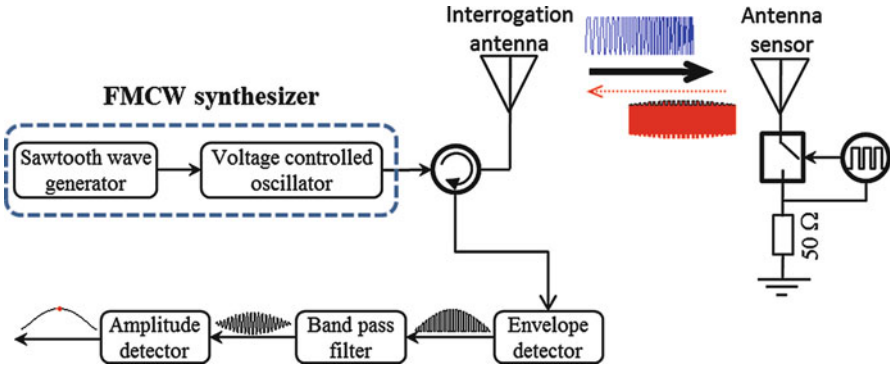


Fig. 14 Dynamic interrogation of wireless antenna strain sensor based on amplitude modulation and FMCW radar

and

$$P^{(0)} = |\Gamma_2|^2 P_{tx} G_r^2 G_s^2 |\rho_r \rho_s|^4 \left(\frac{\lambda}{4\pi R}\right)^4 + P_{\text{structure}}. \tag{24}$$

The power difference between these two states is then

$$\Delta P = P^{(1)} - P^{(0)} = \left(|\Gamma_1|^2 - |\Gamma_2|^2\right) P_{tx} G_r^2 G_s^2 |\rho_r \rho_s|^4 \left(\frac{\lambda}{4\pi R}\right)^4. \tag{25}$$

Again, ΔP is a function of the interrogation frequency and the antenna resonant frequency is determined as the frequency at which ΔP is the largest.

The RFID-enabled approach allows the use of commercially available RFID chips and readers, which makes the sensor fabrication much easier and the cost low. In addition, since each antenna sensor is identified by a unique code, code division sensor multiplexing can be easily implemented. One limitation of such an approach is that the operation frequency of the antenna sensor is limited by the RFID frequency. Most of the published RFID sensors operate at 960 MHz or lower frequencies. As a result, the antenna sensors have relatively large sizes. Another limitation of the RFID-enabled antenna sensors is their low interrogation speed. Because the RFID chip has to harvest sufficient power to be activated, it could take a long time for it to send out the data, especially when there is a large difference between the interrogation frequency and the antenna resonant frequency. Therefore, the RFID-enabled antenna sensor is not suitable for dynamic sensing, such as for vibration measurements that require a fast sensor response.

Amplitude Modulation for Dynamic Wireless Interrogation

Dynamic interrogation of wireless antenna sensors can be achieved based on the principle of amplitude modulation and FMCW radar. As shown in Fig. 14, the interrogation signal is generated using an FMCW synthesizer consisting of a sawtooth wave generator and a voltage-controlled oscillator (VCO). As a result, the FMCW synthesizer generates a chirp signal whose frequency linearly varies with time. By switching the load of the antenna sensor between an open and 50 Ω load using a square wave, the backscattered antenna signal is an amplitude modulated signal whose envelope is a square wave. The amplitude of the envelope is determined by the interrogation frequency; when the interrogation frequency is not very different from the antenna resonant frequency, the amplitude of the envelope is large. In contrast, when the interrogation frequency is very different from the antenna resonant frequency, the amplitude of the envelope is small. The envelope of the amplitude modulated signal can be demodulated

DOI 10.1007/978-981-4560-75-7_86-1

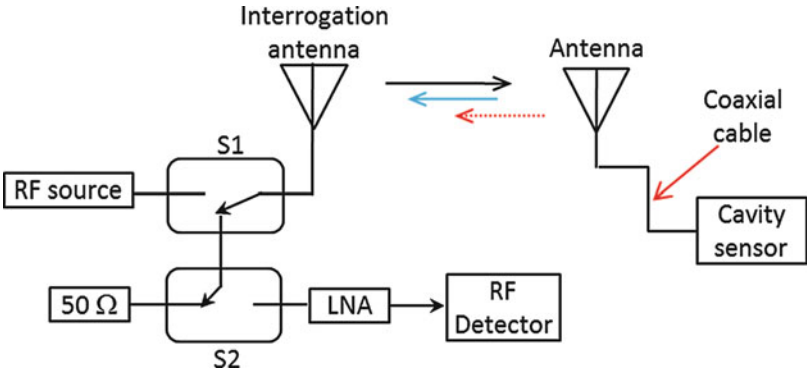


Fig. 15 Time-domain gating approach for wireless interrogation of an EM cavity sensor (Adapted from Thomson et al. 2009)

at the wireless reader using analog components. The envelope of the received backscattered signal is extracted using a microwave envelope detector. Passing the envelope signal through a band pass filter produces a sinusoidal signal with time-varying amplitudes. Finally, a low-frequency amplitude detector can be used to detect the time-amplitude relationship of the demodulated signal. By correlating the frequency-amplitude relationship with the time-frequency relationship of the chirp signal, the antenna resonant frequency can be determined from the filtered envelope signal. Since the FMCW synthesizer can sweep the interrogation frequency at a rate up to a few thousand kilohertz and the demodulation was carried out in an analog fashion, dynamic interrogation of the wireless antenna sensor can be achieved.

Other Wireless Interrogation Schemes

There are other wireless interrogation schemes that have been developed for EM resonant sensors but have not been used for wireless antenna sensors directly. Since an antenna is essentially an EM resonator, these wireless interrogation techniques can be easily adopted for the antenna sensors. Therefore, they are briefly discussed here. A comparison of different wireless interrogation schemes is given at the end of this section.

Gated TDR Based on Frequency Stepping and Delay Switches

To develop a field usable wireless reader for an EM cavity sensor, Thomson et al. (2009) implemented the time-gated TDR using a frequency stepping RF source and delay switches. The cavity sensor has a high quality factor (Q factor) of 1,000 and was connected to an antenna via a coaxial cable. Due to the high Q factor of the cavity sensor, it reradiates the energy in the form of a decaying RF signal after the input signal is switched off. The schematic for wireless interrogation of such a high Q factor cavity sensor is given in Fig. 15. The wireless reader has two operational modes, namely, the transmitting and receiving modes. At the transmitting mode, an interrogation signal is generated using a frequency stepping RF source and broadcasted to the wireless sensor via the reader antenna. It was characterized that the cavity sensor takes about 400 ns to reach equilibrium. 500 ns after the cavity sensor has reached equilibrium, the wireless reader is switched to the receiving mode by connecting the reader antenna to the 50 Ω load using the two switches S1 and S2. After a short delay, the position of S2 is changed so that the reader antenna is connected to a low-noise amplifier (LNA). Due to the time delay, the structural backscattering has largely died out, and thus, the signal amplified by the LNA is dominated by the cavity sensor signal. Detecting the power of the amplified signal using an RF detector at different frequencies therefore enables detecting the resonant frequency of the cavity sensor.

DOI 10.1007/978-981-4560-75-7_86-1

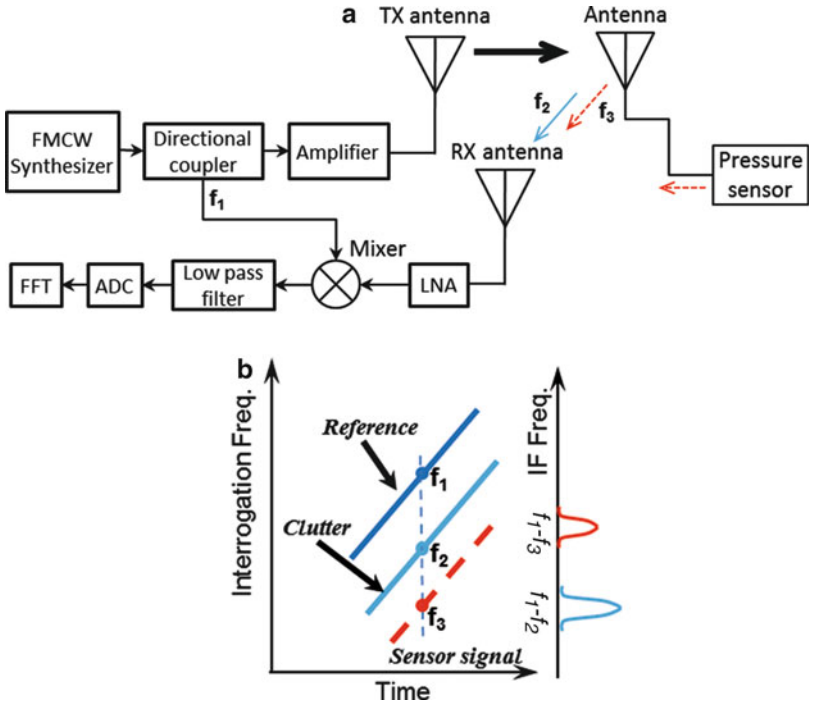


Fig. 16 (**a**) Wireless interrogation of a passive EM pressure sensor based on FMCW radar (Adapted from Jatlaoui et al. 2010); (**b**) the time-frequency relationships of different signals

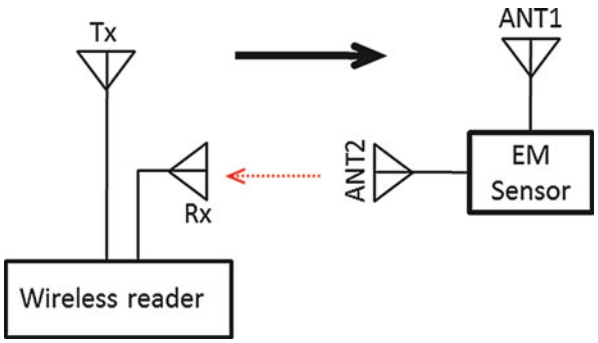


Fig. 17 Wireless interrogation of passive sensor based on orthogonal backscatter scheme

Gated TDR Based on FMCW Radar

The antenna backscattering and the structural backscattering can also be separated based on the principle of FMCW radar as long as the delay between these two backscatterings is sufficiently large (Jatlaoui et al. 2010). As shown in Fig. 16a, the interrogation signal is a chirp signal generated using an FMCW synthesizer. After being amplified by a power amplifier, the chirp signal is broadcasted to the wireless sensor. The time delay between the backscattered sensing signal and the structural backscattering is again introduced by connecting the EM pressure sensor to the antenna using a coaxial cable. Due to the time delay and the linearly varying interrogation frequency, the backscattered signals received by the wireless reader at a given time have different frequencies, as shown in Fig. 16b. A homodyne receiver can then be implemented at the wireless reader to separate the signals with different frequencies. Assuming the frequencies of the structural backscattering and the backscattered sensing signal are f_2 and f_3, respectively, mixing these signals with a reference signal f_1 produces two low-frequency signals $f_1 - f_2$ and $f_1 - f_3$. After

DOI 10.1007/978-981-4560-75-7_86-1

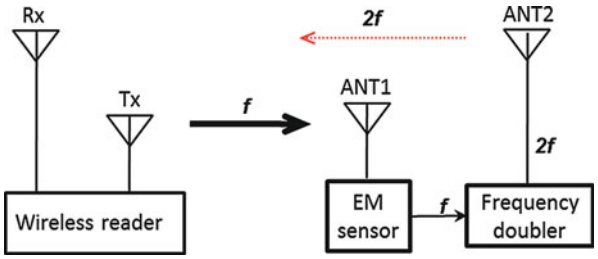


Fig. 18 Wireless interrogation of passive EM sensor based on harmonic radar

filtering the mixer output using a low-pass filter and acquiring the signals using an ADC, the spectrum of the acquired signal is calculated using FFT, based on which the power of the backscattered sensing signal can then be determined. Again, the resonant frequency of the EM pressure sensor is measured from the interrogation frequency-signal power relationship.

Passive Wireless Interrogation Based on Orthogonal Backscatter

The polarization of the antenna radiation pattern can also be exploited to address the self-jamming problem (Mandel et al. 2011; Kubina et al. 2012; Girbau et al. 2012). As shown in Fig. 17, the EM sensor is connected to two antennas with orthogonal polarizations. The interrogation signal is first broadcasted to the antenna sensor along one of the polarization directions. The received signal is input into the EM sensor. By transmitting the sensor output using the second antenna, a 90° difference in the polarizations of the backscattered sensing signal and the incident signal is created. The structural backscattering, on the other hand, will have the same polarization as the incident signal. Due to this polarization mismatch, the structural backscattering will be rejected by the receiving reader antenna. As a result, the signal received by the wireless reader is dominated by the backscattered sensing signal.

Passive Wireless Interrogation Based on Harmonic Radar

A harmonic radar tag receives the interrogation signal at one frequency and transmits the radar signal at twice that frequency (Charvat et al. 2003; Weller and Davidova 2013; Presas et al. 2007). Since the background clutter remains at the interrogation frequency, the radar receiver is able to discriminate the tag return from the clutter in the frequency domain. This principle can be extended for wireless interrogation of passive EM sensors. As shown in Fig. 18, the interrogation signal with a frequency *f* is first transmitted to the wireless sensor by the wireless reader. After interacting with the EM sensor, the sensing signal is input into a passive frequency doubler which outputs a signal having a frequency *2f*. This frequency doubled signal is then transmitted by the second antenna. As such, a wireless reader equipped with a band pass filter centered at *2f* will reject the background clutter and only receives the signal returned by the passive wireless sensor.

Comparisons of Wireless Interrogation Schemes

The comparisons of wireless interrogation schemes discussed above are given in Table 6. It is difficult to provide quantitative comparisons due to the large variations in the configurations, principle, and operating frequencies of the published wireless sensing systems. Therefore, the wireless interrogation schemes are compared in a relative term. Based on the same consideration, one of the important figures of merit, i.e., the distance between the wireless sensor and reader, was left out intentionally. In order to compare the interrogation distances of different wireless sensing systems, the wireless sensing systems should be tested at the same operational frequency and fabricated in a controlled manner, which cannot be achieved without a comprehensive comparison test.

DOI 10.1007/978-981-4560-75-7_86-1

Table 6 Comparisons of wireless interrogation schemes for passive EM resonant sensors

Techniques	Representative reference	Sensor size	Wireless reader		Frequency resolution	Acquisition speed
			Cost	Availability		
Gated TDR with delay line – VNA	(Ramos et al. 2011)	Large+	High	Commercial	Coarse	Medium
Normalized TDR – VNA	(Deshmukh and Huang 2010)	Small	High	Commercial	Coarse	Medium
Gated TDR with delay switches	(Thomson et al. 2009)	Large+	Low	Customized	Fine	Medium
Gated TDR with FMCW	(Jatlaoui et al. 2010)	Large+	Low	Customized	Coarse	Medium
RFID enabled	(Sidén et al. 2007)	Medium	Low	Commercial	Small	Slow
Amplitude modulation	(Yao et al. 2014)	Small	Low	Customized	Small	Fast
Orthogonal polarization	(Mandel et al. 2011)	Large	Low	Customized	Small	Fast
Harmonic radar	(Charvat et al. 2003)	Large	Low	Customized	Small	Fast

+ include delay transmission line

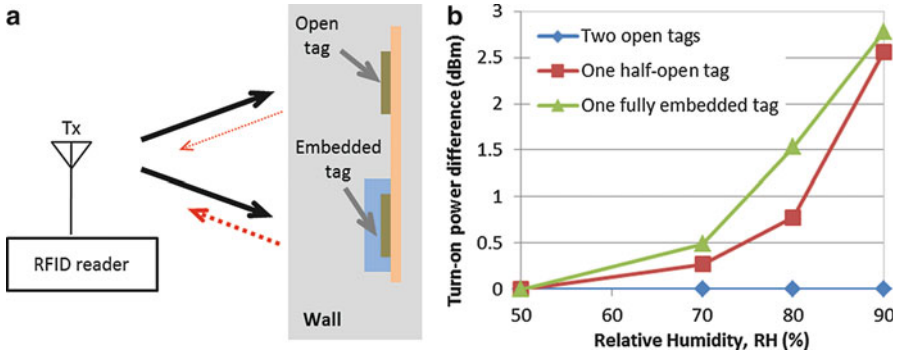


Fig. 19 Wireless interrogation of an RFID-enabled relative humidity sensor (Adated from Sidén et al. 2007); (**a**) system configuration; (**b**) measurement results

Application Examples

RFID-Enabled Moisture Sensor

The configuration of a wireless moisture sensing system using a pair of regular RFID tags is shown in Fig. 19a (Sidén et al. 2007). The sensor label was implemented by placing two passive 865–868 MHz Gen-2 tags from Alien Technology Corp on a sheet of high-density polyethylene (HDPE). The center-to-center distance between the two tag antennas was 206 mm. Ten sheets of blotting paper, acting as water-absorbing material, were stacked and placed in front, behind, and both in front and behind one of the RFID tags. The other RFID tag was uncovered and served as the reference. An RFID reader from SAMsys was placed at a distance of 1.47 m from the sensor label with equal distances from the two RFID tags. Experiments were carried out in a climate room with temperature fixed at 23°C, and the relative humidity increased from 50% to 90%. The blotting paper samples were left in the climate room for 24 hours before taking measurements. The differences in the turn-on powers of the two RFID tags at different relative humidity levels are shown in Fig. 19b. Three sensor labels were tested; the first sensor label had two open tags, which severs as the baseline case; the second sensor label had an half-open tag, i.e., the tag only had the paper in front; a fully embedded tag, i.e., the tag had paper in front and behind it, was implemented in the third sensor label. It was found that the turn-on powers for both tags were about 16 dBm in an

DOI 10.1007/978-981-4560-75-7_86-1

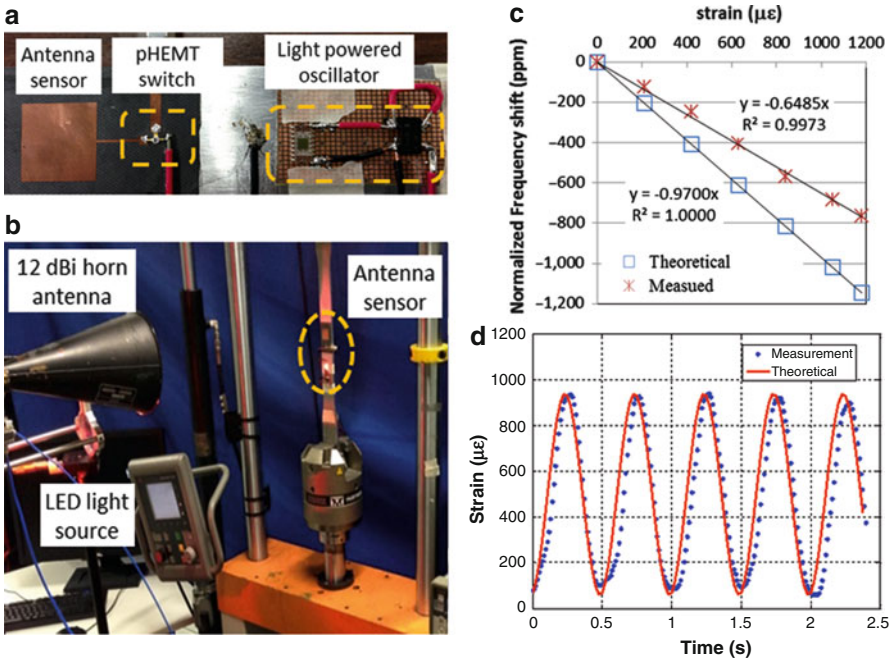


Fig. 20 Dynamic interrogation of a passive wireless antenna strain sensor; (**a**) sensor configuration; (**b**) experiment setup; (**c**) static tensile test results; (**d**) dynamic test results at a loading frequency of 2 Hz

environment with 50% relative humidity at 23°C. As the relative humidity increased, the power differences for the second and third sensor labels increased while the power differences for the first sensor label remained to be zero. The authors suggested the water-absorbing material should be as thick as possible. In addition, they also pointed out that the paper displayed some hysteresis behaviors, which can be a problem since the memory effects can lead to incorrect readings.

Dynamic Strain Sensor Based on Amplitude Modulation and FMCW Radar

The dynamic antenna strain sensing system presented is illustrated in Fig. 20a (Yao et al. 2014). The antenna sensor is a 5.4 GHz microstrip patch antenna fabricated on a high-frequency dielectric substrate (Rogers Corporation, RT/Duroid 5880). For a substrate dielectric constant of 2.2, the length and width of the radiation patch are 18.0 and 21.6 mm, respectively. The patch antenna was connected to a pseudomorphic high-electron-mobility transistor (pHEMT) with its drain terminal connected to a 50 Ω terminator. The gate termination of the pHEMT was connected to a low-power oscillator (Epson Electronics, SG-3040) that produces 32.78 kHz square waves. The oscillator output changes the gate voltage of the pHEMT between 0 and 1 V. As a result, the termination of the patch antenna is switched between an open and a 50 Ω load. A solar cell (Clare, CPC 1832) supplies the power to the oscillator when it is illuminated by a light-emitting diode (LED).

The wireless reader depicted in Fig. 14 was assembled using commercially available discrete components. For the FMCW synthesizer, the periodic sawtooth wave was generated using a signal generator (Agilent 33250A) and supplied to a VCO (Mini-circuits, ZX95-3100-S+) as the control voltage. The frequency of the VCO output was doubled using a frequency multiplier (Mini-circuits ZX90-36-S+) so that the frequency sweep range covers the antenna resonant frequency. The output of the frequency multiplier was amplified using a microwave amplifier (Mini-circuits ZVE-8G+) to reach a power of 25 dBm before it was transmitted to the interrogation antenna for broadcasting. The analog demodulator was implemented using an RF envelope detector board (Analog Devices, ADL5511) and a band pass filter (Krohn-Hite Model 3944) centered at 33 kHz. Since the output of the band pass filter has a low frequency

DOI 10.1007/978-981-4560-75-7_86-1

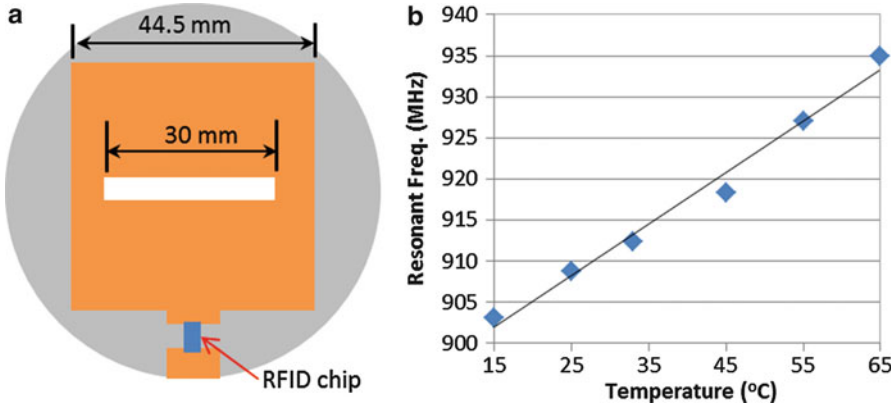


Fig. 21 RFID-enabled antenna temperature sensor (Adapted from Qiao et al. 2013); (**a**) sensor configuration; (**b**) temperature-antenna frequency relationship

of 32.78 kHz, this signal was directly acquired using an oscilloscope (LeCroy, Wavepro 760zi). The acquired signal was processed by a real-time data processing algorithm to extract the antenna resonant frequency.

Mechanical testing of the dynamic strain sensor is shown in Fig. 20b. The antenna sensor was bonded on a tensile specimen and loaded with an Instron mechanical tester. A 12 dBi horn antenna and a 134 lm LED light source (Luxeon Star, MR-H2070-20S) were placed at a distance of 0.9 m from the antenna sensor. The FMCW synthesizer was set to sweep the frequency from 5,325 to 5,430 MHz at a repetition rate of 500 Hz. The sampling rate of the oscilloscope was set to be 1 MHz. Static tensile test of the antenna sensor was first carried out to calibrate the strain sensitivity of the antenna strain sensor. As shown in Fig. 20c, the measured strain sensitivity was 0.661 ppm/με. While this value is about 68 % of the theoretical strain sensitivity, the antenna sensor displayed an excellent linearity. Therefore, a gain factor of 1/68 % = 1.47 was applied to the sensor measurements to account for this discrepancy. The dynamic testing results of the antenna strain sensor are shown in Fig. 20d. The applied load was varied at a frequency of 2 Hz. It was discovered that the measurement results track the theoretical strains very well. Therefore, the performance of the dynamic wireless interrogation system is validated.

RFID-Enabled Temperature Sensor

An RFID-enabled temperature sensor was demonstrated using a 900 MHz patch antenna with an HDPE-BST substrate (Qiao et al. 2013). As shown in Fig. 21a, the 44.5 mm long radiation patch has a 30 mm long and 4 mm wide slot at its center. The slot was introduced to reduce the size of the radiation patch. The radiation patch is fed with a 10 mm wide microstrip transmission line, which was terminated with an Alien Higgs 3 RFID chip. The RFID chip was grounded using another microstrip transmission line. The HDPE-BST substrate material was custom synthesized through extrusion process. The dielectric constant and tangent loss of the HDPE-BST material at different temperatures were characterized at 900 MHz using an Agilent E4991A RF impedance analyzer. It was found that the dielectric constant of the HDPE-BST material decreased from 7.10 to 6.66 while the tangent loss increased from 0.0022 to 0.0034 as the temperature increased from 20 °C to 80 °C. The temperature coefficient of the dielectric constant was therefore estimated to be −1,033 ppm/°C, which is much higher than conventional substrate materials. Wireless measurements of the antenna resonant frequency were carried out in an anechoic chamber. The wireless reader, a Tagformance device, was placed at a distance of 0.5 m from the antenna sensor. The turn-on power of the RFID chip was measured at frequencies ranging from 880 to 960 MHz at different temperature settings. The maximum read range of the RFID sensor was calculated from the measured turn-on powers as

DOI 10.1007/978-981-4560-75-7_86-1

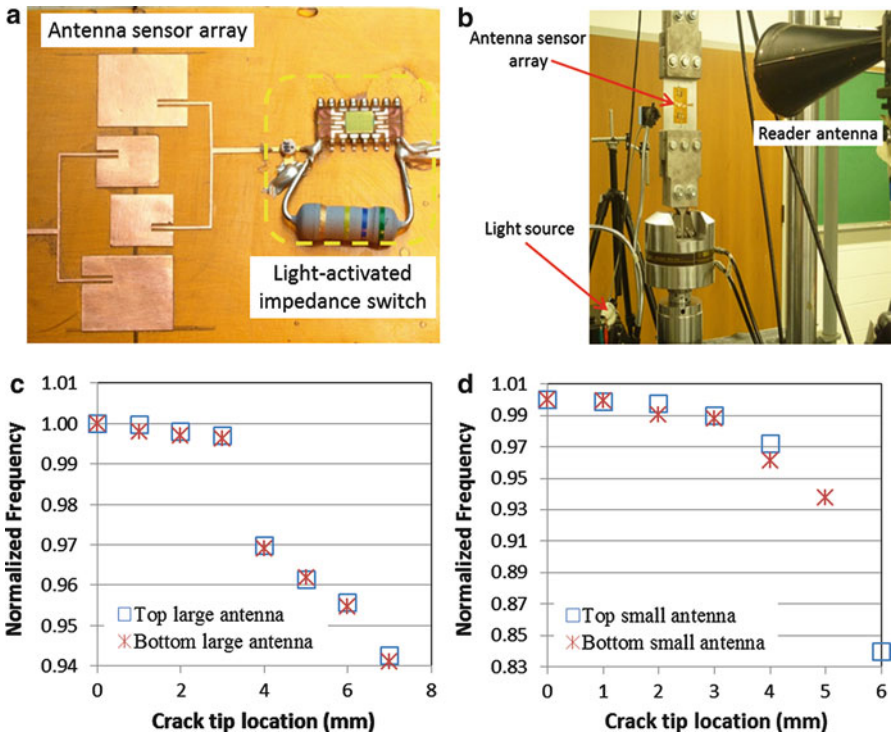


Fig. 22 An antenna sensor array covering a large area with high crack sensitivity (Xu and Huang 2012b); (**a**) realization of antenna sensor array with light-activated impedance switch; (**b**) wireless interrogation of antenna sensor array during fatigue test; (**c**) frequency shift-crack growth relationship of large antennas; (**d**) frequency shift-crack growth relationship of small antennas

$$d_{\max} = \frac{\lambda}{4\pi}\sqrt{\frac{\mathrm{EIRP}}{P_{\text{turn-on}}}}, \tag{26}$$

where the equivalent isotropic radiated power (EIRP) of the RFID reader was 3.28 W. The resonant frequency shifts of the antenna sensor at different temperatures can then be measured from the frequency-read range plots. As shown in Fig. 21b, the antenna resonant frequency displayed a relatively linear performance for temperatures ranging from 15 °C to 65 °C. The temperature sensitivity of the antenna sensor was estimated to be 800 kHz/°C.

Antenna Crack Sensor Array Based on Frequency Division Multiplexing

The design of the antenna crack sensors is a trade-off between the crack sensitivity and the sensor coverage area (Xu and Huang 2012b). In general, a high-frequency antenna is more sensitive to minute crack growth. On the other hand, the sensor coverage of the antenna crack sensor is confined within the physical dimensions of the radiation patch. Since the resonant frequencies of an antenna sensor are inversely proportional to the size of the radiation patch, a high-frequency antenna sensor offers high crack sensitive but small coverage area. One solution to resolve this dilemma is to use antenna arrays based on the principle of frequency division multiplexing (Xu and Huang 2012b) or RFID code division multiplexing (Yi et al. 2013). A picture of the frequency division multiplexed antenna array is shown in Fig. 22a. The four-element antenna array consists of two pair of patch antennas. Each pair has two antennas of different sizes; the large one operates at 8.4 GHz and the smaller one operates at 12 GHz. Since each antenna resonance occupies a narrow bandwidth, this frequency division multiplexing scheme enables the interrogation of these two antennas with one frequency scan. Each pair is connected to an

DOI 10.1007/978-981-4560-75-7_86-1

impedance switch powered by a solar cell, as shown in the Fig. 22a. Similar to the dynamic antenna strain sensor, the impedance switch was implemented using a pHEMT, a 560 kΩ resistor, and a photocell. The antenna pattern was fabricated from a thin copper film using laser machining and bonded on a Kapton film, which in turn was glued on the aluminum alloy specimen that also serves as the ground plane for the antenna sensors. The fatigue testing of the specimen instrumented with the antenna sensor array is shown in Fig. 22b. A horn antenna and a flashlight were placed at a distance of 0.84 m from the antenna sensors. The S_{11} parameters of the horn antenna were acquired at two states, i.e., when the light was turned on and when it was turned off. The acquired S_{11} parameters were then processed following the procedure described in Fig. 13b to extract the resonant frequencies of the antenna sensors. The specimen was fabricated with two slots for crack initiation, which propagate toward each other under fatigue loading. After the cracks reached the two large antenna sensors, the crack growth was detected from the shift of the resonant frequency at around 8.4 GHz. The reader was able to wirelessly measure the resonant frequencies of the large antenna sensors until the crack reached the last 1 mm of their radiation patches. It took 5,000 more cycles for the crack to reach the small antenna sensors; after which the crack growth was detected from the shift of the resonance at around 12 GHz. The small antenna sensors remained functional until the specimen fractured after 190,000 fatigue cycles. The normalized resonant frequency shifts for both the small and large antenna sensors are shown in Fig. 22c, d. As expected, the small antennas displayed higher crack sensitivities than the large antennas. In addition, the results indicated that the fabrication of the antenna sensors can be controlled to achieve repeatable performances.

Future Directions and Open Problems

Passive wireless antenna sensors that consume substantially less power than active wireless sensors and can be interrogated from middle range distances are promising technologies that could have profound impacts on our drive toward ubiquitous sensing. Unlike active wireless sensors, passive wireless antenna sensors are still in its infancy. Even though the fundamental operating principles have been validated and a variety of practical applications have been demonstrated, more fundamental research and development works are needed to realize the full potentials of these passive wireless sensors. Listed below are a few open problems that need further investigations and development.

Sensor miniaturization: a passive wireless antenna sensor consists of only two components, i.e., the antenna and the modulator. The size of the antenna can be easily reduced by increasing its operation frequency. For example, the strain sensor demonstrated by Tata et al. (2009) is only 5.3 mm long and 4 mm wide, which is comparable with the size of a conventional foil strain gauge. However, the modulator has to be implemented using discrete electronic components, which increases the sensor size as well as the cost. Using RFID chips is an attractive option since all the electronic components are integrated into one IC chip. Unfortunately, most commercially available RFID chips operate at around 900 MHz, which dictates the size of the antenna. Despite different approaches to reduce the antenna size at this frequency, the resulting antennas are still too large for most real-life sensor applications. Hopefully, the maturity of the passive wireless antenna sensors will motivate the IC industry to develop RFID chips tailored for the needs of the antenna sensors, especially at high frequencies and with fast interrogation speeds.

Multimodality sensing and data processing: an antenna sensor can be sensitive to multiple physical measurands simultaneously, including strain, temperature, humidity, pressure, etc. This characteristic can be a double-edged sword. On one hand, a single antenna sensor can be used to measure multiple variables. On the other hand, the coupling of different effects can make interpreting the measurement results extremely difficult. Therefore, it is desirable to develop data analysis techniques that can decouple these effects and extract the physical measurands individually. This cannot be achieved using one sensing

DOI 10.1007/978-981-4560-75-7_86-1

parameter, such as the resonant frequency shift. Additional sensing parameters, e.g., the resonant bandwidth, may be exploited for this purpose. Ideally, system identification algorithms can be developed to identify the physical measurands that match the simulated antenna radiation curves with the measurements. This approach, however, requires the development of flexible and efficient antenna simulation models that are currently not available.

Sensor identification and multiplexing: to realize the full benefits of passive wireless antenna sensors, the wireless antenna sensors need to be arranged in a distributed sensor network with a centralized wireless reader communicating with each individual sensor. This arrangement mandates the development of sensor identification and multiplexing strategy. A frequency and spatial division multiplexing scheme was developed and selective interrogation of four antenna sensors was demonstrated (Xu and Huang 2012b). However, the number of sensors that can be interrogated is limited by the allocated frequency bandwidth. Other sensor multiplexing techniques are needed in order to interrogate tens or hundreds of wireless antenna sensors. The RFID-enabled antenna sensors may hold the key to resolve this challenge, but again the size of the RFID antenna has to be reduced to make this option feasible.

Cross-References

- Antenna Measurement/Setups-Introduction
- Commercial Antenna Design Tools
- Impedance Matching and Baluns
- Microstrip Patch Antennas
- Millimeter-Wave Antennas and Arrays
- Mm-Wave Sub-mm-Wave Antenna Measurement
- Numerical Modeling in Antenna Engineering
- Radio Wave Propagation Models
- RF Material Characterization
- Transmission Lines

References

Aberle JT, Pozar DM, Birtcher CR (1991) Evaluation of input impedance and radar cross section of probe-fed microstrip patch elements using an accurate feed model. IEEE Trans Antennas Propag 39(12):1691–1696. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=121589

Bhattacharyya R, Floerkemeier C, Sarma S (2009) Towards tag antenna based sensing – an RFID displacement sensor. In: 2009 I.E. international conference on RFID, Orlando, FL. pp 95–102. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4911195

Bhattacharyya R, Floerkemeier C, Sarma S (2010a) Low-cost ubiquitous RFID tag antenna based sensing. Proc IEEE 98(9):1593–1600

Bhattacharyya R, Floerkemeier C, Sarma S (2010b) RFID tag antenna based temperature sensing. In: 2010 I.E. international conference on RFID (IEEE RFID 2010), Orlando, FL. pp 8–15. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5467239

Bhattacharyya R et al (2011) RFID tag antenna based temperature sensing in the frequency domain. In: 2011 I.E. international conference on RFID, Orlando, FL. pp 70–77. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5764639

Bogosanovich M (2000) Microstrip patch sensor for measurement of the permittivity of homogeneous dielectric materials. IEEE Trans Instrum Meas 49(5):1144–1148. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=872944

Caizzone S, Marrocco G, Member I (2012) RFID-Grids for deformation sensing. In: 2012 IEEE international conference on RFID, Orlando, FL. pp 130–134

Carver KR, Mink JW (1981) Microstrip antenna technology. IEEE Trans Antennas Propag 29(1):2–24

Chang K et al (2006) Patch antenna using synthesized polyimide for RFID sensing. In: Proceedings of the 9th European conference on wireless technology, Manchester, UK. pp 83–86

Charvat GL et al (2003) Harmonic radar tag measurement and characterization. In: IEEE antennas and propagation society international symposium, Columbus, OH, USA. pp 696–699

Daliri A et al (2010) Circular microstrip patch antenna strain sensor for wireless structural health monitoring. In: Proceedings of the World congress on engineering, London, UK. pp 1–6

Daliri A et al (2012) Wireless strain measurement using circular microstrip patch antennas. Sens Actuators A Phys 184:86–92. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0924424712004268. Accessed 16 June 2014

Denoth A (1997) The monopole-antenna: a practical snow and soil wetness sensor. IEEE Trans Geosci Remote Sens 35(5):1371–1375

Deshmukh S, Huang H (2010) Wireless interrogation of passive antenna sensors. Meas Sci Technol 21(3):035201

Deshmukh S et al (2011) Antenna sensor skin for fatigue crack detection and monitoring. Smart Struct Syst 8(1):93–105

Erdmann J, Huang H (2011) Microstrip antenna sensor for fatigue damage monitoring under lap-joints. In: Electromagnetic nondestructive evaluation, XIV, Szczecin, Poland. pp 456–465

Gagnadre I, Gagnadre C, Fenelon JP (1995) Circular patch antenna sensor for moisture content measurement on dielectric material. Electron Lett 31(14):1167–1168

Girbau D et al (2012) Passive wireless temperature sensor based on time-coded UWB chipless RFID tags. IEEE Trans Microw Theory Tech 60(11):3623–3632

Harrington RF (1963) Electromagnetic scattering by antennas. IEEE Trans Antennas Propag 11(5):595–596

Hasani M et al (2013) Implementation of a dual interrogation mode embroidered RFID-enabled strain sensor. IEEE Antennas Propag Lett 12:1272–1275

Jang S, Kim J (2012a) Passive wireless structural health monitoring sensor made with a flexible planar dipole antenna. Smart Mater Struct 21:027001 (6 pp)

Jang S-D, Kim J (2012b) Passive wireless structural health monitoring sensor made with a flexible planar dipole antenna. Smart Mater Struct 21(2):027001. Available at: http://stacks.iop.org/0964-1726/21/i=2/a=027001?key=crossref.acd264e6f077f238e3b2c3e8b0a1f4cd. Accessed 8 July 2014

Jatlaoui MM et al (2010) Original identification technique of passive EM sensors using loaded transmission delay lines. In: Proceedings of the 40th European microwave conference, Paris, France. pp 1106–1109

Jiang H et al (2014) Patch antenna based temperature sensor. In: Proceedings of SPIE, vol 9063, San Diego, p 90631P–1

Kalansuriya P, Bhattacharyya R, Sarma S (2013) RFID tag antenna-based sensing for pervasive surface crack detection. IEEE Sens J 13(5):1564–1570

King DD (1949) The measurement and interpretation of antenna scattering. Proc IRE 37(7):770–777. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1698079

Kubina B et al (2012) A wireless chipless temperature sensor utilizing an orthogonal polarized backscatter scheme. In: 42nd European microwave conference (EuMA), Amsterdam, Netherland. pp 61–64

Mandel C et al (2011) Passive chipless wireless sensor for two-dimensional displacement measurement. In: Proceedings of the 41st European microwave conference, Manchester, UK. pp 79–82

Marrocco G, Amato F, Politecnico V (2009) Self-sensing passive RFID: from theory to tag design and experimentation. In: Proceedings of the 39th European microwave conference, Rome, Italy. pp 1–4

Matsuzaki R, Melnykowycz M, Todoroki A (2009) Antenna/sensor multifunctional composites for the wireless detection of damage. Compos Sci Technol 69(15–16):2507–2513. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0266353809002589. Accessed 3 Sept 2014

McGrath MP, Sabouni RN, Pham A-VH (2004) Development of nano-based resonator gas sensors for wireless sensing systems. In: Islam MS, Dutta AK (eds) SPIE proceedings, Philadelphia, USA. vol 5593, pp 62–72. Available at: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=852132. Accessed 3 Aug 2014

Merilampi S et al (2010) Printed passive UHF RFID tags as wearable strain sensors. In: 2010 3rd international symposium on applied sciences in biomedical and communication technologies (ISABEL), Rome, Italy. pp 8–12

Merilampi S et al (2011) Embedded wireless strain sensors based on printed RFID tag. Sens Rev 31(1):32–40. Available at: http://www.emeraldinsight.com/10.1108/02602281111099062. Accessed 8 July 2014

Mohammad I, Huang H (2010) Monitoring fatigue crack growth and opening using antenna sensors. Smart Mater Struct 19(5):055023. Available at: http://stacks.iop.org/0964-1726/19/i=5/a=055023?key=crossref.eef9836e5bd3b36c41f5b51988993867. Accessed 4 Jan 2013

Mohammad I, Huang H (2012a) Pressure and shear sensing based on microstrip antennas. In: Tomizuka M, Yun C-B, Lynch JP (eds) SPIE, San Diego, USA. vol 8345, p 83451D 1–8. Available at: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1314025. Accessed 24 Apr 2013

Mohammad I, Huang H (2012b) Shear sensing based on a microstrip patch antenna. Meas Sci Technol 23(10):105705

Mohammad I et al (2012) Detecting crack orientation using patch antenna sensors. Meas Sci Technol 23(1):015102. Available at: http://stacks.iop.org/0957-0233/23/i=1/a=015102?key=crossref.5a003d7092eecaa6dac2f762b76f6610. Accessed 4 Jan 2013

Newman EH, Forrai D (1987) Scattering from a microstrip patch. IEEE Trans Antennas Propag AP-35(3):235–251

Nikitin PV, Rao KVS, Martinez RD (2007) Differential RCS of RFID tag. Electron Lett 43(8):431. Available at: http://digital-library.theiet.org/content/journals/10.1049/el_20070253

Occhiuzzi C, Paggi C, Marrocco G (2011) Passive RFID strain-sensor based on meander-line antennas. IEEE Trans Antennas Propag 59(12):4836–4840. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5993515

Penttila K et al (2006) Radar cross-section analysis for passive RFID systems. IEE Proc Microw Antennas Propag 153(1):103–109

Potyrailo RA et al (2011) RFID sensors as the common sensing platform for single-use biopharmaceutical manufacturing. Meas Sci Technol 22(8):082001. Available at: http://stacks.iop.org/0957-0233/22/i=8/a=082001?key=crossref.827350ab642ad91ac4715b68974eb987. Accessed 3 Aug 2014

Pozar DM (1987) Radiation and scattering from a microstrip patch on a uniaxial substrate. IEEE Trans Antennas Propag AP-35(6):613–621

Presas SM et al. (2007) High efficiency diode doubler with conjugate-matched antennas. In: Proceedings of the 37th European microwave conference, Munich, Germany. pp 250–253

Qiao Q et al (2013) Reconfigurable sensing antenna with novel HDPE-BST material for temperature monitoring. IEEE Antennas Wirel Propag 12:1420–1423

Ramos A et al (2011) Time-domain measurement of time-coded UWB chipless RFID tags. Prog Electromagn Res 116:313–331

Reineix A, Paillol J, Jecko B (1993) FDTD method applied to the study of RCS of microstrip patch antennas. Antenna Telecommun 48(11–12):589–593

Salmani Z et al (2011) Application of antenna in strain measurement. In: 2011 international workshop on antenna technology (iWAT), vol 1, pp 336–339. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5752311

Sidén J et al (2007) Remote moisture sensing utilizing ordinary RFID tags. In: 2007 I.E. sensors conference, Atlanta, USA. pp 308–311

Tata U et al (2009a) Bio-inspired sensor skins for structural health monitoring. Smart Mater Struct 18(10):104026. Available at: http://stacks.iop.org/0964-1726/18/i=10/a=104026?key=crossref.732fe0903b517012ae34d3f1fd9ef5e7. Accessed 4 Jan 2013

Tata U et al (2009b) Exploiting a patch antenna for strain measurements. Meas Sci Technol 20(1):015201

Thomson DJ, Card D, Bridges GE (2009) RF cavity passive wireless sensors with time-domain gating-based interrogation for SHM of civil structures. IEEE Sens J 9(11):1430–1438

Verma R et al (2013) Carbon nanotube-based microstrip antenna gas sensor. In: 2013 I.E. 56th international midwest symposium on circuits and systems (MWSCAS), Columbus, USA. pp 724–727. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6674751

Weller T, Davidova R (2013) High-sensitivity, AM-modulated harmonic transceiver for wireless sensing. Electron Lett 49(8):566–567. Available at: http://digital-library.theiet.org/content/journals/10.1049/el.2012.3388. Accessed 27 Aug 2014

Wiwatwithaya S et al (2011) Real-time monitoring glucose by used microwave antenna apply to biosensor. In: 2011 international conference on biomedical engineering and technology, Chiang Mai, Thailand. pp 100–104

Xu X, Huang H (2012a) Battery-less wireless interrogation of microstrip patch antenna for strain sensing. Smart Mater Struct 21(12):125007

Xu X, Huang H (2012b) Multiplexing passive wireless antenna sensors for multi-site crack detection and monitoring. Smart Mater Struct 21(1):015004

Yang F et al (2012) Reconfigurable sensing antenna: a slotted patch design with temperature sensation. IEEE Antennas Wirel Propag Lett 11:632–635

Yao J et al (2014) Dynamic interrogation of passive wireless antenna sensor. In: Proceedings of SPIE, vol 9061, San Diego, p 9061X–7

Yi X et al (2011) Passive wireless smart-skin sensor using RFID-based folded patch antennas. Int J Smart Nano Mater 2(1):22–38. Available at: http://www.tandfonline.com/doi/abs/10.1080/19475411.2010.545450. Accessed 4 June 2014

Yi X et al (2013) Passive wireless antenna sensor for strain and crack sensing – electromagnetic modeling, simulation, and testing. Smart Mater Struct 22(8):085009. Available at: http://stacks.iop.org/0964-1726/22/i=8/a=085009?key=crossref.c95e7f98c482b667224abae5a02d47c1. Accessed 8 July 2014

Yi X et al (2014) A slotted patch antenna for wireless strain sensing. In: 2014 structures congress, Boston, USA. pp 2734–2743. Available at: http://ascelibrary.org/doi/abs/10.1061/9780784413357.239

You KY et al (2010) A rectangular patch antenna technique for the determination of moisture content in soil. In: PIERS proceedings, Cambridge, USA. pp 850–854

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_88-1

Antenna Systems for Cellular Base Stations

Roland Gabriel[a] and Matthias Geissler[b]
[a]Kathrein Werke KG, Rosenheim, Germany
[b]IMST GmbH, Kamp-Lintfort, Germany

Abstract

Base station antenna systems have undergone a dramatic development within the last decades: in the early days of cellular communications, the cells where more or less of similar size and circular shape, having a base station in the center. This base station typically was equipped with a passive omnidirectional antenna illuminating the whole cell. Since then, the cell topology has become more and more diverse, using different cell sizes and also cell sectorization of higher orders. Moreover, the increasing demand on dynamic control and adaptation of the cell size and of the cell capacity results in complex requirements for both the antenna structure and the transceiver system. Therefore, it makes sense to consider the base station antenna system as a whole, comprising the antenna radiators, the transceiver system, and the data preprocessing.

This system design approach is the basic idea of this chapter: it starts with basic antenna considerations, then different system topologies are described including their functionality and their development over the years, and finally, an outlook into future challenges and design aspects is given.

Keywords

Base station antenna; Cellular network; Sector antennas; Mobile communication; Antenna array; Beamforming; Adaptive antennas; Smart antennas; Remote radio heads; Downtilt

Introduction

The antenna is part of a transmitting and receiving system; thus, the properties of this element have to be designed and analyzed together with the other elements of the system. Consequently, the design of the receivers and transmitters and furthermore the signal processing have a big influence on the requirements of the antennas. As an example, the correlation of signals of different antennas will result in a beamforming operation – in the sense of a superposition of antennas – if both signals will be combined using a defined phase relation. Instead, if the same antennas will operate in a MIMO mode (MIMO = multiple input, multiple output), the resulting patterns of the superposition of both signal streams will be different and will depend on the channel properties.

Active antenna systems have significant advantages like reduced noise figures for the receiving (Rx) path and lower losses of the transmission (Tx) path. Beside these advantages, the separate steering of different columns or radiators offers extended possibilities of user- or channel-defined beamsteering and multi-MIMO operation. The decision about the operation mode depends however on the supported standards and the available signal-to-noise ratio (S/N ratio) at the mobile and/or the base station. For two-dimensional antenna arrays, the question arises, which dimension of steering (horizontal, vertical) offers the higher gain in the network. Field trials sometimes deliver remarkably different answers to this

DOI 10.1007/978-981-4560-75-7_88-1

question and moreover are not comparable with each other because of different operation modes and scenarios.

The main challenges for the hardware realization of active antenna systems are reliability, cost, and efficiency: due to the high number of electronic modules, the lifetime decreases 4–10 times compared to an equivalent remote radio head (RRH). Innovative solutions are required to overcome this disadvantage in the future. A higher integration level of the electronic components in IC's is maybe the only way towards high reliability. The overall efficiency is also lowered by the separation of the duplex filters and by the medium-power level amplifiers. The best compromise between size and costs on one hand and attenuation and linearity on the other hand needs to be found. Nevertheless, a disadvantage of current system designs is still the limitation to only one operating frequency band, whereas today's technology of passive antennas offers up to five different bands within one antenna.

Cellular networks use distributed terrestrial base stations to serve the different mobile users. The antenna characteristics determine the coverage of the base station and the statistics of the propagation channel. One of the major requirements of a mobile communication network is the knowledge of the availability and the location of any defined user. The second challenge is – different from the broadcast service – the bidirectional traffic. Thus, powerful and flexible technologies have to be defined to coordinate the transmitting and receiving path.

The downlink and uplink coordination is realized using different duplex technologies like time division duplex (TDD) or frequency division duplex (FDD). Also, the combination of different access technologies is possible. Depending on the technology used for the channel separation, one talks about time division multiple access (TDMA), frequency division multiple access (FDMA), code division multiple access (CDMA), or space division multiple access (SDMA).

Historical Review

The localization of the mobiles today is organized in the home location register (HLR); this makes sure that any mobile can automatically be "found" within the network. At the beginning of the mobile communication system, the localization of the users was a big issue: instead of automatic detection, the mobile user had to give a call to announce his current cell location (see A-network, B-network in Germany, 1954–1970, MJ, MK-network USA). The HLR was already introduced in the C-network in Germany as well as in the NMT 450 in the Scandinavian countries. Figure 1 illustrates the technological milestones in the evolution of the mobile communication standards.

First cellular networks typically use omnidirectional antennas. The oldest technology used to achieve transmitting and receiving coordination was the time division duplexing. Thus, the receiver and transmitter use the same frequency channel, but they communicate during different time slots. At the very beginning of mobile communication, this coordination was done manually: first come, first serve or "push to talk." The collisions had to be solved manually by a certain priority logic. Today's TDD networks employ synchronized time slots in the millisecond range; thus, one will hardly recognize any difference to a parallel transmission.

Concerning the antenna design, the TDD system requirements with respect to passive intermodulation (PIM) are very relaxed, as no parallel transmitting and receiving signals exist within the same time slot. Examples of the TDD systems are the TD-SCDMA or TD-LTE system. TD-SCDMA means time duplexing synchronous code division multiplexing network. In this network the switching point between transmission and reception is synchronized over the whole network.

Base station technology finds itself in a continuous evolution process. In Fig. 2, typical key technological milestones are shown. It reveals that in the last 20 years there has been a clear tendency towards the

DOI 10.1007/978-981-4560-75-7_88-1

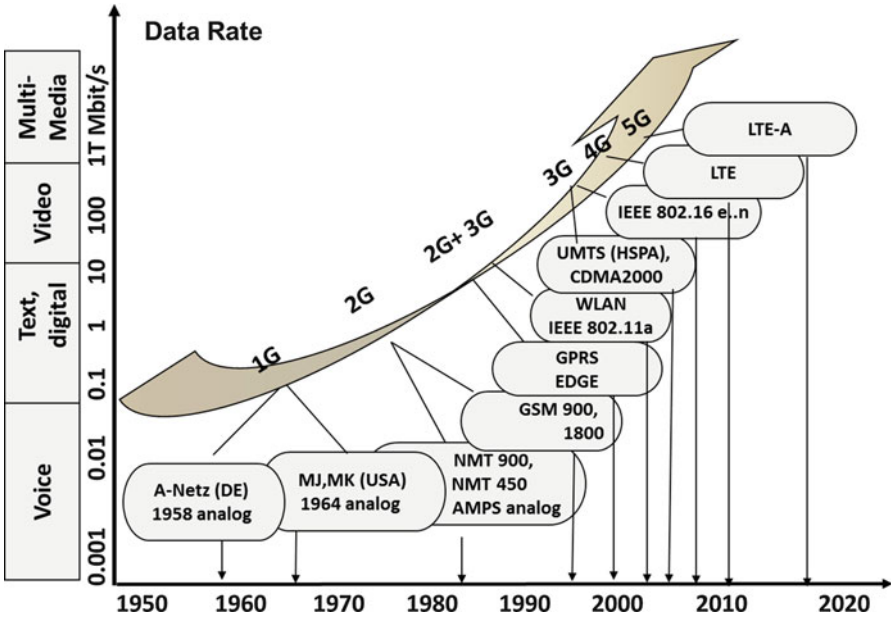


Fig. 1 Historical development of mobile communication systems

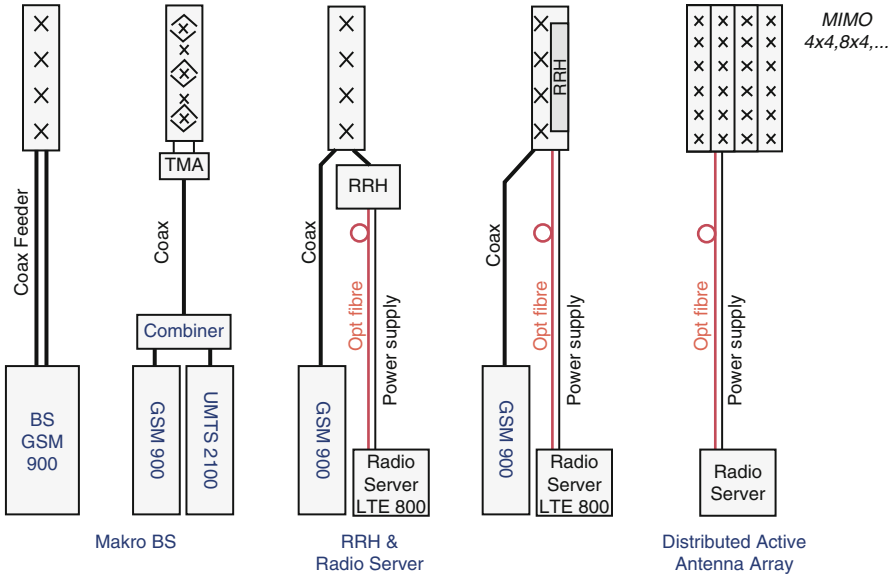


Fig. 2 Technology evolution of base station antenna systems

integration of electronic components into antennas. The first active components near to an antenna were "tower-mounted amplifiers" (TMAs) which amplified the receiving signals of the antenna and bypassed the transmitting signals by means of low-loss filter structures. The acceptance of such electronics near to the antennas is driven by higher reliability figures for the electronic components. In the meantime, the MTBF (mean time between failure) could be increased to values between 100,000 and 500,000 h. The same acceptance problem has prevented for a long time the use of remote radios near to antennas, comprising the receiving as well as the transmitting parts. Nevertheless, today's reliability is high enough,

DOI 10.1007/978-981-4560-75-7_88-1

and the advantage of power saving has forced the breakthrough of the use of RRHs beside the use of TMAs in the fields.

Antenna Design for Base Stations

Background and Definitions

Base stations are used in many different communication standards and can be installed by public authorities, network providers, companies, or private persons. When talking about base stations, two principal types of network philosophies have to be distinguished:

- **Cellular networks** typically use frequency bands which are licensed by a network provider. They consist of several base station antennas serving a network of neighboring cells. The size and shape of the cells, the capacity, and the frequency reuse are coordinated within the whole network. The installation and the service of these networks happen typically under the control of the network provider owning the license.
- **Wireless hot spots** typically use license-free frequency bands; they can be installed by companies, authorities, or private persons to supply a home, an office, or a building area. These base stations are generally not coordinated with other systems in the vicinity in terms of coverage and cell size, nevertheless, in the actual systems the frequency channel allocation is coordinated by means of self organizing features. Therefore, a network provider is not necessarily involved in the installation and the use of this type of base station.

Both types can be used in combination, e.g., a wireless LAN cell within a train wagon, whereas the train itself is connected via GSM to the cellular network. However, both types have to be carefully distinguished when talking about network architecture and corresponding antenna requirements. The main focus of this chapter is on cellular networks; however, the wireless hot spot technology is also described.

Important Communication Standards

Today's wireless and cellular communication use different digital communication standards in several frequency bands, some important are given in Table 1.

UMTS, LTE, and WLAN/WIFI comprise higher frequency bands. Some of the higher bands offer large bandwidth, so they have the potential of a strong increase in capacity. The technical system realization in these higher bands is however much more complex; therefore, they are currently rarely used (see also section "Antenna Related Network Aspects" in this chapter). The allocation of additional frequency bands for cellular applications in the range of 500–800 MHz from the broadcast services as well as in the higher frequency bands are in discussion. The allocations will be decided by the World Radio Conference (WRC).

Cell Topology of the Network

The **mobile communication network is structured** using cells. Each cell is equipped with at least one antenna transceiver system in a fixed mounted installation. The antenna transceiver system can serve the mobile users within the cell by the use of a certain number of channels dedicated to this cell. The effective border of a specific cell is defined by the resulting strength of the signal, the signal-to-noise-and-interference ratio (SNIR), respectively, received by a mobile on the border line of the cell. The mobile phone handover to a neighboring cell is controlled by this received power level: if the received power level of the neighboring cell is higher, the mobile phone switches to that cell.

DOI 10.1007/978-981-4560-75-7_88-1

Table 1 Important communication standards

Comm. standard	Frequency band	Application	Regions
NMT 450/GSM 450	450–470 MHz	Cellular	Scandinavia/Europe
D-AMPS	824–894 MHz	Cellular	North America
GSM	880–960 MHz 1710–1880 MHz 1850–1990 MHz	Cellular	Worldwide
DECT	1880–1920 MHz	Wireless (cordless phones)	Europe, USA, Canada, Japan
UMTS/IMT-2000	824–894 MHz 1710 MHz–2170 MHz Higher bands	Cellular	Worldwide
LTE	700 MHz–960 MHz 1710 MHz–2170 MHz 2500 MHz–2690 Mhz Higher bands	Cellular	Worldwide
Bluetooth	2400 MHz–2480 MHz	Wireless	Worldwide
WLAN/WIFI	2400–2500 MHz 5725–5875 MHz Higher bands	Wireless	Worldwide

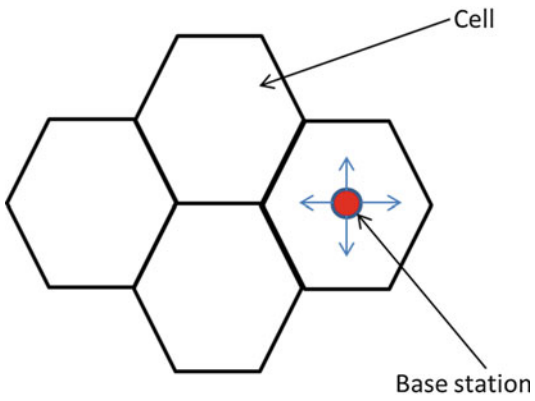


Fig. 3 Cell topology of the mobile communication network

Thus, from the antenna system perspective, the cell border can be defined by controlling the radiated power of the base station and by shaping the radiation pattern of the base station antenna.

In Fig. 3, the cells are displayed as neighboring hexagons, which is an appropriate model for an ideal cell topology, having well-defined borderlines and avoiding any overlapping of cells as well as any coverage gaps between cells. The base station is located at the center of the cell, illuminating the cell by using an omnidirectional antenna.

In order to maximize the capacity as well as the coverage of the mobile communication network, several cells can use the same subset of channels for transmission. In FDMA systems, the different channels are divided by frequency. So a certain subset of frequencies is dedicated to one cell. In order to avoid co-channel interference, these frequencies must not be used in adjacent cells, but they can be used in cells farther apart from each other. So a channel reuse scheme is implemented to make sure that the reuse of the channels happens always with an adequate spatial gap in between.

DOI 10.1007/978-981-4560-75-7_88-1

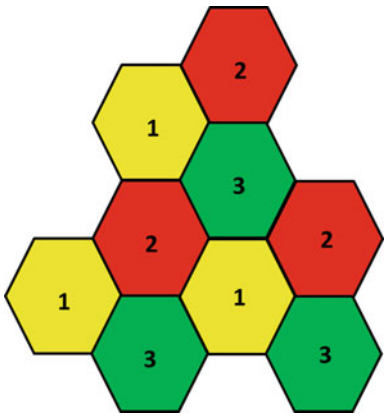


Fig. 4 Example for a frequency reuse scheme in a cellular network

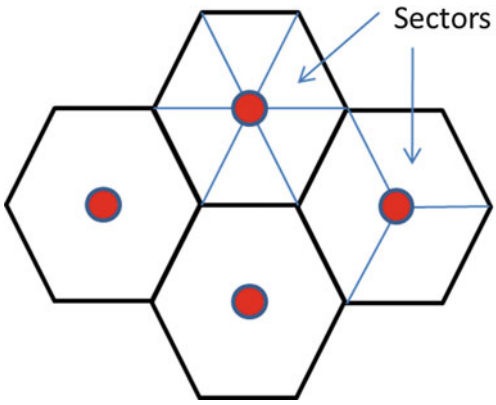


Fig. 5 Cell sectorization for increased capacity

Figure 4 shows the principal: the different cells use different frequency channel subsets 1, 2, and 3. As a result, two cells using the same channel subset always have a minimum distance in between. The reuse factor for a certain channel subset can be described by the value $1/K$, where K is the number of different channel subsets in the network. Typical reuse factors are 1/3, 1/4, 1/7, or 1/12.

In practical implementation, the so-called reuse distance is of importance: this term describes the minimum distance between the borders of two cells using the same frequencies. The propagation loss is typically proportional to the distance. The reuse distance therefore ensures that the signal level of the first cell has fallen down to an acceptable level of interference in the second cell, so that the reuse of the frequencies is possible (Collins 2009).

If the total available bandwidth is B, each cell can only use a number of frequency channels corresponding to a bandwidth of B/K.

In order to increase the capacity, existing cells can be subdivided into sectors. Figure 5 shows an example for sectorization: the cell at the right is subdivided into three sectors, and the cell at the top is subdivided into six sectors. Each of the sectors is treated as a separate cell, having its own subset of frequencies and its own antenna system. If the cell has three sectors, the resulting azimuthal coverage of each base station antenna is 120°, by division into six sectors; the resulting azimuthal coverage is only 60° (Derneryd and Johansson 2009).

Another possibility of adapting the capacity of the network is the use of different cell sizes: if the cell diameter is very large, one can cover a large area with a minimum of antennas. However, this also limits

DOI 10.1007/978-981-4560-75-7_88-1

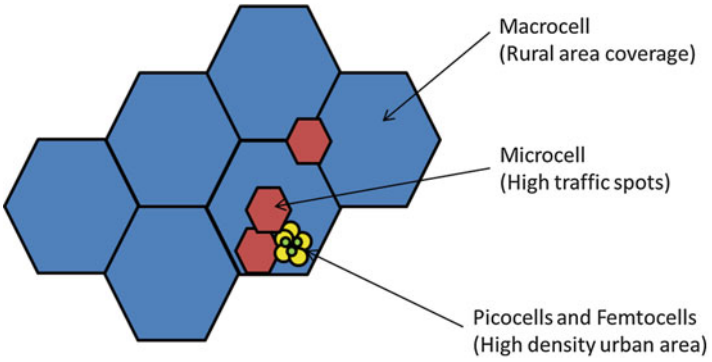


Fig. 6 Cell topology of the mobile communication network (dimensions not to scale)

the capacity, as the number of users in the cell is limited to the number of channels dedicated to this cell. In order to increase the capacity within a specific area, one can subdivide this area into smaller cells, each of them having its own dedicated bandwidth. The following classification related to the cell size has been established:

- **Macrocells** cover mainly rural areas. Their diameter can be up to 30 km or more, depending on the frequency and on the valid power regulation limits.
- **Microcells** have a typical diameter of about 200 m and typically serve hot spot areas like railway stations or inner city spots.
- **Picocells** and **femtocells** can be very small, e.g., in the order of 10–20 m. They are used to cover very small high-density areas like exhibition halls or hotel areas; even the coverage of single meeting rooms is possible (Fig. 6).

Both mechanisms of the adaptation of the cell size and the advanced sectorization can be combined adequately, so that the optimum constellation of coverage versus capacity can be obtained for a certain region.

Antenna Parameters and Definitions

Each cell needs to be illuminated in an appropriate way by the dedicated antenna system. This can be done through shaping the antenna pattern and through the control of the radiated power level. Important antenna parameters are listed in the following section. As basis for the description of the three-dimensional antenna radiation in far field, the following coordinate system shall be used (Linehan and Gianola 2011):

In the coordinate system of Fig. 7, the antenna is located in the origin, and the antenna dimensions are small versus the radius r of the far-field sphere. The x-y plane represents the horizontal plane – parallel to the ground. The horizontal plane is also called the azimuth plane, and the ϕ angle is called the azimuth angle. The z-axis is the vertical axis perpendicular to the ground, the x-z plane and the y-z planes are called the elevation planes, and the θ angle is called the elevation angle.

Power Density

Relevant for the effective range of the antenna pattern is the resulting strength of the transmitted signal at the designated borderline of the cell: the higher the output power, the larger the range. Looking at the antenna far field, where only transversal components of E and H exist and assuming harmonic oscillation, one can describe the power density in the far field as follows:

DOI 10.1007/978-981-4560-75-7_88-1

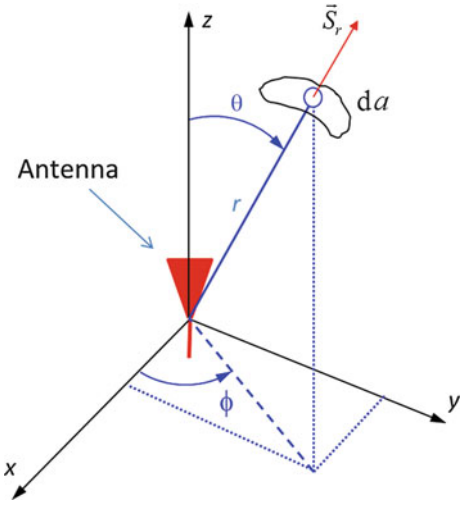


Fig. 7 Coordinate system for the far-field pattern description of an antenna

$$\underline{S} = \underline{S}_r = \frac{1}{2}\left(\underline{E}_\vartheta \underline{H}_\varphi^* - \underline{E}_\varphi \underline{H}_\vartheta^*\right)$$

By integrating the power density over the far-field sphere, one obtains the total radiated power P_{rad} of the antenna:

$$P_{rad} = \int_A S_r df = r^2 \int_{\varphi=0}^{2\pi} \int_{\vartheta=0}^{\pi} S_r \sin\vartheta d\vartheta d\varphi$$

The typical reference antenna is the isotropic radiator for which the power density can be described as follows:

$$S_r = S_i = \frac{P_{rad}}{4\pi r^2} = const_r$$

For a specific antenna device, the power density and the antenna gain are referred to the values of the isotropic radiator.

Far-Field Radiation Pattern

Relevant for the shape of the antenna beam is the antenna pattern, which thus defines the initial shape of the cell or sector that is illuminated. The normalized far-field pattern C of an antenna is defined as follows:

$$C(\varphi,\vartheta) = \left.\frac{\left|\vec{E}\,(\varphi,\vartheta)\right|}{\left|\vec{E}\,(\varphi,\vartheta)\right|_{\max}}\right|_{r=const.,\, r\to\infty} = \left.\frac{\left|\vec{H}\,(\varphi,\vartheta)\right|}{\left|\vec{H}\,(\varphi,\vartheta)\right|_{\max}}\right|_{r=const.,\, r\to\infty}, \text{ with } 0 \le C \le 1$$

Half-Power Beamwidth (HPBW)

The half-power beamwidth (HPBW) describes the angular width of the main beam area, where the power density reduction is 3 dB or less. Please note that for a given sectorial angle (e.g., 120°), the HPBW of the

DOI 10.1007/978-981-4560-75-7_88-1

antenna to be chosen is typically smaller (e.g., 65°) in order to minimize the interference towards the neighboring cells.

Antenna Directivity and Antenna Gain

The antenna directivity is defined as the factor between the power density in the main beam direction versus the power density of the isotropic radiator:

$$D = \frac{S_{r,\max}}{S_i} = 4\pi r^2 \frac{S_{r,\max}}{P_{rad}}$$

The antenna gain G is defined as follows:

$$G = \eta D = 4\pi r^2 \frac{S_{r,\max}}{P_{in}}$$

where D is the directivity and η is the total antenna efficiency of the antenna. The antenna gain is typically given in dBi, whereat the index i indicates the reference to the isotropic radiator ($G = 0$ dBi). Accordingly, the antenna gain can also be calculated by integrating the radiation pattern over the far-field sphere. The following relation is valid:

$$G = \eta D = \eta \frac{4\pi}{\int_{\varphi=0}^{2\pi} \int_{\vartheta=0}^{\pi} C^2(\vartheta, \varphi) \sin\vartheta d\vartheta d\varphi}$$

EIRP

The EIRP is the so-called Effective Isotropic Radiated Power of an antenna. This is a very useful parameter, as it turns the normalized parameter G into an absolute power level. The definition is

$$EIRP = G \cdot P_{in}$$

The EIRP is typically given in dBm or dBW.

Effective Aperture Size

The effective aperture size A_{eff} is a measure for the amount of power that an antenna can extract from a plane wave field. The received power P_R at the antenna depends on the A_{eff} and on the power density S of the incoming field as follows:

$$P_R = A_{eff} S$$

The effective aperture size can be put in the following relation to the gain of the same antenna:

$$A_{eff} = \frac{{\lambda_0}^2}{4\pi} G$$

This relation is generally valid.

DOI 10.1007/978-981-4560-75-7_88-1

Downtilt Angle

Another important parameter for base station antennas is the downtilt angle. In order to illuminate the cell adequately from an antenna mounted on a high tower, the antenna pattern is tilted slightly downwards. The tilt helps to avoid interference with neighboring cells, and it saves transmission power.

The downtilt-angle α is defined as the angle between the horizontal plane and the tilted main beam direction of the antenna. Typical downtilt-angle values are between 2° and 10°. One has to distinguish between "mechanical tilt," realized by a tilted mounting position of the antenna on the mast, and "electrical tilt," realized by an electrical phase shift within a vertically mounted antenna array. Both techniques can be used for a sector antenna; however, the resulting pattern is not identical, as the implementation of a phase shift effectively changes the antenna properties (Collins 2009).

Mechanical and electrical tilt can also be combined within one antenna; moreover, the electrical tilt can be dynamically modified by using remotely controlled phase shifters. Today's phase shifters for implementing electrical tilt are typically realized as electrical delay lines, where the delay can be varied by moving a slider along a set of bended lines (see also section "Antenna Line Devices").

Polarization

Polarization is the orientation of the oscillating field vector of an electromagnetic wave. In far field, the E-field vector is orthogonal to the propagation direction. In most cases of cellular communication, linear polarization is used, that means that the E-field vector is not rotating around the propagation axis. Depending on the field vector's orientation versus the ground, one talks about "vertical polarization" or "horizontal polarization" or "slanted polarization."

Cross-Polarization Ratio

As most of the base station antennas are dual-polarized antennas, there are defined two orthogonal polarizations with two separate antenna inputs. The cross-polarization ratio or cross-polar discrimination defines the ratio of the desired part of polarization to the part of orthogonal polarization in a defined polarization system (e.g., ±45° or horizontal–vertical). This ratio typically is dependent on the direction and can be visualized in a copular and cross-polar pattern. The cross-polar ratio determines the maximum diversity gain in a polarization diversity system; therefore, typically a minimum value of 10 dB is required within the sector of ±60° (see Linehan and Gianola (2011)).

Cross-Polar Isolation

The cross-polar isolation is the measured isolation between the inputs of a dual-polarized antenna and includes the parasitic coupling between the radiators. Values of more than 30 dB are typically required.

Tracking

The port-to-port tracking or port-to-port tracking error describes the gain difference at a given azimuth angle, typically in the ±60° sector, for both ports of the orthogonal polarizations. This is caused by asymmetries of the radiator systems and can be tilt dependent (see also Linehan and Gianola (2011) for more details).

Azimuthal Beam Squint

The azimuthal beam squint determines the difference between the main beam direction derived from the middle between the −3 dB points of the beam and the mechanical bore-side direction of the antenna.

DOI 10.1007/978-981-4560-75-7_88-1

Null Fill and Sidelobe Suppression

Null filling is a measure to avoid coverage gaps and describes the minimum level in the first null below the main beam. The upper sidelobe suppression determines the maximum level of the first or also the second sidelobe above the horizon, typically in a range of $0 \ldots + 20^\circ$.

Front-to-Back Ratio

The front-to-back ratio is important for the interference situation in the rear direction of the antenna. The value is specified exactly in the rear direction or within a given range of angle, typically $\pm 30^\circ$.

Pattern Design Strategy

Based on the preconditions mentioned above, three key aspects have to be fulfilled in order to adapt the antenna pattern optimally to the dedicated cell or sector area: the radiated power, the vertical radiation characteristics, and the horizontal radiation characteristics.

Adaptation and Control of the Radiated Power

By adapting the radiated power of the base station, the effective range can be controlled. This corresponds to the effective radius of the cell or sector.

Based on the known antenna gains G_t of the transmitter and G_r of the receiver and on the distance R between the base station and mobile device, one can calculate the received power at the mobile device using the Friis formula to

$$P_{R,\,mobile} = \left(\frac{\lambda_0}{4\pi R}\right)^2 G_T G_R P_{T,\,basestation}$$

where λ_0 is the wavelength. However, this formula is only valid in ideal scenarios, where both antennas of the base station and of the mobile are pointed perfectly towards each other in line-of-sight (LOS) conditions. In practical cases, the effective field strength within the cell depends on the topology of the area and on the superposition of the reflected signals. Moreover, this formula refers to a free space scenario; it does not take into account the presence of the ground. The reflected ray from the ground interferes with the direct ray from the base station at any point within the cell.

A modified version of this formula using the empirical factor γ can be seen as a more realistic prediction of the received power at the mobile (Beckman and Lindmark 2007):

$$P_{R,\,mobile} = \left(\frac{\lambda_0}{4\pi}\right)^2 \left(\frac{1}{R}\right)^{\gamma} G_T G_R P_{T,\,basestation}$$

The factor γ is found to be in a range between two and five depending on the propagation environment: $\gamma = 2$ is used for very open rural areas, whereas $\gamma = 5$ represents high-density urban areas. So by controlling the radiated power of the base station antennas of neighboring cells, the effective size of the cells can be adapted.

Optimization of the Horizontal Pattern

The horizontal pattern of the antenna defines the adequate illumination of the dedicated sector. The ideal goal would be that the HPBW of the antenna pattern covers the sector with almost constant SNIR at any point of the sector border (which is equal to the minimum acceptable SNIR in this case), avoiding any overlapping with the neighboring sectors.

DOI 10.1007/978-981-4560-75-7_88-1

This theoretical constellation can typically not be realized in practice. However, from practical implementations of 3-sector cells, it turned out that an HPBW of 90° for more rural environments and 65° for high-density urban environments deliver acceptable results in terms of coverage and interference.

Optimization of Vertical Pattern

The optimization of the vertical pattern is important to "guide" most of the radiated power into the cell and to avoid waste of power in unwanted directions. If the base station antenna is mounted at a height similar to the typical height of the user terminals in this cell, the power should be concentrated in the horizontal plane to reduce radiation into the ground or into space. If however the antenna is mounted on a high-elevation position, the additional implementation of a downtilt is necessary, so that the beam is directed downwards into the cell. Moreover, within the illuminated sector, the power density should be as homogeneous as possible, avoiding local maxima or minima.

In order to optimize the vertical pattern, antenna arrays are very often used. The basic principle of antenna arrays is the linear superposition of the single-element radiation based on the relation

$$\underline{E}_{\vartheta} = \underline{F}_{E}\underline{F}_{Ar}$$

where $\underline{E}_{\theta}$ is the resulting electrical field at any point in the far field, $\underline{F}_{e}$ is the element factor, and $\underline{F}_{Ar}$ is the array factor. The element factor describes the spherical geometry of the element radiation as a complex function. The array factor for an arbitrary constellation of antennas of the same type, located in different positions r_n' and radiating with different amplitudes I_0 as well as with different phases δ_n, can be described as follows:

$$\underline{F}_{Ar}(\vartheta,\varphi) = \sum_{n=1}^{N}\left[I_0^{(n)} e^{j\delta_n} e^{jk_0\left(x_n'\sin\vartheta\cos\varphi + y_n'\sin\vartheta\sin\varphi + z_n'\cos\vartheta\right)}\right]$$

Assuming a linear constellation of N elements positioned along one axis having a defined inter-element distance d and a constant inter-element phase shift δ, the array factor can be described as follows:

For x-axis constellations:

$$\underline{F}_{Ar}^{x-axis}(\vartheta,\varphi) = \sum_{n=0}^{N-1} I_0^{(n)} e^{\delta n} e^{jk_0 nd\sin\vartheta\cos\varphi}$$

For y-axis constellations:

$$\underline{F}_{Ar}^{y-axis}(\vartheta,\varphi) = \sum_{n=0}^{N-1} I_0^{(n)} e^{\delta n} e^{jk_0 nd\sin\vartheta\sin\varphi}$$

For z-axis constellations:

$$\underline{F}_{Ar}^{z-axis}(\vartheta,\varphi) = \sum_{n=0}^{N-1} I_0^{(n)} e^{\delta n} e^{jk_0 nd\cos\vartheta}$$

In the case of isotropic radiators, the maximum directivity can be reached with uniform amplitude taper and broadside radiation. The directivity is

DOI 10.1007/978-981-4560-75-7_88-1

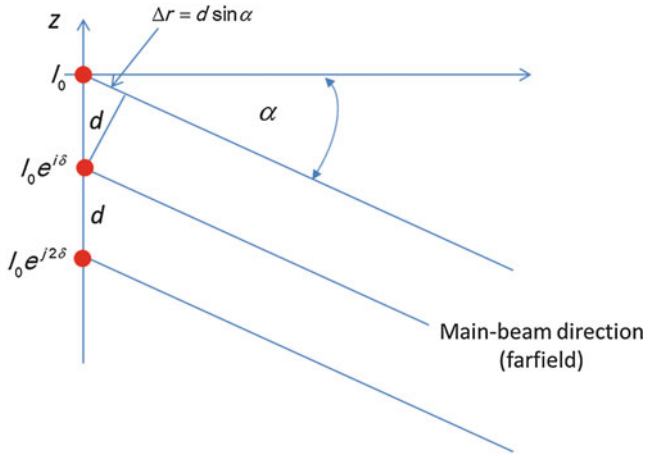


Fig. 8 Beam downtilt realization by introducing a phase shift

$$D(\delta = 0) \approx 2N \frac{d}{\lambda_0}$$

By introducing a phase shift δ, the pattern of the linear array can be tilted (Fig. 8).

Considering the precondition for the main beam direction of

$$\delta - k_0 d \sin \alpha = 0$$

The relation between the tilt angle α and the inter-element phase shift δ is

$$\alpha = \arcsin \frac{\delta}{k_0 d}$$

Besides of the variation of the phase, the amplitude can also differ from element to element. This amplitude taper can be described as discrete values $I_0(n)$, but also as a continuous distribution function $J_F(y)$. Based on the tapering function, the Fourier transform represents the relation between the amplitude distribution and the far-field pattern of the antenna.

$$\underline{J}_F(y) \circ\!\!-\!\!\bullet\, \underline{F}_{Gr}(u)$$

This means that based on the Fourier transform, the antenna pattern can be calculated out of the amplitude distribution and vice versa:

$$\underline{F}_{Gr}(\vartheta, \varphi) = \int_{-\infty}^{\infty} \underline{J}_F(y) e^{juy} \text{ using } u = k_y = k_0 \sin \vartheta \sin \varphi$$

This relation can be used very effectively for the synthesis of a desired pattern shape. The realization of a sector beam, for example, can be done using the Fourier relation of the pedestrian function, which is the sin x/x-function. So the amplitude tapering function in this case goes with

$$J_F(y) = I_0 \sin py / py$$

DOI 10.1007/978-981-4560-75-7_88-1

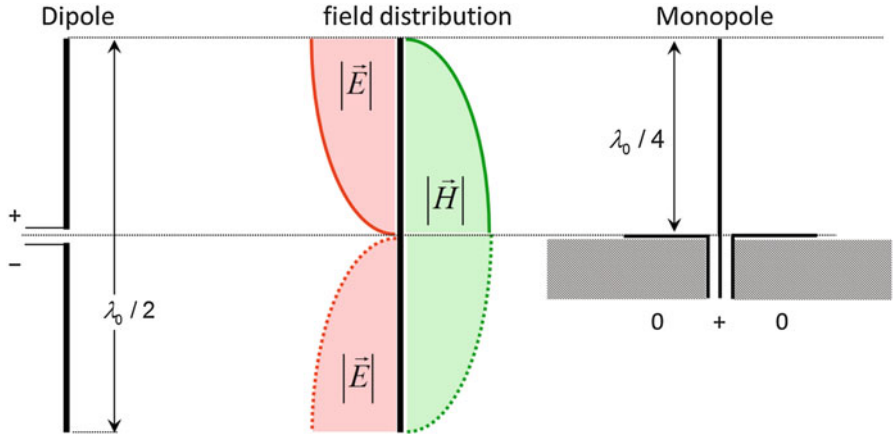


Fig. 9 Basic structure of the dipole and the monopole

for realizing a sector beam. The required discrete amplitude values of the distribution can then be derived by "sampling" the distribution function along the array line. Negative values of the function can be approximated by adding a 180° phase shift to the corresponding array coefficients.

Another very important aspect in vertical pattern synthesis is the filling of nulls in the pattern. Especially the first null just below the main lobe of the vertical pattern can cause a minimum in the sectorial coverage. By modifying the amplitude distribution according to the technique mentioned above, the first null can be filled resulting in a more homogeneous sector illumination.

Antenna Design Concepts

Typical Primary Radiators

Dipole and Monopole

The antenna design for base stations is based on well-known radiator elements: the half-wave dipole, the quarter-wave monopole, the patch antenna, and the planar inverted-F antenna (PIFA) are typical candidates here. A big variety of base station antennas are based on these types (Fig. 9).

The dipole is a wire antenna having an effective length of

$$l_{dipole} = \lambda_0/2$$

It is the smallest unfolded self-resonant structure that is possible. The E-field has its maximum at both ends of the antenna; the H field has its maximum at the center of the antenna. The antenna is typically fed in the center using symmetrical feeding (±). Due to the capacities at both ends of the dipole, the physical length is typically a little shorter than half of a wavelength.

The monopole antenna corresponds to the upper arm of the dipole antenna:

$$l_{monopole} = \lambda_0/4$$

However, this antenna type is not self-resonant; it needs a ground plane as counterpart. If the monopole is mounted on an infinite lossless ground plane, the field distribution along the antenna is similar to that of the upper dipole arm; the field below the ground plane level is zero by definition in this case.

DOI 10.1007/978-981-4560-75-7_88-1

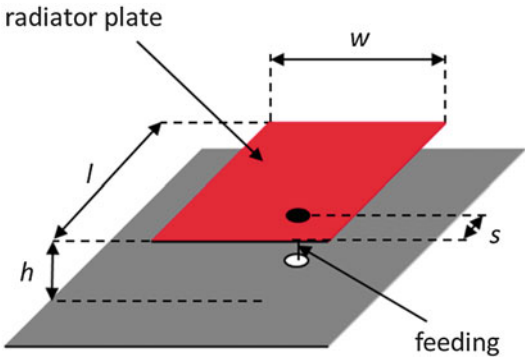


Fig. 10 Basic structure of the patch antenna

Patch

The patch antenna consists of a radiator plate parallel to the ground plane (Fig. 10). The presence of the ground plane defines the antenna properties; however, it is not a functional part of the antenna. The patch is a self-resonant structure; the resonance length is l:

$$\lambda/2 = l_{res}$$

The bandwidth Δf increases with increasing width (w) or increasing height (h) of the patch:

$$\Delta f = f_2(w, h)$$

The input impedance Z_{in} of the patch depends on the location of the feeding point: for $s = l/2$, the input impedance is zero; for $s = 0$ and $s = l$, the input impedance is at its maximum:

$$\underline{Z}_{in} = f_1(s)$$

As the patch is self-resonant and the resonance current is isolated from the ground plane, it is a very good antenna for realizing arrays, for example. Another advantage of the patch is the possibility of realizing two orthogonal polarizations within the rectangular radiator plate, one in l direction and the other one in w direction.

Planar Inverted-F Antenna (PIFA)

The planar inverted-F antenna (PIFA) consists of a radiator plate, which is mounted over a ground plane and a ground contact plate which makes the electrical connection between the radiator plate and the ground (Fig. 11). This antenna is resonant but not self-resonant, as it needs the ground plane as counterpart to function. The design criteria for the PIFA are as follows:

$$h + (w - w') + l = l_{res} \approx \lambda_0/4$$

The antenna is of quarter-wave type; the parameters h, l, and w determine the resonance frequency; and the input impedance can be increased by increasing the feeding distance s from the ground contact plate. More details about the design of PIFA's are given in Taga T. (1992) and Taga and Tsunewage (1987). The bandwidth of the antenna can be increased by increasing the height h.

DOI 10.1007/978-981-4560-75-7_88-1

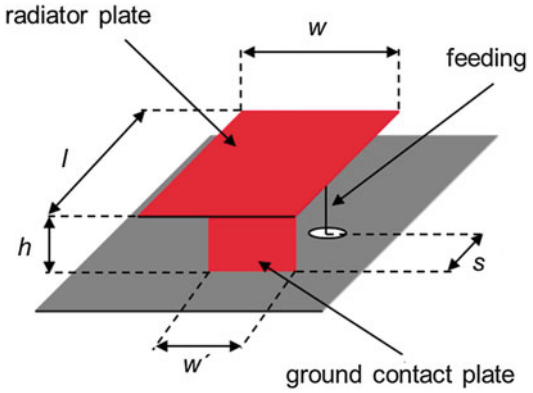


Fig. 11 Basic structure of the planar inverted-F antenna (PIFA)

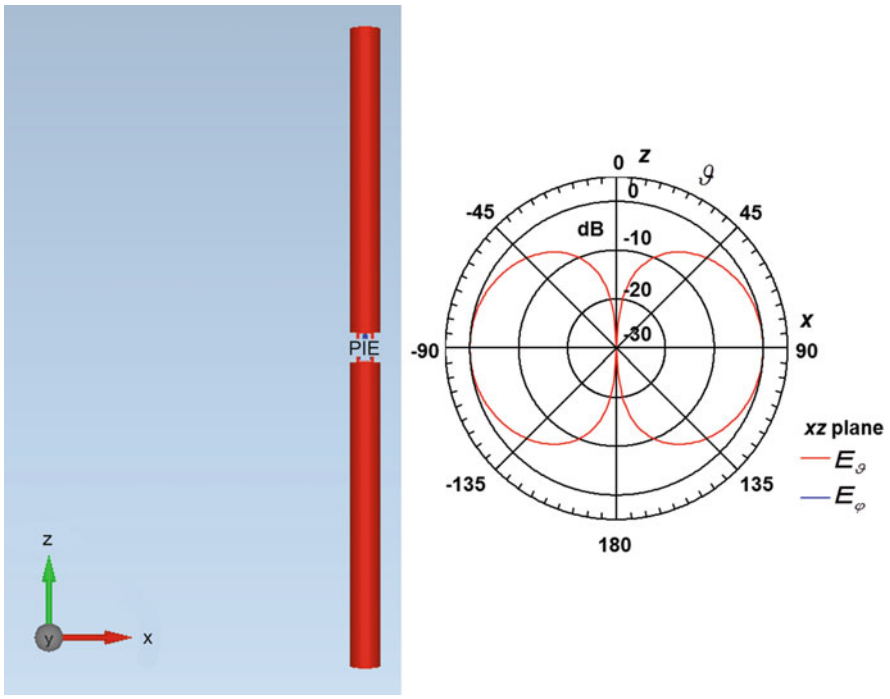


Fig. 12 Half-wave dipole ($f = 2$ GHz, $l_{Dipole} = 60$ mm, $d = 3$ mm, $D = 2.14$ dBi)

As the PIFA is a quarter-wave antenna, it can be designed in a very compact volume. However, the ground plane cannot be decoupled from the resonance current. The polarization direction is defined by the antenna geometry and the orientation; it cannot be changed by modifying the radiator plate only.

In the following sections, typical base station antenna designs using these basic antenna types will be explained.

Single-Element Configurations

Vertical Half-Wave Dipole

The vertical half-wave dipole is a very good antenna for an omnidirectional illumination of a cell. The figure shows a display of a 3D EM simulation model (Empire 2014) (left) of a half-wave dipole and the corresponding far-field pattern results of the 3D numerical simulation (right). The far-field pattern of

DOI 10.1007/978-981-4560-75-7_88-1

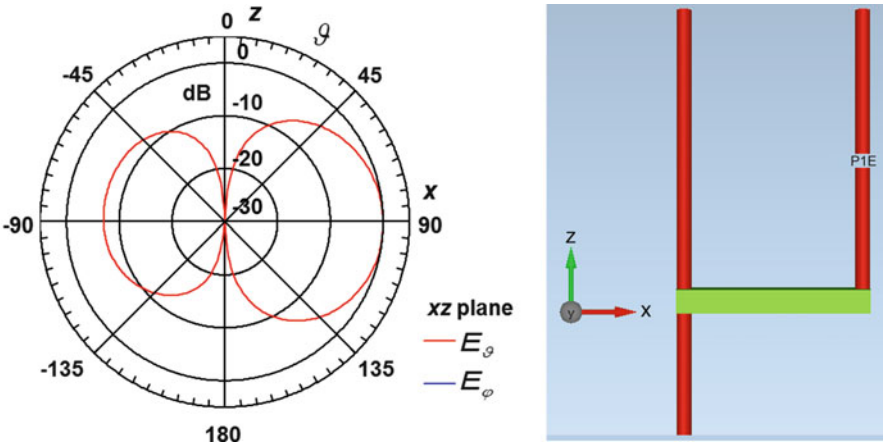


Fig. 13 Dipole (*right side*) in front of a mast (*left side*), both of metal. The dipole is fixed at the mast using a nonmetal support (*green*)

the dipole is a rotation symmetric to the z-axis. In the far-field vertical cut (yz), a null appears along the z-axis, and the maximum gain is obtained within the x-y plane. Due to the rotation symmetry, the horizontal pattern (xy) is circular, so the antenna has its maximum directivity of 2.14 dBi as a constant value within the horizontal plane (Fig. 12).

Looking at these pattern results, it can be seen that this antenna is very good for being positioned within the center of a nearly round cell. The cell should not be too big, as the directivity is low, and the antenna position should not be too high above the users, as there is no downtilt implemented in the antenna beam. So this is a typical antenna for microcells using a central base station positioned at low altitude in the cell center.

Dipole in Front of a Mast

If a dipole is mounted near to a mast, then the radiation characteristic is no longer omnidirectional. Depending on the parameters, this constellation can be used for the adjustment of the coverage of the cell. The mast functions like a passive "director" for the dipole. It absorbs energy out of the reactive near field and starts radiating itself. The far-field pattern results from the linear superposition of the fields of both radiators. Figure 13 shows a dipole in front of a mast and resulting radiation characteristics for a distance of $\lambda/4$ between the dipole and the mast.

Depending on the thickness of the mast, the directivity of the radiation pattern increases, so that it can be adapted to the required cell illumination.

Monopole Antenna on Small Base Station

The monopole antenna is also a very common type for base station applications. However, as the antenna needs a ground plane as counterpart, it is very often directly mounted onto the base station casing.

Typical use cases are the mounting onto picocell base stations which can have a total size of a large book. In the following simulation, a monopole antenna is mounted on a picocell base station, which is modeled as metal-coated box (Fig. 14).

The far-field pattern of this simulation shows a dipole-like pattern in the E-Theta component of the vertical cut (x-z plane). This indicates that the casing of the base station acts as an active counterpole rather than an infinite ground plane. So the monopole together with the casing functions like a dipole antenna.

This happens because the size of the casing is not infinite but in the order of a wavelength. Figure 15 shows the electrical near field in a vertical cut through the whole device: one can see that the field concentration is very high at the dipole, but there is also reactive field around the whole device. The metal

DOI 10.1007/978-981-4560-75-7_88-1

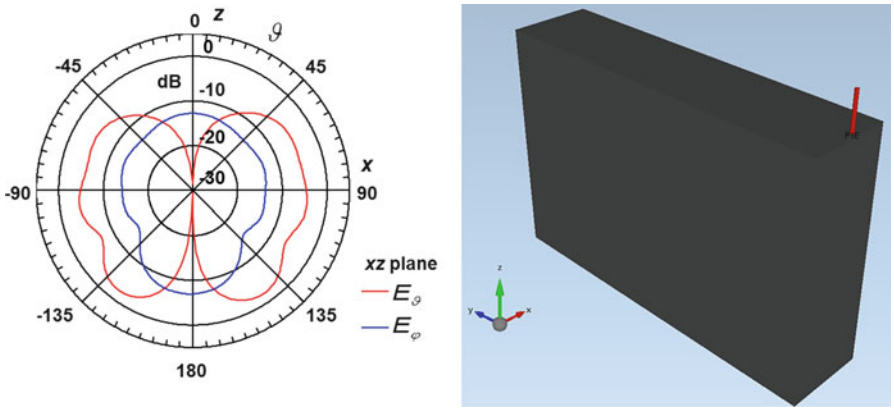


Fig. 14 Monopole on top of a small base station ($f = 2$ GHz, box dimensions $70 \times 200 \times 300$ mm^3, $D = 4.14$ dBi)

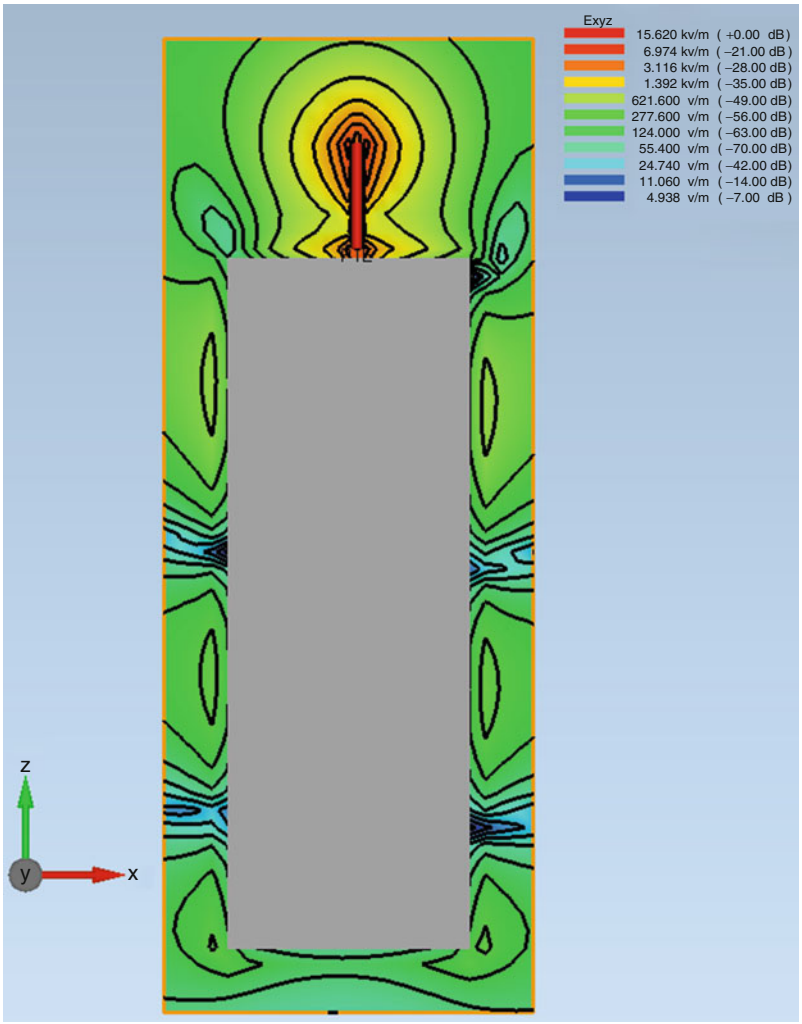


Fig. 15 Monopole on top of a small base station: *E*-field distribution in the vicinity of the device

DOI 10.1007/978-981-4560-75-7_88-1

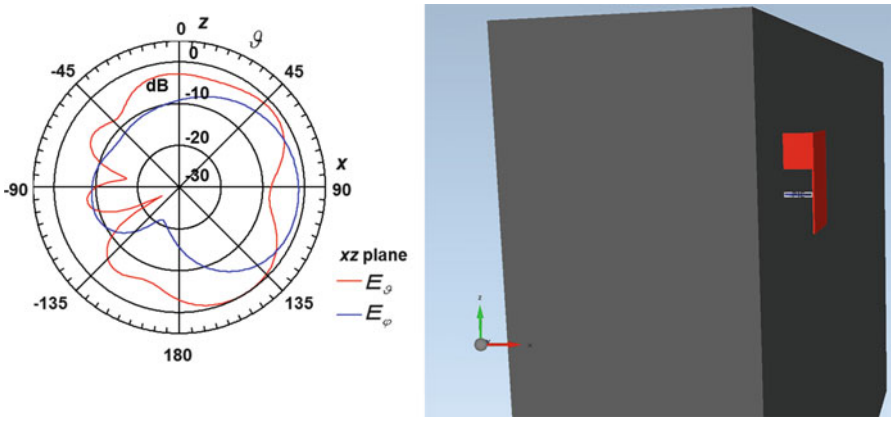


Fig. 16 Planar inverted-F antenna (PIFA) on a small base station ($f = 2$ GHz, box dimensions 70 × 200 × 300 mm^3, dimensions of the PIFA: 8.5 × 20 × 28 mm^3, $D = 2.41$ dBi)

casing acts as a part of the antenna and contributes to the radiation. (The orange frame in the picture limits the rectangular area, in which the near field was calculated).

Planar Inverted-F Antenna on Small Base Station

The planar inverted-F antenna is also very often used for small base stations: this antenna type needs a ground plane as counterpart as well. It can be mounted on the top of a casing, but also on the front side (+x direction) as shown in the simulation setup below. When assuming that the whole base station is mounted on a vertical wall oriented in z direction (wall not implemented in the simulation), the frontside position of the PIFA is advantageous: most of the far-field radiation is directed towards the +x direction away from the wall (Fig. 16). Moreover, the distribution of the reactive near field is blocked by the presence of the casing: this effect can be seen in the near-field cut in Fig. 17, and the reactive near field is concentrated in the vicinity of the antenna, whereas the backside of the casing shows very low field values. Therefore, the backside is less sensitive to the presence and the electrical properties of the wall in this constellation.

Besides of the forward directivity, this front mounting position also avoids a potential detuning of the antenna caused by the presence of the wall. The presence of the casing hides the wall influence from the antenna; it makes the antenna properties less sensitive to the wall effects.

Array Configurations

Dipole Array

If the base station antenna is located in the center of a circular cell, the half-wave dipole antenna can be used as explained above, as it has its maximum directivity in the horizontal plane. The gain of a single half-wave dipole however is relatively low. In order to increase the gain, several dipoles can be arranged in one line, where each of them is radiating at the same amplitude and phase. The resulting pattern has a higher gain in the horizontal plane than a single element, and the HPBW is smaller. So the radiation is directed into the horizontal plane and reduced towards upward and downward elevation directions (Fig. 18).

Sector Antennas

In many cases, cells are subdivided into sectors in order to increase the capacity. In this case, a directive antenna is necessary, and the horizontal pattern has to be adapted to the sector size and shape, whereas the vertical pattern has to be adapted to the cell diameter and the cell-mounting position (Fig. 19).

DOI 10.1007/978-981-4560-75-7_88-1

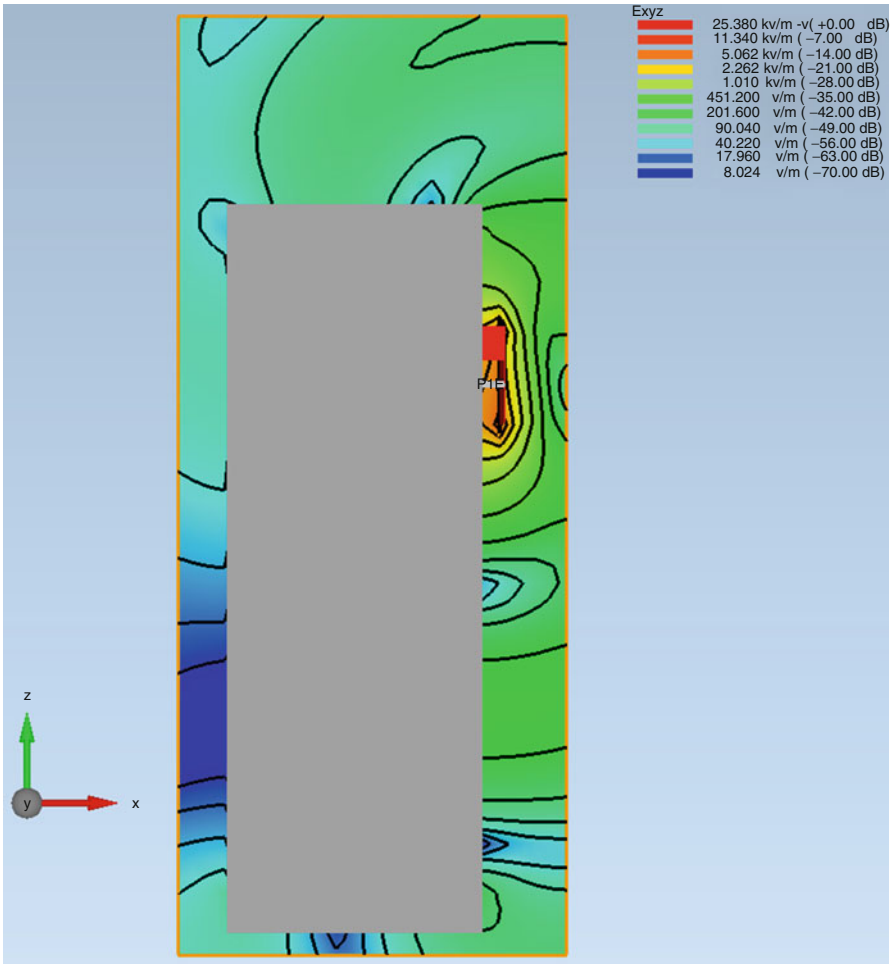


Fig. 17 PIFA on a small base station: *E*-field distribution at the device

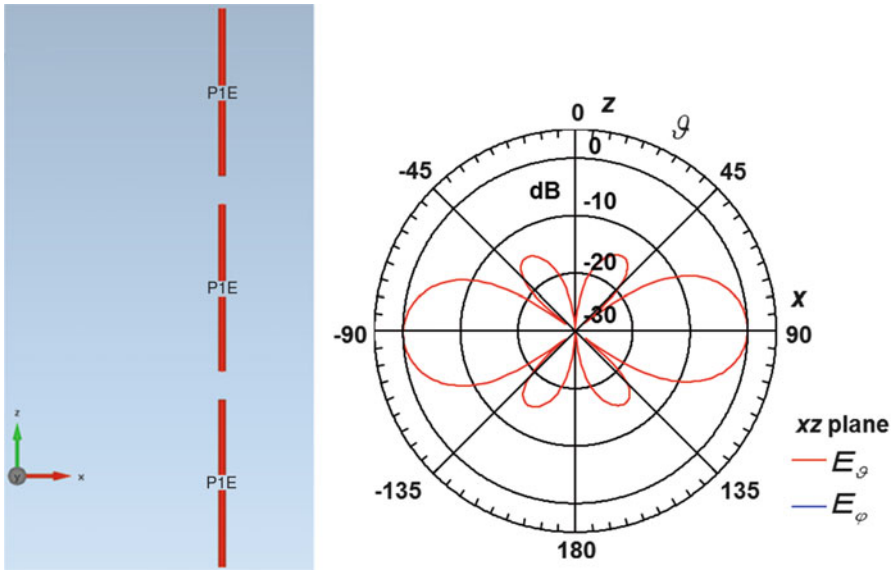


Fig. 18 Array of three half-wave dipoles along the z-axis (f=2 GHz, $l_{Dipole} = 60$ mm, $th_{\text{Dipole}} = 3$ mm, $d = 75$ mm, $D = 6.2$ dBi)

DOI 10.1007/978-981-4560-75-7_88-1

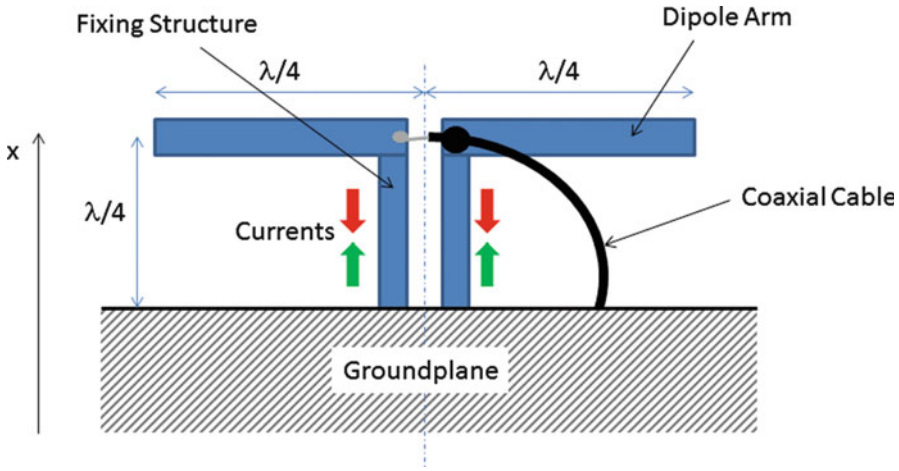


Fig. 19 Antenna element for sector array

A typical basic antenna element for realizing sector antennas is a half-wave dipole mounted over a metal plate. The metal plate acts as a reflector and as a ground plane too. The mounting distance is chosen to λ/4, which has several effects:

1. The mounting distance of *l*/4 transforms the electrical short of the antenna fixing structure into an electrical open at the dipole feeding point, so the fixing structure is electrically "invisible."
2. The ground plane acts as a reflector, and the backward radiation (in –x direction) from the dipole is reflected and interferes constructively with the radiation into +x direction.
3. The fixing structure acts like a balun, converting the unsymmetrical mode from the feeding coaxial cable into the symmetric dipole. Unwanted surface currents resulting from the unsymmetrical feeding (red) are suppressed by reflected currents having the same amplitude but the opposite phase.

So this constellation is a very good basic element for realizing sector antennas. The shaping of the horizontal characteristics can be done by modifying the size and the shape of the reflector as well as the dipole positioning over the reflector. So typically, required HPBW like 90° or 65° can be realized using this type of basic radiator.

The vertical characteristics can be shaped by building an array of dipoles. Figure 20 shows the simulation model of a five-element array. The dipoles are in this case vertically oriented, and the array is also arranged in vertical direction.

If all elements are fed by the same amplitude and phase, the resulting pattern has its maximum in the horizontal plane, and the resulting directivity is 14.11 dBi. Figure 21 shows the vertical pattern of this constellation: the presence of the reflector causes a directivity in +x direction, and the maximum can be found in the horizontal plane. Figure 22 shows the resulting vertical pattern of the five-element array: the directivity is increased versus the single-element case by adding more elements. If all elements are in phase, the resulting maximum remains in the horizontal plane.

In order to realize an electrical downtilt within this array, a phase shift has to be implemented. This phase shift of constant value between neighboring elements causes a declination of the wave front in the E-field, and the propagation direction of the waves is tilted downwards (Fig. 23)

The far-field pattern in Fig. 24 shows the resulting far-field pattern which is tilted downwards. So this kind of antenna array is very well suitable for sector antennas: the horizontal pattern HPBW can be adapted via modifying the reflector shape, the directivity can be increased by the array factor, and the vertical pattern can be adapted by adapting the electrical downtilt.

DOI 10.1007/978-981-4560-75-7_88-1

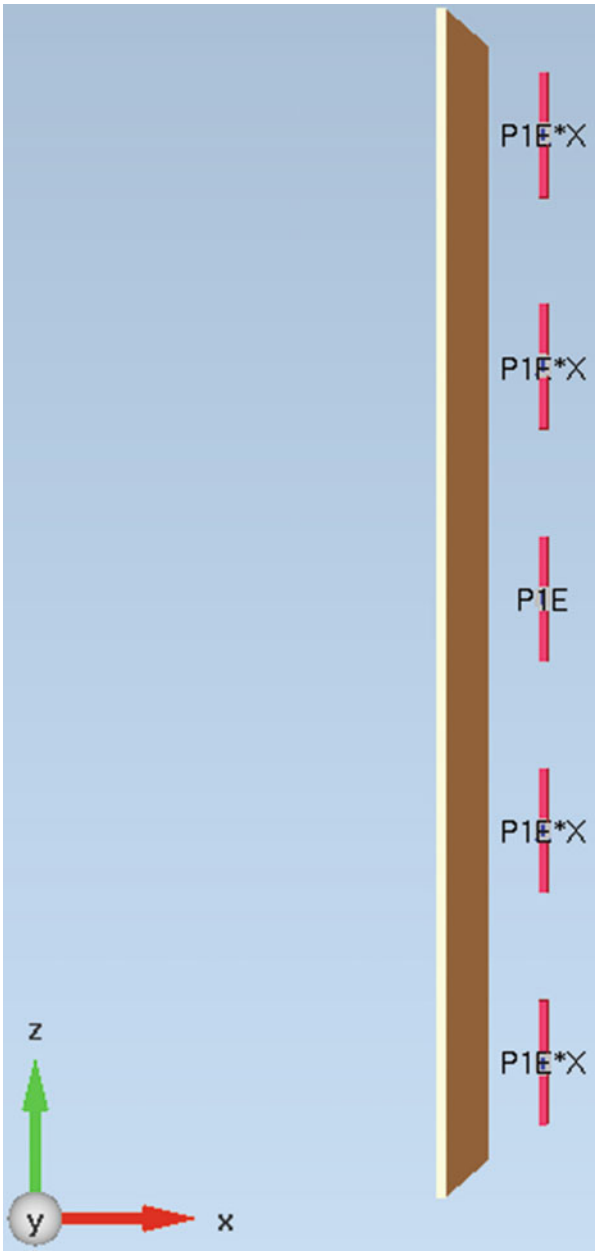


Fig. 20 5-Element-array of dipoles in front of a reflector plane ($f = 950$ MHz, $l_{Dipole} = 140$ mm, $th_{dipole} = 10$ mm, d=260 mm, reflector dimensions 1290 × 250 mm^2, $D = 14.11$ dBi)

Realization of Diversity or MIMO

Diversity is a common technique to improve the signal transmission between a base station and a mobile. The use of two antennas allows the compensation for any fading either by switching between the two antennas or by combining the two signals in the receiver.

Most of today's base station antenna systems offer diversity in order to overcome any fading issues. Very often polarization diversity is used. This means that two antenna radiators having orthogonal polarizations are built within the same volume.

Instead of horizontal–vertical arrangements, slanted arrangements of ±45° are commonly used. In this arrangement, the symmetry of the paths is better compared to the horizontal and vertical polarized arrangements. On the other hand, the isolation between the radiator parts and the resulting systems of

DOI 10.1007/978-981-4560-75-7_88-1

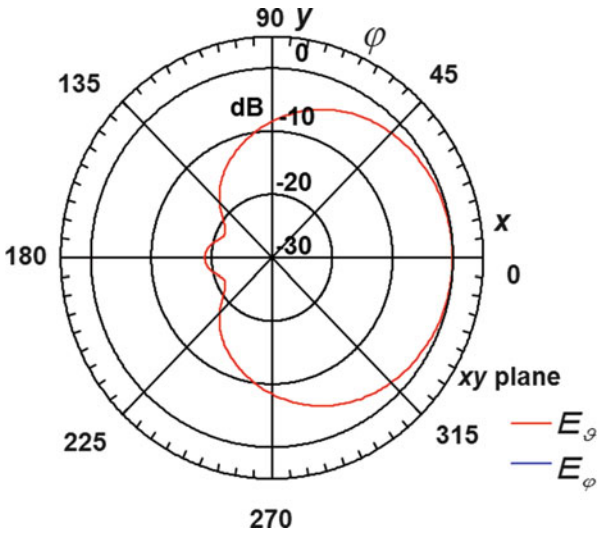


Fig. 21 Horizontal far-field pattern of the 5-element sector array

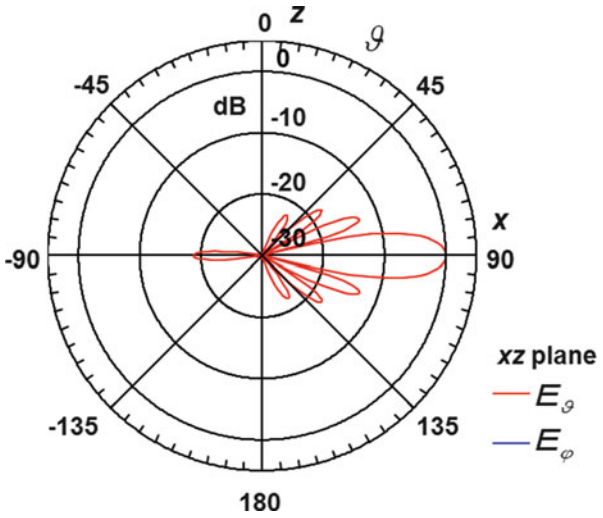


Fig. 22 Vertical far-field pattern of the 5-element sector array

the antennas are great challenges. In Fig. 25, dipole arrangements for different half-power beamwidths are shown schematically. A typical arrangement is based on two orthogonal dipoles, each of them in 45° arrangement versus the reflector plane (Fig. 25). The HPBW can be modified depending on the arrangement.

The orthogonal dipole arrays are then connected via two separate ports, so that the antenna can be used to perform diversity. The combining of the signals is done in the receiver. This constellation allows very good decoupling of the two input ports in the order of 20–25 dB.

Figure 26 shows a realization example of a radiator element which includes diversity: it consists of four square sectors, where two of them are built within one dipole element. The coaxial cables for the two ports are stored in the holes of the shaft and connected to two diagonally oriented elements each. This results in two orthogonal dipoles where the E-field can be oriented according to the red and the blue arrows, respectively.

DOI 10.1007/978-981-4560-75-7_88-1

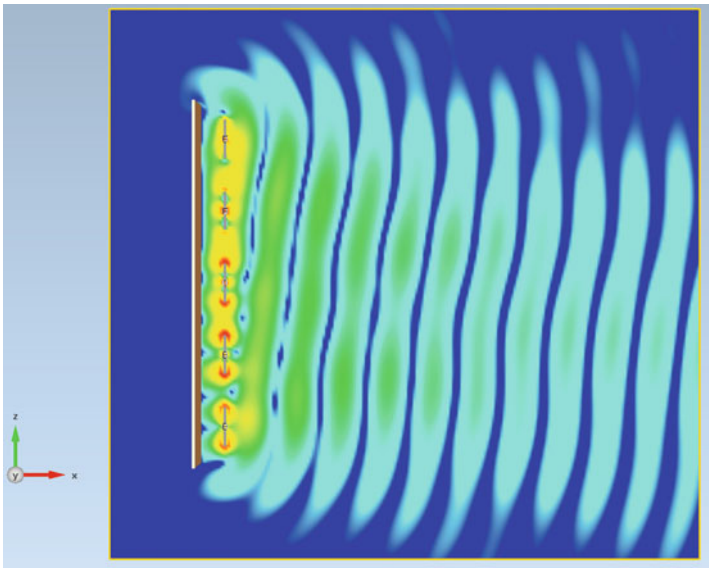


Fig. 23 Sector array with implemented phase shift: E-field distribution in front of the antenna

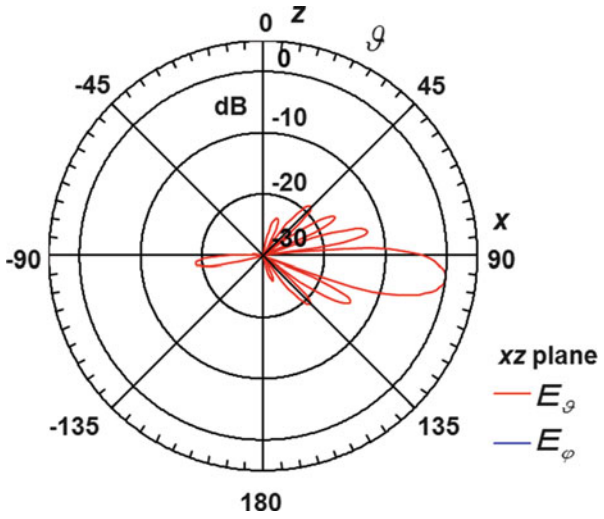


Fig. 24 Far-field pattern of reflector array with electrical downtilt

One challenge to solve is the isolation between both polarizations and both frequency bands within the antenna system. Especially in the slanted configuration, the problem is the projection of the radiating structures; see Fig. 27. This projection together with a non-orthogonal axis to the reflector sides requires additional measures to improve the isolation between the radiators.

This can be done by introducing parasitic coupling elements, which compensate the undesired coupling between the elements. To achieve the desired effect, a phase shift of 180° between the parasitic elements and the undesired coupling needs to be achieved for the operating frequency band. In Göttl (2004) as well as in Gabriel and Göttl (1997), such elements are described in forms of sheets, crosses, or stubs between the single radiators. Effectively, the elements are working only in the near field of the radiators and do not contribute significantly to the far-field radiation.

DOI 10.1007/978-981-4560-75-7_88-1

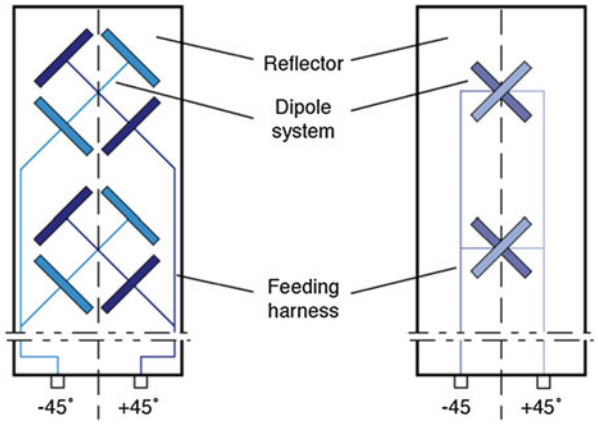


Fig. 25 Principle arrangements of dual-polarized dipole structures for 65° HPBW (*left*) and for 90° HPBW (*right*)

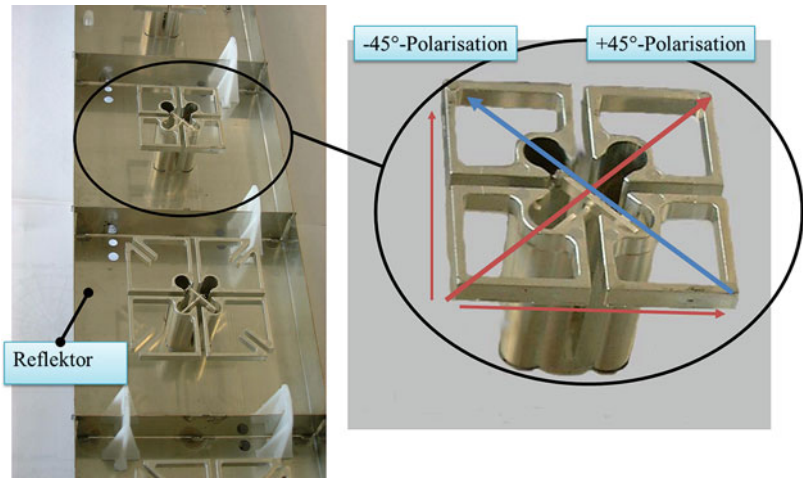


Fig. 26 Realization example of a dipole sector antenna with diversity

Based on these dual polarized antenna structures, MIMO can also be realized: MIMO stands for multiple input, multiple output and describes a technique of using multiple antennas at the transmitter side and at the receiver side to improve the link performance. MIMO configurations can be driven in different transmission modes that perform true MIMO operation, diversity, beamforming, or even spatial multiplex to either improve the S/N, to ensure a certain quality of service, or to maximize the data throughput.

For MIMO realization, the same basic radiator constellations can be used as for diversity: in a dual polarization dipole constellation, both ports can be used to perform a 2×X MIMO. By using several columns in a parallel arrangement, more channels can be provided for higher-order MIMO systems. Today's LTE system uses 2×2 MIMO as basic configuration, and higher order like 4×4 or even 8×8 is already in use.

Realization of Multiband Radiators

As the number of communication standards and the number of available frequency bands rapidly increases, there is an increasing demand for multiband antenna systems. Very often the available integration space at the base station site is very limited, so there is a strong need to implement several frequency bands into one antenna system.

DOI 10.1007/978-981-4560-75-7_88-1

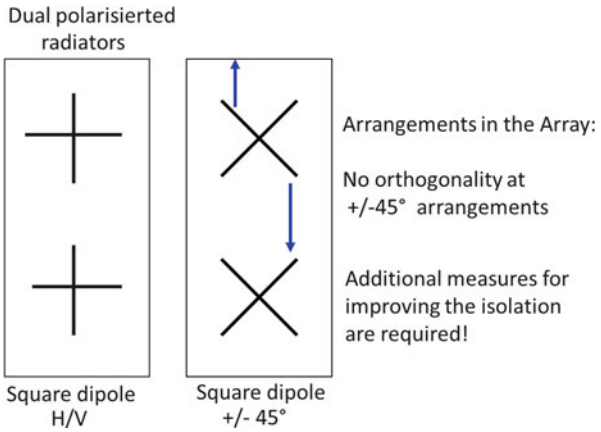


Fig. 27 Reduced isolation in slanted configurations of dual-polarized antennas

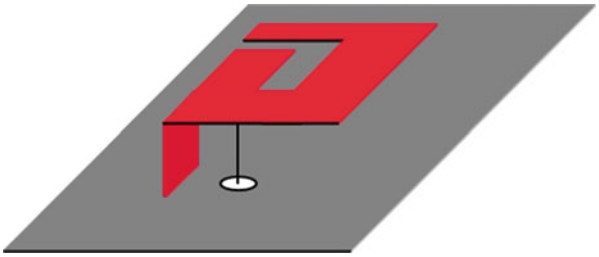


Fig. 28 Multiband PIFA

The integration of bands can happen on the element level and/or at the array level. In Fig. 28, an example for the integration on element level is displayed: the radiator plate of the planar inverted-F antenna is subdivided into two "arms" of different lengths, resulting in two different resonance frequencies. So this kind of antenna can, for example, cover the 900-MHz band and the 1800-MHz band. Based on this principle, three or even four resonances can be realized within one PIFA. This kind of multiband realization can be found especially in picocell base stations, where the PIFA is directly mounted onto the casing of the base station.

In Fig. 29 a dual-band, dual polarization structure is displayed: four dipoles in a square arrangement cover the low frequency, offering two polarizations, whereas a cross-dipole arrangement in the center of the configuration covers the high band. The dipole square operates in parallel as a reflector for the inner cross dipole; therefore, comparable half-power beamwidth for both bands can be obtained. With this kind of arrangement, the integration of the high band into the antenna does not need any more space in comparison to the single low-band antenna. However, the issue of increased mutual coupling between high and low band but also between the two orthogonal polarizations of each frequency band rises. So the whole structure has to be optimized to minimize any mutual coupling. In order to achieve this, special decoupling structures might be integrated as mentioned before. In Fig. 29 the current density of a simulation with the exited cross dipole for the high freqeuncy band, using the simulation tool "CST Microwave Studio" CST (2014) is shown.

Another important primary radiator is the patch structure. The dual-polarized patches use both orthogonal resonances of the patch. An example of a dual-band structure according to Lindmark (2000) is shown in Fig. 30. All patches are working in an orthogonal resonance mode, and the upper patch for the

DOI 10.1007/978-981-4560-75-7_88-1

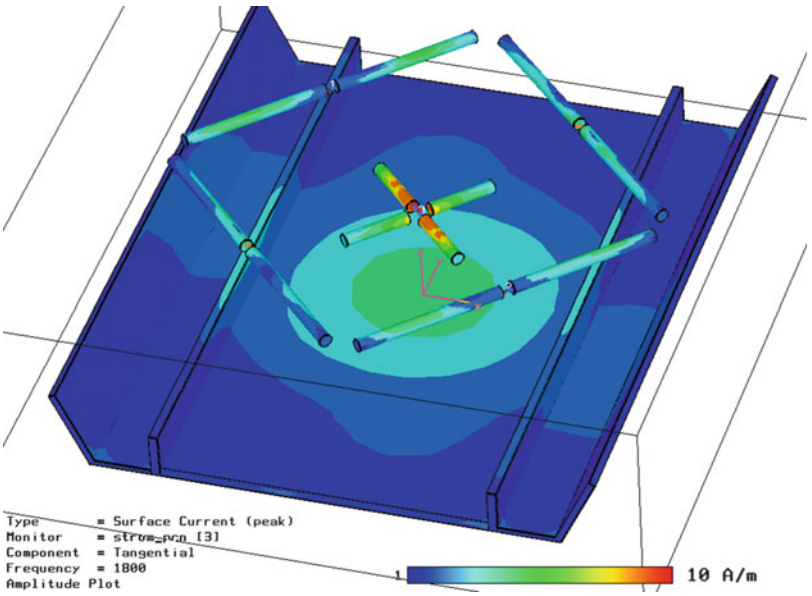


Fig. 29 Dual-band arrangements of dual-polarized dipoles structures

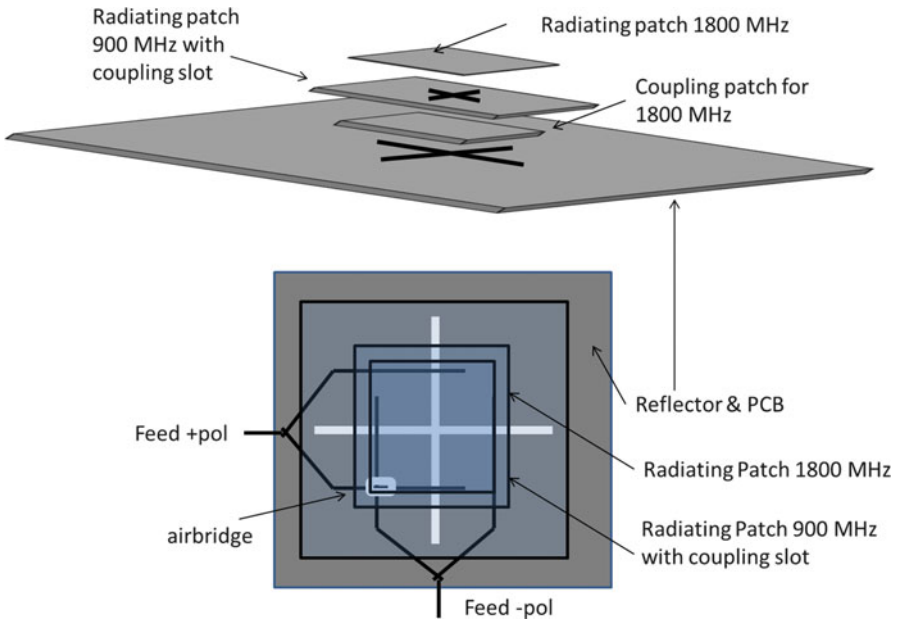


Fig. 30 Dual-polarized, dual-band patch arrangements

1800 MHz band is coupled by the coupling slots in the patch of the 900 MHz band and by means of an additional coupling patch between the radiating patch for the 900 MHz and the reflector.

When realizing multiband antennas, the multiresonant elements are only one issue, and the other issue is the realization of the multiband arrays. As described in the section above, the array performance depends strongly on the distance of the antenna elements, in relation to the wavelength. This leads to the fact that the optimum distance for a low frequency band is different for the optimum distance for a higher frequency band.

DOI 10.1007/978-981-4560-75-7_88-1

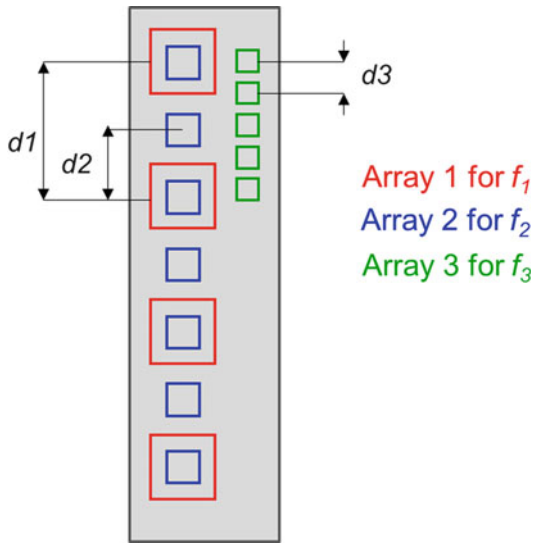


Fig. 31 Multiband sector array, configuration of elements to optimize the array distances

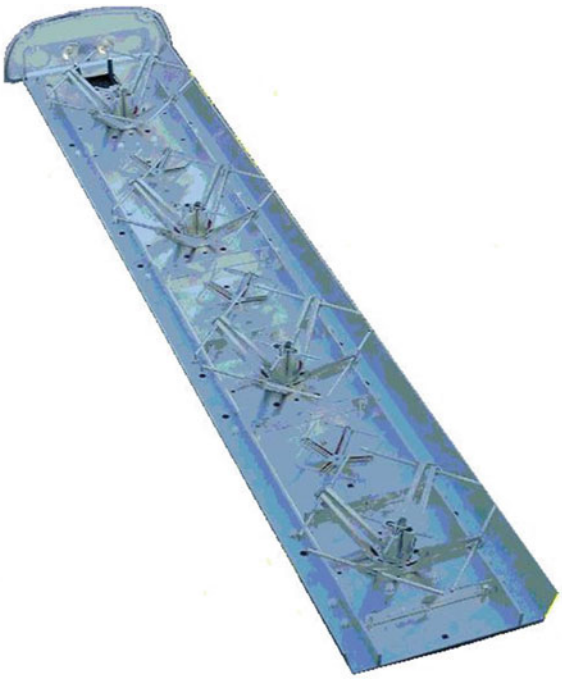


Fig. 32 Dual-band array arrangement

Figure 31 illustrates this issue: assuming that f1 = 900 Mhz, f2 = 1800 MHz, and f3 = 3 GHz, the resulting optimum distances for the arrays within the multiband antenna systems differ significantly. As f2 = f1 × 2, the distance d2 is exactly half of d1. So in this case, array 1 and array 2 can be interleaved as shown in the sketch. Array 3 however does not "fit" into the array grid of d1 and d2, so d3 has to be realized as a separate array.

In Fig. 32, a practical realization of a dual-band sector antenna is displayed (Göttl et al. 1999). One can see the grid of the low-band array and the smaller additional elements of the high-band array in between.

DOI 10.1007/978-981-4560-75-7_88-1

When realizing multiband arrays, not only the issue of different optimum distances has to be solved and integrated but also different tilt angles for the different arrays should be possible, so that the optimum illumination for each frequency band can be realized.

This can be done by using separate phase shifters for each band; however, in this case, the different frequency bands have to be separated by filtering in front of the phase shifters and combined again afterwards (see also section "Antenna Line Devices").

Antenna-Related Network Aspects

The antennas on both sides of the transmission path determine the propagation channel from the base station to the mobile and vice versa. Considering only linear media, the propagation channel is fully reciprocal. Just here the problem occurs that the mobile user transmits at lower-transmitting power compared to the power of the base station, due to the limited battery capacity of the device and due to human exposure limits. This discrepancy is solved by diversity principles. The reuse of resources as time slots or frequencies requires a carefully adjusted beam of the antennas to cover only the desired cell area. The more the coverage of the different sectors overlaps, the higher is the undesired interference of the neighboring sectors, and the reuse of the resources becomes worse. In the networks of the second generation, spatial frequency reuse scenarios of five or seven are typical. So the same frequency carriers can be reused in every seventh sector, and this means the available spectrum has been organized in seven different channel clusters. Networks of the third and fourth generation have a reuse factor of one. Nevertheless, here the different resources of frequency channels, time slots, or resource blocks in the orthogonal frequency-division multiplexing (OFDM) scheme have to be organized in a dynamic manner. So from a statistical point of view, here also neighboring sectors cannot use the same resources for different users, if the level of the undesired channel is still too high (intra-channel interference). On the other hand, the orthogonal channels at other frequencies or orthogonal codes generate an equivalent noise for the desired channel (inter-channel interference). This gives the desired channel to interference ratio, which determines the maximum possible modulation scheme with the maximum data transmission according to the theorem of Shannon–Hartley (Shannon 1949):

$$C = B * \log 2\left(1 + \left(\frac{S}{N}\right)\right)$$

where C is the capacity of the channel in Bit/s, B is the bandwidth in Hz and S/N is the signal to noise ratio. Therefore, the careful design of the radiation characteristics of the antenna and the alignment of the antenna into the desired sector are essential. Especially the adjustment of the mechanical or electrical tilt of the beam influences the C/I in the network. Current technology uses an adjustable tilt mechanism of the antennas, the so-called remote electrical tilt (RET). Here beamsteering of the vertical characteristic of the array is realized by integrated phase shifters, which are driven by motors. Electronically operated phase shifters cannot typically be employed, since most of the systems are sensitive to passive intermodulation and need an extremely high linearity in the components used (Fig. 33).

Influence on Network Planning

Frequency Reuse, Cell Size, and Interference

The mobile communication network is structured into different cells and sectors as mentioned earlier. Typically, every sector uses other resources; thus, an antenna will only serve the desired cell and will not

DOI 10.1007/978-981-4560-75-7_88-1

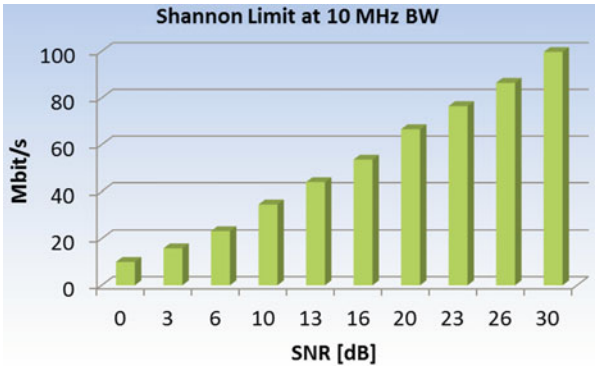


Fig. 33 Maximum theoretical data rate within a 10 MHz bandwidth according to the Shannon theorem

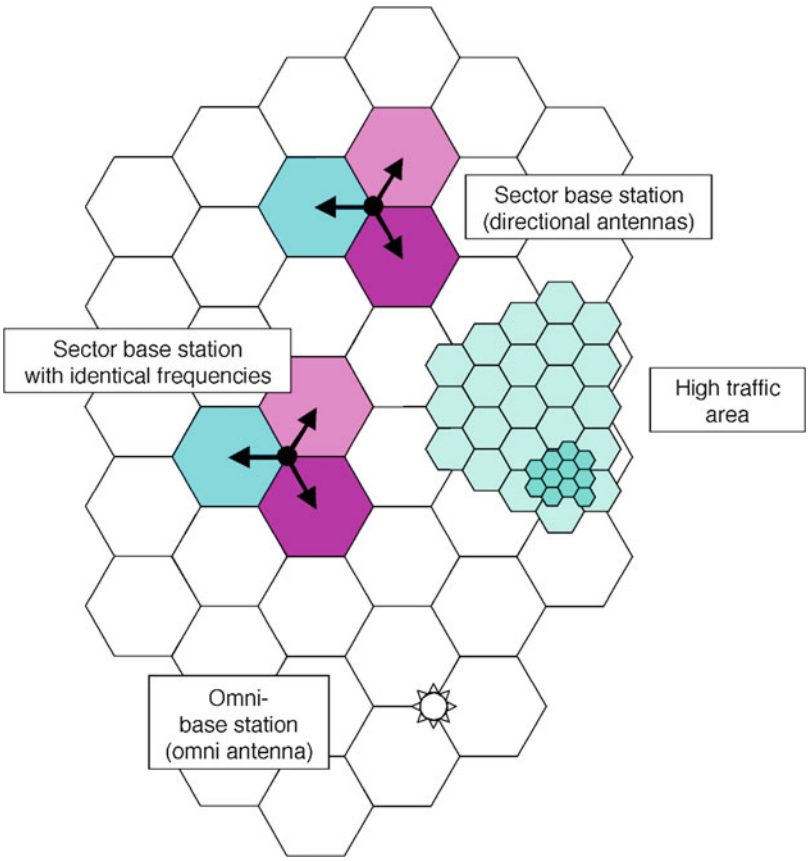


Fig. 34 Different cell topologies

transmit into other cells. This is realized by an optimized design of the horizontal and the vertical shape of the beam. Especially the suppression of upper side lobes and the enhanced front-to-back ratio are essential requirements (Fig. 34).

Thus, the vertical and horizontal beamwidth are important parameters of the antenna as well as the vertical electrical and mechanical tilt. Typically, 65° up to 90° horizontal half-power beamwidth antennas are used for sectored arrangements with three sectors. The theoretical shape of the beam appears in a first approach as 120°; nevertheless, here the overlapping area of neighboring sectors becomes too large, which results in a high interference level. The same adjustment functionality for the interference and

DOI 10.1007/978-981-4560-75-7_88-1

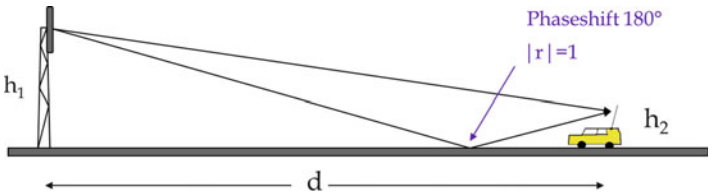


Fig. 35 Two-way model with reflective ground

coverage shows the tilt level of the main beam. Typical values between 2° and 10° can be seen; the smaller the cell, the higher is the tilt level.

Of course, in reality the coverage is not regularly shaped. Therefore, the mobile receives in most locations a mix of different carriers from different cells; the selection of the connected cell will be organized by a handover mechanism. The spreading of pilot carries over different cells causes a lot of handovers and reduces the effective traffic and quality of service.

Path Loss, Link Budget, and Fading Channel

The calculation of the link budget of a channel between base station antenna and mobile depends on the scenario as described in section "Antenna Design for Base Stations." In most cases, no line-of-site connection (LOS) exists, and reflections and scattering lead to a fading of the connecting path between base station and the mobile. The calculation for an LOS connection can be derived from the known gain of the antenna in the main beam direction according to the Friis equation given in section "Antenna Design for Base Stations." The input power P_{in} will be transformed into a radiated transmitting power P_{rad} with certain efficiency of the antenna. The radiated power is distributed over the angle space according to the radiation characteristic of the antenna. In the main beam direction, the maximum power density at a certain distance r can be found to be equal to the product of radiated total power multiplied by the linear gain Gi and divided by the surface of the surrounding sphere.

As already mentioned, the path loss in most of the cases has to include non-line of sight effects as reflections and scattering. This results in a stronger degradation of the received power compared to line-of-sight connections. LOS path loss is proportional to the square of the distance, assuming NLOS connections; the path loss is more proportional to the third up to the fourth order of the distance as described earlier. Derived from measurements, empirical models can be used according to Okumura Hata (Okumura et al. 1968; Hata 1980) where more detailed arrangements can be evaluated, depending on the height of the transmitting and receiving antennas, the distance, and the type of environment (dense urban area or rural area). A simplified model which also yields to a dependency of the receiving level proportional to r^4 is the two-way model, which only takes into account the ground reflection on a flat ground surface (Fig. 35). This gives the relatively simple formula for $d \gg h$:

$$\frac{P_E}{P_S} = G_S G_E \left(\frac{h_1 h_2}{d^2}\right)^2$$

Where P_E and P_S are the the received and transmitted power, G_S and G_E are the gain of the antennas. A typical level dependency of a simulated level according to an LOS path with a dependency of r^2, a simulation with a dependency on distance of r^4, and an equivalent measurement are shown in Fig. 36. The height of the base station antenna is 30 m, and the relevant level is measured or simulated at a height of 2 m.

It can be seen that there is a projection of the radiation characteristic of the antenna on the ground level, but also the nulls of the pattern are filled by reflections and scattering. A comprehensive overview about the used models is given in Walke (2000).

DOI 10.1007/978-981-4560-75-7_88-1

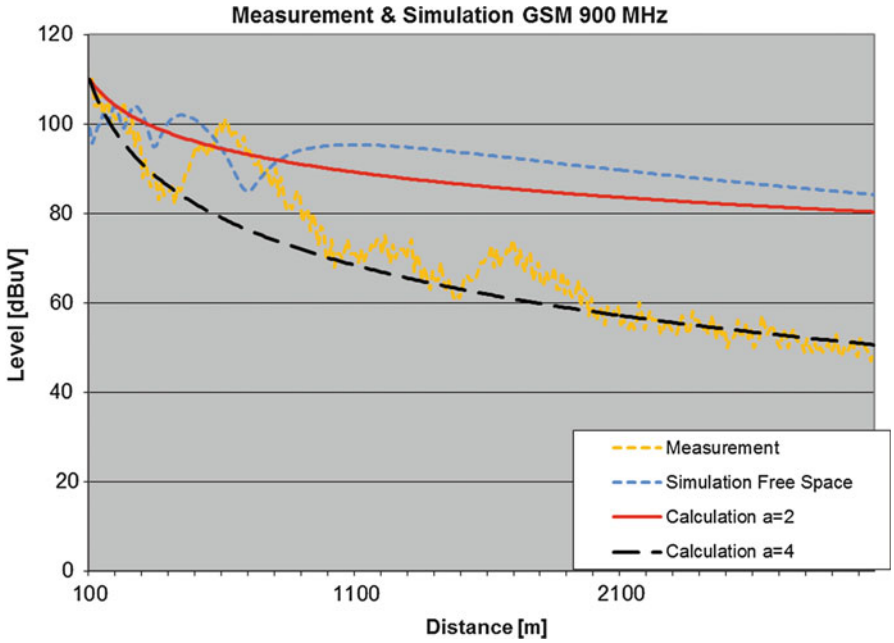


Fig. 36 Measurement and simulation of the signal level with different models

It is obvious that the superposition of more than one path as a direct and reflected wave is strongly dependent on the position and on the environment and also time variant. This causes fading effects and delay spread. The fast change of the phase between different incoming waves results in a fast-fading effect. The fading can happen within the range of some meters or, depending on the location, in a range of some centimeters in the considered frequency bands. The average level of the path is subject to slow fading, which is caused by changing major path properties, e.g., shadowing by buildings or trees. Since the fast fading depends on the phase of incident waves, the superposition will occur in another manner at different frequencies.

The behavior of the path will be different if the transmission frequency changes. The bandwidth, within which the behavior of the path is nearly unchanged or in other words where it is correlated, is called the coherence bandwidth; see Maurer (2005). Typically, the coherence bandwidth is in the range of 20–50 MHz at 900–2000 MHz. Using several frequencies for transmission with a frequency variance higher than the coherence bandwidth leads to a significant reduction of fading. This technique does not only allow high data rates through the large bandwidth used; it also improves the stability of the connection, and it is used in frequency hopping transmission (GSM) as well as in carrier aggregation transmission (LTE) (Fig. 37).

The power level distribution of the signal over time is dependent on the number of incident waves and especially on the strength of the LOS path. Considering a transmission of a short pulse than the incoming signal consists of a sum of delayed pulses, every pulse with a certain power P_i, having a delay τ_i as it is shown in Fig. 38:

$$\tau_{rms} = \sqrt{\frac{1}{\sum_{i=1}^{n} P_i} * \sum_{i=1}^{n} \left(\tau_i^2 P_i\right) - \tau_d^2}$$

DOI 10.1007/978-981-4560-75-7_88-1

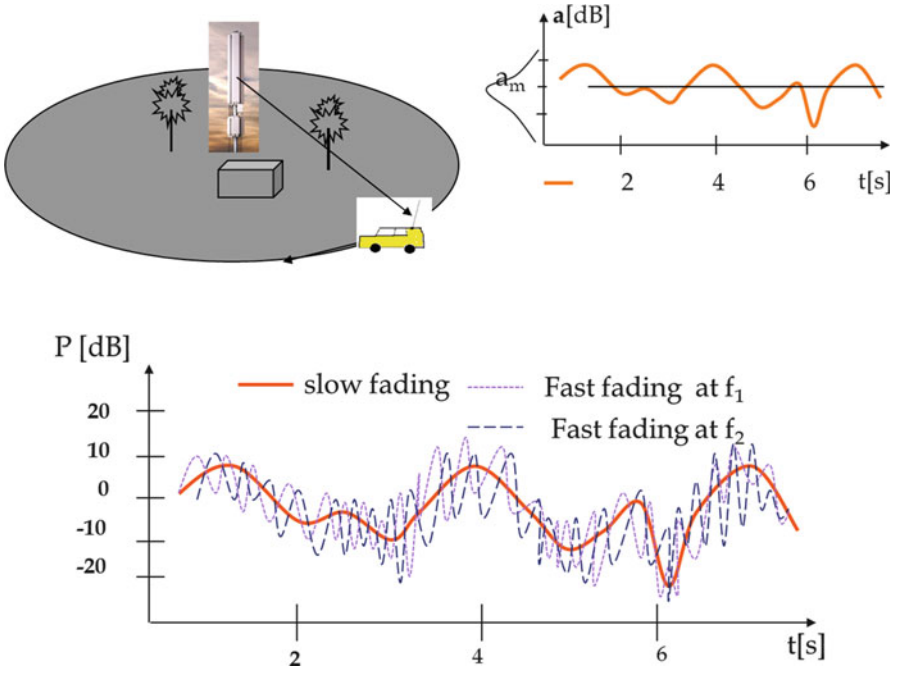


Fig. 37 Slow and fast-fading principle

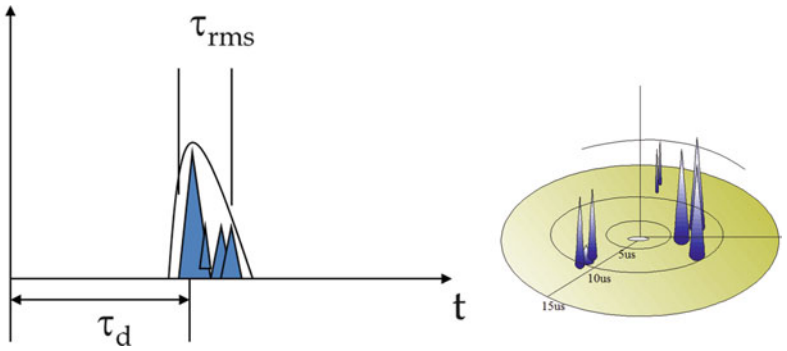


Fig. 38 Delay spread with space dependency

$$\tau_d = \frac{\sum_{i=1}^{n} \tau_i P_i}{\sum_{i=1}^{n} P_i}$$

The delay spread and the average total delay are given by the above expression, whereas the average delay spread is the basis for the coherence bandwidth:

$$B_C = \frac{1}{\tau_{rms}}$$

The given description of the channel properties is important regarding the signal processing implementation used for the matched filtering to restore the original signal: the selection of the proper signaling

DOI 10.1007/978-981-4560-75-7_88-1

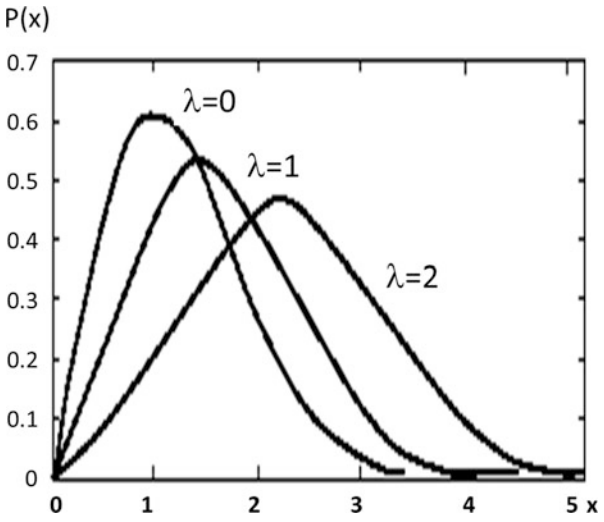


Fig. 39 Rice and Rayleigh distribution

schemes and operation bandwidths is determining significantly the effort in the processing. So the frame length should be long compared to the average delay spread; otherwise, the intersymbol interference is very high. On the other hand, having enough signal processing power available, the multipath propagation can directly be used for transmitting. This might be more effective than having one data stream within a MIMO implementation, as will be described later.

If no LOS path exists and the number of incident waves at a given location is very high, then the resulting signal level distribution is given by a Rayleigh distribution. If on the other hand several waves with an LOS connection exist, the distribution will change to a Rice distribution:

$$\text{Rayleigh distribution} \quad p(x) = \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}$$

$$\text{Rice distribution} \quad p(x) = \frac{x}{\sigma^2} e^{-\frac{x^2+x_s^2}{2\sigma^2}} * I_0\left(\frac{x x_s}{\sigma^2}\right)$$

In these formulas, Io is the Bessel function of first kind, zero order σ is the average variance, x is the normalized magnitude level, and x_s relates to the average LOS part of the magnitude distribution; $\lambda = x_s/\sigma$ is also called the Rice factor. Setting the Rice factor to zero, the Rice distribution is changing to the Rayleigh distribution (Fig. 39).

Diversity

In order to overcome fading effects of the channel and to improve the transmission, diversity arrangements are used. This effectively means that several propagation channels are active. To achieve a certain gain within a rice scenario, the correlation between the different paths should be low, as it can be derived from Fig. 40. Typically, this requires different located antennas or different polarization of the antennas. Space diversity requires a certain distance between the two receiving antennas or antenna elements, typically a minimum of ten wavelengths; see Fig. 42. Polarization diversity, which is the preferred technology today, uses 2 orthogonal polarizations, e.g., ±45° or the combination of horizontal and vertical polarization. The ±45° polarization or the so-called X-polarization shows a better symmetry for transmission than the horizontal–vertical combination.

DOI 10.1007/978-981-4560-75-7_88-1

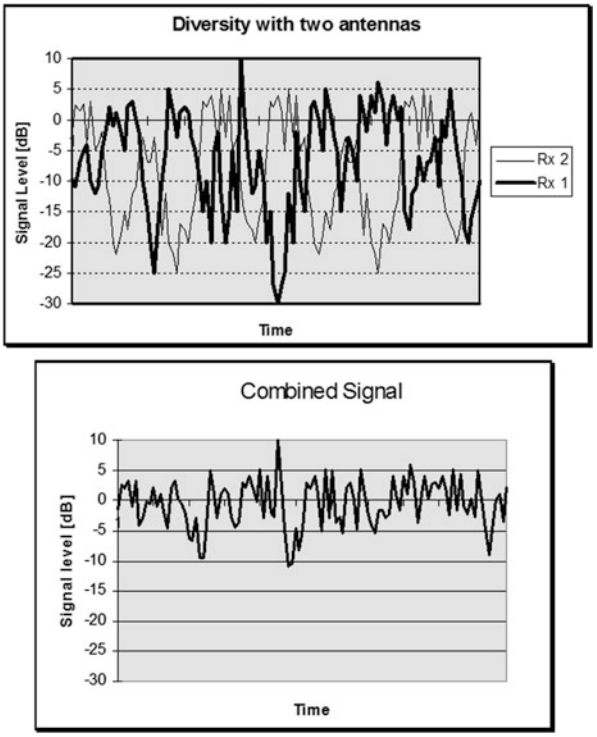


Fig. 40 Combining of fading signals using the diversity principle

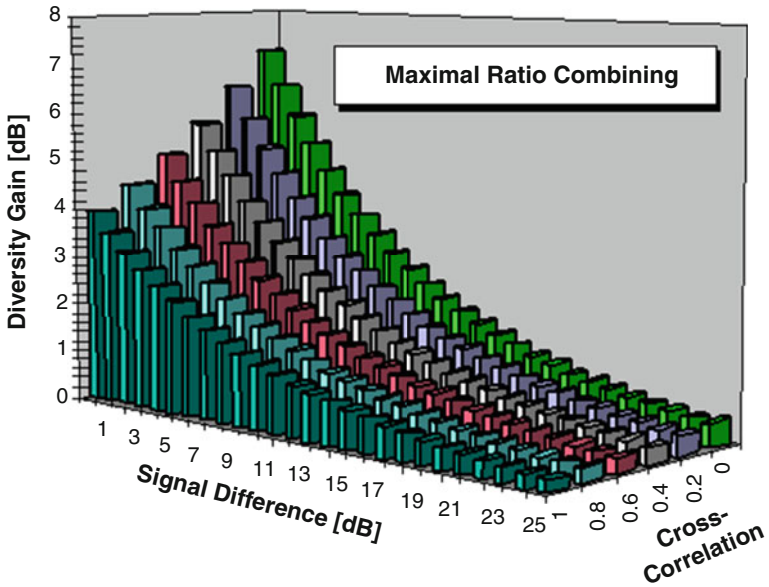


Fig. 41 Dependency of diversity gain on correlation and level difference for a typical GSM system

Within the X-arrangement, the antennas consist of two resulting systems of orthogonally polarized arrays with separate connectors. The challenge is to optimize the isolation of these systems. The typical requirement is a minimum of 30 dB isolation. The diversity gain depends on the level difference of both systems and the correlation between the two levels of the signals over time. In Fig. 41, the dependency of a typical GSM system on both parameters is shown. Besides correlation and power level, the diversity gain

DOI 10.1007/978-981-4560-75-7_88-1

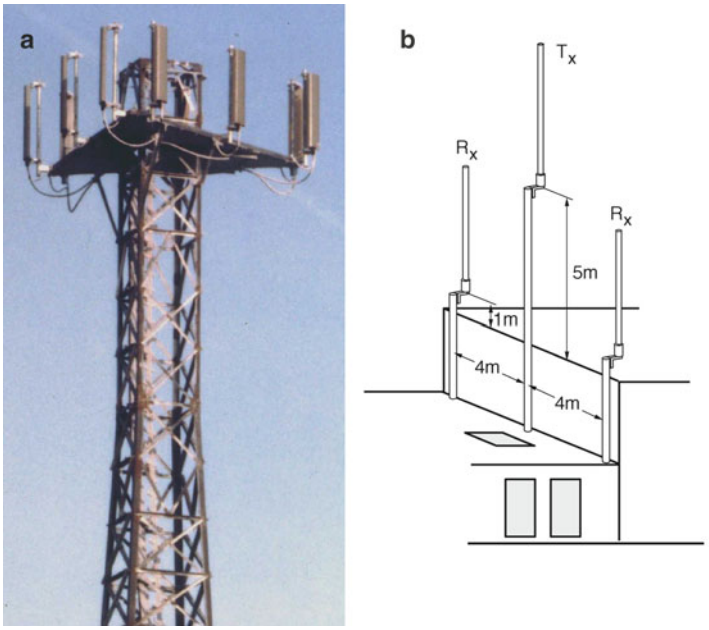


Fig. 42 Space diversity arrangements in a sectored configuration (**a**) and in an omnidirectional configuration (**b**) for 900 MHz

also depends on the system parameters, e.g., the acceptable probability of interruptions or failures. The higher the requirements of availability of the channel, the higher the diversity gain will be. This is obvious since the diversity principle avoids interruptions through fading and will improve signal quality, especially the interruptions of time frames (Fig. 42).

In addition to the gain due to the reduction of failure probability, the diversity also helps to overcome the problem of imbalance between the transmitting power of the mobile and the transmitting power of the base station.

MIMO: Operation

As an extension of diversity, multiple antennas can be operated with more than one signal stream in a MIMO operation (see also section "Historical Review").

The theoretical limits of transmission of multiple data streams are considered in detail in many publications. In Telatar (1995), a theoretical derivation of the maximum capacity of Gaussian channels with Rayleigh fading are considered; in Pontes et al. (2011), implementation-specific scenarios are evaluated and limited knowledge of the channel matrix is considered. A comprehensive work on MIMO systems can be found in Waldschmidt (2004). Here, it should be noted that the multichannel antennas can be used for diversity or MIMO or beamforming operation alternatively, depending on the current situation in the propagation channel. So the beamforming operation can be considered as a special case of MIMO with only one data stream. The capacity of MIMO can be evaluated by the modified Shannon equation with knowledge of the channel matrix H:

$$C = \Delta f \cdot \sum_{k=1}^{n} \log_2 \left(1 + \lambda_k \frac{P_k}{N}\right)$$

where λ_k is the k th eigenvalue of the channel matrix H, which is proportional to the square of the singular values of H, and P_k is the transmitting power of the kth data stream and N the noise. For simplification, the noise is considered to be equal for every data stream in this case. The maximum MIMO capacity can

DOI 10.1007/978-981-4560-75-7_88-1

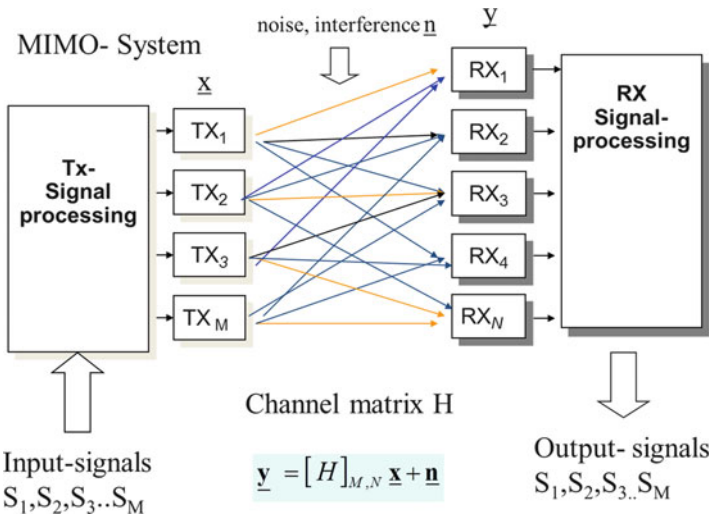


Fig. 43 Principle of an M × N MIMO system

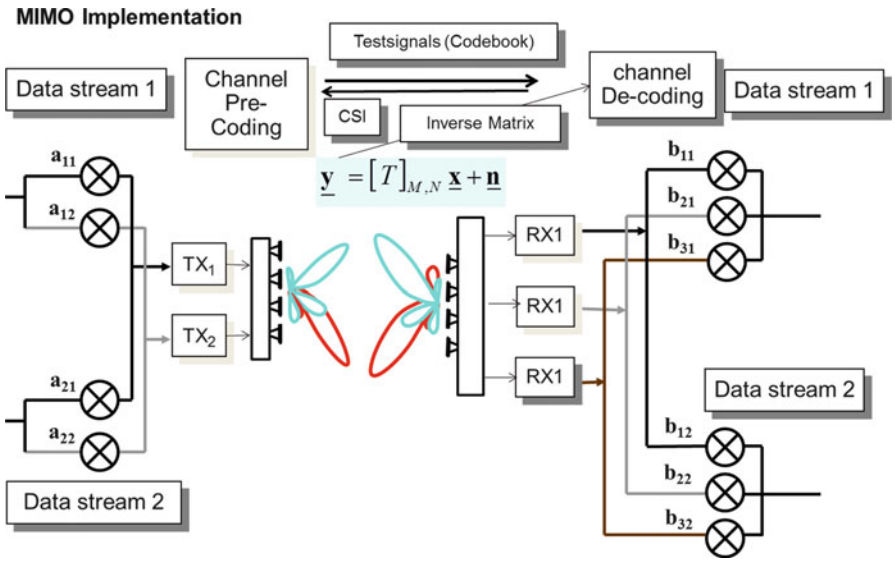


Fig. 44 Schematical 2×3 MIMO implementation

consequently only be achieved having the full knowledge of the channel matrix. The implementation of such adaptive MIMO operation is known as the "waterfill principle"; see also Waldschmidt (2004, pp. 1–17). The maximum capacity therefore depends on the signal-to-noise ratios of the different resulting channels for the connected data streams and cannot exceed the n-times capacity of a single-in-single-out (SISO) channel.

In Fig. 43, the signal model and the basic principle are shown, and in Fig. 44 the more detailed implementation is shown. The challenge is the determination of the channel matrix. This can be carried out by a closed-loop principle; in this operation, the data streams are processed with certain agreed coefficients (training sequences) according to an agreed code book (see also chapter "► Active Antennas"). The best fitting code is selected by evaluation of the best transmission case.

This approach is equivalent to an approximate solution of the channel matrix H inversion and the determination of the eigenvalues by a selection process of limited solutions. Of course, due to the

DOI 10.1007/978-981-4560-75-7_88-1

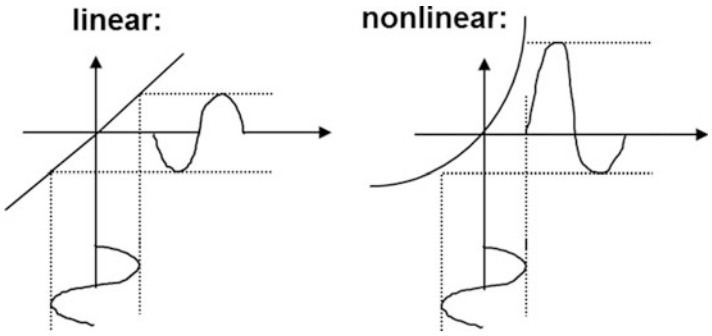


Fig. 45 Linear and nonlinear transfer function and signal

limitation of granularity, the maximum capacity according to the "waterfill principle" cannot be fully achieved.

MIMO operation in the sense of spatial duplex shows good performance if the following provisions are fulfilled: (a) high signal-noise ratio, (b) uncorrelated paths and/or channels, and (c) limited time variance of the channels. The selection of MIMO operation or operation with only one data stream – which is equivalent to a beamforming mode – can be carried out in parallel, while the Rx processing with a rake receiver takes place. Thus, the operation can be switched during a running connection to the best fitting mode. If the signal-to-noise ratio becomes too small, the beamforming operation typically shows better results than the MIMO mode. This is especially true for dynamic modulation schemes: the beamforming gain also increases the signal-to-noise ratio and leads to a higher modulation efficiency.

Passive Intermodulation

Intermodulation is a well-known issue in active component design like amplifiers or mixers. Nevertheless, also in passive components intermodulation becomes relevant if these devices operate within receiving and transmitting signals at the same time. Then the difference of the power levels can be more than 150 dB, which requires extreme linearity of the components in the signal path. If a CW signal is transmitted using a nonlinear device, the sinusoidal function will be changed into a new time-dependent signal, which can be expressed as a series of new sinusoidal functions with higher frequencies.

The mathematical derivation of the intermodulation level is given as follows (Fig. 45):

The output signal $i(t)$ can be written as the response of the two-tone input signal $u(t)$ as

$$u(t) = A \cdot \cos(\omega_1 t) + B \cdot \cos(\omega_2 t)$$
$$i(u) = a_0 + a_1 u + a_2 u^2 + a_3 u^3 + \ldots a_N u^n$$

The third-order term is then

$$a_3 u^3 = a_3 (A^3 \cos^3(\omega_1 t) + 3A^2 B \cdot \cos^2(\omega_1 t) \cdot \cos(\omega_2 t) + \ldots$$
$$\ldots 3AB^2 \cos(\omega_1 t) \cos^2(\omega_2 t) + B^3 \cos^3(\omega_2 t))$$

Considering the frequency $2\omega_1 + \omega_2$,

we find

$$1.5A^2 B\{\cos \omega_2 t + 0.5[\cos(2\omega_1 + \omega_2)t + \cos(2\omega_1 - \omega_2)t]\}.$$

Expressed in logarithmical scale, this is evident to

DOI 10.1007/978-981-4560-75-7_88-1

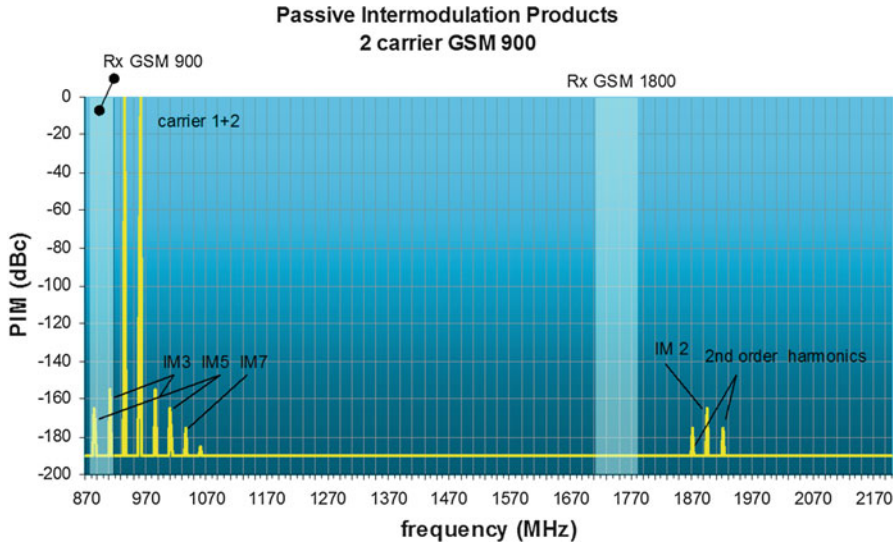


Fig. 46 Two-tone carrier intermodulation products

$$IM_3(2\omega_1 - \omega_2) \sim \log\left(3a_3A^2B/4\right) = \log(3a/4) + 2\log A + \log B$$

$$IM_3(2\omega_2 - \omega_1) \sim \log\left(3a_3AB^2/4\right) = \log(3a/4) + \log A + 2\log B$$

For equal signal levels for both frequencies, this results in

$$IM_3(2\omega_1 - \omega_2) \sim \log\left(3a_3A^3/4\right) = \log(3a/4) + 3\log A$$

For equal signal levels of the transmitting carriers at f1 and f2, it can be seen that the intermodulation level will rise three times faster compared to the input level in logarithmic scale in decibel. This means that if the input level of both transmitting carriers increases by 3 dB, the intermodulation level of the third order will increase by 9 dB!

In Fig. 46, the intermodulation carriers for a two-tone system is shown. The new spectrum of the intermodulation carriers consists of odd and even order spectrum, as the odd order spectrum of the third, fifth, and seventh order can fall into the receiving band of the relevant equipment. So the odd order intermodulation is the most relevant for the distortion of the receiving signals.

The situation becomes more complex when dual-band antennas are considered. Here, the second-order harmonics or second-order intermodulation products can potentially be relevant in the receiving band. This is shown in Fig. 47 for a dual-band antenna. Nevertheless, measurements have shown that the second-order intermodulation products in passive equipment are typically significantly lower compared to the third-order products. A general dependency of the levels of the different intermodulation products is shown in Fig. 48. The noise limited sensitivity of a receiver with a noise figure of 3 dB in 1 MHz bandwidth can be evaluated:

$$\begin{aligned} Noise &= NF \cdot kT\Delta f = NF \cdot 1.38 \times 10^{-23}\text{J/K} \cdot 290\ \text{K} \cdot 1\ \text{MHz} \\ &= NF \cdot 400 \times 10^{-23}\,Ws \cdot \text{MHz} \\ &= NF + (-174\ \text{dBm/Hz} + 60\ \text{dB}) \\ &= -111\ \text{dBm} \end{aligned}$$

DOI 10.1007/978-981-4560-75-7_88-1

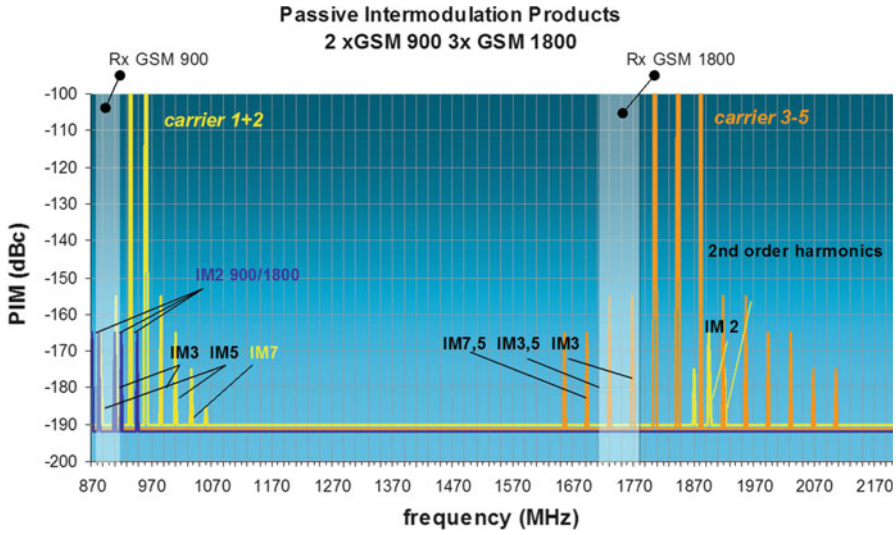


Fig. 47 Intermodulation products of dual-band equipment for 900 and 1800 MHz bands

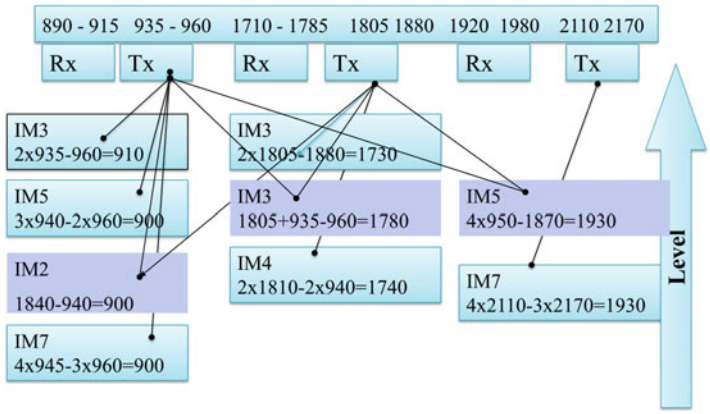


Fig. 48 Schematical level of different intermodulation products

This very hard intermodulation requirements result from the parallel operation for transmitting with typically 20–50 W (which is equivalent to 43–47 dBm) and the receiving sensitivity near to the noise floor with approximately −110 dBm, so the typical level differences are around 150–170 dB.

For an operation bandwidth of 200 kHz within a GSM system, the thermal noise level is −121 dBm (!) and the transmitting level is +43 dBm, which results in a theoretical requirement of −164 dBc. In praxis, the noise level is higher, due to feeder cable losses and jumpers. Assuming therefore that the typical noise figures of the system are approximately 10 dB and there is a higher system bandwidth, the antennas are typically specified with –150 dBc, which is good enough for most applications. Nevertheless, if the intermodulation level exceeds −140 to −130 dBc, this will cause a significant degradation in the sensitivity of the system.

The level increases from one odd order to the next by approximately 8–12 dB; this means if a third-order level of −140 dBc is measured, a fifth-order level of approximately –150 dBc can be expected. The difference depends of course on the special kind of transfer function; nevertheless, these values are typical measured figures. Passive intermodulation is one of the biggest challenges for the production and development practice of base station antennas and related equipment. Careful design of all contacts and

DOI 10.1007/978-981-4560-75-7_88-1

selection of the proper materials are essential to achieve highly linear behavior of the signal path. The sources of the passive intermodulation can be divided into two classes: material-based sources and contact-based sources. The use of the wrong material can include the use of steel or ferromagnetic material, pollution in plastic spacers and surfaces, and the use of nonlinear conductive material. So often, silver-plated surfaces, tin-plated surfaces, copper, and aluminum are used to achieve good intermodulation behavior. However, the use of aluminum is often sensitive with regard to any contacts, as the surface of aluminum oxidizes very fast in contact with air and humidity.

Antenna Line Devices

Antennas are strongly related to the electronic transmitting and receiving equipment. Currently, there is a clear tendency to integrate the electronics into the antenna or to put the electronic equipment, such as amplifiers, close to the antenna. Special control equipment for the steering of the tilt of the antenna, the so-called remote electrical tilt (RET), is commonly used. In state-of-the-art communication systems, the properties of an antenna system cannot be considered for itself anymore, as the parameters of the antenna array influence the system behind it and vice versa. Thus, it is of very high importance for the design of an antenna to basically understand all features and requirements of the system in which the antenna is operating.

Remote Electrical Tilt (RET)

For about 15 years, the remote control of the electrical downtilt has been commonly used for base station antennas. The success of this technology was driven by a worldwide standardization of the control interface in the Antenna Interface Standards Group (AISG 2006). Later, this standard was implemented and extended in certain parameters in the 3rd Generation Partnership Project (3GPP 2006). Whereas in the 3GPP only the RET functionality is standardized, the AISG also supports the interface to other ALD equipment, such as TMAs, sensors for location, and much more. The advantage of the AISG is also that the implementation of specific parameters, such as connectors or supply voltages, is standardized.

The steering of the tilt is provided by passive phase shifters inside the antenna. These phase shifters control the phase of the radiators in such a way that the resulting tilt of the antenna points to the desired direction. The principle for an antenna consisting of five radiators is shown in Fig. 49. It can be seen that for the outer radiators, the phase shift is twice as big as for the inner radiators, while the middle radiator will remain at a constant phase. Due to intermodulation requirements, no electronically operated phase shifters are used, but only passive phase-shifting devices. The principle construction of an example of such passive phase shifters can be seen in Fig. 50.

As the picture shows, the phase shifter is constructed using a microstrip or triplate structure, typically with a nominal impedance of 50 Ω. If the impedance at the end of the arcs is well matched to 50 Ω, then the impedance at the connection of the lever and the arc will remain constant at 25 Ω. The lever, together with the capacities at the connection points at the arc and the rotation center, realizes the transformation as a kind of quarter-wave transformer to achieve again the 50 Ω matching at the input of the phase shifter. In a first development, a combination of two 3-way phase shifters was used. Here, a gear for the rotation speed of the first phase shifter for the outer radiators and for the inner radiators has to be used. A schematic realization is shown in Fig. 51.

In the last few years, new kinds of phase shifters have been developed. In a first step, the 3-way phase shifter was extended to an n-way phase shifter, consisting of more than one arc (Göttl et al. 1999). An example is shown in Fig. 52. This phase shifter realizes, in addition to the phase shift in the right relation

DOI 10.1007/978-981-4560-75-7_88-1

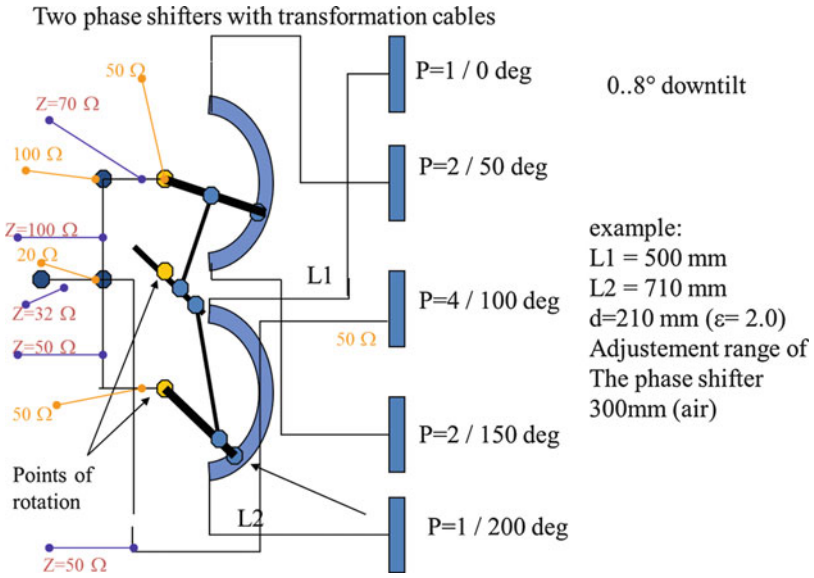


Fig. 49 Phase shifter arrangement for electrical tilt

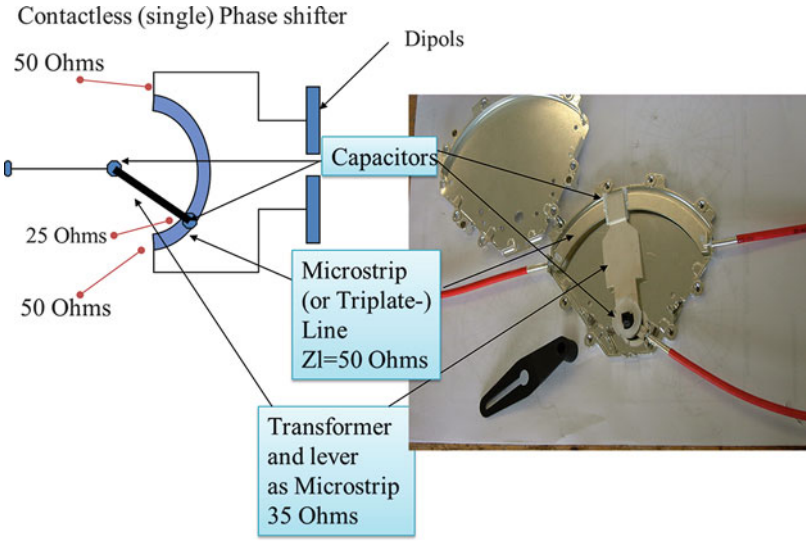


Fig. 50 Principle of contactless phase shifter for three radiators

for the inner and outer radiators, a matched power distribution for the radiators. This can be adjusted by the shape of the lever, the coupling sections to the arc, and the distance of the spacers.

In the meantime, phase shifters are now used with 3, 4, and 5 arcs which can serve up to 11 radiators, working under the same principle.

To drive these phase shifters, motors with control electronics are used. These motors can be integrated into the antenna; however, at the beginning of the use of RET systems, the motors were mounted as upgrade equipment at the bottom of the antenna. Such an upgrading device from the KATHREIN company is shown in Fig. 53. This device is compliant to the AISG specifications, see AISG (2006).

The RET devices have to fulfill a number of requirements: lifetime and failure probability, which is expressed in MTBF values; number of driving cycles; lightning protection; and the operating temperature

DOI 10.1007/978-981-4560-75-7_88-1

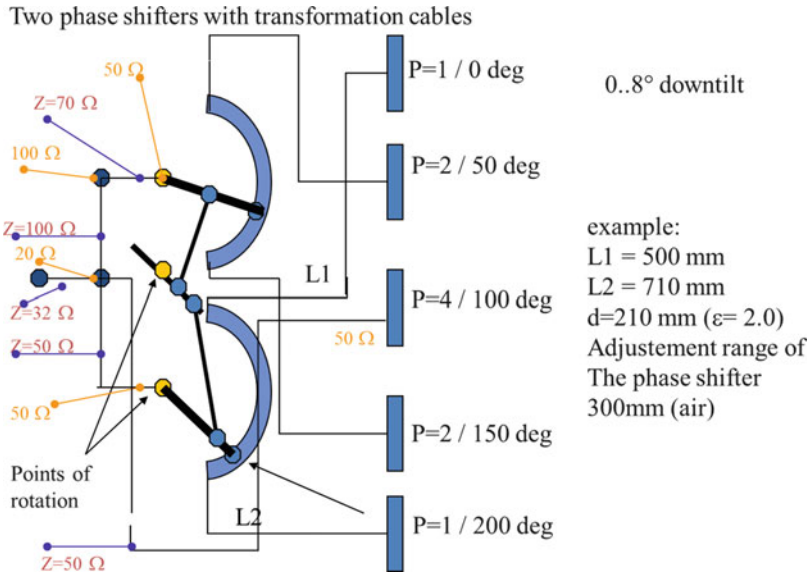


Fig. 51 Phase shifter arrangement with two 3-way phase shifters

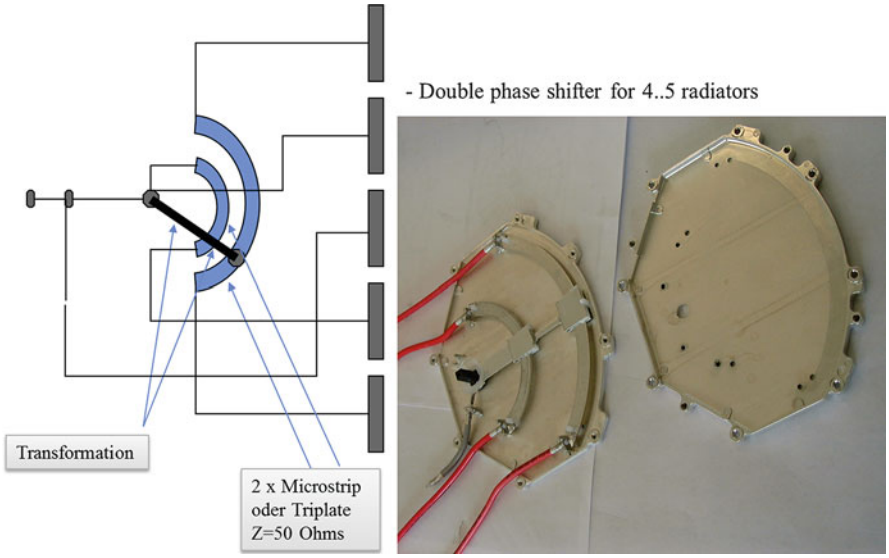


Fig. 52 Double phase shifter design

range from −40° to +50° ambient air temperature. The required number of driving cycles is 10,000–50,000. Nevertheless, the RET functionality is often used to adjust the tilt of the sites in the network, based on measured field results or statistical results in the quality figures of the network.

The beamsteering of the antennas can also be extended to the adjustment of the azimuthal direction. The optimization of both tilt and azimuth gives the best results; however, the misaligning in the azimuthal direction seems not to be so significant, and this is why the RET functionality is commonly used, and the azimuthal remote adjustment is only used in special applications.

Two possibilities exist for implementing the control of the AISG interface: one is the use of separate cables for the power supply and control signal and the other is to use the existing feeder cables for the DC power supply and the control signal. A principal arrangement of such control is shown in Fig. 55. In this

DOI 10.1007/978-981-4560-75-7_88-1

Fig. 53 Device for upgrading remote electrical tilt (RET)

case, smart bias tees are used, which transform the signal onto the twisted pair cable. The RS 485 interface is added to the pilot signal with on/off modulation scheme on the feeder cables (AISG 2006; 3GPP 2006).

Tower-Mounted Amplifiers (TMAs)

Tower-mounted amplifiers are widely used equipment in base station antenna applications. These amplifiers are implemented into the feeder cable between the base station and the antenna. The device amplifies the receiving signals within the relevant band of operation. Different arrangements are shown in Fig. 56. If the feeders are using Tx and Rx in combination, then the amplifier needs duplex filters to separate the transmitting and receiving signals by means of filters as sketched in Fig. 57.

As typically two separate polarizations for diversity purposes are used, the TMAs are realized as double units, so-called double tower-mounted amplifiers (DTMAs). The challenges here are low insertion loss, combined with the required isolation for the Rx path. To not overdrive the Rx amplifier, the linearity of the Rx low noise amplifier (LNA) has to be designed in a matched manner as the Rx filter stop attenuation in the Tx range has to be sufficient, to reduce the rest of the Tx signal at the Rx amplifier so that any intermodulation products are suppressed. An additional requirement in the Rx range is the stop attenuation in the Tx chain regarding the Rx frequencies, to prevent any ripple in the Rx band by the closed loop of the amplifier and to prevent oscillation in any worst case. So typical isolation requirements of 50–70 dB have to be realized in the input of the Rx amplifier and 40–60 dB in the output of the Rx amplifiers. The insertion loss is in the range of 0.2–0.5 dB, and the noise figure is typically in a range between 1.0–2.0 dB.

The TMAs also had to fulfill passive intermodulation requirements of typically −160 dBc at 2 × 20 W transmitting power. The design of the filter components of the TMAs as well as the other ALD devices depends on the following major requirements: (a) the insertion loss, (b) the isolation, (c) the maximum power, and (d) the linearity. Therefore, so far the power requirements have only been fulfilled with coaxial air-filled cavities, ceramic block cavities, or wave guide types. SAW or FBAR filters are not usable due to

DOI 10.1007/978-981-4560-75-7_88-1

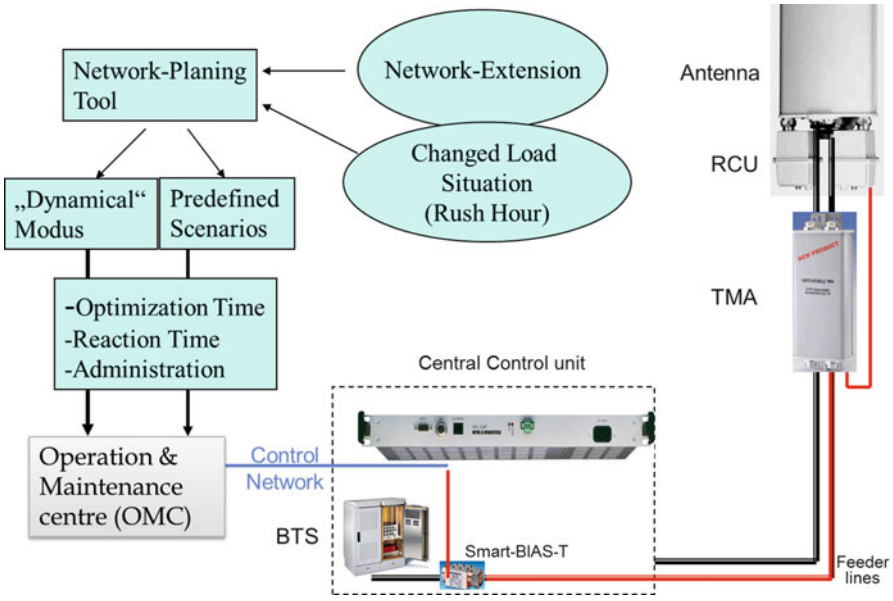


Fig. 55 RET control arrangement for different scenarios

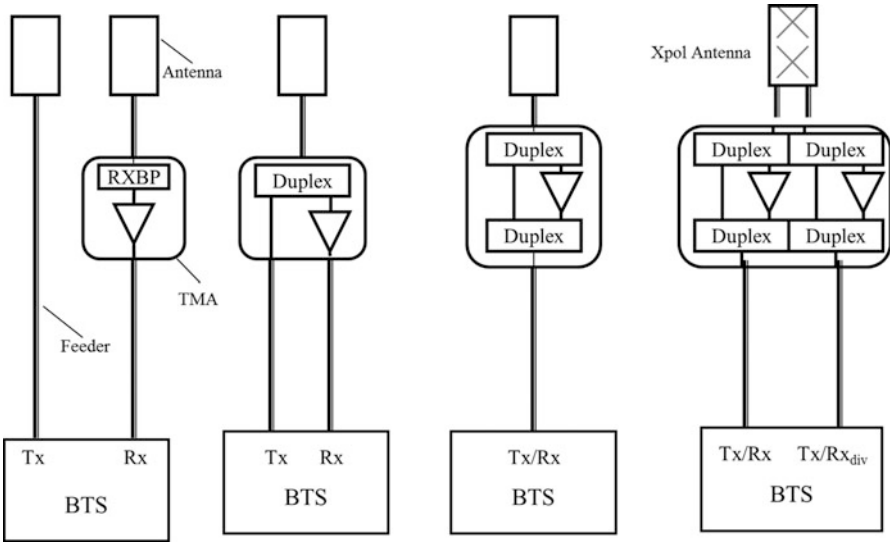


Fig. 56 Different tower-mounted amplifier arrangements

insertion loss, linearity, and maximum power requirements. Monoblock ceramic filters are currently under discussion; nevertheless, the insertion loss is still a little higher compared to air-filled types or block ceramic types (as puck or sphere). A typical filter in a coaxial realization as a part of a TMA is shown in Fig. 58.

The number of cavities that has to be used depends on the required Rx–Tx isolation; the duplex frequency distance, especially the gap between the Rx and the Tx band; and the required insertion loss. In order to fulfill typical requirements of today's TMAs, 6–8 cavities for every part of the Rx and Tx filtering are necessary.

DOI 10.1007/978-981-4560-75-7_88-1

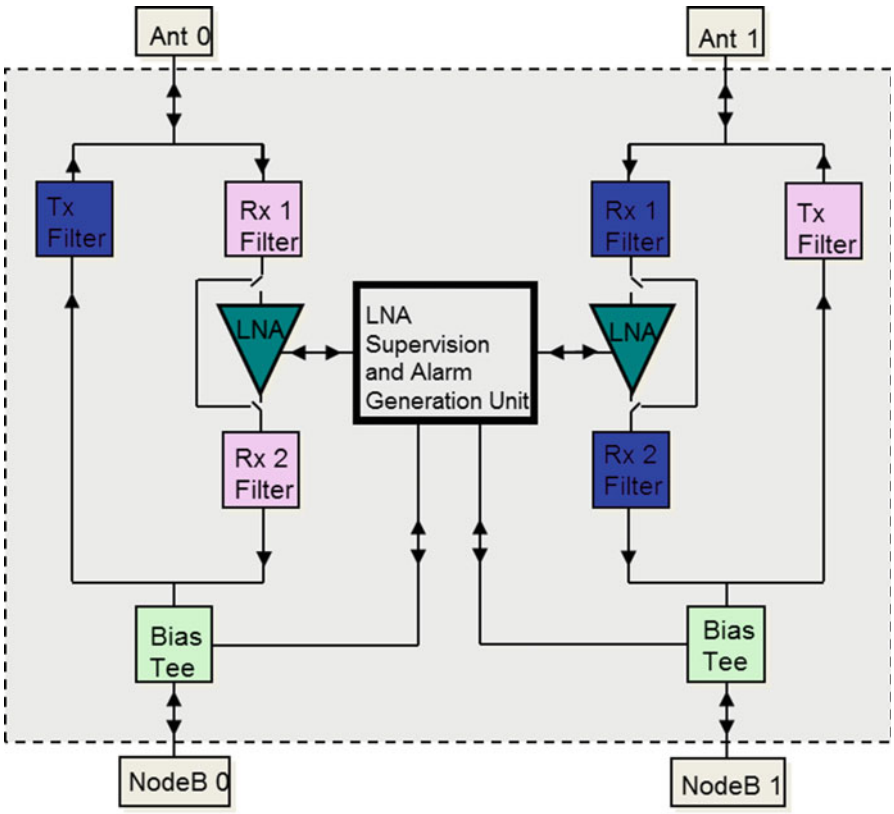


Fig. 57 Principle design of a DTMA

Every cavity of the filter has to be tuned to the right resonance frequencies. Since the manufacturing tolerances do not allow a production of the cavities for just the right resonance frequency, special tuning elements are required for every cavity and also for the coupling between those cavities. In the example shown in Fig. 58, these tuning elements consist of metallic rods, which are moved with a thread in the lid of the cavity. Here, the design of the contact between the rod and the lid is essential regarding the intermodulation parameters.

Filter Combiner

Often, one antenna is used for more than one base station. This can be the case to combine different frequency bands, as well as to combine different operators at the same band or at different bands. Here, frequency selective combiners are used; an example is shown in Fig. 59. Combiners are typically used in the same environment as the antennas; therefore, the same strong parameters are valid regarding the environmental resistance. The combiners have to carry the transmitting and receiving signals and therefore have to fulfill the typical requirements of a low passive intermodulation design. Nevertheless, the isolation can be reduced compared to the isolation of duplex filters. Here, no selection of receiving and transmitting signals is required, but rather the selection of two different frequency bands. Typical isolation requirements should be comparable to the use of two different antennas, as an alternative solution. Thus, 40–60 dB is reasonable for the isolation. For the use of high power up to 200 W, the design is based on standard coaxial resonators as shown in Fig. 58.

DOI 10.1007/978-981-4560-75-7_88-1

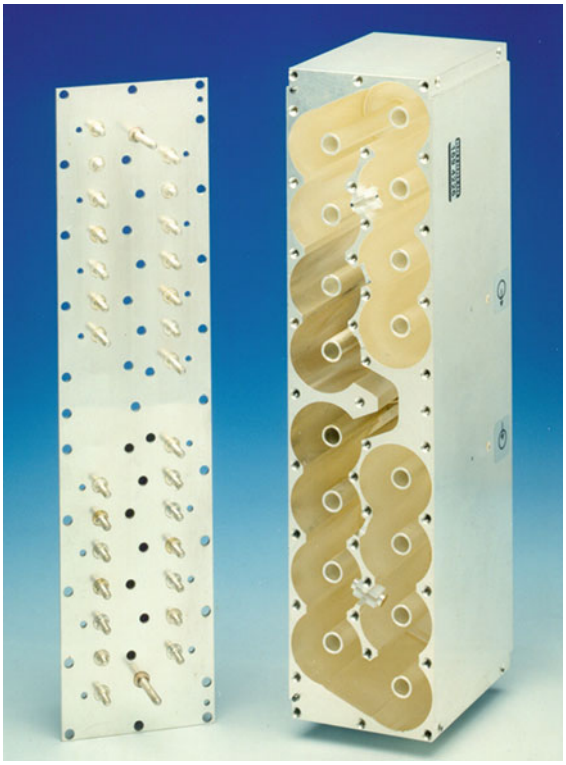


Fig. 58 Coaxial duplex filter with 8 Rx and 8 Tx cavities

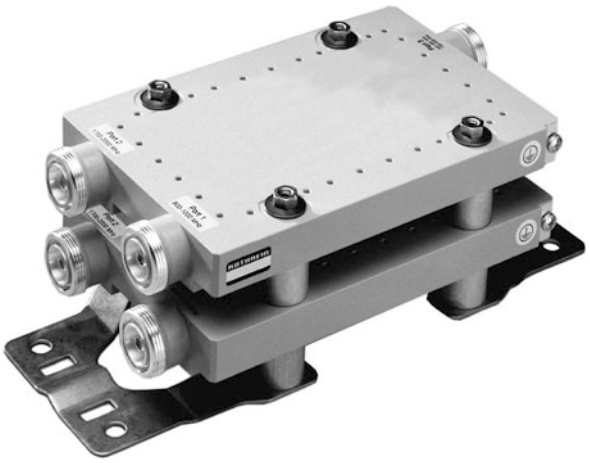


Fig. 59 Dual-band combiner for 800–1000 MHz and 1700–2000 MHz

Active Antennas

Remote Radio Head (RRH)

Remote radio heads are used to place the transmitter and the receiver near to the antenna and therefore to avoid additional losses in the feeder cables. This increases the available transmitting power at the antenna, the efficiency, and the sensitivity of the whole system. Challenges are the required reliability and the environmental conditions. Compared to the base station equipment, which is typically located in

DOI 10.1007/978-981-4560-75-7_88-1

Fig. 60 RRH 1×60 W WCDMA 2100 MHz

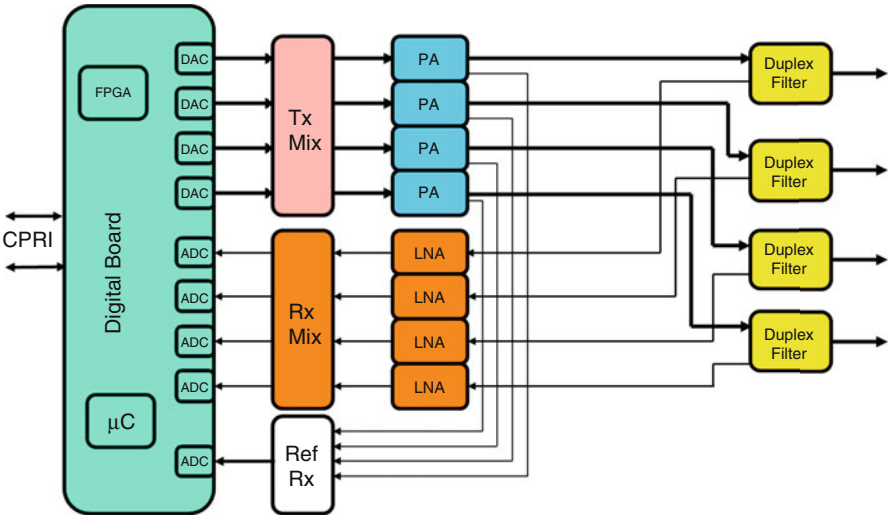


Fig. 61 Principle design of RRH

air-conditioned shelters, the RRHs have to withstand temperature range of −40 °C to 60 °C, vibrations, rain, salt mist, and all other difficult issues within the outdoor base station environment. Due to the limited efficiency, the dissipation power has been dissimilated, and this is one of the biggest design challenges regarding the size of the RRHs (Fig. 60).

In Fig. 61, a principle scheme of an RRH with four Tx and Rx chains is shown. The digital input is typically based on a fiber optical connection. One or two optical lines can be used for the transmitting and receiving signals, depending on the protocol used. Certain protocol standards are used, such as CPRI (Common Public Radio Interface), OBSAI (Open Base Station Architecture Initiative), and ORI (Open Radio Interface). The ORI standard is supported by ETSI (European Telecommunications Standards Institute), whereas the others are widely used industrial standards. The most commonly used standard today is the CPRI interface.

The digital board in Fig. 61 with the microcontroller and FPGA has the following main functions: (a) de-multiplexing of the incoming and multiplexing of the receiving signals into the CPRI data stream, (b) level adjustment and crest factor reduction (CFR) of the transmitting signal, (c) linearization and digital predistortion (DPD), and filtering of the spectrum. The signals then will be transferred into the

DOI 10.1007/978-981-4560-75-7_88-1

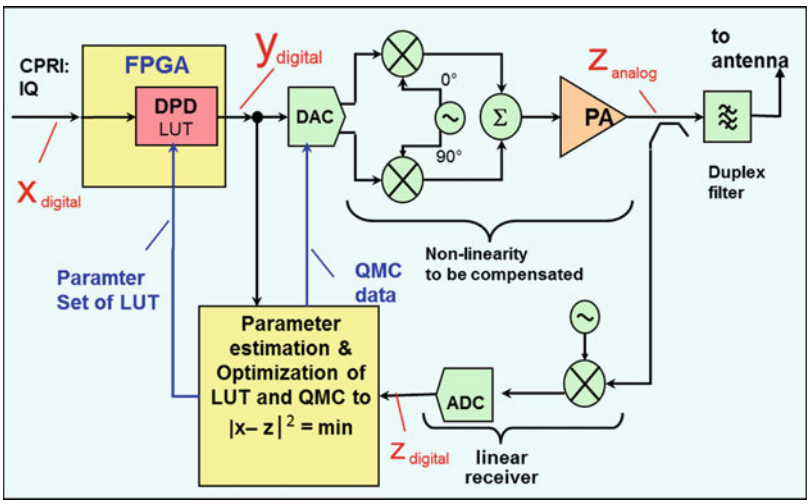


Fig. 62 Principle structure of a digital predistortion

analogue domain by digital analogue converters and mixed into the original frequency band by means of a mixer. Typically, the signals are divided into an inphase and quadrature phase part and mixed by means of a quadrature mixer. The inverted principle is applied for the Rx chain, with the exception that no CFR and DPD are needed. For the DPD, a reference signal is needed from the output of the power amplifier to determine the correction values.

In Fig. 62, the principle of a DPD is shown. Samples of the signal of the power amplifier output are taken, transferred into the digital domain by an analogue digital converter to a signal y, and these samples will be compared with the original input signal x. Different principles are then used to solve the optimization problem for the minimum difference of x and y. One possibility is a "Look-Up Table," LUT, in the FPGA, by means of which the original input signal x is predistorted to then achieve the desired output signal. It is important to know that the problem is a nonlinear problem as the behavior of the amplifier also depends on the history of the signal. For an effective correction, this time-dependent transfer function of the power amplifier has to be taken into account. Often, Volterra sequences are used to describe the transfer function. Other approaches use a nonlinear model of the amplifier to compare the calculated nonlinear signal with the real output signal. Then the minimum deviation between x and y is achieved by tracking the determined coefficients.

In addition, the correction of the DC part and symmetry of the mixer is required, which also causes a disturbed signal. This is corrected separately by adjusting the quadrature mixer correction (QMC).

RRHs increase the power efficiency and the sensitivity of the communication system. Table 2 shows typical figures. Here, it can be seen that the total power efficiency is dramatically higher, and the sensitivity of the site shows significantly better performance. For the calculation of the figures, an attenuation of the feeder cables of 3 dB is assumed, the jumper cables to the antennas have an attenuation of 1 dB, and the noise figures of the TMA and RRH have been determined to be 2 dB. The attenuation of the TMA is considered to be approximately 0.3 dB and the noise figure of the base station to be about 2–4 dB. The higher noise figure of the system with a TMA is caused by the limited amplification of the TMA, so the other noise contributions in the signal chain still increase the noise.

It can be seen that the RRH shows the best performance for the system regarding the sensitivity and power efficiency. That is one driver for the integration of the RRH into the antenna.

DOI 10.1007/978-981-4560-75-7_88-1

Table 2 Comparison of sites with TMA, RRH, and legacy technology

Parameter	Conventional system	System with TMA	System with RRH	System with antenna-integrated RRH
Transmit power at device output	2×60 W	2×60 W	2×40W	2×40 W
Transmit power at antenna	2×24 W	2×22W	2×32W	2×40 W
Input power (DC/AC)	∼700 W	∼710 W	∼400 W	∼400 W
Total power efficiency	6.9 %	6.2 %	16 %	20 %
System noise figure	6..8 dB	3.5 dB	3 dB	2 dB

Antenna Integrated RRH

The steps towards the integration of electronics into the antenna are the logical consequence of improving the efficiency and sensitivity of the system. Nevertheless, the acceptance of electronic integration was strongly related to the reliability of the integrated components. This is true regarding all the already mentioned active antenna line devices, such as TMAs, RET, or RRHs. In the last few years, an acceptable level of reliability has been achieved by careful design of the electronic parts, including a higher integration level and increased power efficiency of the power amplifiers. Therefore, today's equipment shows MTBF values of at least 100,000–200,000 h for RRHs and 500,000 h for DTMAs. Of course, the calculated MTBF values do not exactly reflect the expected lifetime, but they are an important indicator. A great influence is the operating temperature, which is of course much higher in the RRHs than in the TMAs due to the higher dissipation power. Therefore, careful design with regard to the thermal flow is required. In Fig. 63, an antenna with an integrated radio is shown (Antenna Integrated Radio, AIR).

Adaptive Active Antenna (AAS)

For many years adaptive antennas with beamsteering technologies have been discussed. These antennas are called "Smart Antennas" or "Adaptive Antennas." The principle consists of the flexible control of the magnitude and phase of every radiator or of groups of radiators. The principle is well known and is already used in other fields of antennas. For instance, in RADAR systems active antennas with beamforming technologies have been used for many years.

However, the first commercially available realizations for mobile communication systems with FDD had only limited success due to the lack of acceptance in the market. This was mainly caused by costs, reliability, and low efficiency of the antennas realized. For TDD systems, AAS has been used for a couple of years; here, the advantage of the TDD for the exact determination of the direction of arrival (DoA) and the parallel use of the calculated coefficients for Rx and Tx are helpful.

In Fig. 64, a principle structure of such an AAS for FDD operation is shown.

It can be seen that every radiator has its own receiver and power amplifier, together with a duplex filter. For simplification of the mixers, PLLs are not shown in detail; the structure of every RF path is very similar to the structure of an RRH. Of course, the power is smaller compared to the RRH, as the total transmitting power is the sum of the transmitting power of each single radiator. The main functions of the digital part consist in the synchronization and calibration of the phase of the signals at every radiator. This can typically be combined with the signal for the linearization of the power amplifiers. The reference signal for linearization and calibration is selected by means of a coupler from the output of the power amplifier. The correction of the phase, the synchronization, and the digital predistortion for the amplifier linearization can be implemented by central signal processing, if a common board for all the different amplifiers is used – or alternatively separately at the amplifier modules. Nevertheless, to achieve coherent signals at the output of the modules, at least a synchronization of the PLLs and the control of the phase of different modules have to be implemented.

DOI 10.1007/978-981-4560-75-7_88-1

Fig. 63 Active antenna with integrated radio Ericsson-Kathrein 2012

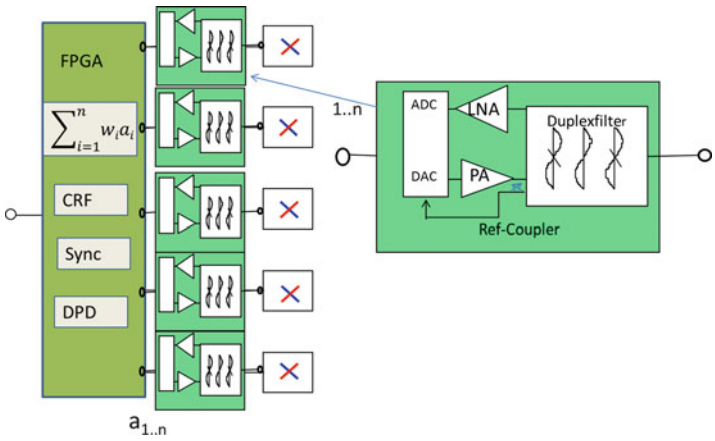


Fig. 64 Principle structure of an AAS

DOI 10.1007/978-981-4560-75-7_88-1

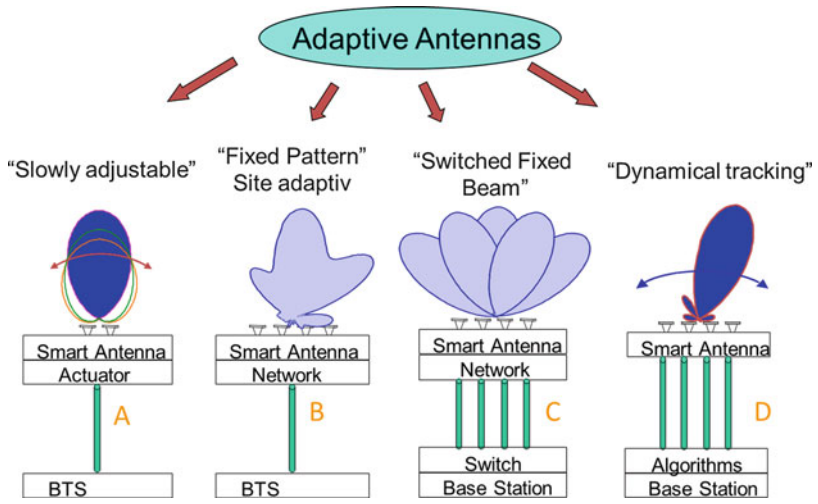


Fig. 65 Different realizations of adaptive antennas

Beamsteering Technologies

The big advantage of adaptive active antennas (AAS) is the ability of user-specific beamsteering to receive and transmit signal. The implementation of beam control can be realized in open- or closed-loop configurations. Open-loop configuration realizes the determination of the direction of arrival (DoA) and approximates the direction of the transmission beam from the calculated coefficients for the DoA of the incident waves. Different possibilities of beamsteering possibilities which are considered as beamsteering antennas or "smart antennas" are shown in Fig. 65.

Simple forms are slowly adjustable antennas, where adjustable phase shifters for multi-column antennas or moveable parts of antenna reflectors can be used. Smart antenna networks use site-specific distribution networks inside antenna arrays. METAWAVE has proposed special solutions in this field (Feuerstein 2001). Nevertheless, the use of such antennas was not widely accepted. Switched fixed beam antennas were demonstrated in successful arrangements. The beamforming procedure is simplified compared to dynamic beamsteering, whereas the results in gain of capacity are not so far from the gain with dynamic beamsteering implementation. The dynamic tracking operation is based on a DoA estimation and a following well-defined beam. All these algorithms are based on a maximization of the desired signal. What is neglected here is the influence of undesired interferers. Taking this into account, the solution may differ from only maximizing the signal. The principle is shown in Fig. 66.

It is obvious that the optimization problem consists in maximizing the signal-to-noise ratio; this is to use a maximum ratio combiner for the adaptive antenna. The problem is well described in literature (Viberg 2005). For isotropically distributed noise, the solution just results in the optimum coefficients for the desired direction of the incident signal. The solution becomes simple for the simplified problem of only one incident planar wave. Assuming that in Fig. 67 the distance r is big enough, then the incident angle will be the same for all antenna elements. In an even more simplified approach, it is assumed that the characteristics of the antenna elements are the same, and the path loss is normalized to one; it can be denoted for the incoming signals at the antenna elements:

$$y = \sum_{i=1}^{n} w_i(x_i + n_i) \quad = s \cdot \sum_{i=1}^{n} w_i C_i(\varphi) \ \exp(-jkr_i) + \sum_{i=1}^{n} w_i n_i$$

If a_s is considered as the normalized vector of incident waves, w is the steering vector of the antenna array, and n is the noise vector, the optimum solution is just given simply by

DOI 10.1007/978-981-4560-75-7_88-1

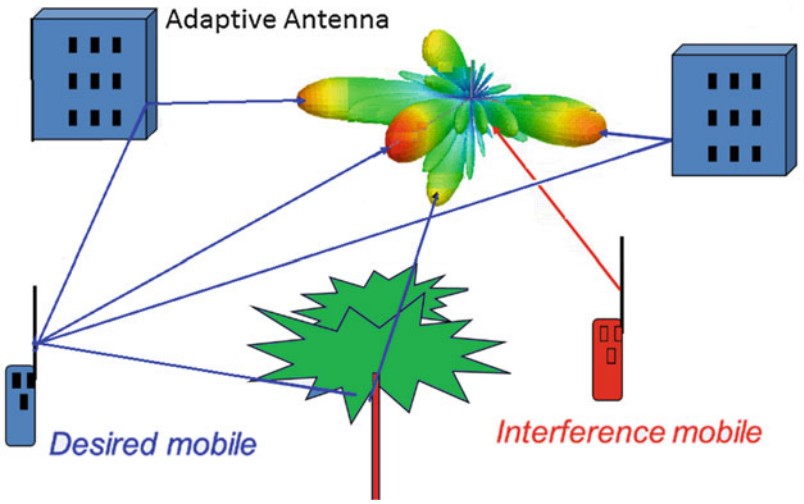


Fig. 66 Adaptive antenna with interferer

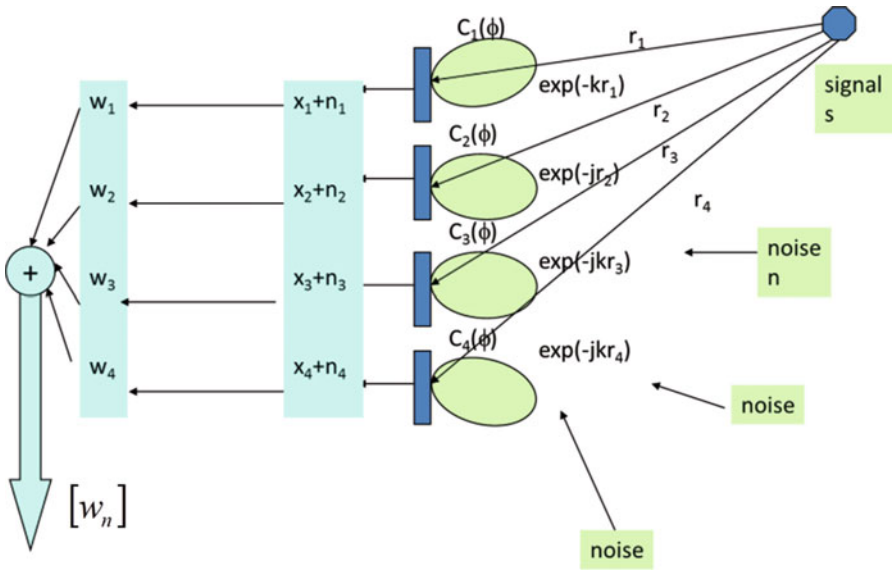


Fig. 67 Schematic beamforming arrangement of an adaptive antenna

$$[w_{opt}] = [x]^*$$

This means that the conjugate complex steering vector of the signals of the incident wave front is required. This is evident, as the resulting beam points exactly in the direction of the incident wave. The realized steering vector $[w]$ realizes a matched spatial filter.

For non-isotropic distributed interferers, the optimum solution depends on the angle-dependent interference. A matched filter then is realized as

$$[w_{opt}] = c \cdot \left\| R_N^{-1} \right\| [x]^*$$

Where R_N is the covariance matrix of the interference signals, $[x]^*$ is the optimum steering vector for the desired incoming signal, and c is a scalar coefficient. This equation is also well known as Wiener–Hopf

DOI 10.1007/978-981-4560-75-7_88-1

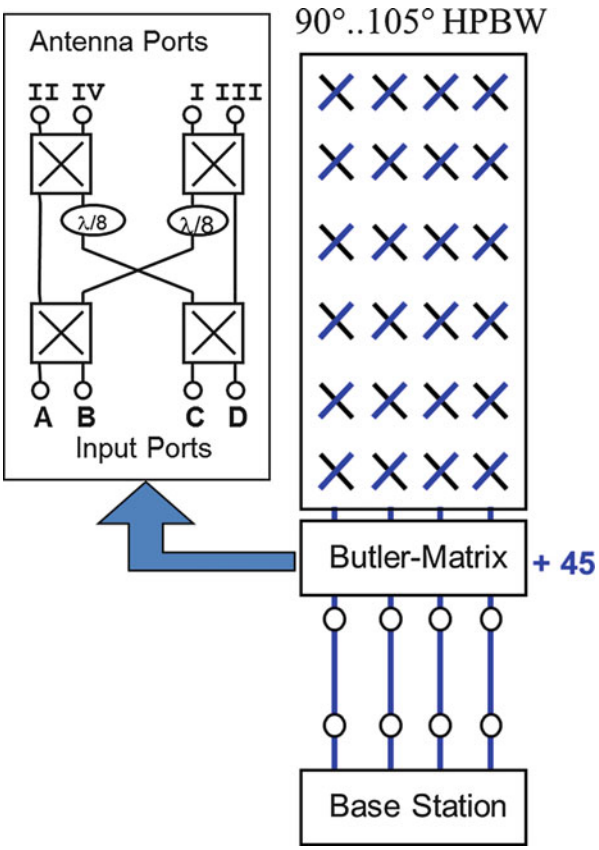


Fig. 68 Beamforming antenna with Butler matrix

equation; see also Konrad (2003), Vega and Rey (2013). The implementation of this algorithm takes the interferer into account and tries to shift a null in the resulting characteristic into the direction of the interferer. This algorithm is also called zero-forcing beamsteering.

Nevertheless, as the incident waves and the interferer are not known per se, different optimization algorithms have to been employed.

Another possibility is the implementation of switched fixed beams. This is typically based on the approach with isotropical interferers. Nevertheless, some interesting hardware implementations exist, e.g., through hardware butler matrixes. In Fig. 68, an antenna with a butler matrix is shown; the input ports and antenna ports produce 4 different beams, whereas the input ports are decoupled. The butler matrix consists of four 3 dB 90° couplers which are connected as shown in Fig. 68. Two additional 45° phase shifters are required in two connection lines, and then four independent beams can be obtained. Nevertheless, the isolation of the ports is limited by the 3 dB couplers used and the reflection of the input ports of the antenna.

MIMO Implementation

The realization of MIMO can be based in principal on the same antenna configurations as for diversity and/or beamforming systems. However, for MIMO implementation-specific restrictions have to be considered:

(a) FDD systems use different frequencies for receiving and transmitting. Therefore, the coefficients for transmitting operation have to be approximated from the coefficients of the DoA at the Rx frequency. This principle only works for a dominant path from the mobile to the base station which has similar

DOI 10.1007/978-981-4560-75-7_88-1

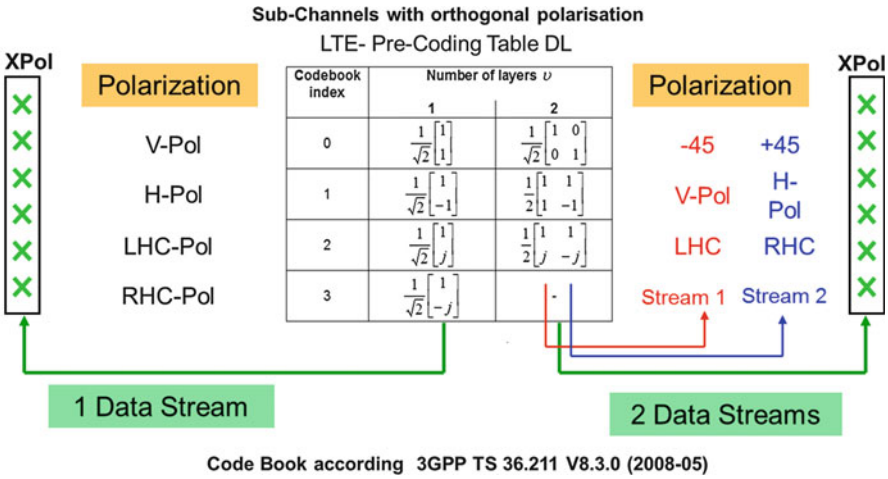


Fig. 69 Code book for two transmission paths resulting in a polarization control

properties for both frequencies. This excludes, for instance, reflections, scatterers, or indirect paths, which have strong frequency-dependent behavior. Also, the duplex distance should not be too big. The phases for transmitting have to be derived from the receiving coefficients, typically by a calculation of a direction of the main incident waves. Based on the determined direction, in a second step the coefficients for the transmitted waves are calculated, which also require an exact calibration of the antenna array.

(b) For TDD systems, the procedure is simplified, as the determined coefficients of the optimum received configuration can be used directly for the transmission. Here, only the time variation of the channel changes path; nevertheless, for short switching times in the ms-range, the constancy of the path is typically good enough.

(c) Closed-loop configurations overcome the problem of calibration and determine the optimum transmission and receiving coefficients by a feedback from the receiver for a given test sequence. The transmitters transmit agreed combinations of coefficients, a so-called code book, and the receiver gives feedback about which code book should be used for the following transmission sequence. The advantage is that by means of the test sequences, the whole transmission matrix from every transmitter to every receiving element can be determined, and the optimum capacity of the connection can be calculated by determining and selecting the right code. The procedure includes the option of MIMO transmission, as the code book can include the transmission of more than one signal stream. The disadvantage here is the limitation of the codes of the code book. So only a certain defined granularity of the code space can be used. In the 3GPP standard for LTE, an example is shown of the code book for a 4×4 system with corresponding beams for beamforming transmission and MIMO transmission (Figs. 69 and 70).

Future Trends

Today's mobile communication has become a key element of modern society, and smart devices are omnipresent: almost everybody communicates over mobile devices at home, in the office, and on travel. Most people use several devices in parallel – e.g., a smart phone, a laptop computer, and a tablet – to cover their needs of voice communication, data exchange, and Internet use (infotainment).

For example, cloud computing requires reliable connections and high data rates to ensure adequate access to this service. Moreover, machine-to-machine communication (M2M) is rapidly growing: new concepts like "Industry 4.0" describe the massive wireless interconnection of machines in a production

DOI 10.1007/978-981-4560-75-7_88-1

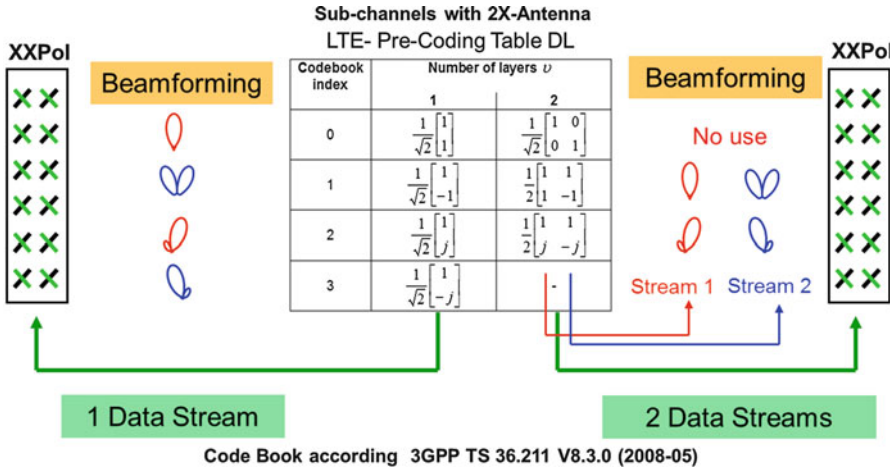


Fig. 70 Code book for two transmission paths resulting in a beamforming for 1 and 2 data streams

facility to enable intelligent, self-organizing production processes. All these aspects will lead to a rising demand in mobile communication coverage as well as throughput capacity enhancement.

Requirements and Challenges

The demand for mobile connectivity will dramatically increase in near future: people want to be connected all the time everywhere and want to use voice and data services to a large extend. These demands result in an ongoing increase of capacity worldwide: from 2008 to 2013, the global mobile data traffic grew by a factor of 45; a factor of 24 is projected for the period from 2010 to 2015. If these developments go on in this pace, the global mobile traffic might increase to extremely high levels, and a growth factor of even 1000 is expected in the 2020s in comparison to 2010 (NTT Docomo 2014; CISCO 2014).

The Vision of 5G

Based on these predictions, mobile communication experts are currently discussing possible architectures regarding a fifth generation network (5G) as follow up for LTE and LTE advanced (which are commonly labeled as 3.9G and 4G, respectively). In order to manage the expected growth and to define a well-suited network for the future demands, the following elements and strategies are discussed:

- **Network densification:** the number of mobile communication cells shall be dramatically increased, whereas the average cell size shall be reduced, so that the total capacity of the network significantly increases.
- **Active antennas:** active (or adaptive) antenna systems shall be an integral part of the cell architecture, so that the cell capacity and coverage can be dynamically adapted to the demand at hand.
- **Spectral efficiency**: the introduction of adaptive coding and higher-order modulation schemes together with smart antennas can lead to a better spectral efficiency and higher reuse factors in the network. This includes also new intelligent receiver architectures and better dynamic methods of interference suppression.
- **Massive MIMO**: in current LTE networks, 2×2 MIMO is widely used; 4×"X" and 8×"X" are already available on the base station side, but very seldom in use. Ongoing projects investigate if a massive use of MIMO with very high number of antennas can lead to a significant improvement of the average data rate in very small cells.

DOI 10.1007/978-981-4560-75-7_88-1

- **New frequency bands**: one of the key items in 5G investigations is the allocation of new frequency resources. As the frequency spectrum below 5 GHz is already occupied (and often congested) by many services, the main focus is on higher frequency bands: especially 30 and 60 GHz are promising candidates, as these bands offer potentially new very high capacity.

The so-called 60 GHz band is an unlicensed band reaching from 57 to 64 GHz, which theoretically sustains data rates up to 50 GBit/s. With respect to this number, it is no surprise that the allocation of higher frequency bands has become a key aspect for all investigations of 5G networks. At 60 GHz the wavelength is, however, only 5 mm, and this results into completely different network requirements and antenna architectures as compared to the established mobile communication standards (NTT Docomo 2014; CISCO 2014).

Antenna Systems for MM-Waves

Requirements

In order to start the discussion on mm-wave antennas, it is very helpful to look again at the Friis formula (see also section "Antenna Design for Base Stations"):

$$P_{R,\,mobile} = \left(\frac{\lambda_0}{4\pi R}\right)^2 G_T G_R P_{T,\,basestation}$$

The power received at the mobile depends on the transmitted power and on the gains of both the base station and the mobile, but also on the distance R and on the wavelength. When changing the frequency from 2 to 60 GHz, the wavelength changes by a factor of 30. Assuming that all parameters in the formula above would not change except for the wavelength, the power received at the mobile when going from 2 to 60 GHz would consequently be reduced by a factor of 30^2=900!

So it is obvious that achieving "reasonable" transmitting distances in the 60 GHz band is a demanding task. The key parameters for achieving an adequate link budget are the distance R on one hand and the antenna gains G_T and G_R on the other hand: if highly directive antennas are used, and the transmitting distance is limited to a short range, a sufficient level for the received power can be obtained. Therefore, most of today's scenarios at higher frequency bands foresee the use of short range only and the implementation of directive antennas on the transmitter as well as on the receiver side.

When the antennas are not omnidirectional but directive, it is essential in the case of moving users that the antenna beams dynamically follow the user's position. So looking at the antenna system requirements, the higher frequency bands do not only require highly advanced RF technology, but they also require the implementation of beamforming and automatic tracking capability on the base station side as well as on the mobile side.

The use of higher frequency bands results in the following challenges to be solved:

- **Use of directive antennas**: typical candidates are antenna arrays or reflector antennas.
- **Active smart antennas**: due to the high cable losses at the higher frequencies, the transceiver electronics have to be mounted rather closely to the antenna aperture.
- **High-density packaging technologies**: due to the short wavelength, the array grid becomes very small. Therefore, advanced high-density packaging technologies are necessary to realize active antenna front ends.
- **Beamforming and tracking**: because of the dynamic user scenario, the base station antennas (and the mobile antennas) need an automatic beamforming and tracking functionality. For regulatory issues,

DOI 10.1007/978-981-4560-75-7_88-1

high power transmitters require a beam with a very small HPBW, which dramatically increases the tracking accuracy requirements.
- **Thermal issues**: due to the high packaging density and the low efficiency of RF components at higher frequencies, the heat dissipation is a key topic in front-end design. In order to avoid liquid cooling, advanced technology concepts have to be developed to ensure an adequate temperature of the front end in function.

Besides these issues, currently the most realistic mobile transmission scenario is the line-of-sight connection over short and medium distances (up to several hundred meters). Non-line-of-sight signals will suffer from high losses at each reflection point of the signal, so that the received power will be quite low. Moreover, any obstacle, which is placed within the link, like a human body or any other objects will lead to significant losses, too. Also, the rain attenuation at 30 and 60 GHz can be very high.

So, the use of higher frequencies brings up demanding challenges in the area of antenna design and RF integration. However, the throughput potential in the higher frequency bands is so enormous that it makes sense to follow that path and to investigate the possibilities hereof. For specific scenarios in very small cells (picocells), the high bands have the potential to fulfill the high-throughput demands of the future. Another advantage within the higher frequency bands is the security aspect: the high path attenuation in combination with the use of directive antennas functions like a "spatial filtering": an unwanted listener can be excluded from the communication just by the dramatic reduce of power level in unwanted directions (Stevens and Graften 2012).

Beamforming Concepts

In order to realize a beamforming antenna, one could think of mounting a reflector antenna onto a mechanical multi-axis platform, so that the antenna beam can be mechanically aligned with any position of the receiver. However, this seems unrealistic in the case of 60 GHz base stations, as the mechanical systems suffer from high maintenance, resulting in limited lifetime due to the mobile user scenarios. Also, the required beam agility – the angular tracking speed that is needed to keep the narrow beam locked onto the desired target – is very hard to fulfill with a mechanical system. Moreover, the integration of a mechanical tracking platform in any mobile device like a laptop is simply unrealistic.

So the focus is on electronic beamforming, on the base station side, and on the mobile terminal side. Electronic beamforming antennas can be designed using different concepts: they can be realized as direct radiating active arrays or alternatively as active feed arrays in combination with a reflector or lens. In order to implement the electronic beamforming, three basic concepts can be used: switched beam, phased array, and digital beamforming (DBF) (Fig. 71).

- **Switched beam** means that beamforming networks are used, each representing a particular beam pointing into a certain direction. One can then switch between the different networks in order to point the antenna beam. As long as the number of beams is low, this method can be applied. One should however not forget that switches at higher frequencies can cause high RF losses.
- **Phased arrays** use analogue or digital phase shifters. These phase shifters allow modifying the phase of any antenna element or group, so that the beam of the array can be steered electronically. This concept is generally more complex than switched beam concepts, but the beamsteering is more flexible.
- **Digital beamforming arrays** consist of antenna elements, each of them having its "own" T/R electronic section, including up/down conversion. In this way, each antenna element signal is available at digital level, and the beamforming can be performed on digital baseband level. This technique is quite complex and requires both a large amount of active RF components and also significant processing power. However, this concept offers the highest beamforming flexibility.

DOI 10.1007/978-981-4560-75-7_88-1

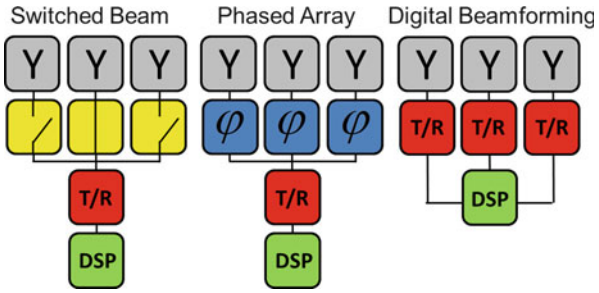


Fig. 71 Different concepts for electronic beamforming (T/R=transmit/receive Module, DSP = digital signal processor)

Before selecting one or the other technique, the requirements have to be carefully analyzed. Factors like the expected angular tracking range, the HPBW, the available size for the antenna, and the costs have to be analyzed. Very often, the different techniques are combined to arrive at an optimum solution in terms of performance, size, complexity, and cost for a given application.

Front-End Examples

Two examples of front-end realizations are presented here: the first example is a front end at 30 GHz. The RF front end of the system is built in LTCC (low temperature co-fired ceramics), which is a material suited for high-density multilayer electronics (Kulke and Uhlig 2010). The front end in Fig. 72 was designed within a German research activity (Holzwarth et al. 2009), and the design requirements were derived from future satellite systems. The front end consists of 64 antenna elements that are part of a 16-layer-LTCC module, including 64 active Tx channels and a calibration network. Liquid cooling channels are integrated within the ceramic material to ensure an adequate operating temperature. The front end is capable of full DBF transmit only at 30 GHz.

In Fig. 73, the measured far-field patterns are shown: the beamsteering of the module covers the region from boresight down to 60° in elevation. The beamforming shows excellent performance at boresight; towards lower elevation angles, the performance degrades due to the reduction of the effective antenna aperture and the characteristic of the antenna element.

Figure 74 shows another antenna front end that was realized in the 60 GHz band. In this case, the chosen technology was also LTCC (Sanadgol et al. 2010). The antenna module consists of 4×6 antenna elements, and electronic beamsteering was implemented in one plane only. The technology applied was substrate integrated wave guide.

The front end was designed for high data rate wireless links over short distances for consumer devices. The measurement results showed excellent beamforming performance and an antenna efficiency of nearly 80 %. Figure 75 shows the simulation model of one array column. The wave guide was formed by metal and vertical planes via fences embedded in the material. The display of the vertical cut shows the field distribution in the network as well as at the antenna element level.

Summary

Base station antenna systems find themselves in an ongoing dynamic evolution process. The first cellular communication systems offer only relatively low data rates for a limited number of users. The main focus of the first cellular networks was to establish a basic connection and to provide basis coverage. The relatively simple passive antenna structures were typically mounted on a tower and had the transceiver electronics stored in a small building at the bottom of the base station.

DOI 10.1007/978-981-4560-75-7_88-1

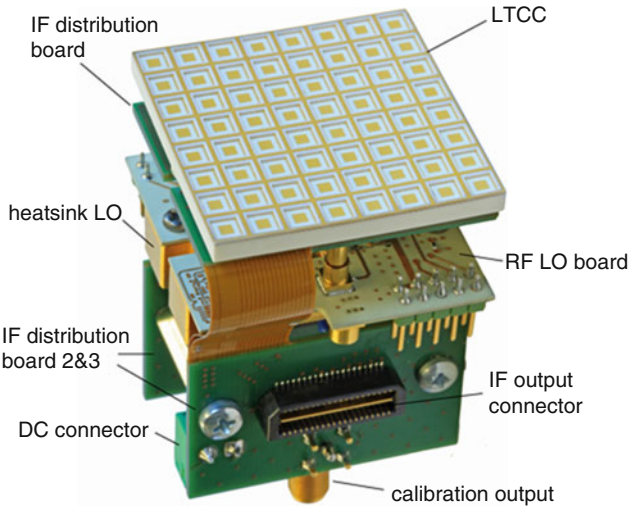


Fig. 72 Tx front end of a digital beamforming array at 30 GHz (IMST)

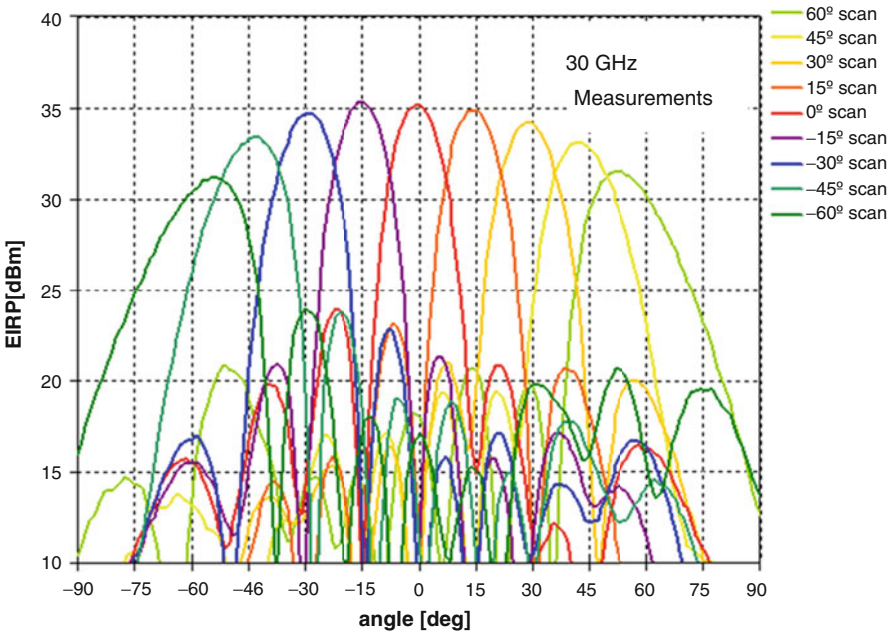


Fig. 73 Measured antenna beams of the DBF array at 30 GHz

DOI 10.1007/978-981-4560-75-7_88-1

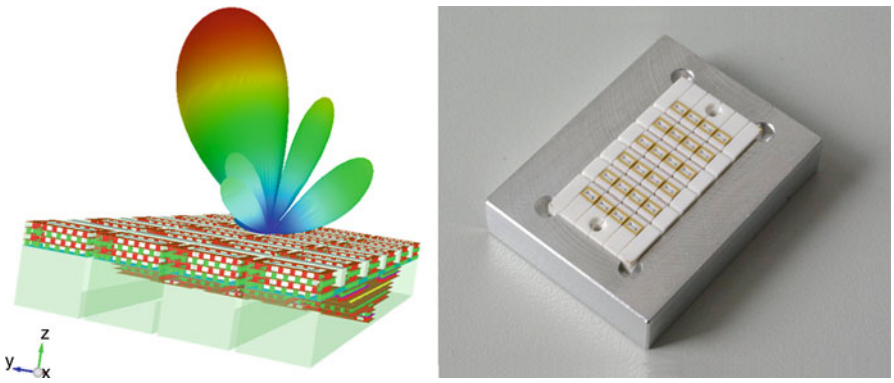


Fig. 74 Antenna front end for 60 GHz: simulation model and front-end hardware within test fixture (Beam Networks, IMST)

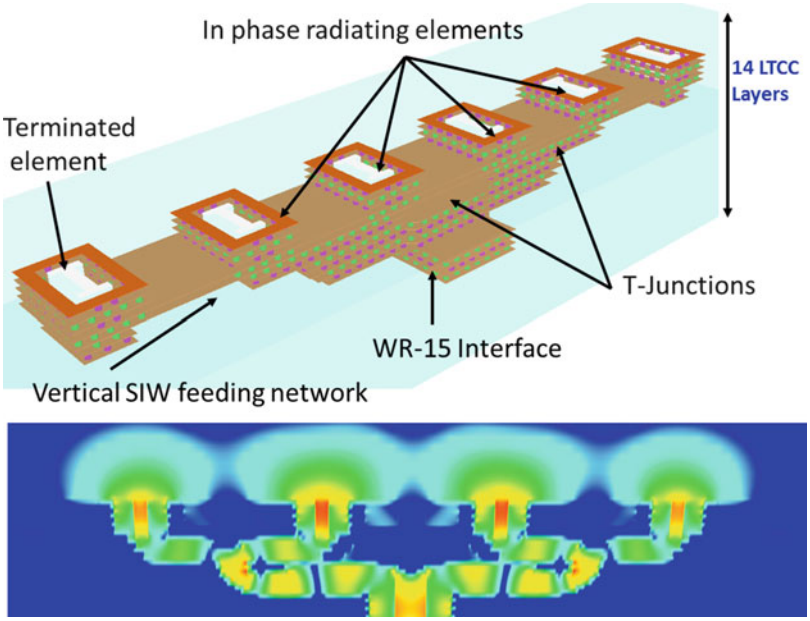


Fig. 75 Substrate-integrated wave guide concept of the 60 GHz front end: simulation model of one column and electric field in a vertical cut plane

Since the early days of cellular communications, a dramatically increasing demand in terms of capacity, coverage, and availability of mobile communication has taken place. One of the first changes in mobile networks was the replacement of large cells by a set of smaller cells to increase the total network capacity. In parallel, new standards and frequency bands have been allocated, higher modulation schemes have been implemented, and special carrier aggregation techniques have been realized to increase the number of users in the network and to allow much higher data throughput.

The antenna technology changed from very simple omnidirectional radiators via sector arrays towards active beamforming antennas. Innovative complex arrays were developed to allow a very exact pattern and gain control well suited for a specific cell or sector. Multiband antenna systems were also designed to combine several communication standards within the same form factor. The transceiver electronics moves more and more close to the antenna to avoid cable losses and to improve the signal quality. The remote control of the antenna beam shape allows the dynamic adaptation of the cell size and the alignment of the cell's capacity to the user's demand.

DOI 10.1007/978-981-4560-75-7_88-1

The worldwide communication and data traffic faces a dramatic increase within the coming years. Therefore, the experts currently investigate new concepts for a 5th generation mobile network. One key issue in these concepts is the allocation of higher frequency bands at, e.g., 30 and 60 GHz. The implementation of communication links in those bands will require highly directive antennas, including electronic beamforming. The required technology is complex and quite demanding; however, the successful design of powerful front ends is the key to access these very interesting frequency resources.

Cross-References

- ▶ Antenna Design for Diversity and MIMO Application
- ▶ Antennas in Access Points of WLAN/Wi-Fi
- ▶ Antennas in Microwave Wireless Power Transmission
- ▶ Broadband and Multiband Planar Antennas
- ▶ Linear Wire Antennas (Including Yagi-Uda Antennas)
- ▶ Loop Antennas
- ▶ Low-Profile Antennas
- ▶ Microstrip Patch Antennas
- ▶ Millimeter-Wave Antennas and Arrays
- ▶ Phased Arrays
- ▶ Small Antennas (PIFA/PILA/Loading Antenna/etc.)
- ▶ Substrate-Integrated Wave guide Antennas

References

3GPP (2006) 3GPP TS 25.463 UTRAN Iuant Interface Remote Electrical Tilting (RET) Release 6. 3GPP

AISG (2006) Antenna Interface Standards Group. Standard No. AISG v2.0. AISG Ltd., London. www.aisg.org.uk

Beckman C, Lindmark B (2007) The evolution of base station antennas for mobile communications. IEEE, 1-4244-0767-2/07

CISCO (2014) Cisco visual networking index: global mobile data traffic forecast. 2009–2014 reports. CISCO, USA

Collins B (2009) Base station antenna systems. In: Zhi Ning Chen, Kwai Man Luk (eds) Antennas for base stations in wireless communications. McGraw Hill, New York

CST (2014) 3D EM Simulation software. www.cst.com

Derneryd A, Johansson M (2009) Advanced antennas for radio base stations. In: Zhi Ning Chen, Kwai Man Luk (eds) Antennas for base stations in wireless communications. McGraw Hill, New York

Docomo (2014) 5G radio access: requirements, concepts and technologies. White paper NTT Docomo, Japan

Empire (2014) 3D EM Simulation tool based on finite difference time domain. www.empire.de

Feuerstein M (2001) The evolution of smart antennas to 3G. Presentation on the CDG Technology Forum, Costa Mesa, pp 1–22

Gabriel R, Göttl M (1997) Antennenarray. European Patent EP 0848862

Göttl M (2004) US Patent 6734829 B1

DOI 10.1007/978-981-4560-75-7_88-1

Göttl M, Gabriel R, Markof M (1999) Hochfrequenz- Phasenschieberbaugruppe. European Patent EP 1208614

Hata M (1980) Empirical formula for propagation loss in Land Mobile Radio Services. IEEE Trans Vecicular Technol VT-29(3):317–325

Holzwarth S et al (2009) SANTANA: advanced electronically steerable antennas at Ka-Band. In: Eucap: 3rd European conference on antennas and propagation, Berlin, Germany, 23–27 March 2009, pp 471–478

Konrad W (2003) Design of an adaptive antenna demonstratorfor digital beamforming. Dissertation, TU Wien, Fakultät für Informations – und Elektrotechnik, Wien

Kulke R, Uhlig P (2010) LTCC technology for RF applications. In: EEEfCOM, electrical and electronic engineering for communication, workshop, Hochschule Ulm, 17 June 2010

Lindmark B (2000) A dual polarized dual band microstrip antenna for wireless communications. In: Analysis and design of base station antennas for mobile communications (Thesis). Chalmers University, Göteborg, pp attachment 1–7

Linehan K, Gianola B (2011) Recommendation on base station antenna standards. White paper of the NGMN alliance

Maurer J (2005) Kapitel 5 Systemtheoretische Beschreibung des Funkkanals. In: Strahlenoptisches Kanalmodell für die Fahrzeug-Fahrzeug-Funkkommunikation. IHE Universität Karlsruhe, Karlsruhe, pp 96–114, ISSN: 0942–2935

Okumura Y et al (1968) Field strength and its variability in VHF and UHF land-mobile service. Rev Electric Commun Lab 16:825–873

Pontes J, Corcoles J, Gonzales MA, Zwick T (2011) Modal Network Model for MIMO antenna in-system optimization. IEEE Trans Antennas Propag 59(2)

Sanadgol B, Holzwarth S, Milano A, Popovich R (2010) 60 GHz substrate integrated waveguide fed steerable LTCC antenna array. In: 4th European conference on antennas and propagation, EuCAP 2010, Barcelona, 12–16 April 2010

Shannon C (1949) The mathematical theory of information. University of Illinois Press, Urbana

Stevens M, Graften G (2012) The benefits of 60 GHz unlicenced wireless communications. White paper from Sub10 Systems Limited, Newton Abot, United Kingdom

Taga T (1992) Analysis of planar inverted-F antennas and antenna design for portable radio equipment. In: Hirasawa K, Haneishi M (eds) Analysis, design and measurement of small and low profile antennas. Verlag Artech House, Boston, pp 161–180

Taga T, Tsunekawa K (1987) Performance analysis of a built-in planar inverted F antenna for 800 MHz band portable radio units. IEEE J Sel Areas Commun SAC-5(5):921–929

Telatar I (1995) Capacity of multi-antenna Gaussian channels. Technical report, June 1995. AT&T Bell Laboratories, Murray Hill

Vega L, Rey H (2013) A rapid introduction to adaptive filtering. Springer, New York

Viberg M (2005) Direction of Arrival estimation, chapter 16. In: Kaiser et al. (ed) Smart Antennas State of the Art, Hindawi

Waldschmidt C (2004) Systemtheoretische und experimentelle Charakterisierung integrierbarer Antennenarrays. IHE der Universität Karlsruhe (TH), Forschungsberichte Bd. 41, Karlsruhe, ISSN: 0942–2935

Walke B (2000) Modelle der Funkfeldberechnung. In: Mobilfunknetze und ihre Protokolle. Teubner Verlag, Stuttgart, pp 47–51

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_91-1

Antennas in Wireless Charging Systems

Qiang Chen[a]* and Qiaowei Yuan[b]
[a]Tohoku University, Sendai, Japan
[b]National Institute of Technology, Sendai College, Sendai, Japan

Abstract

The antenna is one of the most critical components in wireless charging systems since the power transfer efficiency of the system is largely dependent on the antenna performance. In this chapter, the modeling and analysis of antennas are introduced with a focus on increasing the power transfer efficiency. The relationship between wireless power transfer efficiency and antenna parameters like antenna geometry, ohmic loss, impedance matching circuits, and distance between transmitting and receiving antennas is studied analytically and numerically by using circuit theory and full-wave electromagnetic analysis, in order to provide some fundamental and theoretical knowledge to develop antennas for highly efficient wireless charging systems. Finally, challenges and recent studies of antennas for wireless power transfer systems, as well as standards on this field, are briefly introduced.

Keywords

Near-field coupling; Wireless power transfer; Wireless charging; Wireless power transmission; Antenna; Impedance matching; Efficiency; Equivalent circuit; Scattering parameters; Scattering matrix; Impedance matrix

Introduction

Research on wireless power transfer (WPT) started more than 100 years ago. Nikola Tesla developed the first practical WPT system in 1989 (Tesla 1904) and revealed the system configuration in his U.S. patent in 1914 (Tesla 1914). Reinhold Rudenberg described the fundamental theory of electric power absorption by antennas early in 1908 (Reinhold 1908). Since then, there have been numerous researches of both fundamental and application studies on the WPT. This technology has been widely applied to many applications from wirelessly charging electronic devices to the solar power satellite (SPS) system (McSpadden and Mankins 2002). However, this technology still faces challenges. One of the biggest challenges is keeping high power transfer efficiency (PTE) over a long transmission distance.

Wireless power transfer can be classified into two categories based on how the electromagnetic field is used to transfer power. One is far-field radiation, and the other is the near-field coupling. The near-field coupling approach is also called the inductive coupling method or induction method. This method is effective for a short-range power transfer with a relatively high power transfer efficiency. The near-field coupling approach has already been applied to many practical applications such as wireless charging for cordless phone, electric toothbrush, and passive radio-frequency identification (RFID) tags. On the other hand, the far-field radiation approach is applicable to power transfer over a long distance. But the power transfer efficiency is extremely low, depending on the transmission distance. One of the most well-known

*Email: chenq@ecei.tohoku.ac.jp

DOI 10.1007/978-981-4560-75-7_91-1

applications is the space solar power system (SPS), which is a big project to collect sunlight in geostationary orbit, convert it to electromagnetic power, and transfer the power to the earth in the form of electromagnetic waves. This method requires directive antennas with high gain to increase the power transfer efficiency, as well as a direction tracking system to track the movements of the antennas. Therefore, the far-field radiation approach is usually limited to the applications in the environment of line-of-sight propagation.

A very efficient near-field power transfer was experimentally demonstrated by introducing a concept called evanescent resonant coupling (Kurs et al. 2007). In the paper, two self-resonant coils were used as receiving and transmitting antennas, and a power transfer for lighting up a 60-W light bulb with 40 % efficiency over distances in excess of 2 m was demonstrated experimentally. This magnetically resonant system was critical to realize power transfer over a relatively long distance while keeping a good power transfer efficiency compared with the conventional induction coupling method. There have been many research papers and reports soon after this publication (for example, Karalis et al. 2008; Cannon et al. 2009; Tak et al. 2009; Ishizaki et al. 2010; Yuan et al. 2010). It was shown that the evanescent resonant coupling method can transmit the energy for longer distance than the previous near-field induction method (Murakami et al. 1996; Hatanaka et al. 2002) and was more efficient than the far-field radiation method where the vast majority of energy was wasted due to the transmission loss (Brown 1984; Matsumoto 2002; Rodenbeck and Chang 2005). It was also found that the electric coupling was effective in wireless power transfer with a high efficiency (Kim and Ling 2007). It was stated in Chen et al. (2012) that the conjugate impedance matching for both transmitting and receiving antennas could bring about the maximum power transfer efficiency instead of using the concept of antenna resonance or magnetic resonance. It was demonstrated in the paper that an efficiency up to 100 % could be achieved in the near-field region when the antennas were perfectly conducting and electrically small if the conjugate-matching condition was satisfied. It was also described that, in practice, loop antennas were superior to the dipole antennas in achieving high efficiency, if the conducting losses of both the antennas and the matching circuits were considered.

In this chapter, modeling and analysis of antennas in wireless charging systems based on the near-field coupling approach by using circuit theory and full-wave model electromagnetic analysis is introduced from a viewpoint of increasing wireless power transfer efficiency for the wireless charging systems. Circuit theory can provide a clear image of how electromagnetic power is coupled through the antennas, and it can give some basic knowledge on finding the key factors to increase power transfer efficiency. Full-wave analysis can give an electromagnetic simulation in a practical situation where antenna characteristics including antenna impedance, radiation, and conducting loss can be evaluated accurately. Because these antenna characteristics play an important role in determining the power transfer efficiency, the full-wave analysis is required to design antennas for a practical wireless charging system. Full-wave analysis of conical models where two types of antennas, dipole antennas and loop antennas, are used for near-field coupling show some general principles for antenna design to achieve the maximum power transfer efficiency for the wireless charging systems. Finally, some recent application developments and standards of the wireless charging technology are introduced.

Configuration of Wireless Charging System and Definition of Power Transfer Efficiency

A typical wireless charging system is generally composed of a source, impedance matching circuits, transmitting and receiving antennas, a rectifier, and a load impedance, as shown in Fig. 1. At the transmitting side, the electromagnetic power is generated by a voltage source V_0 with internal impedance

DOI 10.1007/978-981-4560-75-7_91-1

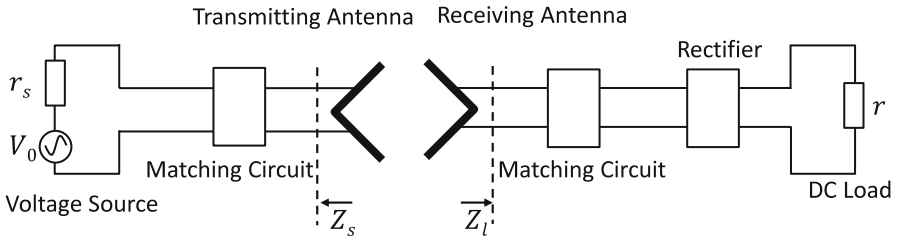


Fig. 1 Configuration of a typical wireless charging system

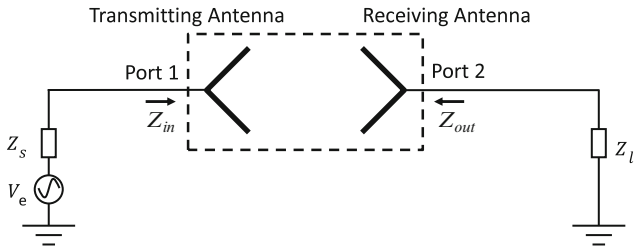


Fig. 2 Transmitting and receiving antennas in wireless charging system

r_s and is transferred through the impedance matching circuit to the transmitting antenna. The electromagnetic field is wirelessly coupled to the receiving side through the receiving antenna, impedance matching circuit, rectifier circuit, and is finally consumed by the resistance r. This configuration is also valid for the far-field radiation approach if the antennas are located in the far-field region.

Because antennas are the main topic in this chapter, it is adequate to consider a more simplified configuration model shown in Fig. 2. In this model, Z_s is the equivalent impedance looking toward the matching circuit and power source, and Z_l is the impedance looking to rectifier circuit and load. Although rectifier is a nonlinear circuit, it is assumed to be a linear parameter which is independent of the received power in this chapter. It is important to note that characteristics of a single antenna, such as the resonant frequency, input impedance, and gain, are useless for evaluating the antenna performance in the wireless charging system based on the near-field coupling approach, because the transmitting and receiving antennas are coupled strongly with each other.

The power transfer efficiency, one of the most important parameters to evaluate performance of the wireless charging system, is defined as the ratio of the absorbed power in the load impedance Z_l to the incident power into the transmitting antenna:

$$\eta = \frac{P_l}{P_{inc}}, \tag{1}$$

In this case, P_{inc} is the incident power and P_l is the absorbed power. The incident power is the maximum power available to the transmitting antenna when the condition of impedance matching is satisfied at Port 1 of Fig. 2.

It will be shown that the power transfer efficiency η reaches the maximum value when the impedances Z_s and Z_l are complex conjugate matching to the input impedance of Port 1 and Port 2, respectively. That is,

$$Z_{in} = Z_s^*, \tag{2}$$

DOI 10.1007/978-981-4560-75-7_91-1

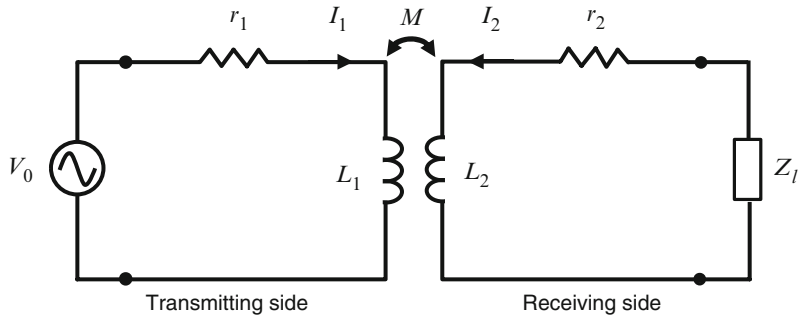


Fig. 3 Equivalent circuit for magnetic field coupling

at Port 1, and

$$Z_{out} = Z_l^*, \tag{3}$$

at Port 2, where Z_{in} is the input impedance looking into Port 1 and Z_{out} is the input impedance looking into Port 2. It should be noted again that Z_{in} is not equal to the input impedance of the transmitting antenna and Z_{out} is not the input impedance of the receiving antenna because the mutual coupling cannot be ignored between the transmitting and the receiving antennas when the near-field power coupling is discussed.

In the following sections, the relationship between the impedance matching condition and the condition for obtaining the maximum power transfer efficiency as well as the relationship between wireless power transfer efficiency and antenna parameters are investigated theoretically and numerically by using circuit theory and full-wave electromagnetic analysis.

Circuit Theory

When wireless charging systems operate at lower frequencies, the radiation from the system can be ignored, and antennas in the system can be approximately modeled using lumped circuit elements. Analysis using lumped circuit elements can be carried out based on circuit theory, which is very simple and provides a straightforward model demonstrating the operating principle of antennas in wireless charging systems.

At low frequency, the electromagnetic field can be assumed to be a quasistatic field, and electromagnetic coupling can be classified into induction through the magnetic field and coupling through the electric field. In this section, the power charge system using the magnetic induction is focused on and discussed.

Magnetic induction can be performed using two inductors, also called coils, as a transmitting antenna and a receiving antenna because coils can produce strong magnetic fields in the near-field area. The coil can be either a single-turn coil or a multiturn coil, but the total length of the wire should be much smaller than the wavelength. Otherwise, electromagnetic power would be radiated from the system, and the power charging efficiency, or power transfer efficiency, would be largely degraded.

An equivalent circuit for describing the basic phenomena of magnetic field induction is shown in Fig. 3. Two inductors for transmitting and receiving are equivalent to two inductors L_1 and L_2 and are coupled to each other with mutual inductance M. r_1 and r_2 represent the ohmic loss of the transmitting and receiving circuits, respectively, mainly caused by the inductors. A voltage source with voltage V_0 is fed to the transmitting circuit. Z_l is the load impedance where the power provided by the voltage source is absorbed.

DOI 10.1007/978-981-4560-75-7_91-1

In the following, it is shown how to obtain the maximum power transfer efficiency by optimizing the value of Z_l. This analysis provides a fundamental knowledge to develop a highly efficient wireless charging system by designing the load impedance.

When Kirchhoff's voltage law is applied to the electric circuit in Fig. 3, the equations

$$(r_1 + j\omega L_1)I_1 + j\omega M I_2 = V_0 \tag{4}$$

$$j\omega M I_1 + (Z_l + r_2 + j\omega L_2)I_2 = 0 \tag{5}$$

are obtained. By solving Eqs. 4 and 5 simultaneously, current at the transmitting side becomes

$$I_1 = \frac{\begin{vmatrix} V_0 & j\omega M \\ 0 & Z_l + r_2 + j\omega L_2 \end{vmatrix}}{\begin{vmatrix} r_1 + j\omega L_1 & j\omega M \\ j\omega M & Z_l + r_2 + j\omega L_2 \end{vmatrix}} = \frac{(Z_l + r_2 + j\omega L_2)V_0}{(r_1 + j\omega L_1)\ (Z_l + r_2 + j\omega L_2) + \omega^2 M^2}. \tag{6}$$

The relationship between magnitude of the currents at the transmitting and receiving sides can be derived as

$$|I_2|^2 = \frac{\omega^2 M^2}{r_2^2 + (X_2 + \omega L_2)^2}|I_1|^2. \tag{7}$$

The input impedance looking into the circuit from the voltage source is derived from Eq. 6 as

$$Z_{in} = \frac{V_0}{I_1} = r_1 + j\omega L_1 + \frac{\omega^2 M^2}{Z_l + r_2 + j\omega L_2}. \tag{8}$$

Then, the real part of Z_{in} becomes

$$R_{in} = r_1 + \frac{\omega^2 M^2 (R_l + r_2)}{(R_l + r_2)^2 + (X_l + \omega L_2)^2}, \tag{9}$$

where R_l and X_l are the real and imaginary parts of Z_l, respectively.

The total power consumed in the circuit generated by voltage V_0 is

$$P_{in} = \frac{1}{2} R_{in} |I_1|^2, \tag{10}$$

and the power consumed in the load Z_l is

$$P_l = \frac{1}{2} R_l |I_2|^2. \tag{11}$$

Therefore, the power transfer efficiency is evaluated as

DOI 10.1007/978-981-4560-75-7_91-1

$$\eta = \frac{P_l}{P_{in}} \tag{12}$$

$$\frac{\omega^2 M^2 R_l}{r_1\left((R_l + r_2)^2 + (X_l + \omega L_2)^2\right) + \omega^2 M^2 (R_l + r_2)}. \tag{13}$$

Now, the load resistance R_l and the load reactance X_l are optimized to achieve maximum power transfer efficiency. For this purpose, the partial derivatives of the efficiency with respect to R_l and X_l should satisfy the conditions

$$\frac{\partial \eta}{\partial X_l} = 0, \tag{14}$$

and

$$\frac{\partial \eta}{\partial R_l} = 0. \tag{15}$$

The optimized value of R_l is obtained by solving the above two equations as

$$R_{lm} = r_2\sqrt{1+\alpha}, \tag{16}$$

and the optimized value of X_l is also obtained as

$$X_{lm} = -\omega L_2, \tag{17}$$

where α is defined as

$$\alpha = \frac{\omega^2 M^2}{r_1 r_2}. \tag{18}$$

Therefore, when the load impedance Z_l satisfies both Eqs. 16 and 17, the power transfer efficiency η reaches to the maximum, expressed as

$$\eta_m = 1 - \frac{2}{1+\sqrt{1+\alpha}} = \frac{\left(1-\sqrt{1+\alpha}\right)^2}{\alpha}. \tag{19}$$

This analysis indicates that the maximum power transfer efficiency can be achieved by the load impedance which satisfies both Eq. 16 for real part and Eq. 17 for imaginary part. Because Eq. 17 is the condition of resonance in receiving circuit, this analysis indicates that the maximum power transfer efficiency can be achieved if both Eqs. 16 and 17 are satisfied simultaneously, while the resonance condition is necessary but not enough. It is also noted that the optimized load resistance is related not only with the resistance of inductor in the receiving side but also with that in the transmitting side through the parameter α which indicates the mutual coupling effect between these two inductors. α can be also expressed as

DOI 10.1007/978-981-4560-75-7_91-1

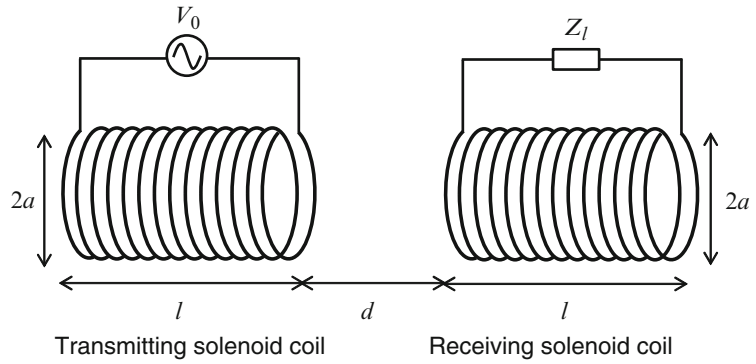


Fig. 4 Analysis model of wireless charging system by using two solenoid coils

$$\alpha = k^2 Q_1 Q_2,$$

where k is the coupling coefficient,

$$k = \frac{M}{L_1 L_2}, \tag{20}$$

and Q_1 and Q_2 are the quality factors

$$Q_1 = \frac{\omega L_1}{r_1}, \tag{21}$$

and

$$Q_2 = \frac{\omega L_2}{r_2}. \tag{22}$$

Therefore, the maximum power transfer efficiency can be expressed as

$$\eta_m = 1 - \frac{2}{1 + \sqrt{1 + k^2 Q_1 Q_2}}, \tag{23}$$

which shows that the maximum power transfer efficiency is dependent on coupling factor k and the quality factors Q_1 and Q_2. A large quality factor and a strong coupling between the coupled inductors will increase the value of the maximum efficiency.

The condition for obtaining the maximum efficiency is not given in the transmitting circuit in the above discussion. It is because the power transfer efficiency is normalized by the input power to the circuit, rather than the incident power in this section. If the efficiency is defined as the absorbed power over the incident power, the efficiency optimization should be made by defining an internal impedance of the power source and then optimizing the internal impedance in the same way as discussed for the receiving circuit.

As a numerical example, a wireless power transfer by two solenoid coils is analyzed by using circuit theory. The analysis model is shown in Fig. 4. Two solenoid coils are used as the antennas for wireless power transfer. The solenoid coils have a length l, N turns per meter, and radius a. The coils are made of conducting wire with radius R_a and conductivity σ. The two coils are separated over a distance d.

DOI 10.1007/978-981-4560-75-7_91-1

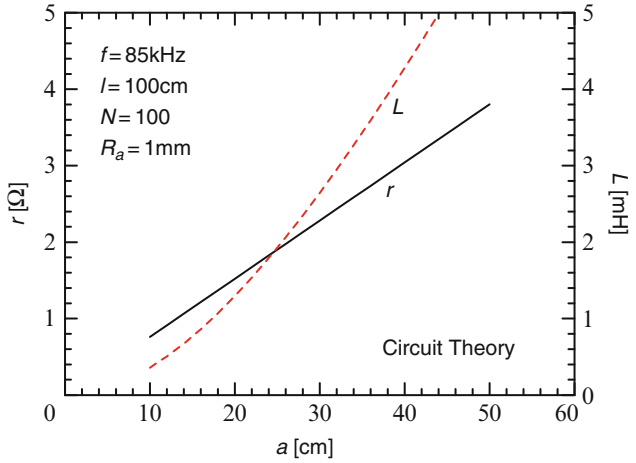


Fig. 5 Equivalent resistance and inductance of the solenoid coils

When the solenoid coil is extremely small compared with wavelength, it can be assumed the current on the solenoid coil is uniformly in-phase distributed on the surface of the conducting wire. The inductance L of the solenoid coil is then given in a closed form as

$$L = \mu_0 n^2 \pi a^2 \left(\sqrt{a^2 + l^2} - a \right), \tag{24}$$

where μ_0 is permeability in vacuum. The resistance r of the coil is derived as

$$r = \frac{2\pi a N}{2\pi R_a \delta \sigma} = \frac{aN}{R_a \delta \sigma}, \tag{25}$$

where δ is skin depth of the wire expressed as

$$\delta = \sqrt{\frac{2}{\omega \mu_0 \sigma}}. \tag{26}$$

It is difficult to give a closed form to express exactly the mutual inductance M between the coils. For simplicity, it is assumed here that the distance between coils is small enough that the coupling factor becomes $k = 1$. Therefore, the mutual inductance M is equal to L.

Consider the case that $f = 85$ kHz, $l = 100$ cm, $N = 100$, $R_a = 1$ mm, and $\sigma = 5.8 \times 10^7$ S/m. Figure 5 shows the equivalent inductance and resistance of the solenoid coils which are calculated using Eqs. 24 and 25. As the radius of the solenoid coil increases, both the inductance and resistance increase as well. Corresponding to the change of radius, the optimized load impedance to achieve the maximum power transfer efficiency can be obtained by using Eqs. 16 and 17, as shown in Fig. 6. The maximum power transfer efficiency, when the load impedance is optimized, is shown in Fig. 7, which is calculated by using Eq. 23. It is shown that the maximum power transfer efficiency is as high as nearly 100 % because the coupling factor between coils is assumed to be 1. Based on this assumption, the maximum power transfer efficiency is independent of the distance between coils. Therefore, the numerical analysis using the circuit theory is limited to simple analysis models where the lumped circuit elements can be accurately evaluated. In practice, because geometry of both transmitting and receiving antennas would be very complicated, it is usually impossible to obtain the exact characteristics of those lumped circuit elements such as the

DOI 10.1007/978-981-4560-75-7_91-1

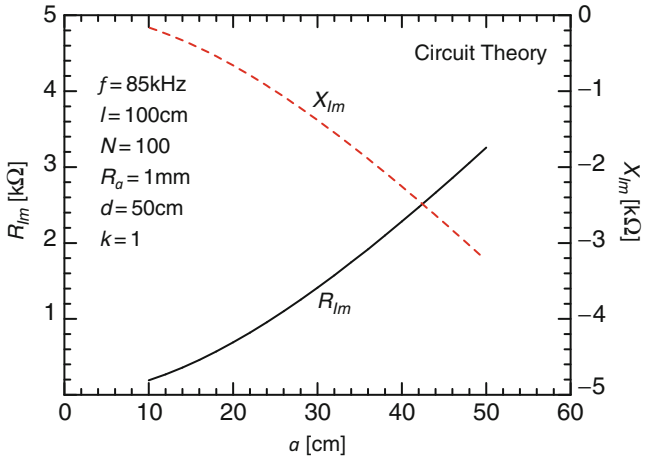


Fig. 6 Optimized load impedance to achieve the maximum power transfer efficiency

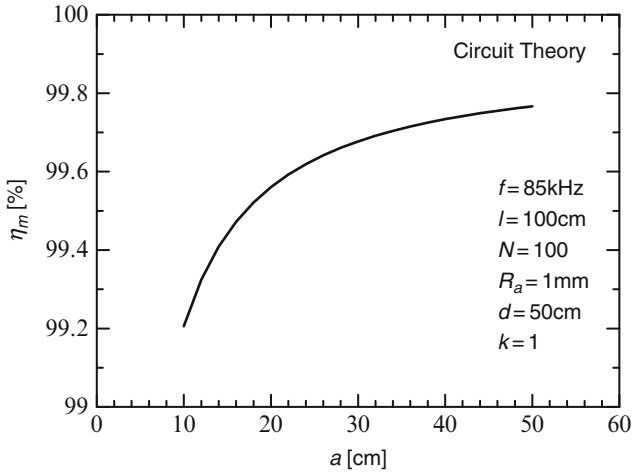


Fig. 7 The maximum power transfer efficiency when the optimized impedance is loaded

inductance and resistance of the coils, as well as the mutual coupling factor in the equivalent circuit. In the following sections, some approaches based on a full-wave electromagnetic analysis are introduced which is capable of accurately analyzing the antennas with complicated geometry and moderate electrical size in the wireless charging system.

Impedance Matrix Approach

In this section, the near-field coupling is modeled by using the impedance matrix approach. The wireless charging system can be expressed in a form of two-port network shown in Fig. 8. Here, Z_s is the internal impedance of the voltage source and Z_l is the load impedance. Both of them have a complex value, expressed as $Z_s = R_s + jX_s$ and $Z_l = R_l + jX_l$, respectively. The relation between the port currents I_1, I_2 and port voltages V_1, V_2 is given in terms of the impedance matrix,

DOI 10.1007/978-981-4560-75-7_91-1

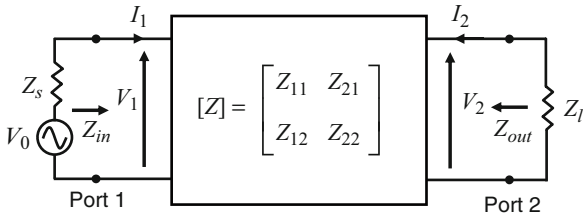


Fig. 8 Two-port network for modeling WPT using magnetic field coupling

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \tag{27}$$

If the geometry of the wireless charging system is known, the matrix impedance in Fig. 8 can be obtained by using some numerical methods based on full-wave electromagnetic analysis.

If Z_{in} is the impedance looking into Port 1 and Z_{out} is the impedance looking into Port 2, they are expressed in the form of impedance parameters (Z-parameters) as

$$Z_{in} = Z_{11} - \frac{Z_{12}Z_{21}}{Z_l + Z_{22}}, \tag{28}$$

and

$$Z_{out} = Z_{22} - \frac{Z_{12}Z_{21}}{Z_s + Z_{11}}. \tag{29}$$

The condition of conjugate impedance matching applied to both ports is expressed as

$$Z_{in} = Z_s^*, \tag{30}$$

and

$$Z_{out} = Z_l^*. \tag{31}$$

By solving these two equations simultaneously, the matched load impedance becomes

$$R_{lmat} = R_{22}\sqrt{(1-\alpha_1)\ (1+\alpha_2)}, \tag{32}$$

$$X_{lmat} = R_{22}\sqrt{\alpha_1\alpha_2} - X_{22}. \tag{33}$$

The matched internal impedance becomes

$$R_{smat} = R_{11}\sqrt{(1-\alpha_1)\ (1+\alpha_2)}, \tag{34}$$

$$X_{smat} = R_{11}\sqrt{\alpha_1\alpha_2} - X_{11}, \tag{35}$$

where $\alpha_1 = R_{12}^2/(R_{11}R_{22})$ and $\alpha_2 = X_{12}^2/(R_{11}R_{22})$.

When the power transfer efficiency is definted as the absorbed power by the load impedance over the input power (not incident power!) to the Port 1, it can be expressed as

DOI 10.1007/978-981-4560-75-7_91-1

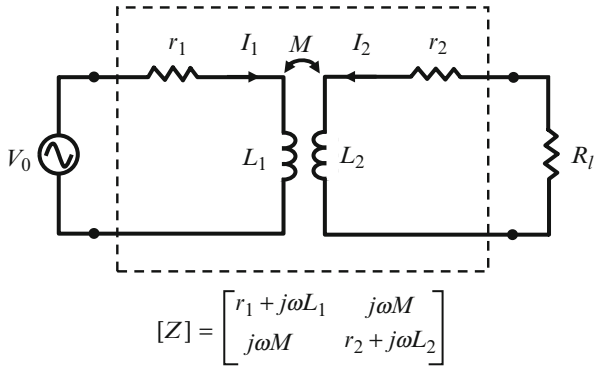


Fig. 9 Two-port impedance matrix network for modeling wireless charging system using magnetic field coupling

$$\eta = \frac{R_l |I_2|^2}{R_{in} |I_1|^2}. \tag{36}$$

By using relation between currents $|I_2|^2$ and $|I_1|^2$

$$|I_2|^2 = \frac{|Z_{12}|^2}{|Z_l + Z_{22}|^2} |I_1|^2, \tag{37}$$

the power transfer efficiency can be calculated as

$$\eta = \frac{R_l |Z_{12}|^2}{R_{in} |Z_l + Z_{22}|^2}. \tag{38}$$

If impedance matching condition is satisfied at Port 2, substituting the matched load impedance R_{lmat} in Eq. 32 and X_{lmat} in Eq. 33 for Z_l, the power transfer efficiency in terms of the input power becomes

$$\eta_{mat} = \frac{\alpha_1 + \alpha_2}{\left(\sqrt{1 - \alpha_1} + \sqrt{1 + \alpha_2}\right)^2}. \tag{39}$$

When the impedance matrix approach is applied to the analysis model in Fig. 3 in section "Circuit Theory," as shown in Fig. 9, the impedance matrix for the two-port network becomes

$$[Z] = \begin{bmatrix} r_1 + j\omega L_1 & j\omega M \\ j\omega M & r_2 + j\omega L_2 \end{bmatrix}. \tag{40}$$

Then, $\alpha_1 = 0$ and $\alpha_2 = \omega^2 M^2/(r_1 r_2) = \alpha$. If it is assumed that the impedance matching condition is satisfied at Port 2, the load impedance can be derived by using Eqs. 32 and 33 as

$$R_{lmat} = r_2 \sqrt{1 + \alpha}, \tag{41}$$

and

$$X_{lmat} = -\omega L_2. \tag{42}$$

DOI 10.1007/978-981-4560-75-7_91-1

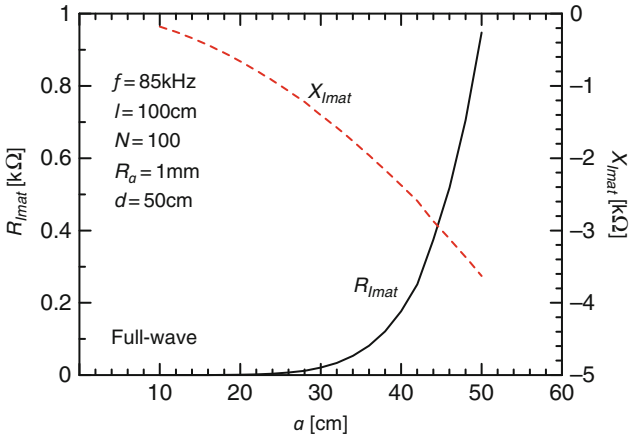


Fig. 10 Matched load impedance versus coil radius for the model using two solenoid coils

The power transfer efficiency in terms of the input power can be derived as

$$\eta_{mat} = \frac{\left(1 - \sqrt{1+\alpha}\right)^2}{\alpha}. \tag{43}$$

Equations 41, 42, and 43 agree completely with Eqs. 16, 17, and 19, demonstrating the fact that the impedance matching condition is nothing but the condition to achieve the maximum power transfer efficiency. Therefore, in developing a wireless charging system based on the near-field coupling approach, the biggest challenge is to keep the impedance matching condition being satisfied when the transmitting antenna and receiving antenna are moving relative to each other. It is required in the practical system design to adaptively change the antenna geometry or the impedance matching network to keep satisfying the impedance matching condition at both transmitting and receiving ports.

Here, the analysis model shown in Fig. 4 is again solved by using the impedance matrix approach, but the impedance matrix is numerically calculated by using the method of moments, one of the most effective full-wave electromagnetic analysis methods.

Corresponding to radius of the solenoid coils, the optimized load impedance to achieve the maximum power transfer efficiency can be obtained by using Eqs. 41 and 42, as shown in Fig. 10. Compared with the results shown in Fig. 6, the reactance is almost the same, but the resistance is very different. The maximum power transfer efficiency, which is calculated by using Eq. 43 under the impedance match condition, is shown in Fig. 11. Efficiency drops greatly when radius of the coil becomes small, which is very different from that obtained in the previous section. The reason for these large differences in calculated efficiency is due to the fact that mutual coupling between coils, radiation loss, as well as conducting loss of the coils are approximately calculated by using the analytical forms in section "Circuit Theory" but accurately calculated by the method of moments in this section. When the distance between the coils changes, the matched load impedance is shown in Fig. 12, and the maximum efficiency under the condition of impedance matching is shown in Fig. 13. These analysis results provide important knowledge to design antenna load impedance and evaluate the maximum power transfer efficiency of the wireless charging system.

DOI 10.1007/978-981-4560-75-7_91-1

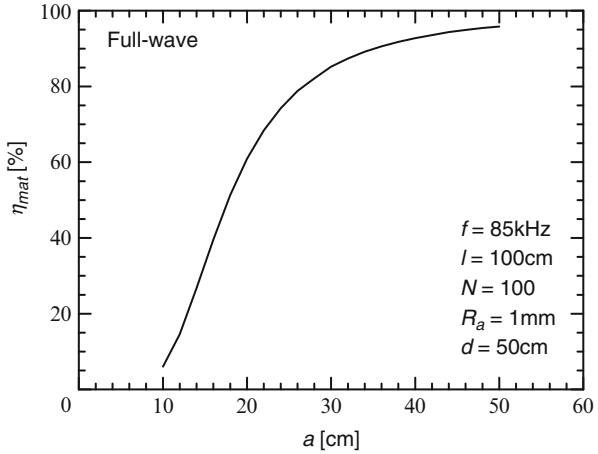


Fig. 11 Maximuam power transfer efficiency versus coil radius when load impedance is matched

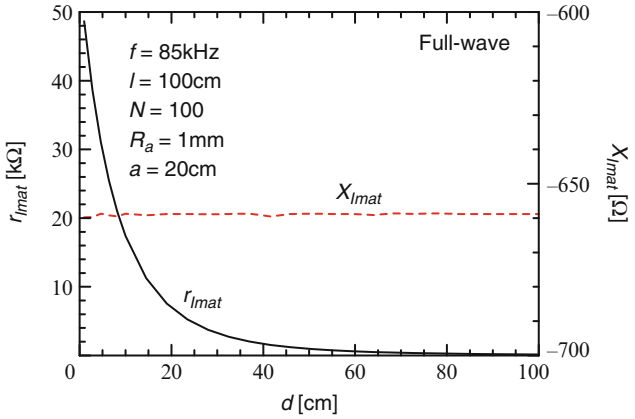


Fig. 12 Matched load impedance versus coil distance

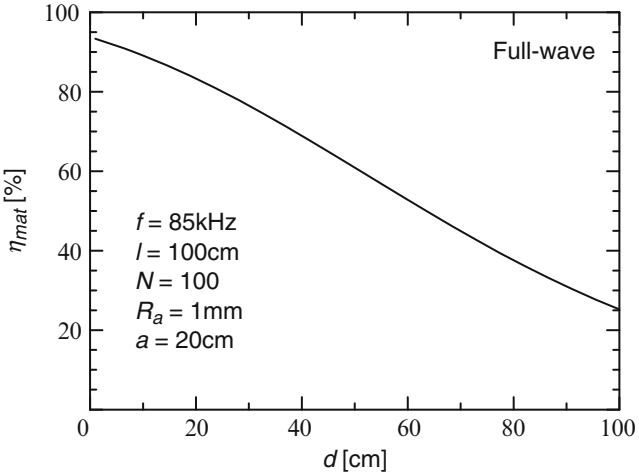


Fig. 13 Maximum power transfer efficiency versus coil distance when load impedance is matched

DOI 10.1007/978-981-4560-75-7_91-1

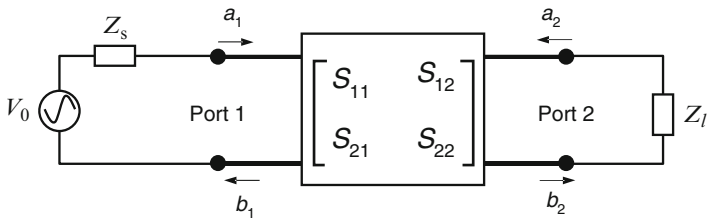


Fig. 14 Two-port network for modeling antennas in wireless charging systems

Scattering Matrix Approach

The antennas in wireless charging system can be also analyzed by using the scattering matrix approach as shown in Fig. 14, where the system consisting of transmitting port and receiving port is again expressed as a two-port network as in the last section. Port 1 is the transmitting port including the transmitting antenna and its load impedance Z_s, and Port 2 is the receiving port including receiving antenna and its load impedance Z_l. Here, instead of impedance matrix, the scattering matrix S is used to express the two-port network properties. The scattering parameters are very effective in evaluating the performance of multiport network, which are easily measured using a vector network analyzer and calculated by many commercial full-wave electromagnetic simulators.

The reflection coefficients of the internal impedance Z_s and load impedance Z_l are

$$\Gamma_s = \frac{Z_s - Z_0}{Z_s + Z_0}, \tag{44}$$

and

$$\Gamma_l = \frac{Z_l - Z_0}{Z_l + Z_0}, \tag{45}$$

respectively, where Z_0 represents the reference impedance of 50 Ω, the same with that in the scattering matrix. The condition of conjugate impedance matching becomes

$$\Gamma_s = \Gamma_{in}^*, \tag{46}$$

$$\Gamma_l = \Gamma_{out}^*, \tag{47}$$

where Γ_{in} and Γ_{out} are the reflection coefficient at Port 1 and Port 2, respectively, and are expressed as

$$\Gamma_{in} = S_{11} + \frac{S_{12}S_{21}\Gamma_l}{1 - S_{22}\Gamma_l} \tag{48}$$

and

$$\Gamma_{out} = S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \tag{49}$$

The optimal reflection coefficients for maximum power transfer efficiency can be obtained by solving Equations 46 and 47 as

DOI 10.1007/978-981-4560-75-7_91-1

$$\Gamma_{sm} = \frac{B_1 \pm \sqrt{{B_1}^2 - 4|C_1|^2}}{2C_1}, \tag{50}$$

and

$$\Gamma_{lm} = \frac{B_2 \pm \sqrt{B_2^2 - 4|C_2|^2}}{2C_2}, \tag{51}$$

where

$$B_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2, \tag{52}$$

$$C_1 = S_{11} - \Delta S_{22}^*, \tag{53}$$

$$B_2 = 1 + |S_{22}|^2 - |S_{11}|^2 - |\Delta|^2, \tag{54}$$

$$C_2 = S_{22} - \Delta S_{11}^*, \tag{55}$$

$$\Delta = S_{11}S_{22} - S_{12}S_{21}. \tag{56}$$

The optimal impedance of the voltage source and load impedance can be obtained using Eqs. 44 and 45 from those optimal reflection coefficients given by Eqs. 50 and 51.

On the other hand, the power transfer efficiency can be expressed in terms of the S parameters and the reflection coefficients at the input and output port,

$$\eta = \frac{|b_2|^2 - |a_2|^2}{|a_1|^2} = \frac{\left(1 - |\Gamma_l|^2\right)\ \left(1 - |\Gamma_s|^2\right)|S_{21}|^2}{|(1 - S_{11}\Gamma_s)(1 - S_{22}\Gamma_l) - S_{12}S_{21}\Gamma_s\Gamma_l|^2} \tag{57}$$

where the efficiency can be calculated if the values of Z_s and Z_l are given while the maximum efficiency is available when Z_s and Z_l are the optimal impedance, namely, complex conjugate matched impedance.

Antenna Modeling and Analysis Using Scattering Matrix Approach

Antenna Models

Two kinds of antennas are studied; dipole antenna and loop antenna. When these antennas are used for near-field coupling, a small dipole antenna is used for electric field coupling while a small loop antenna is used for magnetic field coupling. Although these conical models of the antennas are used in the study, these results and observations from the numerical simulations provide theoretical insight into how to design antennas for wireless charging systems based on the near-field coupling approach.

Figure 15 shows a dipole–dipole model. The length of the wire dipole is $2l$; the antenna distance is d. The two antennas are located parallel with the z axis. The internal impedance of the voltage source is Z_s, and the load impedance connected with the receiving antenna is Z_l. Figure 16 gives a loop–loop model. The circular loop antenna has a diameter of D separated at distance of d. The axes of two loop antennas coincide with x axis. All the antennas are made of conducting wire of radius a.

DOI 10.1007/978-981-4560-75-7_91-1

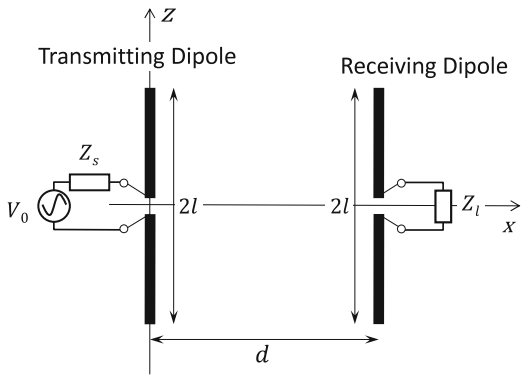


Fig. 15 Dipole–dipole model for wireless charging system

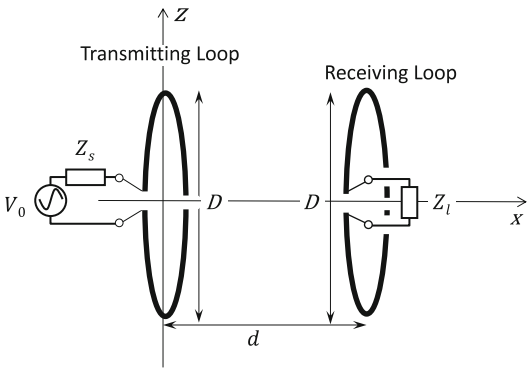


Fig. 16 Loop–loop model for wireless charging system

In this chapter, the method of moments is used to numerically analyze the antennas to calculate the S parameters of the two-port network.

Maximum Power Transfer Efficiency Without Consideration of Loss in Antennas and Matching Circuits

When the antenna size is fixed, the maximum power transfer efficiency under the condition of conjugate impedance matching at the two ports is shown by changing the antenna distance d. The numerical results are shown in Fig. 17. Here, the dipole length $2l$ and loop diameter D are 0.1λ, and it is assumed that the antenna is made of electrically perfect conductor. It is found that the power transfer efficiency rises up to 100 % when transmitting and receiving antennas are coming close to each other. However, in the practical wireless charging system, the power transfer efficiency is degraded by the power loss which includes antenna conducting loss, radiation loss, and return loss of antennas. When the antennas are electrically small and antenna distance is electrically small, the radiation power is much less than the coupled power. Therefore, if the condition of impedance matching at the two ports is satisfied and antenna conducting loss is not considered, it is reasonable that the power transfer efficiency should approach to 100 %. The efficiency decreases when the distance becomes large because the radiation power from the transmitting and receiving antennas increases when the two antennas are moved far away from each other.

In Fig. 18, the relationship between the maximum power transfer efficiency and antenna size is shown. It is found that the maximum power transfer efficiency could be increased by reducing the antenna size to decrease the radiation power. Theoretically speaking, the result is true if the conducting loss is ignored. However, as shown in the following discussion, the conducting loss of the antenna, as well as the

DOI 10.1007/978-981-4560-75-7_91-1

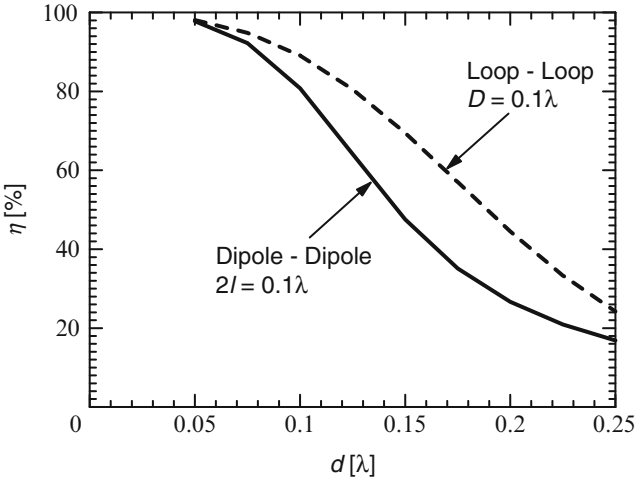


Fig. 17 The maximum power transfer efficiency as a function of antenna distance for dipole–dipole and loop–loop models

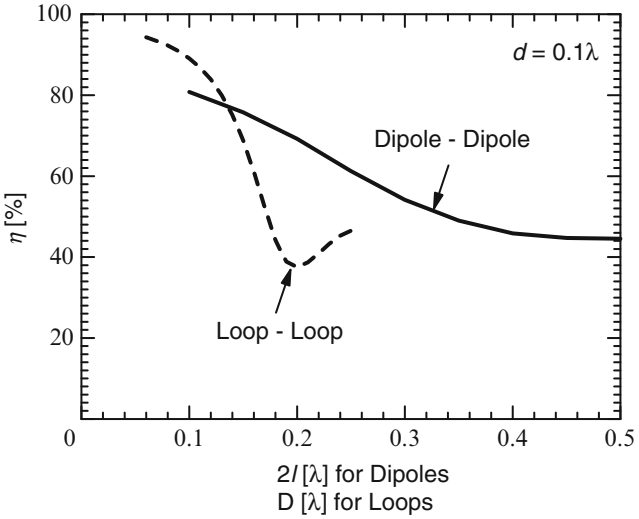


Fig. 18 The maximum power transfer efficiency as a function of antenna size for dipole–dipole and loop–loop models without conducting loss in antennas

conducting loss in the matching circuits of antennas, may dominate the loss when the antenna size becomes extremely small and decrease the maximum power transfer efficiency on the contrary. The efficiency of loop–loop drops more than that of dipole–dipole because the loop has a larger antenna effective aperture and radiates more power than the dipole in the case of $D = 2l$.

Maximum Power Transfer Efficiency with Consideration of Loss in Antennas and Matching Circuits

It was known that antennas in the wireless charging system should be electrically small to reduce the radiation loss as discussed above. However, the conducting loss of the antennas cannot be ignored any more as the antenna is electrically small. Furthermore, the ohmic loss in the matching circuit should be considered in analyzing the maximum power transfer efficiency. The following numerical results show the importance of considering ohmic loss in antennas as well as in impedance matching circuits.

DOI 10.1007/978-981-4560-75-7_91-1

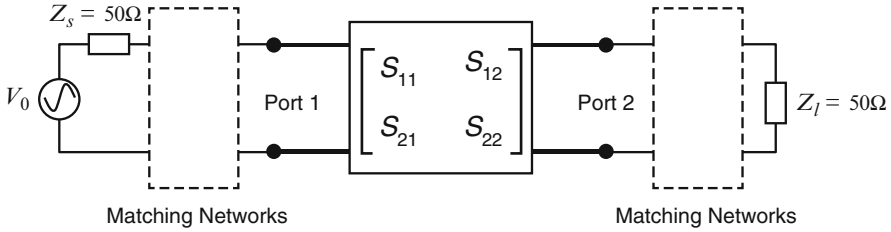


Fig. 19 Two-port network combined with impedance matching circuits for modeling WPT system

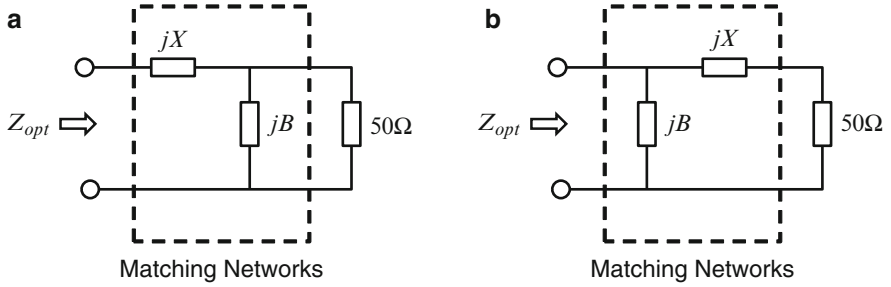


Fig. 20 L-section impedance matching circuits

The impedance matching networks are added into the two-port network to change the optimal impedance to 50 Ω at both sides as shown in Fig. 19. There are many types of the impedance matching networks such as *L*-section matching, quarter wave transformer, and single-stub tuning. *L*-section matching uses two reactive elements to match an arbitrary impedance to an arbitrary desired impedance. This technique is very effective in low-frequency application. In this chapter, the *L*-section matching networks as shown in Fig. 20 are used to match the optimized impedance Z_{opt}, which is required to maximize the efficiency. In this case, the internal impedance of voltage source and the load impedance are assumed to be 50 Ω. There are two configurations for an *L*-section matching network as shown in Fig. 20. Assume that Z_{opt} has a real part of R_{opt}. If the normalized value of $R_{opt}[\Omega]/50$ is larger than unit, the configuration (a) should be selected while if the $R_{opt}[\Omega]/50$ is smaller than unit, the configuration (b) should be used. The reactive elements X and B may be either an inductance L or a capacitance C, depending on the value of normalized Z_{opt}. In the wireless charging applications, because the frequency is usually very low and the matching circuit size is electrically small, lumped element inductors and capacitors can be used. The analytic solutions for the reactive elements X and B are available in reference (Pozar 1998).

The ohmic loss of these lumped elements in the impedance matching circuits is expressed in terms of Q value. It is found that the Q value of a capacitor is generally higher than that of an inductor. A typical Q value of a commercial inductor and capacitor is assumed to be $Q_L = 100$ and $Q_C = 200$, respectively, in this chapter. The antenna conductor is assumed to be copper with its conductivity of $\sigma_{Cu} = 5.8 \times 10^7$ S/m.

Figures 21 and 22 show the relationship between the maximum efficiency and antenna distance when antenna size is fixed for cases of dipole–dipole and loop–loop transmission, respectively. In these figures, the results are compared when only copper loss is considered and when both copper loss and loss of matching circuits are considered. It is found that the copper loss does not degrade the efficiency much, while loss of matching network degrades the efficiency significantly. When transmission distance is the same and the two types of antennas have approximately the same electrical size, the maximum efficiency of loop–loop transmission is larger than that of dipole–dipole transmission. Therefore, the loop–loop

DOI 10.1007/978-981-4560-75-7_91-1

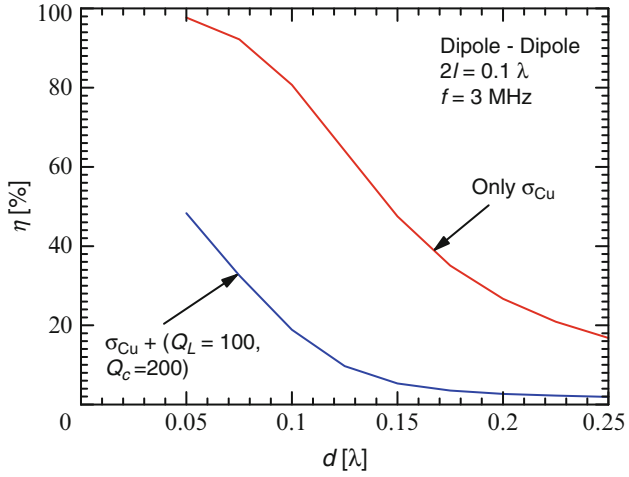


Fig. 21 The maximum power transfer efficiency as a function of antenna distance d for dipole–dipole model when ohmic loss of antennas and impedance matching circuits are considered

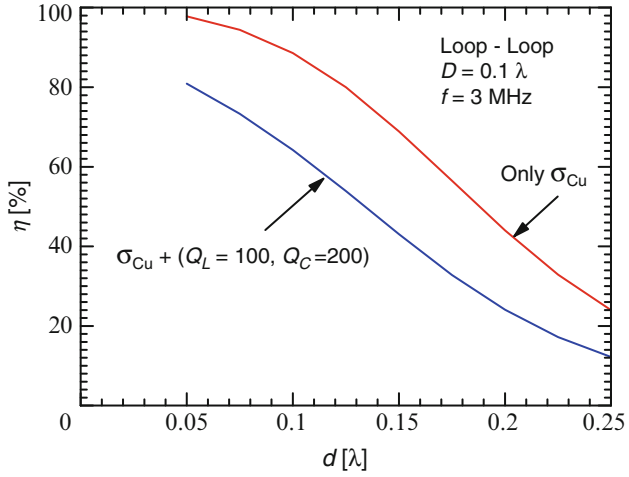


Fig. 22 The maximum power transfer efficiency as a function of antenna distance d for loop–loop model when ohmic loss of antennas and impedance matching circuits are considered

transmission is usually superior to the dipole–dipole transmission in terms of efficiency if the impedance matching condition is satisfied.

Figures. 23 and 24 show the relationship between the maximum efficiency and antenna size when antenna distance is fixed to be 0.1 λ for dipole–dipole and loop–loop transmission, respectively. When the ohmic loss of impedance matching circuits is considered, the efficiency of the dipole system decreases rapidly as the antenna size becomes small. However, for the loop–loop configuration, there is an optimal loop size, in the range of $D = 0.1\ \lambda$ to 0.15 λ. This is because the efficiency is reduced due to the increase of the radiation loss when D becomes large but is also reduced due to the increase of the ohmic losses when D is small. Thus, an optimal loop size exists in the middle. Numerical results of investigation on the loss of power transfer system are shown in Figs. 25 and 26, which give the fraction of loss as a function of antenna size for dipole–dipole and loop–loop models, respectively. Here, a little mismatch loss is observed because the impedance matching circuit was designed when ohmic loss in the matching circuits was omitted. It is found that the loss in the matching circuits dominates the loss and is the most important

DOI 10.1007/978-981-4560-75-7_91-1

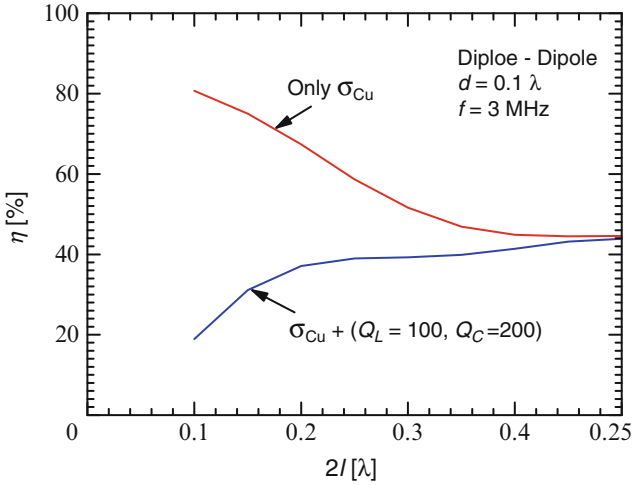


Fig. 23 The maximum power transfer efficiency as a function of antenna size for dipole–dipole model when ohmic loss of antennas and impedance matching circuits are considered

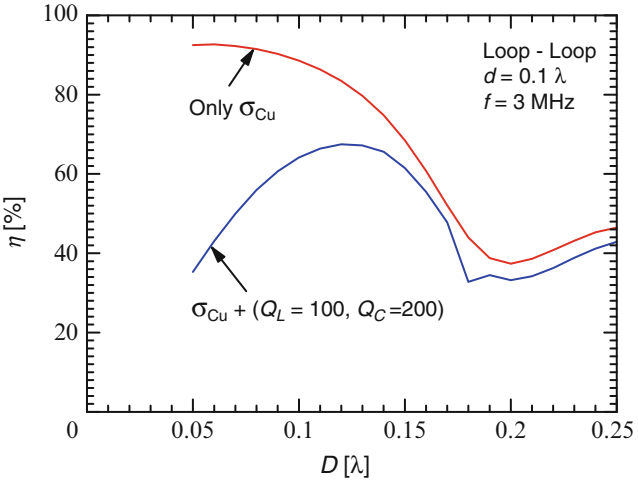


Fig. 24 The maximum power transfer efficiency as a function of antenna size for loop–loop model when ohmic loss of antennas and impedance matching circuits are considered

factor degrading the power transfer efficiency when antennas are electrically small. When antenna size becomes large, the radiation loss increases rapidly and becomes a dominant factor to decrease the efficiency.

Recent Application Developments, Standards, and Challenges

Development and Challenges on Antenna Technology for WPT

As described above, the impedance matching for both transmitting and receiving antennas is required to obtain the maximum value of efficiency of the WPT. However, it is not easy to keep the impedance matching condition for mobile devices where the transmitting antenna and receiving antenna are moving relative to each other. In near-field WPT, because the transmitting antennas and receiving antennas are strongly coupled with each other, a little variation of antenna positions and direction of antenna

DOI 10.1007/978-981-4560-75-7_91-1

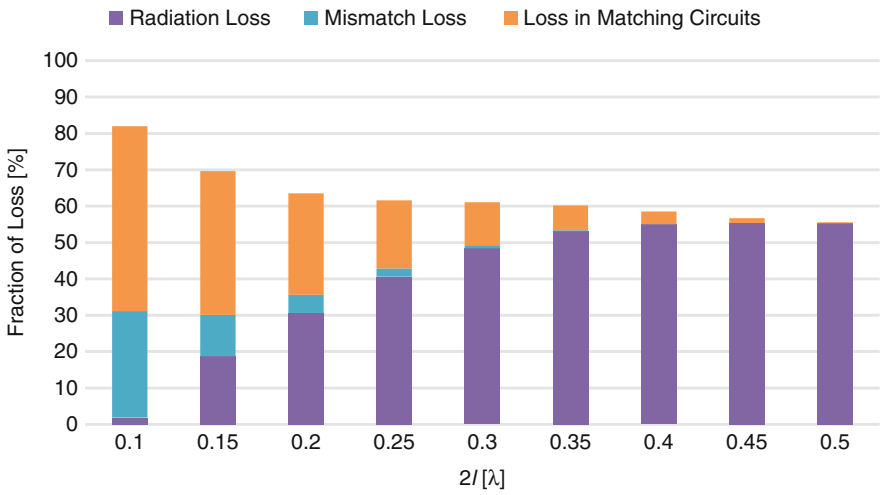


Fig. 25 Fraction of loss as a function of antenna size for dipole–dipole model when ohmic loss of antennas and impedance matching circuits are considered

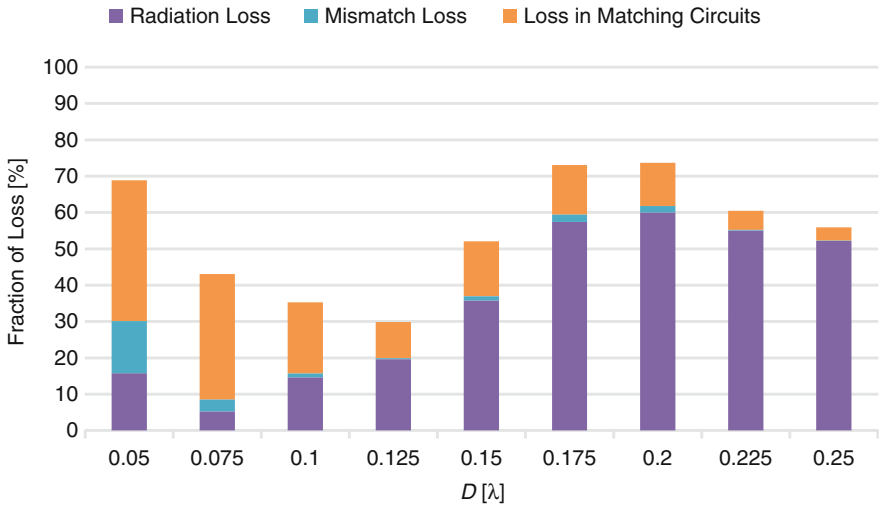


Fig. 26 Fraction of loss as a function of antenna size for loop–loop model when ohmic loss of antennas and impedance matching circuits are considered

polarization may cause a large variation of input impedance of these antennas, significantly decreasing the efficiency. Therefore, there are some researches focused on how to solve the problem of antenna misalignment.

There are two technical approaches to solve this problem. One is to use an impedance matching circuit to adaptively change the impedance at the input and output ports to compensate the impedance mismatching. For example, in Lee et al. (2014), a WPT system with a switchable capacitor array circuit has been presented, for compensating the variation of mutual coupling between two coils. It was demonstrated that the return loss can be improved greatly due to the operation of the capacitor array circuit when the distance between the two coils is changed. This approach was also introduced in Lee et al. (2012) and Aldhaher et al. (2014).

Another technical solution is to design an antenna that performs well even if it is misaligned. For example, a 3D loop with orthogonal polarizations is proposed as a WPT antenna to compensate the

DOI 10.1007/978-981-4560-75-7_91-1

efficiency drop due to the angular and lateral misalignment (Jonah et al. 2013). Also, a coil array was applied to a secondary coil instead of typical single coil to secure a stable power transmission efficiency for implantable devices (Ahn et al. 2014).

It is a big challenge to design antennas and AC-DC conversion circuits with high transmission efficiency for mobile devices. More studies and researches are expected to make a practical near-field WPT system.

Some WPT Standards

It is very important to the wireless charging industry to create a wireless power charging standard on a global scale that can allow electronic products and charging stations to be universally compatible with each other. Currently, several global standards for wireless power transfer have been developed. Some of them are introduced in the following.

Wireless Power Consortium (WPC) developed a standard for wireless charging technology, called "Qi" (Hui 2013). The WPC, established in 2008, is an open-membership cooperation of Asian, European, and American companies in diverse industries, including electronics manufacturers and original equipment manufacturers. The WPC published the Qi for products with an output power of 5 W or less in 2009. The WPC began to extend Qi specifications for medium power-consuming devices delivering power less than 120 W. The Qi specifications are available as free public download from the WPC website.

The Alliance for Wireless Power (A4WP) developed an interface standard called Rezence for wireless electrical power transfer based on the near-field coupling. The A4WP was formed in May 2012 with a goal to create a WPT system using Rezence technology, which consists of a single power transmitter unit and one or more power receiver units for wireless power transfer. The A4WP announced an expansion of the Rezence standard up to 50 W in June 2014, which is expected to expand the range of wireless power transfer for those products including laptops, tablets, and other consumer electronics.

The Power Matters Alliance (PMA), founded in March 2012, is a trade association whose mission is to create a better power paradigm for battery-equipped devices using wireless charging technology. The membership of the PMA is made up of companies across a diverse set of industries including telecommunication, consumer devices, automotive, retail, furniture, surfaces, and more, in order to guarantee consumers interoperable devices that employ wireless power technology. The PMA interface standard describes analog power transfer (inductive and resonant), digital transceiver communication, cloud-based power management, and environmental sustainability. The IEEE established the Wireless Power and Charging Systems Working Group (WPCS-WG) in October 2013, to develop the IEEE P2100.1 Standard Specifications for Wireless Power and Charging Systems following the creation of the PMA. The A4WP and the PMA have signed an agreement aimed at establishing global interoperability of the two standards in February 2014.

Activities of WPT standardization are also very active in Japan in recent years. IEICE Technical Committee on Wireless Power Transfer provides a platform for exchanging new ideas at the frontier of WPT researches. A consortium called Wireless Power Transfer Consortium for Practical Application (WiPoT) was founded in 2013 to accelerate the development of practical applications from the fundamental studies of WPT.

Summary

It was described how to investigate antenna parameters from the viewpoint of antenna theory to obtain the maximum power transfer efficiency in the wireless charging system based on the near-field coupling approach. The antenna parameters include geometry of transmitting and receiving antennas, electrical size

of antennas, impedance matching for antennas, distance between transmitting and receiving antennas, conductor loss (ohmic loss) of antennas, ohmic loss of matching circuits, and so on, in order to clarify the relationship between the maximum efficiency and these parameters. The approaches introduced in this chapter would be helpful to design and develop antennas in the wireless charging systems, and the simulation results can provide theoretical insight to how to optimize antenna design to improve the power transfer efficiency.

Acknowledgments

Authors are grateful for assistance given by Mr. Brock Delong of Ohio State University and Mr. Shun Maruyama of Tohoku University.

Cross-References

- ▶ Antennas in Microwave Wireless Power Transmission
- ▶ Near-Field Antenna Measurement Techniques
- ▶ Numerical Modeling in Antenna Engineering
- ▶ Small Antennas (PIFA/PILA/Loading Antenna/etc)

References

Ahn HR, Kim MS, Kim YJ (2014) Inductor array for minimizing transfer efficiency decrease of wireless power transmission components at misalignment. Electron Lett 50(5):393–394. doi:10.1049/El.2013.4087

Aldhaher S, Luk PCK, Whidborne JF (2014) Electronic tuning of misaligned coils in wireless power transfer systems. IEEE Trans Power Electron 29(11):5975–5982. doi:10.1109/Tpel.2014.2297993

Brown WC (1984) The history of power transmission by radio-waves. IEEE Trans Microw Theory Tech 32(9):1230–1242. doi:10.1109/Tmtt.1984.1132833

Cannon BL, Hoburg JF, Stancil DD, Goldstein SC (2009) Magnetic resonant coupling as a potential means for wireless power transfer to multiple small receivers. IEEE Trans Power Electron 24(7):1819–1825. doi:10.1109/Tpel.2009.2017195

Chen Q, Ozawa K, Yuan QW, Sawaya K (2012) Antenna characterization for wireless power-transmission system using near-field coupling. IEEE Antennas Propag Mag 54(4):108–116

Hatanaka K, Sato F, Matsuki H, Kikuchi S, Murakami J, Kawase M, Satoh T (2002) Power transmission of a desk with a cord-free power supply. IEEE Trans Magn 38(5):3329–3331. doi:10.1109/Tmag.2002.803120

Hui SY (2013) Planar wireless charging technology for portable electronic products and Qi. Proc IEEE 101(6):1290–1301. doi:10.1109/Jproc.2013.2246531

Ishizaki T, Komori T, Ishida T, Awai I (2010) Comparative study of coil resonators for wireless power transfer system in terms of transfer loss. IEICE Electron Express 7(11):785–790. doi:10.1587/Elex.7.785

Jonah O, Georgakopoulos SV, Tentzeris MM (2013). Orientation insensitive power transfer by magnetic resonance for mobile devices. In: IEEE Wireless Power Transfer (WPT) conference, pp 15–16, May 2013, Perugia, Italy

Karalis A, Joannopoulos JD, Soljacic M (2008) Efficient wireless non-radiative mid-range energy transfer. Ann Phys 323(1):34–48. doi:10.1016/j.aop.2007.04.017

Kim Y, Ling H (2007) Investigation of coupled mode behaviour of electrically small meander antennas. Electron Lett 43(23):1250–1252. doi:10.1049/El:20072165

Kurs A, Karalis A, Moffatt R, Joannopoulos JD, Fisher P, Soljacic M (2007) Wireless power transfer via strongly coupled magnetic resonances. Science 317(5834):83–86. doi:10.1126/science.1143254

Lee SG, Hoang H, Choi YH, Bien F (2012) Efficiency improvement for magnetic resonance based wireless power transfer with axial-misalignment. Electron Lett 48(6):339–340. doi:10.1049/El.2012.0012

Lee J, Lim YS, Yang WJ, Lim SO (2014) Wireless power transfer system adaptive to change in coil separation. IEEE Trans Antennas Propag 62(2):889–897. doi:10.1109/Tap.2013.2290795

Matsumoto H (2002) Research on solar power satellites and microwave power transmission in Japan. IEEE Microw Mag 3(4):36–45. doi:10.1109/Mmw.2002.1145674

McSpadden JO, Mankins JC (2002) Space solar power programs and microwave wireless power transmission technology. IEEE Microw Mag 3(4):46–57. doi:10.1109/Mmw.2002.1145675

Murakami J, Sato F, Watanabe T, Matsuki H, Kikuchi S, Harakawa K, Satoh T (1996) Consideration on cordless power station – contactless power transmission system. IEEE Trans Magn 32(5):5037–5039. doi:10.1109/20.539482

Pozar DM (1998) Microwave engineering, 2nd edn. Wiley, New York

Reinhold R (1908) Der Empfang Elektrischer Wellen in der Drahtlosen Telegraphie (The receipt of electric waves in the wireless telegraphy). Annalen der Physik IV 25:446–466

Rodenbeck CT, Chang K (2005) A limitation on the small-scale demonstration of retrodirective microwave power transmission from the solar power satellite. IEEE Antennas Propag Mag 47(4):67–72. doi:10.1109/Map.2005.1589875

Tak Y, Park J, Nam S (2009) Mode-based analysis of resonant characteristics for near-field coupled small antennas. IEEE Antennas Wireless Propag Lett 8:1238–1241. doi:10.1109/Lawp.2009.2036133

Tesla N (1904) The transmission of electrical energy without wires. Electr World Eng, pp 429–431

Tesla N (1914) Apparatus for transmitting electrical energy, US Patent 1,119,732

Yuan QW, Chen Q, Li L, Sawaya K (2010) Numerical analysis on transmission efficiency of evanescent resonant coupling wireless power transfer system. IEEE Trans Antennas Propag 58(5):1751–1758. doi:10.1109/Tap.2010.2044321

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_94-1

Implanted Antennas in Biomedical Telemetry

Asimina Kiourti[a]* and Konstantina S. Nikita[b]
[a]ElectroScience Laboratory, Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, USA
[b]School of Electrical and Computer Engineering, National Technical University of Athens, Zografos, Athens, Greece

Abstract

Biomedical telemetry permits the measurement of physiological signals at a distance, through either wired or wireless communication technologies. One of the latest developments in wireless biomedical telemetry is in the field of implantable medical devices (IMDs). Such devices are implanted inside the patient's body by means of a surgical operation and can be used for a number of diagnostic, monitoring, and therapeutic applications. Implantable antennas, i.e., antennas which are integrated into RF-enabled IMDs, exhibit numerous challenges in terms of design, fabrication, and testing and are, therefore, currently attracting significant research attention. Contributions from researchers of various disciplines build a rich pool of background information, while highlighting future prospects.

Keywords

Biocompatibility; Biomedical telemetry; Industrial scientific and medical (ISM) applications band; Medical implant communication service (MICS) band; Miniaturization; Implantable antennas; In vitro; In vivo; Specific absorption rate

Introduction

Biomedical telemetry permits the measurement of physiological signals at a distance, through either wired or wireless communication technologies. Physiological signals are obtained by means of appropriate transducers, post-processed, and eventually transmitted to exterior monitoring/control equipment. One of the latest developments in wireless biomedical telemetry is in the field of implantable medical devices (IMDs). Such devices are implanted inside the patient's body by means of a surgical operation and can be used for a number of diagnostic, monitoring, and therapeutic applications (Greatbatch and Homes 1991; Chow et al. 2013; Kiourti et al. 2014a).

Nowadays, millions of people worldwide depend upon IMDs to support and improve the quality of their lives. Advances in biological, chemical, electrical, and mechanical sensor technologies as well as in microelectromechanical systems (MEMS) have led to a wide range of IMDs. Specifically, wireless IMDs are already in use for a wide variety of applications, including temperature monitors (Scanlon et al. 1997); pacemakers and cardioverter defibrillators (Wessels 2002); functional electrical stimulators (FES) (Guillory and Normann 1999); blood glucose sensors (Shults et al. 1994); cochlear (Buchegger et al. 2005), gastric, and bladder controllers (Sani et al. 2009); and retinal (Gosalia et al. 2004) implants.

*Email: akiourti@biosim.ntua.gr
*Email: kiourti.1@osu.edu

DOI 10.1007/978-981-4560-75-7_94-1

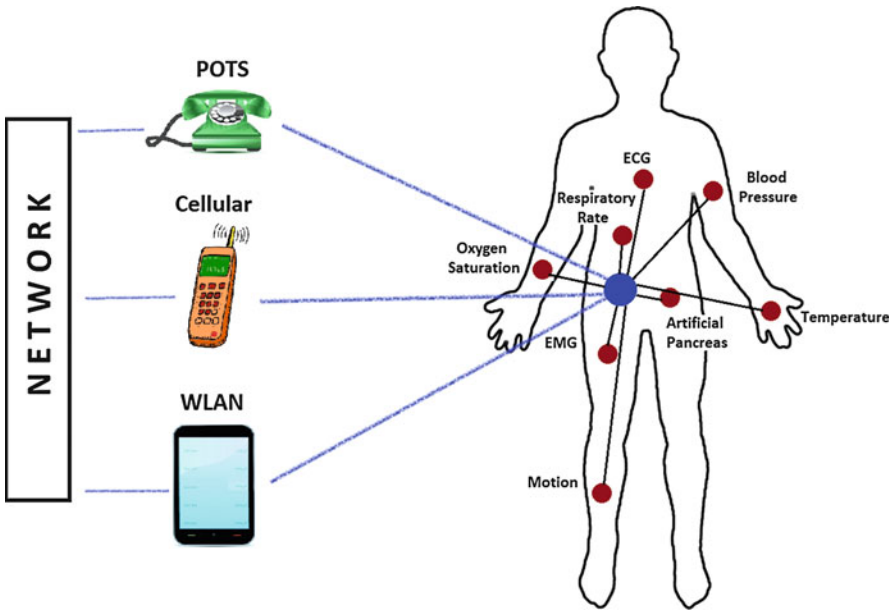


Fig. 1 Schematic of a typical wireless biomedical telemetry system with example IMD applications

As technology continues to evolve, new IMDs are being developed, and their use is expected to rapidly increase from this already large base.

Traditionally, low-frequency (tens to hundreds kHz) inductive links have been employed for the wireless telemetry of IMDs (Tang et al. 1995; Valdastri et al. 2004). However, inductive links suffer from: (a) low data rates (1–30 kbps), (b) restricted communication range (<10 cm), and (c) increased sensitivity to inter-coil positioning. To overcome these limitations, research is recently oriented toward radio-frequency (RF)-linked IMDs. Design of such RF telemetry systems is facilitated by the rapid advances in wireless communications and electronics. As shown in Fig. 1, a key and critical component of RF-linked IMDs is the integrated implantable antenna. In principle, picked-up physiological signals are amplified, digitized, and fed to a transceiver, which will code and modulate the data, and finally lead them to the antenna. The latter enables the IMD's bidirectional communication with the exterior monitoring/control equipment. In a realistic scenario, these implantable antennas are mounted on the existing hardware of the IMD.

Implantable antennas, i.e., antennas which are integrated into RF-enabled IMDs, exhibit numerous challenges in terms of design, fabrication, and testing and are, therefore, currently attracting significant research attention (Kiourti and Nikita 2012a; Chow et al. 2013). Specifically, numerical design of implantable antennas needs to be performed fast and in a way which optimally addresses issues related to operation frequency selection, miniaturization, biocompatibility, patient safety, high-quality communication with exterior equipment, and intersubject variability. Furthermore, prototype fabrication of such miniature antenna structures is highly challenging given their critical tolerance to potential experimental versus numerical inconsistencies. Finally, in vitro and in vivo testing of implantable antennas is highly intriguing given the requirements for (a) phantom formulation that matches the theoretical electrical properties and (b) implantation inside living model animals, respectively. Contributions from researchers of various disciplines build a rich pool of background information, while highlighting future prospects.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_94-1

Design Considerations for Implantable Antennas

Considerations related to the design and performance of implantable patch antennas include (a) selection of the frequency band of operation, (b) miniaturization of the antenna's occupied volume/size, (c) biocompatibility of the antenna structure, (d) conformance of the implantable antenna to international guidelines which preserve patient safety against radiated electromagnetic (EM) fields (IEEE 1999, 2005), (e) high-quality communication of the implantable antenna with exterior monitoring/control equipment, and (f) intersubject variability. These considerations are discussed in detail in the following.

Frequency Bands of Operation

Selection of the operation frequency of implantable antennas, or equivalently IMDs, is receiving considerable attention from the scientific community as attributed to a number of competing factors.

Until recently, no globally accepted frequency band had been dedicated to the biomedical telemetry of IMDs. The situation changed with the ITU-R Recommendation SA.1346 (ITU-R 1998) which outlined the use of the 402.0–405.0 MHz frequency band for medical implant communication services (MICS). The most recent contribution in the field is the Institute of Electrical and Electronics Engineers (IEEE) 802.15.6 standard (IEEE 2012) which deals with short-range, wireless communications in the vicinity of, or inside, the human body. The standard refers to existing industrial, scientific, and medical (ISM) bands as well as frequency bands approved by national medical and/or regulatory authorities. According to this latest standard, an IMD shall be able to support transmission and reception in at least one of the following frequency bands: 402.0–405.0, 420.0–450.0, 863.0–870.0, 902.0–928.0, 950.0–958.0, 2,360.0–2,400.0, and 2,400.0–2,483.5 MHz. Ultrawide band (UWB) IMDs which implement low band (3,494.4–4,492.8 MHz) or high band (6,489.6–9,984.0 MHz) channels are also supported. For example, Table 1 summarizes recent research studies on implantable antennas operating within various frequency bands.

The ISM band of 2,400.0–2,500.0 MHz (ITU-R 2008) is appearing as one of the most promising solutions. This is because it is already well developed in terms of technology (Bluetooth, Wi-Fi, and WLAN), antennas, integrated circuits, and embedded systems. Furthermore, higher operation frequencies allow the use of smaller-sized antennas and components. It is for this purpose that implantable antennas operating at much higher frequencies (e.g., 5.85 and 31.5 GHz) have also been reported in the literature (see Table 1). However, a high number of operating services are colocated in the aforementioned bands. Therefore, interference issues constitute a limiting factor. Interference may cause harmful effects in terms of false IMD activation, link unavailability, and data corruption.

Table 1 Operation frequencies of implantable antennas reported in the literature

Frequency	References for implantable MDs
402 MHz	Kim and Rahmat-Samii 2004; Soontornpipit et al. 2004, 2005; Abadia et al. 2009; Chen et al.2009; Sani et al. 2009; Karacolak et al. 2009; Sánchez-Fernández et al. 2010; Gemio et al. 2010; Huang et al. 2011; Vidal et al. 2012; Kiourti and Nikita 2012b
433 MHz	Weiss et al. 2009; Gemio et al. 2010; Huang et al. 2011; Kiourti and Nikita 2012b
868 MHz	Sani et al. 2009; Sani et al. 2010; Kiourti and Nikita 2012b
915 MHz	Scanlon et al. 2000; Gemio et al. 2010; Kiourti and Nikita 2012b
1,575 MHz	Azad and Ali 2009
2,400 MHz	Kawoos et al. 2008; Karacolak et al. 2009; Xia et al. 2009; Sánchez-Fernández et al. 2010; Gemio et al. 2010; Huang et al. 2011; Scarpello et al. 2011
31.5 GHz	Ahmed et al. 2008

DOI 10.1007/978-981-4560-75-7_94-1

To deal with interference issues, focus for IMDs is mainly on the 402.0–405.0 MHz band, which has been exclusively allocated for medical implant communication services (MICS). The MICS band is currently regulated by the United States Federal Communications Commission (FCC) and the European Radiocommunications Committee (ERC). Its spectrum of 3 MHz allows for 10 channels (bandwidth of 300 KHz each) to operate simultaneously (i.e., multiple IMDs in the same area). It also limits potential interferences from the colocated Meteorological Aids Service band (401–406 MHz). Additional ways of enhancing interference tolerance include (a) automatic repeat request (ARQ) and forward error correction (FEC) techniques to mitigate the effects of impulsive noise (noise which is very short in duration and often of greater amplitude than the IMD signal levels) and (b) the use of frequency agility and channelization as a means of avoiding narrowband interferers (sources with bandwidths comparable to the IMD signal waveform). The MICS band is internationally available and feasible with low power circuits, falls within a relatively low noise portion of the spectrum, and allows for acceptable propagation through human tissue. A review on the regulatory standards for IMDs and the characteristics of MICS transceivers is performed in Savci et al. (2005).

The effect of operation frequency upon the performance of IMDs has been addressed in the literature. Single-cell excitation used to simulate vaginal (Scanlon et al. 2000) and gastric/bladder/cardiac (Sani et al. 2009) implants has been shown to exhibit increased power absorption, higher net body losses, and reduced penetration depths with increasing frequency. Recently, implantable antennas at higher frequencies were found to achieve enhanced gains (a 10.7 % increase in the maximum far-field gain at 915 MHz, compared to the gain at 402 MHz), increased maximum allowable net input power levels (10.1 % and 1.3 % increases imposed by the IEEE C95.1-1999 (IEEE 1999) and IEEE C95.1-2005 (IEEE 2005) safety standards, respectively), and more expanded specific absorption rate (SAR) distributions (Kiourti and Nikita 2012b). Results were attributed to the authors' choice of keeping the antennas' physical dimensions identical and modifying their effective size. Furthermore, the selection of the operation frequency is directly related to the bit rate: higher operation frequencies allow for an increase in bandwidth and enable higher bit rates, which are favorable for high data rate applications. For example, the maximum channel capacity for a band-limited additive white Gaussian noise (AWGN) channel is given by

$$C = BW \log_2\left(1 + \frac{S}{N}\right) \tag{1}$$

where BW is the channel bandwidth, S is the mean signal power, and N is the mean noise power. Therefore, the MICS band (BW 300 kHz) enables low bit rates, and the WMTS (BW 8.5 kHz to 6 MHz) and ISM (selectable BW) bands enable medium bit rates, whereas the UWB (BW $>$ 500 MHz) is appearing as the most promising solution for high data rates. It is remarked that according to the recent IEEE 802.15.6 standard (IEEE 2012), data rates of typically up to 10 Mbps are required to satisfy evolutionary health-care services.

Miniaturization

Recent advances in the technology of IMD electronics lead to ultrasmall designs for IMDs. For instance, implantable retinal prostheses should be small enough to be inserted inside the eyeball (radius of ~12.5 mm), while intracranial pressure monitors should be small enough to fit in standard 12 mm burr holes in the skull (Warty et al. 2008). However, dimensions of the traditional half-wavelength (λ/2) or quarter-wavelength (λ/4) antennas at the frequency bands allocated for medical implants and especially at the low-frequency MICS band make them useless for implantable applications. As an example, the free-space wavelength at 402, 433, 868, and 915 MHz can be computed as 74.6, 69.3, 34.6, and 32.8 cm,

DOI 10.1007/978-981-4560-75-7_94-1

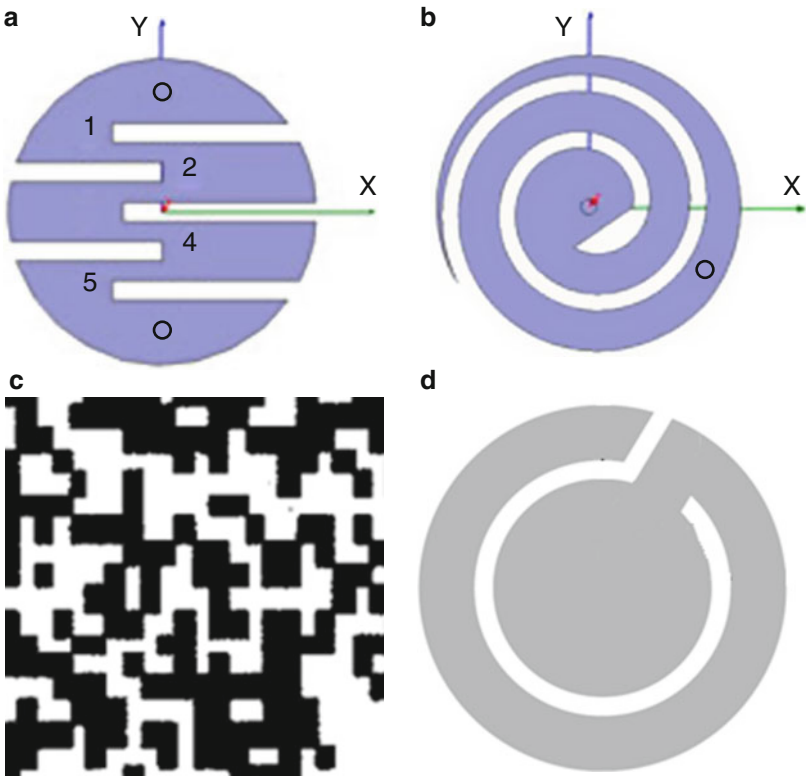


Fig. 2 Lengthening of the current flow path for miniature implantable antennas: (**a**) meandered (Kiourti and Nikita 2011), (**b**) spiral (Kiourti and Nikita 2011), (**c**) waffle-type (Soontornpipit et al. 2005), and (**d**) hook-slotted (Liu et al. 2008a) shaped patches

respectively. Therefore, miniaturization becomes one of the greatest challenges in implantable antenna design.

Fortunately, human tissues in which implantable antennas are intended to operate exhibit high permittivity (e.g., the permittivity of skin at 402 MHz is 46.7 (Gabriel et al. 1996a, b, c)) or, equivalently, reduced wave propagation velocity. This, in turn, increases the effective dielectric constant of the antenna and works to advantageously miniaturize its physical size. As would be expected, additional antenna miniaturization techniques can be concurrently employed. Specifically, miniaturization techniques that have been proposed in the literature for implantable antennas include:

- **Use of high-permittivity dielectric materials**. High-permittivity dielectrics are selected for implantable antennas (e.g., ceramic alumina ($\varepsilon_r = 9.4$) (Kiourti et al. 2011a) or Rogers RO3210 ($\varepsilon_r = 10.2$) (Kiourti and Nikita 2012b)) because they shorten the effective wavelength and result in lower resonance frequencies, thus assisting in antenna miniaturization.
- **Lengthening of the current flow path on the antenna surface**. Longer effective current flow paths excited on the antenna can reduce the resonance frequency and achieve a more compact size for the implantable antenna. For this purpose, meandered (Kiourti and Nikita 2011), spiral (Kiourti and Nikita 2011), waffle-type (Soontornpipit et al. 2005), and hook-slotted (Liu et al. 2008a) shaped implantable antennas have been suggested, as shown in Fig. 2.
- **Addition of shorting pins**. In the case of implantable patch antennas, inserting a shorting pin between the ground and patch planes increases the effective size of the antenna. This, in turn, reduces the required physical dimensions, given a specific operation frequency scenario. The technique works in much the same way a ground plane doubles the height of a monopole antenna, i.e., it typically produces

DOI 10.1007/978-981-4560-75-7_94-1

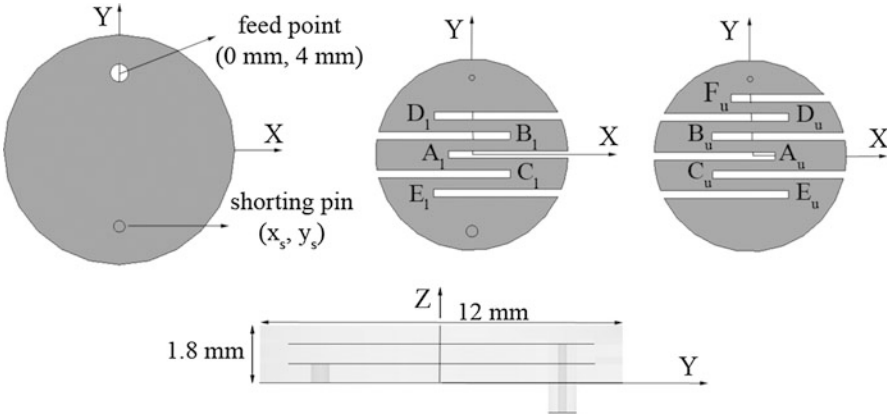


Fig. 3 Geometry of a stacked skin-implantable PIFA with meandered patches (Kiourti and Nikita 2012b)

a planar inverted-F antenna (PIFA) with the same resonance performance as a double-sized antenna without the shorting pin (Soontornpipit et al. 2004).

- **Patch stacking**. In the case of implantable patch antennas, vertically stacking two radiating patches reduces antenna size by increasing (nearly doubling) the length of the current flow path (Kiourti et al. 2011a; Kiourti and Nikita 2012b).

Implantable antennas reported in the literature combine some (or all) of these miniaturization techniques in order to reduce size. Of course, all antenna design parameters, including the location of the coaxial feed, have to be appropriately selected (optimized) for a good 50 Ω match at the desired operation frequency. For example, the skin-implantable antenna of Fig. 3 adapts a stacked PIFA structure of meandered patches built on Rogers RO3210 ($\varepsilon_r = 10.2$) substrate to achieve a miniaturized structure (volume of 214.9 mm^3) resonating in the MICS band (Kiourti and Nikita 2012b).

Nevertheless, it is important to emphasize that implantable antenna miniaturization comes in expense of its radiation and patient safety performance. Specifically, generic results for a skin-implantable antenna placed inside a tissue-simulating cube have indicated degraded gain and SAR performance with a reduction in size. Antenna miniaturization by 32 % and 65 % has been found to reduce the maximum far-field gain values by 5 % and 19 %, respectively, and the maximum allowable input powers imposed by the IEEE C95.1-1999 safety standard (IEEE 1999) by 21 % and 44 %, respectively. The significance of application-specific rather than miniaturization-oriented implantable antenna design is, thus, highlighted.

Biocompatibility

Implantable antennas must be biocompatible in order to preserve patient safety and prevent rejection of the implant. Another consideration is that human tissues are conductive and would short-circuit the implantable antenna if they were allowed to be in direct contact with its metallization. Biocompatibility and prevention of undesirable short circuits are especially crucial in the case of antennas which are intended for long-term implantation. In the literature, there have been reported two approaches for preserving the biocompatibility of implantable antennas and separating their metallic parts from the surrounding biological tissues: (a) covering the antenna structure with a biocompatible superstrate dielectric layer and (b) insulating the antenna with a thin layer of low-loss biocompatible coating.

Specifically, the most widely used approach for preserving the biocompatibility of an implantable antenna, while at the same time separating its metal radiator from the human tissues, is to cover the structure by a superstrate dielectric layer (e.g., Fig. 4a (Karacolak et al. 2008)). Commonly used biocompatible materials include Teflon (permittivity, $\varepsilon_r = 2.1$; dielectric loss tangent, $\tan\delta = 0.001$),

DOI 10.1007/978-981-4560-75-7_94-1

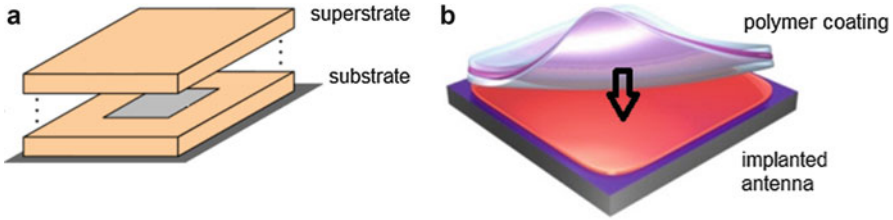


Fig. 4 Biocompatibility issues for implantable antennas: (**a**) addition of a superstrate (Karacolak et al. 2008) and (**b**) thin-layer encapsulation (Karacolak et al. 2010)

MACOR® ($\varepsilon_r = 6.1$, $\tan\delta = 0.005$), and ceramic alumina ($\varepsilon_r = 9.4$, $\tan\delta = 0.006$) (Soontornpipit et al. 2004). It is important to highlight, however, that ceramic substrates do not lend themselves easily to drilling and round cuts (Warty et al. 2008).

Insulating the implantable antenna with a thin layer of low-loss biocompatible coating is another reported approach (e.g., Fig. 4b; Karacolak et al. 2010). Materials proposed for biocompatible encapsulation include zirconia ($\varepsilon_r = 29$, $\tan\delta \approx 0$) (Skrivervik and Merli 2011), polyetheretherketone (PEEK) ($\varepsilon_r = 3.2$, $\tan\delta = 0.01$) (Abadia et al. 2009), polydimethylsiloxane (PDMS) ($\varepsilon_r = 3$, $\tan\delta = 0.005$), Silastic MDX4-4210 Biomedical Grade Base Elastomer ($\varepsilon_r = 3.3$, $\tan\delta \approx 0$), and Parylene ($\varepsilon_r = 2.95$, $\tan\delta = 0.005$) (Karacolak et al. 2010). Because of its electrical properties, zirconia is a better candidate material for biocompatible insulation from an electromagnetic point of view. High permittivity and low-loss tangent values allow the near fields of the antenna to concentrate inside the low-loss encapsulation layer, thus mitigating power loss. However, easiness of preparation and handling must also be taken into account. For example, PEEK and Silastic MDX4-4210 Biomedical Grade Base Elastomer are much easier to prepare and handle. Furthermore, thickness of the biocompatible insulation layer is an important factor in antenna design. Computation of its optimum thickness is, thus, considered to be highly significant for lowering power loss without aimlessly increasing antenna size.

It is remarked that addition of an insulation coating or superstrate layer significantly affects the performance of the antenna and, thus, needs to be taken into account within the design. Furthermore, a biocompatible hermetic package is necessary for housing the electronics, which may be made of low-temperature co-fired ceramic (LTCC), Parylene, liquid crystal polymer (LCP), silicon, or alumina (Chow et al. 2010).

Patient Safety and Specific Absorption Rate (SAR)

It was not until recently that research on the biological effects of IMDs started being carried out. Specifically, issues related to patient safety limit the maximum allowable power incident to the implantable antenna. The specific absorption rate (SAR) (rate of energy deposited per unit mass of tissue) is generally accepted as the most appropriate dosimetric measure, and compliance with international guidelines is assessed. For example, the IEEE C95.1-1999 standard restricts the SAR averaged over any 1 g of tissue in the shape of a cube to less than 1.6 W/kg ($SAR_{1g,max} \leq 1.6$ W/kg) (IEEE 1999). The ICNIRP's basic restrictions limit the SAR averaged over 10 g of contiguous tissue to less than 2 W/kg (ICNIRP 1998). To harmonize with the ICNIRP guidelines, the IEEE C95.1-2005 standard restricts the SAR averaged over any 10 g of tissue in the shape of a cube to less than 2 W/kg ($SAR_{10g,max} \leq 2$ W/kg) (IEEE 2005).

DOI 10.1007/978-981-4560-75-7_94-1

Specifically, the power absorbed by the human body in the presence of an incident electromagnetic field is given by

$$P_{\text{abs}} = \frac{1}{2}\int \sigma |E|^2 dV, \tag{2}$$

where σ is the conductivity of the human tissues and $|E|$ is the intensity of the electric field inside the body (Kim and Rahmat-Samii 2004). Equation 2 indicates that the absorbed power is related to the electric field, so that maximum SAR values are recorded in the areas where maximum electric field intensities occur.

Based on the deduction that peak averaged SAR values are generated from high near fields, novel implantable patch antennas can be designed, which aim at lower electric field intensities. For example, in Kim and Rahmat-Samii (2006), the radiation mechanism of an implantable antenna was discussed, in an attempt to modify its design for reducing the spatial-averaged SAR in human tissue. Replacing the uniform-width spiral radiator of an implantable MICS PIFA with a nonuniform-width radiator was found to decrease the electric field intensity and, in turn, $\text{SAR}_{1\text{g,max}}$. The simulated near electric field distribution showed that the high electric field area of the PIFA employing the nonuniform-width radiator was much smaller than that of the original PIFA. The value of $\text{SAR}_{1\text{g,max}}$ was, thus, limited from 310 to 210 W/kg, considering a net input power of 1 W.

Several safety evaluation studies have been performed for implantable antennas. Of importance is that the IEEE C95.1-1999 standard has been found to be much stricter than the recent IEEE C95.1-2005 standard (Kiourti and Nikita 2012b). The latter has shown to be almost insensitive to changes in the tissue model properties (anatomical features and dielectric parameters) (Kiourti and Nikita 2013a).

Radiation Performance

Biomedical telemetry systems for IMDs are comprised of the IMD and an exterior monitoring/control device, which is placed at some distance (typically 2 m) away from the body. Biotelemetry links may be used for device parameter adjustment, transmission of stored information, as well as real-time transmission of vital monitoring information. Therefore, the implantable antenna should provide a signal that is strong enough to be picked up by the exterior device, regardless of any power limitations. It is important to highlight that apart from patient safety, interference issues also limit the maximum allowable power incident to the implantable antenna. For example, a strict limit of −16 dBm (25 μW) has been set on the effective radiated power (ERP) of IMDs operating in the MICS band in order to prevent interference to the colocated Meteorological Aids Service band (ITU-R 1998).

Given the SAR and ERP power limitations, far-field gain of the implantable antenna indicates the desired receiver sensitivity for achieving reliable biotelemetry communication. In order to increase the range of biotelemetry communication, implantable antennas with enhanced gain are solicited. However, reduced-size antennas exhibit degraded electromagnetic performance: miniaturization degrades gain, while high-gain antennas exhibit relatively increased size. Low values of gain imply poor radiation efficiencies; however, compromises on the system performance are inevitable given the miniaturized antenna dimensions.

Of importance is that (a) the symmetry of the implantation tissue model affects the symmetry of the antenna's far-field radiation pattern, accordingly. Omnidirectional, monopole-like radiation is observed inside symmetrical tissue models (see Fig. 5a; Kiourti et al. 2011a; Kiourti and Nikita 2012b), whereas asymmetrical radiation is recorded within anatomical tissue models which are irregular and inhomogeneous (see Fig. 5b; Kiourti et al. 2011a; Kiourti and Nikita 2012b).

DOI 10.1007/978-981-4560-75-7_94-1

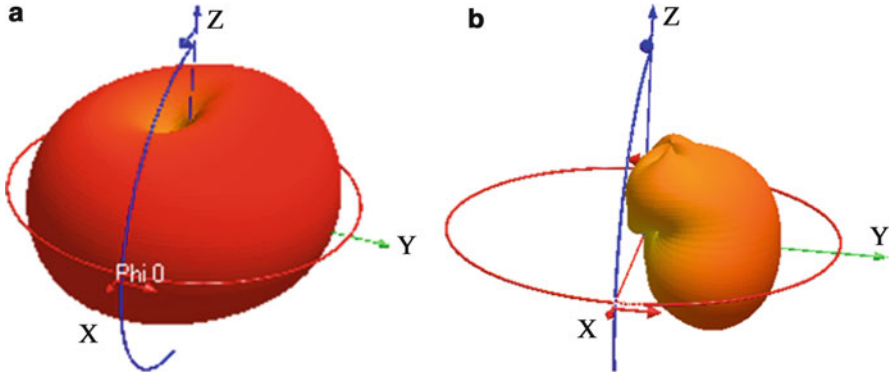


Fig. 5 Far-field gain radiation pattern of the skin-implantable antenna proposed in Kiourti et al. (2011a) inside (**a**) a 100 mm-edge skin cube and (**b**) the skin tissue of an anatomical human head model

Powering Considerations

All the components of an IMD, including the implantable antenna, require power. Integrated power supplies, such as batteries, are only suitable for applications where the IMD lifespan is short (e.g., IMDs which are intended to be implanted for months or a few years, depending on the duty cycle). To elongate the battery life of IMDs, external power transmission has been proposed for recharging purposes (Kendir et al. 2005). The technique is based on electromagnetic (EM) induction between an exterior-transmitting and an implantable receiving coil, which is placed in close distance and is often wound around a dielectric or ferrite core to improve the efficiency. Furthermore, power scavenging sources including motion, vibration, air flow, temperature difference, light, and infrared radiation have been suggested. For example, a vibration-based generator for implantable IMDs, capable of delivering 2 mJ/cycle, has been designed, while ambient EM energy harnessing has recently been investigated (Mitcheson et al. 2004). Such solutions are solicited for IMDs which are intended for an implantation period of several years or even a lifetime (cochlear implants for the deaf or retina implants for the blind).

To meet potential longevity requirements of the IMDs and guarantee their on demand availability, power conservation techniques can additionally (or instead) be applied. Suggested ideas include pre-configured on-off function of the IMD (Furse 2009) and transmission/detection of a "wake-up" alarm signal (Karacolak et al. 2008). In the first case, devices spend most of their time in an ultraefficient sleep mode followed by short bursts of data transmission. Data mining or compression techniques may be used to reduce the actual bits of data to be transmitted. In the second case, the system uses two frequency bands, one for "wake up" and one for transmission. The transceiver stays in "sleep mode" with low power consumption (1 μW) until a "wake-up" signal is sensed in the 2,450 MHz ISM band. In the normal mode, the IMD is fully powered and exchanges data in the MICS band. Following the data transfer, the IMD transceiver returns back to the "sleep mode." To do so, a transceiver with dual-band operation may be used, such as the commercially available Zarlink ZL70101 Transceiver (Zarlink 2006). The system uses two frequency bands, one for "wake up" and one for transmission. The exterior device may be programmed to wake up the implanted device according to a physician-defined schedule or only when a patient event is detected (Savci et al. 2005).

It is remarked that the employment of "wake-up" techniques requires the design of multiband antennas, i.e., a separate band for data biotelemetry and a separate band for transmission/detection of the "wake-up" alarm signal. For example, a dual-band (MICS and ISM) implantable antenna has been proposed in the literature for continuous glucose monitoring (Karacolak et al. 2008). A meandered antenna configuration was considered for optimizing the antenna surface area, and particle swarm optimization was applied to

DOI 10.1007/978-981-4560-75-7_94-1

achieve the desired resonance characteristics. The simulated and measured bandwidths were found to be 82 and 142 MHz in the MICS band and 103 and 174 MHz in the ISM band, respectively. An innovative dual-band (MICS and ISM) patch antenna with a multilayer configuration and electromagnetic coupling-based feeding has further been proposed for implantation inside the left subpectoral region (Sánchez-Fernández et al. 2010). Recently, a novel antenna design was suggested using a π-shaped radiator with stacked and spiral structure, to support triple-band operation with data telemetry (402 MHz), wireless power transmission (433 MHz), and wake-up controller (2,450 MHz) (Huang et al. 2011). The simulated and measured bandwidths were 86 and 114 MHz in the MICS band and 60 and 70 MHz in the ISM band, respectively.

Intersubject Variability

Since implantable antennas are intended to operate inside human tissue, their performance strongly depends on the surrounding tissue environment. The latter includes (a) the anatomical features of the individual and (b) the dielectric parameters (permittivity, ε_r, and conductivity, σ) of the biological tissues.

For example, smaller-sized (female and low body mass index male) anatomical models have been found to exhibit higher radiated power levels and far-field gain values (Sani et al. 2009). Simplified, single-cell excitation was considered in this study to mimic gastric, bladder, and cardiac implants at 402 and 868 MHz. Anatomical differences considering a realistic model of an implantable antenna have also been assessed (Vidal et al. 2012). In this study, two implantable antennas were designed to operate at 403 MHz within mean head and mean body tissues, respectively, and the exhibited resonance performance was studied. Simulations for head, arm, and abdomen implantation were carried out within four anatomical models (two adults and two children), indicating a maximum detuning of 14 MHz from the reference frequency of 403 MHz.

Tissue dielectric parameters (permittivity and conductivity) also affect the design and performance of implantable antennas. Therefore, variations due to uncertainties in dielectric properties and intersubject variability have to be taken into account (Virtanen et al. 2006). It is remarked that maximum standard deviations of 16 % have been reported in the dielectric parameter values of rat brain tissue (Bao et al. 1997), while age dependency has repeatedly been emphasized (Gabriel 2005; Conil et al. 2008). Decrease in permittivity and conductivity values by 4 % and 10 % has been recorded, respectively, for pig tissue within 4 h after death (Schmid et al. 2003).

Recently, the performance of a MICS scalp-implantable antenna was assessed with respect to variations in head properties, i.e., anatomy and dielectric parameters (Kiourti and Nikita 2013a, b). Five head models (3- and 5-layer spherical and 6-, 10-, and 13-tissue anatomical; see Fig. 6) and seven dielectric value scenarios (variations by ±20 % in the reference permittivity and conductivity values) were considered. Compared with the reference dielectric parameter scenario within the 3-layer spherical head model, maximum variations of −19.9 %, +3.7 %, −55.1 %, and −39.2 % were overall recorded in IEEE C95.1-1999 (IEEE 1999), IEEE C95.1-2005 (IEEE 2005), RL, and G_{max} at 403.5 MHz. More specifically, the performance of implantable antennas was found to considerably depend on the anatomy of the tissue model in the area that is immediately surrounding the implant. In particular, the 6-tissue anatomical head model, which exhibited a bump of skin tissue at the top side of the head (implantation site), demonstrated the highest deviations. As long as anatomical features around the implant are relatively similar, overall anatomy and tissue composition/distribution of the numerical model were found to insignificantly influence the antenna performance. On the other hand, tissue dielectric parameters were found to insignificantly affect the implantable antenna performance. As compared to the reference dielectric value scenarios, variations in IEEE C95.1-1999 (P_{1999}) and IEEE C95.1-2005 (P_{2005}) ranged, on average, between −0.4 % and +0.1 % and between −0.4 % and +0.8 %, respectively. RL varied between −33.1 % and +17.6 % and G_{max} between −0.6 % and +3.1 %.

DOI 10.1007/978-981-4560-75-7_94-1

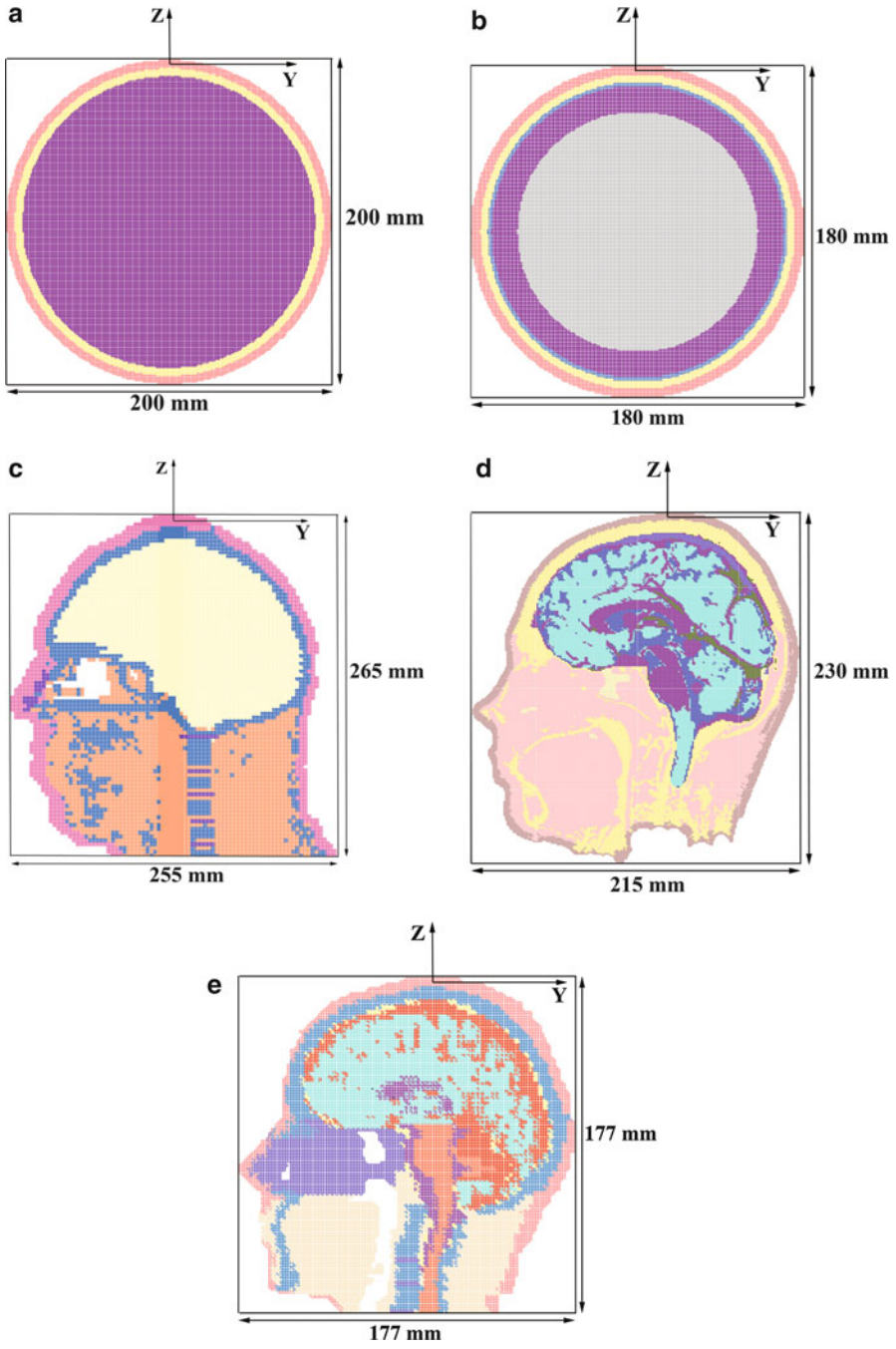


Fig. 6 Antenna implanted inside five numerical human head models: (**a**) 3-layer spherical, (**b**) 5-layer spherical, (**c**) 6-tissue anatomical, (**d**) 10-tissue anatomical, and (**e**) 13-tissue anatomical (Kiourti and Nikita 2013a)

In summary, taking tissue anatomy and dielectric parameter uncertainties into account has been shown to be important for implantable antenna design and performance evaluation. As far as compliance with safety guidelines is concerned, uncertainties inherent to variations in tissue anatomy and dielectric parameters must be taken into account. Nevertheless, it is important to highlight that in contrast to the IEEE C95.1-1999 guidelines, compliance with the recent IEEE C95.1-2005 guidelines occurs to be almost insensitive to tissue properties. Furthermore, results indicate the need for designing implantable antennas with enhanced bandwidth in order to compensate for detuning and impedance mismatch

DOI 10.1007/978-981-4560-75-7_94-1

inherent to intersubject variability. A conservative evaluation of the link budget between the implantable antenna and exterior equipment is also required in order to account for potential degraded gain values and deteriorated symmetry in radiation.

Numerical Design and Performance Assessment of Implantable Antennas

Tissue Models and Design Strategies for Implantable Antennas

The fact that implantable antennas are intended to operate inside biological tissue rather than in free space affects their design and performance in a number of ways. Specifically, given the presence of the human body, design of implantable antennas should be performed either (a) inside free space and further refined for tissue implantation or (b) directly inside an environment surrounded by human tissue. It is noted that in numerical simulations, biological tissues are analyzed as inhomogeneous lossy mediums that have their own permittivity (ε_r), conductivity (σ), and mass density values. These can be either approximated as constant within a narrow frequency range (Kiourti et al. 2011a; Kiourti and Nikita 2012b) or described by means of a Cole–Cole formulation for the complex relative permittivity, according to

$$\varepsilon_c(\omega) = \varepsilon_\infty + \sum_n \frac{\Delta\varepsilon_n}{1 + (j\omega\tau_n)^{(1-\alpha_n)}} + \frac{\sigma_i}{j\omega\varepsilon_0} \tag{3}$$

where ω is the angular frequency, n is the order of the Cole–Cole model, ε_∞ is the high-frequency permittivity, τ_n is the relaxation time, $\Delta\varepsilon_n$ is the pole amplitude, α_n is the parameter that allows for the broadening of the dispersion, and σ_i is the static ionic conductivity (Karacolak et al. 2009; Noroozi and Hojjat-Kashani 2012). Canonical tissue models are often used to speed up simulations and ease the design of implantable antennas. These may be either single layer (e.g., Fig. 7a; Kiourti and Nikita 2012b) or multilayer (e.g., Fig. 7b; Karacolak et al. 2008). To obtain more realistic results, anatomical tissue models (e.g., Fig. 7c; Kiourti and Nikita 2012b) produced by the combination of magnetic resonance imaging (MRI) or computed tomography (CT) data can also be applied.

Suggested design strategies for implantable antenna design include:

- **Strategy #1.** Antenna design in free space and further refinement inside an anatomical model of the intended implantation site. For example, in Rucker et al. (2007), a MICS patch antenna was designed in a free-space environment and further implanted inside the skin tissue of an anatomical head model. Resonance frequency detuning was observed, as attributed to the capacitive loading effect of the surrounding tissues. To refine the resonance, a varactor diode with tuning capabilities was subsequently integrated.
- **Strategy #2.** Antenna design in free-space targeting at high-gain values and further refinement inside a single-layer tissue model. For example, in Abadia et al. (2009), a MICS antenna was designed in free space aiming at high gain (>-20 dB) in order to account for subsequent body absorption losses. The antenna was optimized in free space to minimize size and further covered by a biocompatible layer and placed inside tissue material. Design modifications were performed to account for the frequency shift induced by the presence of encapsulation and human tissue.
- **Strategy #3.** Antenna design directly inside a canonical single-layer tissue model (cubical, rectangular parallelepiped, or cylindrical) of the intended implantation tissue. As would be expected, use of a single-layer tissue model is the simplest and fastest option when designing implantable antennas directly inside tissue material. Following this design strategy, antennas are designed for a "generic" tissue-implantation scenario. Simplified tissue models in the shape of a cube (Kiourti et al. 2011a, b), a

DOI 10.1007/978-981-4560-75-7_94-1

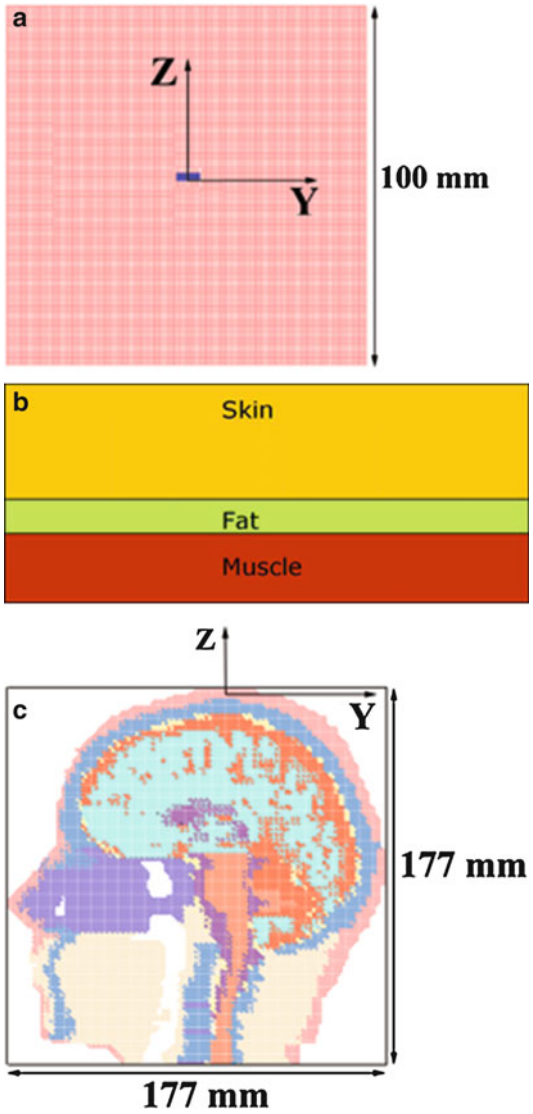


Fig. 7 Tissue models: (**a**) single-layer canonical (skin cube) (Kiourti and Nikita 2012b), (**b**) three-layer (skin/fat/muscle) canonical (Karacolak et al. 2008), and (**c**) anatomical human head (Kiourti and Nikita 2012b)

rectangular parallelepiped (Liu et al. 2008a; Kim and Rahmat-Samii 2004, 2006), and a cylinder (Liu et al. 2008b, 2009) have been used for this purpose. Design is performed by selecting the dielectric material and subsequently optimizing all antenna design parameters to refine tuning at the desired operation frequency.

- **Strategy #4.** Antenna design directly inside a canonical multilayer tissue model (cubical, rectangular parallelepiped, or cylindrical) of the intended implantation tissue. This strategy intends to design the antenna for a specific implantation site by taking into account a specific region of the body. A multilayer tissue model, with either finite or infinite dimensions, is selected. For example, implantable antennas intended for trunk (Karacolak et al. 2008) and chest (Kim and Rahmat-Samii 2004) implantation have been directly designed inside three-layer planar tissue models consisting of skin, fat, and muscle tissues.
- **Strategy #5.** Approximate antenna design inside a cube filled with the intended tissue material and further quasi-Newton optimization inside a canonical model of the intended implantation site (Kiourti and Nikita 2012b). This strategy emphasizes on design speedup and optimized resonance performance

DOI 10.1007/978-981-4560-75-7_94-1

of the implantable antenna inside a specific implantation site. It is remarked that, as far as antenna design is concerned, multilayer canonical models have been shown to provide an acceptable model for the human body. Specifically, highly similar return loss characteristics have been found for implantable patch antennas inside a three-layer planar geometry and a realistic model of the human chest (Kim and Rahmat-Samii 2004), as well as inside a three-layer spherical and an anatomical model of the human head (Kiourti and Nikita 2012b).

- **Strategy #6.** Antenna design inside a small-sized single-layer or multilayer tissue box (Kiourti and Nikita 2012c). Numerical results inside small-sized tissue boxes have been found to be almost identical to those inside canonical models of the intended implantation site, thus, rendering design directly into the latter unnecessary and inadequately slow. The aim is to minimize the use of required computational resources or, equivalently, the required simulation time toward designing implantable antennas optimized for specific medical implantation scenarios. Incorporation of the dielectric loading of the surrounding tissues and exterior air in the design and use of canonical, small-sized tissue models have been found to form the optimum solution for design purposes.

Regarding the numerical performance analysis of implantable antennas, it is noted that analytical methods can only be applied for simplified implantable antennas placed inside canonical tissue models. For example, a spherical dyadic Green's function (DGF) code has been implemented in the literature to characterize a MICS dipole antenna implanted inside a multilayer spherical human head model (Kim and Rahmat-Samii 2004). As a result, emphasis is mainly on numerical methods implemented on commercial electromagnetic simulation platforms. The electromagnetic solvers which are most commonly used in the literature for implantable antenna design are based on the finite element (FE) method (e.g., Ansoft HFSS) (Kiourti et al. 2011b; Liu et al. 2008a, b; Huang and Kishk 2011). The finite difference time domain (FDTD) method is also applied in some studies, because it exhibits simplicity in the implementation of inhomogeneous media and assessment of bioelectromagnetic interactions, while enabling efficient modeling of detailed anatomical human body parts (e.g., CST Microwave Studio, Remcom XFdtd) (Soontornpipit et al. 2004; Kiourti and Nikita 2012b; Kim and Rahmat-Samii 2004, 2006). In all cases, the computational cost heavily depends on the complexity of the tissue and antenna models. Absorbing boundaries (e.g., Mur (Soontornpipit et al. 2004)) or perfectly matched layer boundaries (Soontornpipit et al. 2005) are placed at some distance away from the setups to truncate the simulation domain while extending radiation infinitely far.

A Step-by-Step Example Tutorial to Implantable Antenna Design

In this section, a simple, yet analytical and complete, step-by-step example tutorial is provided on implantable antenna design (Kiourti and Nikita 2014). Simulations are carried out within the framework of a MICS implantable patch antenna for intracranial pressure (ICP) monitoring applications. Nevertheless, the same steps may be applied for any implantable antenna design that the designer may have on hand. It is remarked that the design strategy followed hereafter is the strategy #6 that was discussed in the previous section. As compared to the other strategies, strategy #6 has been shown to result in the fastest design of implantable antennas with optimized resonance characteristics within the medical application on hand. The reason is that it incorporates dielectric loading of both the surrounding tissues and exterior air on the antenna, while employing a canonical (parallelepiped) miniature tissue model, which can be meshed and solved in a relatively easy and fast way (Kiourti and Nikita 2012c). Equivalently, the tissue-simulating box considered in this methodology has been found to be the simplest and smallest tissue model in which the implantable antenna exhibits almost identical reflection coefficient frequency response as it would exhibit inside a canonical or anatomical tissue model of the intended implantation site (Kiourti and Nikita 2012c).

DOI 10.1007/978-981-4560-75-7_94-1

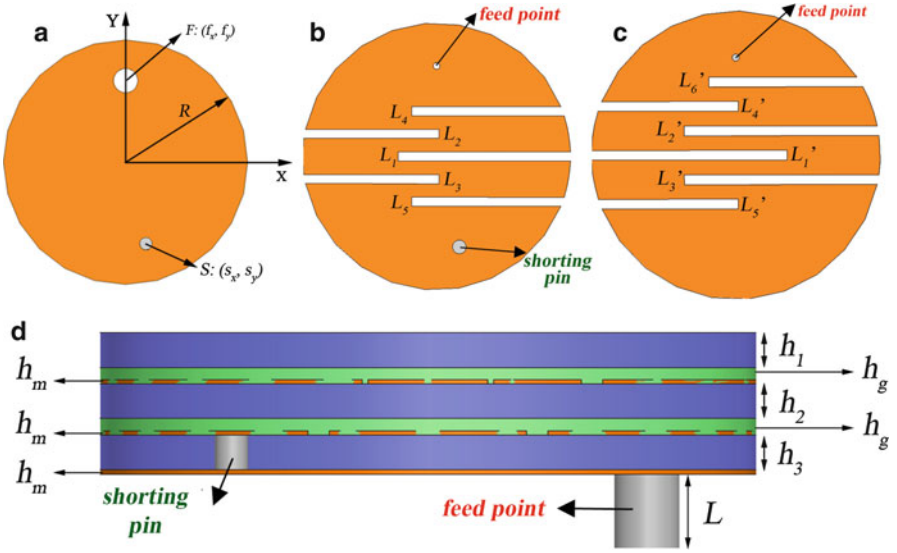


Fig. 8 Parametric implantable antenna model: (**a**) ground plane, (**b**) lower patch, (**c**) upper patch, and (**d**) side view (Kiourti and Nikita 2014)

The parametric implantable antenna model of Fig. 8 is employed to serve the goals of this tutorial. The model consists of a ground plane (radius of R) and two vertically stacked patches (radius of $R = 0.1$ mm each), printed on dielectric substrates (permittivity of ε_{rd} and thicknesses of h_1 and h_2, respectively). Origin of the coordinate system is considered to be located at the center of the antenna ground plane. A dielectric superstrate (permittivity of ε_{rd} and thickness of h_3) covers the structure for biocompatibility purposes. Meanders of variable lengths (L_i, $i = 1$–5, 1–$6'$) and identical widths (0.4 mm) are inserted into the patches to assist in miniaturization. A shorting pin (S: (s_x, s_y)) connects the ground plane to the lower patch, while a 50 Ω coaxial cable of variable type and length (L) excites both patches (F: (f_x, f_y)). Copper sheets (thickness of h_m) are considered for the ground plane and patches, while glue layers (permittivity of ε_{rg} and thickness of h_g) bond the dielectric layers together. It is remarked that, in fabrication, glue layers are to be inserted between multiple substrate layers and/or between substrate and superstrate layers of the antenna for bonding purposes (Kiourti and Nikita 2012a; Abadia et al. 2009). Gluing has been found to be a very critical factor for implantable antenna design and has to be accounted for the following: low-permittivity glue layers isolate the high-permittivity substrate layers, thus decreasing the effective permittivity and electrical length of the antenna, while increasing its resonance frequency. Importantly, due to the miniature size of implantable antennas, inconsistencies between the numerical antenna model and the fabricated prototype might result in a nonfunctional prototype. Therefore, gluing considerations have to be taken into account within simulations.

Once the parametric implantable antenna model has been selected, the next step is to appropriately tune its design parameters using an electromagnetic (EM) modeling and simulation program. The goal is to quickly calculate those parameter values which will optimize the antenna design in terms of impedance matching as well as exhibited radiation and patient safety performance at the desired operation frequency. The flowchart of the design strategy #6 is shown in Fig. 9 (Kiourti and Nikita 2012c). It is hereafter applied within the framework of tuning the parametric implantable antenna model of Fig. 8 for ICP monitoring (Warty et al. 2008) at 402 MHz (MICS band) (Kiourti and Nikita 2012b; Kiourti et al. 2012).

- **"*Fabrication-related*" *parameters*:** Initially, the "fabrication-related" parameters of the antenna are set to the values dictated by the intended fabrication procedure (Table 2). As part of this tutorial, Rogers

DOI 10.1007/978-981-4560-75-7_94-1

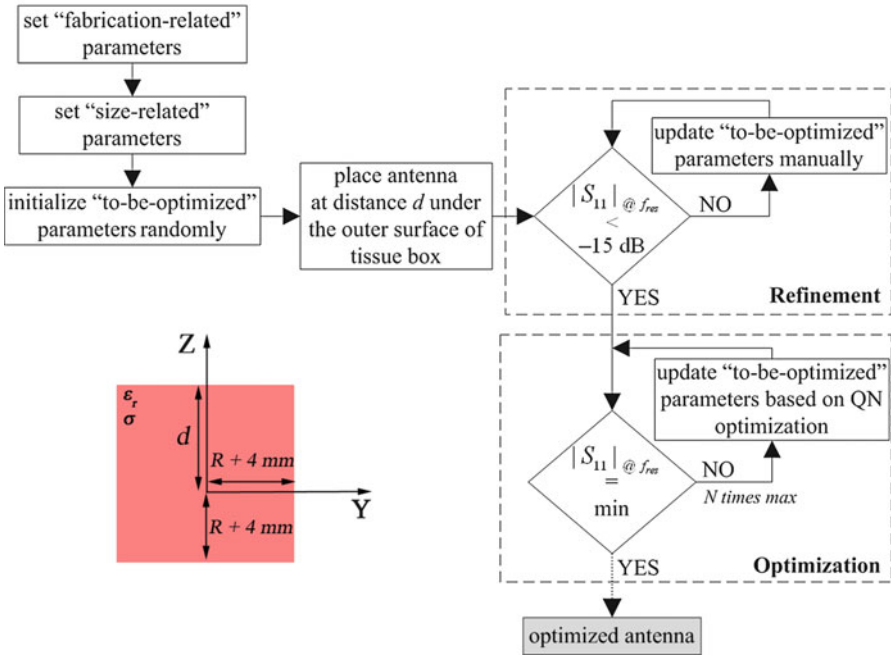


Fig. 9 Flowchart of the methodology for numerical design of implantable antennas

RO3210 ($\varepsilon_{rd} = 10.2$) dielectric sheets with a thickness of 0.635 mm ($h_1 = h_2 = h_3 = 0.635$ mm) are considered. The aforementioned dielectric sheets come premetallized with a 0.017 mm-thick electrodeposited copper foil ($h_m = 0.017$ mm). Sprayable glue 3 M 77 is used to bond the layers ($\varepsilon_{rg} = 2.0$), which has been found to exhibit an average thickness of 0.3 mm ($h_g = 0.3$ mm) for the fabrication process to be followed. The antenna is to be fed by means of a 50 mm-long ($L = 50$ mm) EZ-47 (center conductor diameter of 0.29 mm, PTFE dielectric with a diameter of 0.93 mm, outer conductor diameter of 1.19 mm) semirigid coaxial cable.

- **"*Size-related*" *parameters*:** Next, the "size-related" parameters of the antenna have to be selected, i.e., the parameters which determine the outer dimensions (physical size) of the antenna. In the parametric implantable antenna model considered in this tutorial, these are dictated by the antenna radius, R. Selection of the outer dimensions relies on the expertise and knowledge of the designer and must be performed based on the following two considerations. Firstly, size of the implantable antenna needs to take into account the desired implantation site and medical application scenario, as well as the size of the IMD in which it will be integrated. Secondly, miniaturization should not be set as the sole goal of the design. Previous studies have demonstrated degraded radiation and patient safety performance with size reduction for implantable antennas and have quantified this degradation as a function of size (Kiourti and Nikita 2012d). Given these considerations, a radius of $R = 6$ mm is selected for the ICP monitoring antenna under study (see Table 2).
- **"*To-be-optimized*" *parameters*:** The rest of the design parameters are considered as dimensions in the solution space and have to be tuned for an optimized 50 Ω impedance match at the desired operating frequency ("to-be-optimized"). Design is performed by: (a) setting the "fabrication-related" and "size-related" parameters to the values selected in the previous steps, (b) initializing the "to-be-optimized" parameters to random values, and (c) placing the antenna at a distance d under the outer surface of the tissue-simulating box shown in the inset of Fig. 3 (Kiourti and Nikita 2012c). The distance d corresponds to the actual air-to-antenna separation distance for the desired medical application scenario (implantation depth). The tissue-simulating box extends by $R + 4$ mm in the x- and y-directions (R is

DOI 10.1007/978-981-4560-75-7_94-1

Table 2 Parameter values selected for optimally tuning the implantable antenna model of Fig. 8 at 402 MHz (MICS band)

	Parameters	Values
"Fabrication-related"	ε_{rd}	10.2
	ε_{rg}	2.0
	h_1	0.635 mm
	h_2	0.635 mm
	h_3	0.635 mm
	h_m	0.017 mm
	h_g	0.3 mm
	L	50 mm
	Coaxial type	EZ-47
"Size-related"	R	6 mm
"To-be-optimized"	L_1	7.597 mm
	L_2	10.146 mm
	L_3	10.146 mm
	L_4	3.019 mm
	L_5	3.019 mm
	$L_{1'}$	11.397 mm
	$L_{2'}$	11.146 mm
	$L_{3'}$	11.146 mm
	$L_{4'}$	10.519 mm
	$L_{5'}$	10.519 mm
	$L_{6'}$	8.993 mm
	s_x	1 mm
	s_y	−4 mm
	f_x	0 mm
	f_y	4 mm

the maximum dimension of the antenna in the positive y-axis) and simulates the electrical properties of the intended implantation tissue.

In the "refinement" step of the design strategy, an approximate design is performed for the antenna. The "to-be-optimized" parameters are manually updated in an iterative way, until the magnitude of the reflection coefficient ($|S_{11}|$) at the desired operating frequency (f_{res}) satisfies:

$$|S_{11}|_{@f_{res}} < -15\ dB \tag{4}$$

Manual update relies on the skills and expertise of the designer, who is considered to be aware of the theoretical background related to antenna miniaturization (e.g., longer meanders are expected to increase the length of the current flow and result in lower resonance frequencies (Dey and Mittra 1996).

In the "optimization" step of the design strategy, antenna design is optimized. The "to-be-optimized" parameters are initialized to the values of the refinement step and are optimized based on a software-integrated optimization algorithm. The optimization process terminates when:

$$|S_{11}|_{@f_{res}} = \min \tag{5}$$

DOI 10.1007/978-981-4560-75-7_94-1

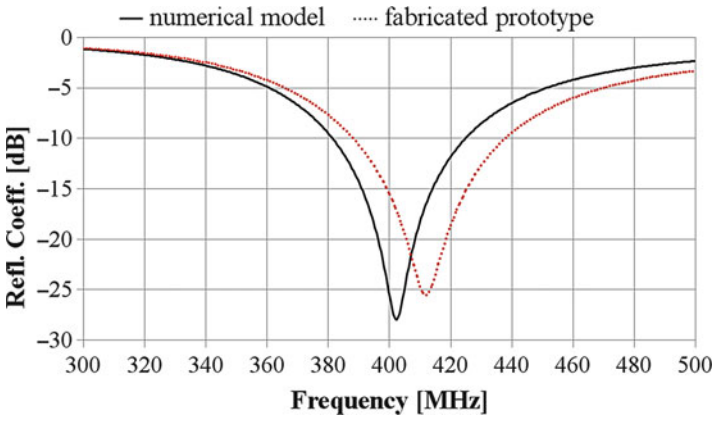


Fig. 10 Numerical and in vitro measured reflection coefficient frequency response of the proposed implantable antenna for ICP monitoring

or when the number of iteration exceeds a pre-defined maximum number.

As part of this tutorial, simulations are performed at a distance of $d = 5$ mm under the outer surface of a small ($R = 6$ mm) tissue-simulating box, which corresponds to the actual average implantation depth of an ICP monitor inside the human scalp. The tissue-simulating box represents skin tissue (scalp) electrical properties at $f_{res} = 402$ MHz ($\varepsilon_r = 46.7$, $\sigma = 0.69$ S/m), which are approximated as constant inside the 300 to 500 MHz range. Using this approximation, the maximum errors of ε_r and σ are given by 6.59 % and 8.89 %, respectively. The "to-be-optimized" parameters of the antenna are optimized based on quasi-Newton optimization (the maximum number of iterations is set to 300), due to its speed and accuracy in cases of insignificant numerical noise (Sun and Yuan 2006). Optimal parameter values are given in Table 2, whereas the reflection coefficient frequency response of the designed antenna is shown in Fig. 10 ("numerical model"). The antenna resonates at 402 MHz with a reflection coefficient of −27.9 dB and a wide 10 dB bandwidth of 44 MHz, which covers the MICS band.

Example Implantable Antennas

Relative position and orientation between an IMD and its exterior monitoring/control equipment counterpart is known a priori. As a result, patch designs are most commonly chosen for implantable antennas because they exhibit directive radiation patterns, lend themselves easily to a number of miniaturization techniques, and are highly flexible in design, conformability, and shape. In a realistic scenario, the implantable patch antenna will be mounted on the existing hardware of the IMD, which will also serve as its ground plane. Table 3 compares the volume occupied by MICS implantable antennas reported in the literature with respect to the applied miniaturization techniques. The bands of operation covered are also included in Table 3. When the number of bands of operation is increased, the size of the antenna is typically increased to cover them. Circular shape is generally preferred in order to avoid sharp edges that could cause injury. The performance of these antennas is further compared in Table 4 in terms of their 10 dB bandwidth (BW), maximum allowable input power levels imposed by the IEEE C95.1-1999 (1 g-avg SAR $\leq$ 1.6 W/kg (IEEE 1999)) (P_{1999}) and IEEE C95.1-2005 (10-g-avg SAR $\leq$ 2 W/kg (IEEE 2005)) (P_{2005}) safety guidelines, and maximum far-field gain (G_{max}). In general, increased-size implantable antennas exhibit more uniform distributions of the electric field and current density across an increased patch surface area, so that lower SAR values are obtained.

Nevertheless, it is remarked that dipole (Kim and Rahmat-Samii 2004), loop (Chen et al. 2009), monopole (Weiss et al. 2009), modified dipole (Scarpello et al. 2011), and 3D spiral (Abadia et al. 2009) antennas have also been reported for implantation purposes.

DOI 10.1007/978-981-4560-75-7_94-1

Table 3 A size comparison of implantable patch antennas reported in the literature

References	Bands [MHz]	Miniaturization technique				Volume [mm^3]
		Dielectric permittivity	Patch shape	Shorting pin	Patch stacking	
Kim and Rahmat-Samii 2004	402–405	10.2	Spiral	No	No	10,240.0
Soontornpipit et al. 2005	402–405	2.94	Waffle	Yes	No	6,480.0
Kim and Rahmat-Samii 2004	402–405	10.2	Spiral	Yes	No	6,144.0
Soontornpipit et al. 2004	402–405	6.1	Spiral	Yes	No	3,457.4
Sánchez-Fernández et al. 2010	402–405 2,400–2,800	6.1	SRR coupled to spiral	Yes	No	1,375.4
Karacolak et al. 2008	402–405 2,400–2,800	10.2	Meandered	Yes	No	1,265.6
Kim and Rahmat-Samii 2006	402–405	10.2	Spiral	Yes	No	1,200.0
Kim and Rahmat-Samii 1996	402–405	10.2	Meandered	Yes	No	1,200.0
Huang and Kishk 2011	402–405	9.4	Spiral	Yes	No	823.0
Lee et al. 2009	402–405	10.2	π-shaped	Yes	No	790.9
Vidal et al. 2013	402–405	6.7	Folded square	Yes	Yes	448.0
Lee et al. 2006	402–405	10.2	Hook-slotted	Yes	Yes	335.8
Permana et al. 2011	402–405	10.2	Spiral	Yes	Yes	273.6
Huang et al. 2011	402–405 433–435 2,400–2,480	10.2	Comb- and π-shaped	Yes	Yes	254.0
Permana et al. 2013	402–405	10.2	Spiral	Yes	Yes	254.0
Kiourti and Nikita 2012b	402–405	10.2	Meandered	Yes	Yes	203.6
Liu et al. 2008b	402–405	10.2	Spiral	Yes	Yes	190.0
Liu et al. 2008a	402–405	10.2	Hook-slotted	Yes	Yes	149.2
Liu et al. 2009	402–405	10.2	Hook-slotted	Yes	Yes	121.6
Kiourti et al. 2011b	402–405	10.2	Meandered	Yes	Yes	110.4
Kiourti et al. 2011a	402–405	9.4	Meandered	Yes	Yes	32.7

Fabrication and Testing of Implantable Antennas

Experimental investigations are required in order to confirm the validity of numerical simulations for implantable antennas. Of course, it is not possible to carry out measurements inside real operating scenarios (i.e., inside the human body). Therefore, experimental testing is performed by measuring the fabricated antenna prototypes inside either tissue-equivalent mediums (phantoms) or animal tissue.

Fabrication of Implantable Antennas

Fabrication of implantable antenna prototypes needs to deal with all challenges related to the fabrication of miniature antenna structures. For example, glue layers used to affix all components together strongly affect the antenna performance by shifting its resonance frequency and degrading its impedance matching. Furthermore, the coaxial cable feed used to connect the antenna with the network analyzer may give rise to radiating currents on the outer part of the cable, which, in turn, may deteriorate the

DOI 10.1007/978-981-4560-75-7_94-1

Table 4 A performance comparison of implantable patch antennas reported in the literature with respect to their occupied volume: 10 dB bandwidth (BW), maximum allowable input power levels imposed by the IEEE C95.1-1999 (P_{1999}) and IEEE C95.1-2005 (P_{2005}) standards, and maximum far-field gain (G_{max}) (N/A denotes that this information is not available)

References	Volume [mm^3]	BW [MHz]	P_{1999} [mW]	P_{2005} [mW]	G_{max} [dBi]
Kim and Rahmat-Samii 2004	10,240.0	20	8.791	N/A	N/A
Soontornpipit et al. 2005	6,480.0	16	N/A	N/A	N/A
Kim and Rahmat-Samii 2004	6,144.0	25	7.656	N/A	N/A
Soontornpipit et al. 2004	3,457.4	28	N/A	N/A	N/A
Sánchez-Fernández et al. 2010	1,375.4	12	N/A	N/A	−6
Karacolak et al. 2008	1,265.6	142	N/A	N/A	−25
Kim and Rahmat-Samii 2006	1,200.0	28	5.161	N/A	N/A
Kim and Rahmat-Samii 1996	1,200.0	40	5.442	N/A	N/A
Huang and Kishk 2011	823.0	25	5.820	N/A	N/A
Lee et al. 2009	790.9	120	5.714	N/A	−27
Vidal et al. 2013	448.0	110	3.7	N/A	N/A
Lee et al. 2006	335.8	50	4.798	N/A	−26
Permana et al. 2011	273.6	39	N/A	N/A	−24
Huang et al. 2011	254.0	113	4.692	N/A	−7
Permana et al. 2013	254.0	5	N/A	60.6	−40
Kiourti and Nikita 2012b	203.6	27	4.928	30.030	−37
Liu et al. 2008b	190.0	50	4.762	N/A	−26
Liu et al. 2008a	149.2	84	2.235	N/A	N/A
Liu et al. 2009	121.6	122	1.778	N/A	−38
Kiourti et al. 2011b	110.4	50	1.932	20.704	−46
Kiourti et al. 2011a	32.7	40	2.354	24.390	−45

measurements. Therefore, fabrication of implantable antennas is highly intriguing. In the following, the fabrication of implantable antennas will be discussed as part of fabricating the prototype of the design shown in Fig. 8. The latter incorporates several fabrication challenges that any prototype fabrication might have to deal with (e.g., multiple layers, inclusion of a shorting pin, etc.).

Specifically, three key aspects of the fabrication are considered to mainly influence the final antenna behavior: (a) substrate cutting, (b) substrate gluing, and (c) layer alignment. One of the problems is that these three steps are not necessarily independent. In fact, because the substrate material is relatively stiff, it cannot (or is hard to) be cut after the antenna has been assembled: micro-soldering of the coaxial cable and shorting pin are very fragile and cannot withstand the vertical pressure and torsion of the cutting tool. Furthermore, external alignment points are proved to be required for the assembling, to be removed after fabrication. Therefore, a mounting base (Fig. 11a) is suggested to be fabricated in order to help with the antenna's assembly. This base ensures the correct alignment between the three layers, while serving as the antenna support for the different soldering procedures. Based on the above, a typical fabrication methodology includes the following steps (Kiourti and Nikita 2014):

- **Photolithography masks**. Photolithography masks are prepared and printed, as shown in Fig. 12. The masks include: (a) a circular circumference which is used to guide the antenna cutting, (b) four circular marks which indicate the position of the holes that match the four pins of the mounting base (Fig. 11a) during the assembly procedure, (c) a square frame which matches the dimensions of the mounting base (Fig. 11a), and (c) complementary alignment marks to help in the alignment of the two sides of the bottom substrate layer (i.e., ground plane and lower patch).

DOI 10.1007/978-981-4560-75-7_94-1

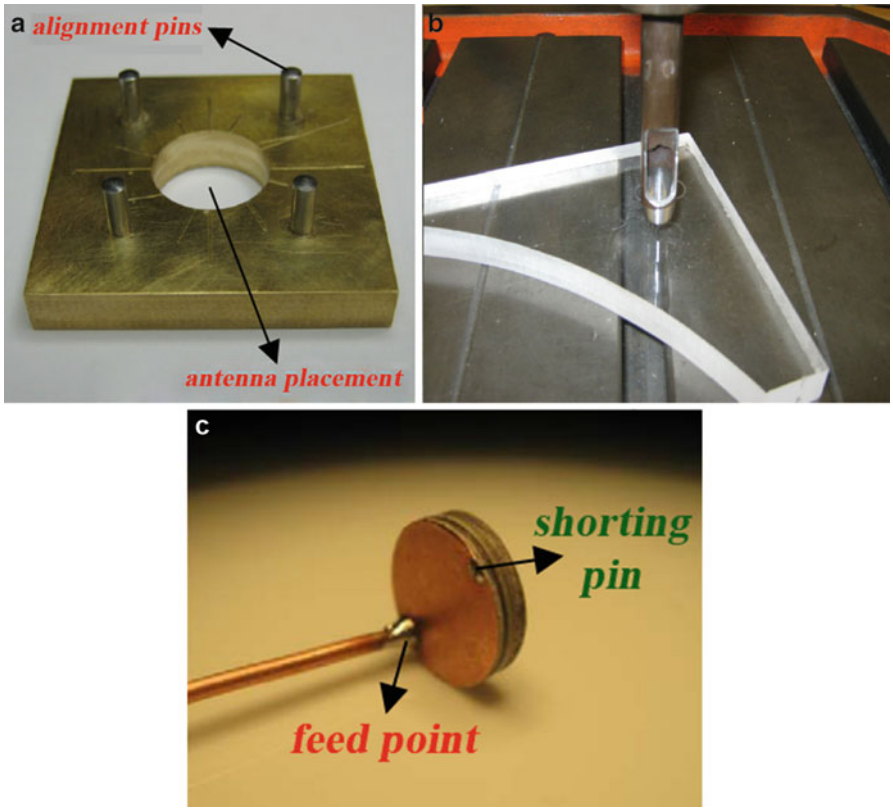


Fig. 11 Fabrication of implantable antennas: (**a**) mounting base, (**b**) circular cutting tool, and (**c**) fabricated antenna prototype

- **Photolithography**. The acquired Rogers RO3210 dielectric layers ($\varepsilon_{rd} = 10.2$, $h_1 = h_2 = h_3 = 0.635$ mm, $h_m = 0.017$ mm) are etched by means of a photolithographic process which makes use of the photolithography masks of the previous step. The lower substrate layer contains the ground plane and the lower patch, the upper substrate contains the upper patch, while the superstrate has no metallization. It is highlighted that due to the unavailability of biocompatible materials in some laboratories, other dielectrics with similar electrical properties may be selected for prototype fabrication. For instance, Rogers RO3210 is often used because it has similar properties as the biocompatible ceramic alumina ($\varepsilon_r = 9.4$, $\tan\delta = 0.006$) (Karacolak et al. 2008; Kiourti and Nikita 2012b; Liu et al. 2008a).
- **Cutting of the layers**. A circular cutting tool is then used to cut the antenna layers, as shown in Fig. 11b. The cutting tool exhibits a nominal diameter of 12 mm, which corresponds to the diameter of the intended antenna prototype ($R = 6$ mm). The adopted strategy is to precut the substrate down to a critical depth, just enough to keep the alignment points solidary with the patch, but weak enough to allow easy detaching without much mechanical stress to the antenna.
- **Antenna assembly**. The antenna is further assembled by making use of the mounting base of Fig. 11a. Layers are aligned and glued (3 M 77 glue: $\varepsilon_{rg} = 2.0$, $h_g = 0.3$ mm), while the shorting pin is set to connect the ground plane to the lower patch through a via. The outer conductor of the coaxial cable gets connected to the antenna ground plane, while the inner conductor gets simultaneously soldered to the lower and upper patches through vias. Nevertheless, detaching the antenna from the excess alignment material has been shown to be relatively hard, thus resulting in some stress to the fragile antenna.

DOI 10.1007/978-981-4560-75-7_94-1

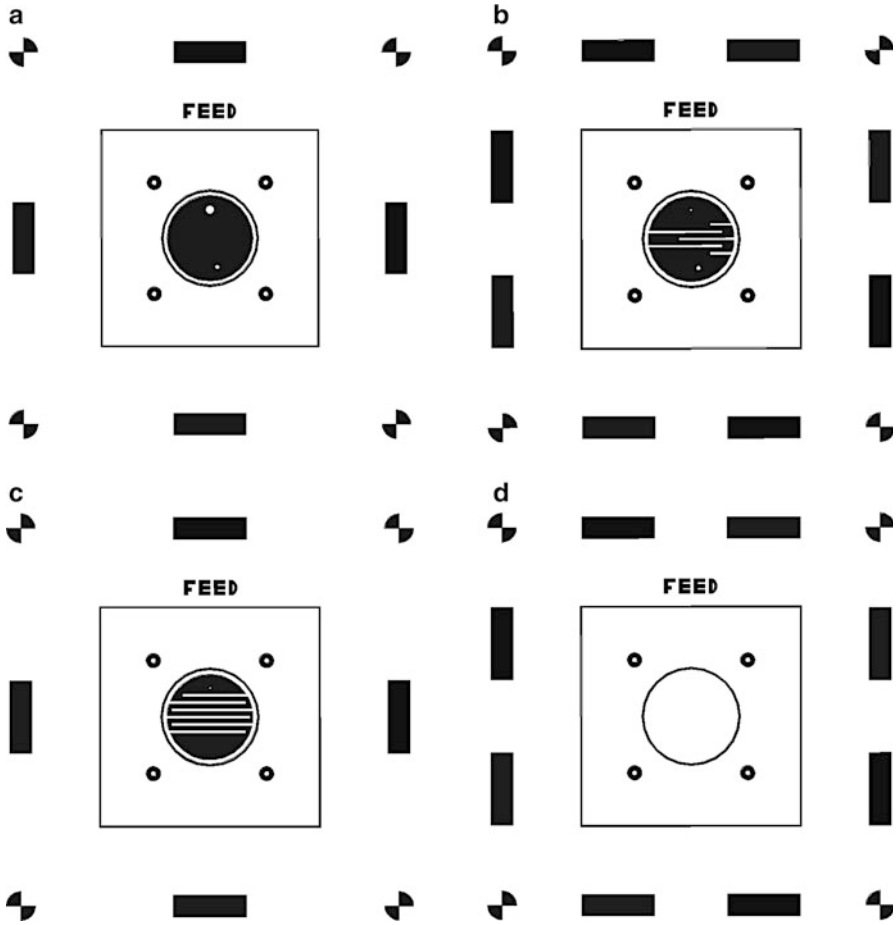


Fig. 12 Photolithography masks for printing the implantable antenna shown in Fig. 8: (**a**) ground plane, (**b**) lower patch, (**c**) upper patch, and (**d**) superstrate

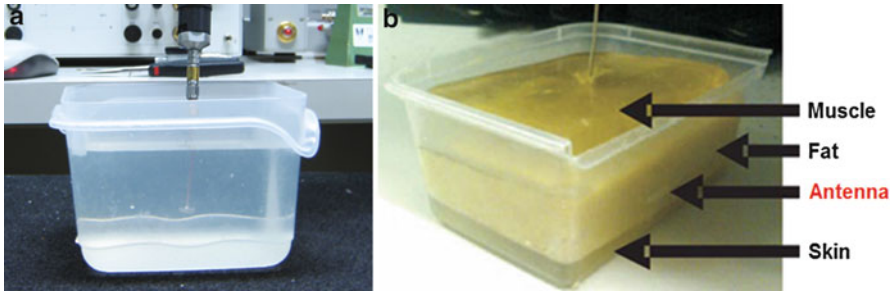


Fig. 13 Canonical phantoms used for testing of implantable patch antennas: (**a**) liquid (Kiourti and Nikita 2012b) and (**b**) multilayer gel (Sani et al. 2010)

The assembled antenna is shown in Fig. 11c. It is remarked that the most critical aspect regarding the fabrication of such miniature implantable antennas occurs to be the control of the glue layer thickness. This is impaired not only by the glue itself but also by the slight bump of the micro-solder near the coaxial cable and the shorting pin that prevents perfect contact between the layers.

DOI 10.1007/978-981-4560-75-7_94-1

Table 5 Phantoms used in the literature for testing of implantable antennas

Phantom			Measured electrical properties			
Tissue(s)	State	Ingredients	f [MHz]	ε_r	σ [S/m]	References
Skin	Liquid	Deionized water, sugar, salt	402	46.7	0.69	Kiourti and Nikita 2012b
	Liquid	Deionized water/sugar/salt/cellulose	402	49.6	0.51	Kim and Rahmat-Samii 2004
	Liquid	Deionized water, fruit sugar, salt, cellulose	402	46.7	0.69	Liu et al. 2008a
	Liquid	–	402	49.6	0.51	Kim and Rahmat-Samii 2004
	Gel	Deionized water, sugar, salt, agarose	402	46.7	0.69	Karacolak et al. 2008
	Gel	Deionized water, sugar, agarose	2,450	38.1	2.27	Karacolak et al. 2008
2/3 muscle	Liquid	Water, sugar, salt, TX-151 powder	402	48.9	0.71	Soontornpipit et al. 2004
	Liquid	Water, sugar, salt, cellulose cetylpyridinium chloride	403	41.3	–	Huang and Kishk 2011
Scalp	Gel	Water, salt, acrylamide, TMEDA, ammonium persulfate	2,450	50	2.2	Warty et al. 2008
Rat tissue	Gel	Deionized water, salt, DGBE	402	0.78	1.3	Karacolak et al. 2009
	Gel	Deionized water, DGBE, Triton X-100	2,450	0.73	1.27	Karacolak et al. 2009
(1) Skin (2) Fat (3) Muscle	(Multilayer) Gel	(1) Deionized water, sugar (2) Deionized water, salt, vegetable oil, flour (3) Deionized water, sugar, salt	868	(1) 38.7 (2) 4.9 (3) 53.0	(1) 0.77 (2) 0.04 (3) 0.92	Sani et al. 2010

In Vitro Testing of Implantable Antennas

Testing inside phantoms is relatively easy and practical to implement. The fabricated prototype is immersed inside a tissue phantom (i.e., a container filled with a liquid or gel material which mimics the electrical properties of biological tissue) and tested.

- **Phantom formulation**. Canonically shaped phantoms have, most commonly, been used for implantable antenna testing (see Fig. 13). In this case, the main challenge lies in the formulation (and, subsequently, characterization) of the tissue-emulating materials. In the literature, there have been several recipes presented for emulation of biological tissues at various frequencies (see Table 5). Gels rather than liquids are to be preferred in cases where multilayer phantoms and, thus, increased realism in experimental modeling are solicited. Deionized or distilled water usually acts as the base ingredient of the phantoms. Addition of sugar or glycerol reduces the permittivity, without, almost, affecting the conductivity. Salt increases the conductivity and, slightly, increases the permittivity (Karacolak et al. 2008). Solidification is usually made possible with agar. Other ingredients may also be used in order to vary the viscosity, preserve the mixture, and further control the permittivity and conductivity (Kiourti et al. 2014b; Ito et al. 2001). To prevent the formation of air bubbles and/or gaps, the mixture must be carefully heated and stirred and slowly poured inside the container of the phantom. Since it is not possible to produce a valid approximation to human tissue for a broad frequency spectrum using a single formula, separate recipes are given for different frequency bands (Karacolak et al. 2008).
- **Measurement of the phantom electrical properties**. In vitro testing of implantable antennas inside phantoms requires experimental measurement of the exhibited electrical properties in order to ensure

DOI 10.1007/978-981-4560-75-7_94-1

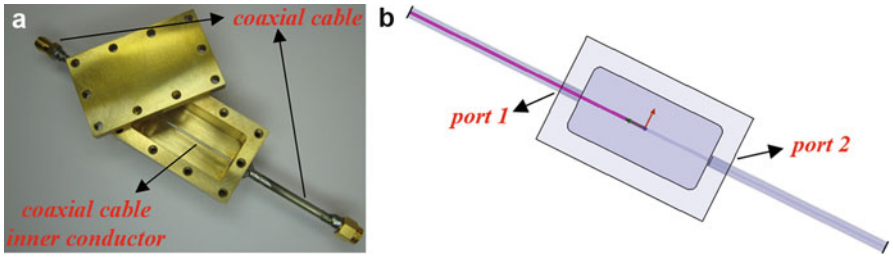


Fig. 14 Complex permittivity measurement of liquid and semisolid phantoms: (**a**) coaxial container and (**b**) numerical model (Kiourti et al. 2012)

conformance with the corresponding theoretical values. Recently, an in-depth analysis has been provided for the measurement of the electrical properties of biological media (Hofmann et al. 2013). In the market, there exist some commercial complex permittivity measurement systems, such as the Agilent Technologies 85071E (Agilent Technologies, Santa Clara, California, USA) or the SPEAG Dielectric Assessment Kit (SPEAG, Switzerland). However, alternative approaches are further solicited for laboratories that are not equipped with such commercial systems. For example, a low-cost and reliable complex permittivity measurement technique has recently been proposed (Kiourti et al. 2012). The measurement setup consists of a parallelepiped container intercepted by the inner conductor of a coaxial cable, as shown in Fig. 14a (exterior container size of 52mm × 32 mm × 32.2 mm, interior cavity size of 40 mm × 20 mm × 20 mm). The coaxial container is filled with the liquid or gel dielectric material under investigation, and once the lid is closed, it represents a transition between coaxial guides with a step characteristic impedance discontinuity. The transfer function between the two coaxial connectors outside the container depends upon the complex permittivity of the container's filling material. This can be de-embedded by comparing the measured scattering matrix (S-matrix) with simulation results for the same structure (Fig. 14b).

- **In vitro measurement of the implantable antenna**. Most studies for implantable antennas are limited to reflection coefficient (S_{11}) measurements (e.g., Kiourti and Nikita 2012b; Soontornpipit et al. 2005; Liu et al. 2008a; Sánchez-Fernández et al. 2010). As part of the experimental setup, prototype antennas are connected to a network analyzer through a coaxial cable, immersed inside the tissue-emulating phantom, and measured. Transmission coefficient (S21) measurements have also been reported. For example, in Warty et al. (2008), a 2,450 MHz PIFA was implanted inside a gel scalp phantom for intracranial pressure monitoring, and a linearly polarized 2,450 MHz chip antenna was used as the receiving (probing) antenna. Recently, in Kiourti et al. (2014b), a miniature broadband implantable antenna and a dual-band on-body antenna were presented along with the transmission performance between the two. The antennas were intended for integration into IMDs and on-body repeaters, respectively.

In Vivo Testing of Implantable Antennas

In vitro verification of an implantable antenna does not guarantee its proper functioning when implanted inside actual biological tissues (Karacolak et al. 2010). Therefore, once functionality of an implantable antenna prototype has been verified in vitro, in vivo testing is recommended to be further performed. Specifically, in vivo testing can be performed either (a) by embedding the implantable antenna inside tissue samples from donor animals (b) or by surgically implanting the antenna inside live model animals.

Use of animal tissue samples provides an easy approach to mimicking the frequency-dependency characteristic of the electrical properties of tissues. This can prove highly advantageous when carrying out measurements for multiband implantable antennas. In this case, the electrical properties of the test tissue

DOI 10.1007/978-981-4560-75-7_94-1

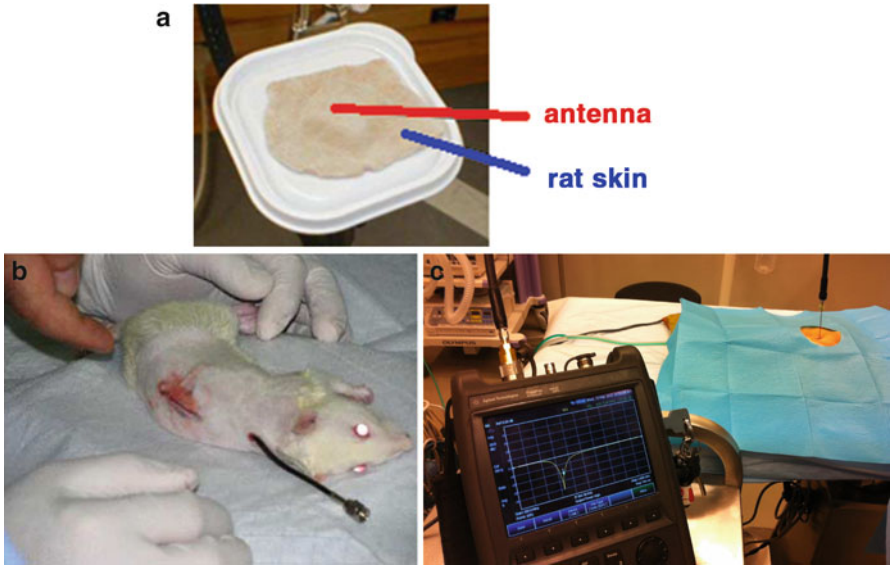


Fig. 15 In vivo testing of implantable antennas: (**a**) Karacolak et al. (2009), (**b**) (Karacolak et al. (2010), and (Kiourti et al. (2013)

can be measured using a dielectric probe kit and a network analyzer. In the literature, an implantable patch antenna with dual resonances at 380 and 440 MHz has been tested inside test tissue obtained by grinding the front leg of pork (Lee et al. 2009). The electrical properties of the adopted pork were found to be between those of human skin and muscle in the MICS band. A dual-band skin-implantable patch antenna operating in the MICS and 2,450 MHz ISM bands has also been tested in real animal skin (Karacolak et al. 2009). Skin samples with dimensions of $50 \times 50 \times 5$ mm^3 were extracted from the dorsal area of three donor rats to cover the designed antenna, and measurements were performed within 30 min of euthanization (Fig. 15a). Finally, a triple-band implantable patch antenna has been tested inside a minced front leg of pork (Huang et al. 2011). Electrical properties of the minced pork were measured and found to correspond to those of human skin and muscle between 100 MHz and 3 GHz.

Investigations inside living animals are also vital in order to investigate the effects of live tissue on the performance of implantable antennas, while providing valuable feedback for antenna design and analysis. The goal is to assess the effects of the following factors which are not accounted for in in vitro experimentation: (a) air gaps between the implanted antenna and the surrounding tissues, (b) presence of multiple types of tissues around the antenna, (c) dependence of tissue electrical properties upon frequency, (d) intersubject variability (anatomy and dependence of tissue electrical properties upon each rat's age, size, sex, internal body temperature, etc.), and (e) variations in the surgical procedures followed (implantation depth, implantation site, length of the wound, closure of the wound with sutures, etc.). In such measurements, the first step is the development of an experimental measurement protocol. The protocol has to be developed in cooperation with an experimental surgery unit, take into account legal requirements regarding the care and use of laboratory animals and address issues related to the type and number of model animals, implantation site of the antenna, anesthesia, surgical procedure, measurements, and post-surgery treatment. As an example, an in vivo experimentation protocol was recently developed for implantable antenna testing (Kiourti et al. 2013). This protocol takes into account the legal requirements regarding the care and use of laboratory animals and can be summarized as follows:

- **Type and number of model animals**. Implantation and measurements are carried out inside rats, which have long been used in the literature as model animals (Karacolak et al. 2010). Wistar outbred

DOI 10.1007/978-981-4560-75-7_94-1

rats (HsdOla:WI) are employed, which exhibit a mean and standard deviation (SD) body weight of 331.3 ± 9.2 g. In order to assess intersubject and surgical procedure variability, each antenna is implanted and measured inside three different rats.

- **Implantation site of the antenna**. Since the antennas under study have been designed for operation inside soft tissues, implantation is carried out within the subcutaneous tissue of the rats' abdomen.
- **Anesthesia**. Each rat is first anesthetized with an intraperitoneal (i.p.) injection of 70 mg/kg ketamine (Ketaset, Fort Dodge, Iowa, USA) and 5 mg/kg xylazine (Rompun, Bayer, Leverkusen, Germany).
- **Surgical procedure**. A wound is further made in the rat's abdomen area, and the antenna is implanted within the abdominal subcutaneous tissue. Following implantation, the wound is closed with 0/4 silk sutures (Silkam, Braun, Aesculap, Tuttlingen, Germany), leaving a 2 mm opening for the feeding coaxial cable to exit the skin.
- **Measurements**. Right after surgery, the feeding coaxial cable is connected to a network analyzer which measured the exhibited reflection coefficient frequency response.
- **Post-surgery treatment**. Once measurement is completed, the implanted antenna is removed, and the rat is euthanized in a CO2 chamber. Time lapse from the start of the surgical procedure to euthanasia of each rat does not exceed 8 min.

In vivo studies reported in the literature are very limited. The return loss frequency response of a skin-implantable antenna has been measured using rats as model animals (Karacolak et al. 2010; Fig. 15b). In this study, the antenna was implanted by means of a surgical operation inside the dorsal midline of three rats (for validation purposes), and euthanasia was applied after the measurements (approximately 13–15 min after the surgery). Canine studies for trans-scalp evaluation of a scalp-implantable antenna at 2,450 MHz have also been presented (Kawoos et al. 2008). Canine models were selected to ensure a large head size, and an intracranial pressure monitoring device with an integrated antenna was fixed to the skull. The monitor was tested while the dog was still under anesthesia. After the measurements, the animal was allowed to emerge from anesthesia and taken to the recovery area. Recently, implantable antennas were tested inside three different rats to assess intersubject variability considerations (Fig. 15c; Kiourti et al. 2013). In this case, numerical and experimental results were found to exhibit quite good agreement. Compared to numerical simulations, percentage changes in the exhibited resonance frequency (f_{res}), reflection coefficient at this frequency ($|S_{11}|@f_{res}$), and 10 dB bandwidth (BW) are found to equal +6.9 %, +51.9 %, and +30.2 %, respectively. Maximum deviations in f_{res}, $|S_{11}|@f_{res}$, and BW recorded among the three measurements in different rats were found to equal 43 MHz, 12.6 dB, and 23 MHz, respectively.

Full Solutions of Commercial IMDs

A number of commercial IMDs have already been reported, as shown in Table 6. Defibrillators and pacemakers are the most common examples of IMDs. Biotronik has recently proposed a small battery-powered electrical impulse generator to be implanted in patients who are at risk of sudden cardiac death due to ventricular fibrillation and ventricular tachycardia (Lumax 540 DR-T) (Biotronik 2012). The Medtronic Adapta with MVP pacing system offers managed ventricular pacing, atrial therapy, ventricular capture, and remote cardiac telemetry (Medtronic 2010a), whereas the Medtronic Revo MRI SureScan pacing system is magnetic resonance (MR) conditional designed to allow patients to undergo magnetic resonance imaging (MRI) under the specified conditions of use (Medtronic 2011a). VeriChip (renamed to PositiveID in 2010) was the only human-implantable microchip to receive approval by the Food and Drug Administration (FDA) in 2004 (PositiveID 2004). The device was typically implanted between the shoulder and elbow of the individual, and, once scanned, it replied with a unique 16-digit number,

DOI 10.1007/978-981-4560-75-7_94-1

Table 6 Example full solutions of commercial IMDs

Commercial IMD	Function of the IMD
Biotronik Lumax	Defibrillator
Medtronic Adapta with MVP	Pacing system
Medtronic Revo MRI SureScan	Pacing system for magnetic resonance imaging
VeriChip/PositiveID	Identity verification
Nucleus Freedom cochlear	Cochlear implant
Medtronic SynchroMed Pump	Drug infusion system
Second Sight Argus II	Retinal prosthesis
Medtronic Guardian REAL-Time	Glucose monitor
Abbott FreeStyle Navigator	Glucose monitor
Dexcom SEVEN PLUS	Glucose monitor
Medtronic MiniMed Paradigm Veo	Glucose monitor

which could be used to retrieve personalized information, for example, identity verification, medical records, etc. However, privacy concerns generated controversy and debate, and marketing was discontinued in 2010. The Nucleus Freedom cochlear implant includes a sound processor which is worn behind the ear and a cochlear implant which is placed under the skin, behind the ear (Nucleus Freedom 2010). The sound processor captures sounds, digitizes them, and sends the digital code to the implant. The implant converts the digitally coded sound to electrical impulses and sends them along an electrode array to further stimulate the cochlea's hearing nerve. The Medtronic SynchroMed Pump is a drug infusion system which provides precise drug delivery for chronic therapy of severe spasticity (Medtronic 2012). In addition to the implanted pump, the system uses a catheter to deliver programmed amounts of intrathecal baclofen to the intrathecal space and cerebrospinal fluid. The Argus II retinal implant was approved by FDA's Ophthalmic Devices Advisory Panel in 2012 (Second Sight 2012). It includes a video camera, a transmitter mounted on a pair of eyeglasses, a video processing unit, and a 60-electrode implanted retinal prosthesis that replaces the function of degenerated cells in the retina. Although it does not fully restore vision, this setup can improve a patient's ability to perceive images and movement. Finally, implantable glucose monitoring systems appear as a promising treatment for diabetes on a continuous basis (e.g., Medtronic Guardian REAL-Time (Medtronic 2010b), Medtronic MiniMed Paradigm Veo (Medtronic 2011b), Dexcom SEVEN PLUS (Dexcom 2008), Abbott FreeStyle Navigator (Abbott 2011)). A tiny sensor is inserted under the skin to measure glucose levels and further transmits this information to an exterior monitor via radio waves.

Conclusion

IMDs with wireless telemetry functionalities appear as a highly promising option toward improving the patients' quality of life and providing medical systems with constant availability, context awareness, reconfigurability, and unobtrusiveness. As such, a key and critical component of RF-linked IMDs is the integrated implantable antenna.

In this chapter, issues related to the design, simulations, and experimental investigations of implantable antennas for biomedical telemetry were discussed. Specifically, design mainly emphasizes on miniaturization issues and biocompatibility. However, electrically small antennas present poor radiation performance and relatively narrow bandwidth. Even though gain enhancement is considered crucial, compromises on the system performance are generally inevitable. Conserving energy to extend the lifetime of the IMD is also significant. Multiband antennas are being designed for this purpose, which

DOI 10.1007/978-981-4560-75-7_94-1

"wake up" the IMD only when there is a need for information exchange. Several methodologies have been proposed for implantable antenna design, all of which need to take into account the host body. Simplified tissue models are proved to be able to substitute for complex anatomical tissue models, thus speeding up simulations. Although a homogenous model is sufficient for basic antenna design, a more realistic model is needed to refine the final antenna design and provide accurate results. Using efficient and accurate simulation tools and tissue models is a key issue for both design and performance analysis. Regarding experimental investigations, implantable antennas exhibit tight fabrication tolerances as attributed to their miniature size. Testing inside tissue-emulating phantoms mainly needs to deal with the formulation and characterization of the tissue-mimicking liquid or gel. To benefit from frequency-dependent tissue electrical properties, testing in animal tissue samples can additionally be performed. The highest challenge, however, lies in measurements within living animals, in which careful consideration is required for developing the optimal testing protocol.

IMDs are a growing technology with a high potential for improving the patients' life and the quality of healthcare. RF technology for IMDs promises many benefits for both patients and caregivers. With rising health-care costs, an aging population, a growing acceptance of home-based medical monitoring, and advances in supporting technology, IMDs are gaining a continually increasing interest in both academia and industry.

References

Abadia J, Merli F, Zurcher JF, Mosig JR, Skrivervik AK (2009) 3D spiral small antenna design and realization for biomedical telemetry in the MICS band. Radioengineering 18(4):359–367

Abbott (2011) FreeStyle navigator, Alameda. https://www.abbottdiabetescare.com/. Last day accessed 4 Aug 2014

Ahmed Y, Hao Y, Parini C (2008) A 31.5 GHz patch antenna design for medical implants. Hindawi Int J Antennas Propag 2008:1–6

Azad MZ, Ali M (2009) A miniature implanted inverted-F antenna for GPS application. IEEE Trans Antennas Propag 57(6):1854–1858

Bao JZ, Lu ST, Hurt WD (1997) Complex dielectric measurements and analysis of brain tissues in the radio and microwave frequencies. IEEE Trans Microw Theory Techn 45(10):1730–1741

Biotronik (2012) Lumax, Berlin. http://www.biotronik.com/wps/wcm/connect/en_de_web/biotronik/sub_top/healthcareprofessionals/products/tachyarrhythmia_therapy/. Last day accessed 4 Aug 2014

Buchegger T, Obberger G, Reisenzahn A, Hochmair E, Stelzer A, Springer A (2005) Ultra-wideband transceivers for cochlear implants. EURASIP J App Signal Process 18:3069–3075

Chen ZN, Liu GC, See TSP (2009) Transmission of RF signals between MICS loop antennas in free space and implanted in the human head. IEEE Trans Antennas Propag 57(6):1850–1853

Chow EY, Chlebowski AL, Chakraborty S, Chappell WJ, Irazoqui PP (2010) Fully wireless implantable cardiovascular pressure monitor integrated with a medical stent. IEEE Trans Biomed Eng 57(6):1487–1496

Chow EY, Morris MM, Irazoqui PP (2013) Implantable RF medical devices. IEEE Microw Mag 14(4):64–73

Conil E, Hadjem A, Lacroux F, Wong MF, Wiart J (2008) Variability analysis of SAR from 20 MHz to 2.4 GHz for different adult and child models using finite-difference time-domain. Phys Med Biol 53:1511–1525

Dexcom (2008) Seven plus, San Diego. http://www.dexcom.com/seven-plus. Last day accessed 4 Aug 2014

Dey S, Mittra R (1996) Compact microstrip patch antenna. Microw Opt Technol Lett 13(1):12–14
Furse CM (2009) Biomedical telemetry: today's opportunities and challenges. In: IEEE international workshop on antenna technology
Gabriel C (2005) Dielectric properties of biological tissue: variation with age. Bioelectromagnetics 26(Suppl 7):S12–S18
Gabriel C, Gabriel S, Corthout E (1996a) The dielectric properties of biological tissues: I. Literature survey. Phys Med Biol 41:2231–2249
Gabriel S, Lau RW, Gabriel C (1996b) The dielectric properties of biological tissues: II. Measurements in the frequency Range 10 Hz to 20 GHz. Phys Med Biol 41:2251–2269
Gabriel S, Lau RW, Gabriel C (1996c) The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues. Phys Med Biol 41:2271–2293
Gemio J, Parron J, Soler J (2010) Human body effects on implantable antennas for ISM bands applications: models comparison and propagation losses study. Prog Electromgn Res 110:437–452
Gosalia K, Lazzi G, Humayun M (2004) Investigation of microwave data telemetry link for a retinal prosthesis. IEEE Trans Microwave Theory Tech 52(8):1925–1932
Greatbatch W, Homes CF (1991) History of implantable devices. IEEE Eng Med Biol Mag 10(3):38–41
Guillory K, Normann RA (1999) A 100-channel system for real time detection and storage of extracellular spike waveforms. J Neurosci Methods 91:21–29
Hofmann M, Fischer G, Weigel R, Kissinger D (2013) Microwave-based noninvasive concentration measurements for biomedical applications. IEEE Trans Microwave Theory Tech 61:2195–2204
Huang W, Kishk AA (2011) Embedded spiral microstrip implantable antenna. Hindawi Int J Antennas Propag 2011:1–6
Huang FJ, Lee CM, Chang CL, Chen LK, Yo TC, Luo CH (2011) Rectenna application of miniaturized implantable antenna design for triple-band biotelemetry communication. IEEE Trans Antennas Propag 59(7):2646–2653
Institute of Electrical and Electronics Engineers (IEEE) Std 802.15.6. 2012 (2012) IEEE standard for local, metropolitan area networks: wireless body area networks. International Committee on Electromagnetic Safety, The Institute of Electrical and Electronics Engineers, New York
Institute of Electrical and Electronics Engineers (IEEE) Std 95.1-1999 (1999) IEEE standard for safety levels with respect to human exposure to radiofrequency electromagnetic fields, 3 kHz to 300 GHz. International Committee on Electromagnetic Safety, The Institute of Electrical and Electronics Engineers, New York
Institute of Electrical and Electronics Engineers (IEEE) Std 95.1-2005. 2005 (2005) IEEE standard for safety levels with respect to human exposure to radiofrequency electromagnetic fields, 3 kHz to 300 GHz. International Committee on Electromagnetic Safety, The Institute of Electrical and Electronics Engineers, New York
International Commission on Non-Ionizing Radiation Protection (ICNIRP) (1998) Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz). Health Phys 74:494–522
International Telecommunications Union- Radiocommunication Sector (ITU-R) (1998) Radio regulations, SA.1346. http://itu.int/home
International Telecommunications Union- Radiocommunication Sector (ITU-R) (2008) Radio regulations, section 5.138 and 5.150. http://itu.int/home
Ito K, Furuya K, Okano Y, Hamada L (2001) Development and characteristics of a biological tissue-equivalent phantom for microwaves. Electron Commun Jpn 84:67–77

Karacolak T, Hood AZ, Topsakal E (2008) Design of a dual-band implantable antenna and development of skin mimicking gels for continuous glucose monitoring. IEEE Trans Microw Theory Tech 56(4):1001–1008

Karacolak T, Cooper R, Topsakal E (2009) Electrical properties of rat skin and design of implantable antennas for medical wireless telemetry. IEEE Trans Antennas Propag 57(9):2806–2812

Karacolak T, Cooper R, Butler J, Fisher S, Topsakal E (2010) In vivo verification of implantable antennas using rats as model animals. IEEE Antennas Wirel Propag Lett 9:334–337

Kawoos U, Tofighi MR, Warty R, Kralick FA, Rosen A (2008) In-vitro and in-vivo trans-scalp evaluation of an intracranial pressure implant at 2.4 GHz. IEEE Trans Microwave Theory Tech 56(10):2356–2365

Kendir GA, Liu W, Wang G, Sivaprakasam M, Bashirullah R, Humayun MS, Weil JD (2005) An optimal design methodology for inductive power link with class-E amplifier. IEEE Trans Circuits Syst 52(5):857–866

Kim J, Rahmat-Samii Y (1996) Planar inverted F antennas on implantable medical devices: meandered type versus spiral type. Microw Opt Technol Lett 48(3):567–572

Kim J, Rahmat-Samii Y (2004) Implanted antennas inside a human body: simulations, designs, and characterizations. IEEE Trans Microw Theory Tech 52(8):1934–1943

Kim J, Rahmat-Samii Y (2006) SAR reduction of implanted planar inverted F antennas with non-uniform width radiator. In: IEEE international symposium on antennas and propagation, Albuquerque

Kiourti A, Nikita KS (2011) Meandered versus spiral novel miniature PIFAs implanted in the human head: tuning and performance. In: 2nd ICST international conference on wireless mobile communication and healthcare (MobiHealth 2012), Kos Island

Kiourti A, Nikita KS (2012a) A review of implantable patch antennas for biomedical telemetry: challenges and solutions. IEEE Antennas Propag Mag 54(3):210–228

Kiourti A, Nikita KS (2012b) Miniature scalp-implantable antennas for telemetry in the MICS and ISM Bands: design, safety considerations and link budget analysis. IEEE Trans Antennas Propag 60(6):3568–3575

Kiourti A, Nikita KS (2012c) Accelerated design of optimized implantable antennas for medical telemetry. IEEE Antennas Wirel Propag Lett 11:1655–1658

Kiourti A, Nikita KS (2012d) Miniaturization vs gain and safety considerations of implantable antennas for wireless biotelemetry. In: International symposium on antennas and propagation, Chicago

Kiourti A, Nikita KS (2013a) Numerical assessment of the performance of a scalp-implantable antenna: effects of head anatomy and dielectric parameters. Bioelectromagnetics 34(3):167–179

Kiourti A, Nikita KS (2013b) Design of implantable antennas for medical telemetry: dependence upon operation frequency, tissue anatomy, and implantation site. IGI Global Int J Monit Surveillance Technol Res (IJMSTR) 1(1):16–33

Kiourti A, Nikita KS (2014) Implantable antennas: a tutorial on design, fabrication, and in vitro/in vivo testing. IEEE Microw Mag 15(4):77–91

Kiourti A, Christopoulou M, Nikita KS (2011a) Performance of a novel miniature antenna implanted in the human head for wireless biotelemetry. In: IEEE international symposium on antennas and propagation, Spokane

Kiourti A, Tsakalakis M, Nikita KS (2011b) Parametric study and design of implantable PIFAs for wireless biotelemetry. In: 2nd ICST international conference on wireless mobile communication and healthcare, Kos Island

Kiourti A, Costa JR, Fernandes CA, Santiago AG, Nikita KS (2012) Miniature implantable antennas for biomedical telemetry: from simulation to realization. IEEE Trans Biomed Eng 59(11):3140–3147

DOI 10.1007/978-981-4560-75-7_94-1

Kiourti A, Psathas KA, Lelovas P, Kostomitsopoulos N, Nikita KS (2013) In vivo tests of implantable antennas in rats: antenna size and inter-subject considerations. IEEE Antennas Wirel Propag Lett 12:1396–1399

Kiourti A, Psathas KA, Nikita KS (2014a) Implantable and ingestible medical devices with wireless telemetry functionalities: a review of current status and challenges. Wiley Bioelectrom 35(1):1–15

Kiourti A, Costa JR, Fernandes CA, Nikita KS (2014b) A broadband implantable and a dual-band on-body repeater antenna: design and transmission performance. IEEE Trans Antennas Propag 62(6):2899–2908

Lee CM, Yo TC, Luo CH (2006) Compact broadband stacked implantable antenna for biotelemetry with medical devices. In: IEEE Annual conference on wireless and microwave technology, Clearwater

Lee CM, Yo TC, Huang FJ, Luo CH (2009) Bandwidth enhancement of planar inverted-F antenna for implantable biotelemetry. Microw Opt Technol Lett 51(3):749–752

Liu WC, Chen SH, Wu CM (2008a) Implantable broadband circular stacked PIFA antenna for biotelemetry communication. J Electromagn Waves Appl 22(13):1791–1800

Liu WC, Yeh FM, Ghavami M (2008b) Miniaturized implantable broadband antenna for biotelemetry communication. Microw Opt Technol Lett 50(9):2407–2409

Liu WC, Chen SH, Wu CM (2009) Bandwidth enhancement and size reduction of an implantable PIFA antenna for biotelemetry devices. Microw Opt Technol Lett 51(3):755–757

Medtronic (2010a) Adapta with MVP Pacing System, Minneapolis. http://www.medtronic.com/for-healthcare-professionals/products-therapies/cardiac-rhythm/pacemakers/adapta-with-mvp-pacing-system/.Last day accessed 4 Aug 2014

Medtronic (2010b) Guardian REAL-Time, Minneapolis. http://www.medtronic-diabetes-me.com/Guardian-REAL-Time.html. Last day accessed 4 Aug 2014

Medtronic (2011a) Revo MRI SureScan, Minneapolis. http://www.medtronic.com/for-healthcare-professionals/products-therapies/cardiac-rhythm/pacemakers/revo-mri-pacing-system/. Last day accessed 4 Aug 4 2014

Medtronic (2011b) MiniMed Paradigm Veo, Minneapolis. http://www.medtronic-diabetes.co.uk/product-information/paradigm-veo/index.html. Last day accessed Aug 4 2014

Medtronic (2012) SynchroMed II Pump, Minneapolis. http://www.medtronic.com/patients/severe-spasticity/therapy/itb-therapy/synchromed-ii-pump/. Last day accessed 4 Aug 2014

Mitcheson PD, Green TC, Yeatman EM, Holmes AS (2004) Architectures for vibration-driven micropower generators. IEEE J Microelectromech Syst 13(3):429–440

Noroozi Z, Hojjat-Kashani F (2012) Three-dimensional FDTD analysis of the dual-band implantable antenna for continuous glucose monitoring. Prog Electromagn Res Lett 28:9–21

Nucleus Freedom (2010) Cochlear Nucleus System. http://www.cochlear.com/wps/wcm/connect/us/recipients/recipient-support/nucleus. Last day accessed 4 Aug 2014

Permana H, Fang Q, Cosic I (2011) 3-Layer implantable microstrip antenna optimized for retinal prosthesis system in MICS band. In: IEEE international symposium on bioelectronics and bioinformatics

Permana H, Fang Q, Rowe WST (2013) Hermetic implantable antenna inside vitreous humor simulating fluid. Prog Electromagn Res 133:571–590

PositiveID (2004) PositiveID/VeriChip, Delray Beach. http://www.positiveidcorp.com/index.html. Last day accessed 4 Aug 2014

Rucker D, Al-Alawi A, Adada R, Al-Rizzo HM (2007) A miniaturized tunable microstrip antenna for wireless communications with implanted medical devices. In: 2nd international ICST conference on body area networks, Brussels

DOI 10.1007/978-981-4560-75-7_94-1

Sánchez-Fernández CJ, Quevedo-Teruel O, Requena-Carrión J, Inclán-Sánchez L, Rajo-Iglesias E (2010) Dual-band microstrip patch antenna based on short-circuited ring and spiral resonators for implantable medical devices. IET Microw Antennas Propag 4(8):1048–1055

Sani A, Alomainy A, Hao Y (2009) Numerical characterization and link budget evaluation of wireless implants considering different digital human phantoms. IEEE Trans Microw Theory Tech 57(10):2605–2613

Sani A, Rajab M, Foster R, Hao Y (2010) Antennas and propagation of implanted RFIDs for pervasive healthcare applications. Proc IEEE 98:1648–1655

Savci HS, Sula A, Wang Z, Dogan NS, Arvas E (2005) MICS transceivers: regulatory standards and applications. In: Proceedings of the IEEE international southeast conference

Scanlon WG, Evans NE, McCreesh ZM (1997) RF performance of a 418 MHz radio telemeter packaged for human vaginal placement. IEEE Trans Biomed Eng 44(5):427–430

Scanlon WG, Burns JB, Evans NE (2000) Radiowave propagation from a tissue-implanted source at 418 MHz and 916.5 MHz. IEEE Trans Biomed Eng 47:527–534

Scarpello ML, Kurup D, Rogier H, Ginste DV, Axisa F, Vanfleteren J, Joseph W, Martens L, Vermeeren G (2011) Design of an implantable slot dipole conformal flexible antenna for biomedical applications. IEEE Trans Antennas Propag 59(10):3556–3564

Schmid G, Neubauer G, Illievich UM, Alesch F (2003) Dielectric properties of porcine brain tissue in the transition from life to death at frequencies from 800 to 1900 MHz. Bioelectromagnetics 24:413–422

Second Sight (2012) Argus II retinal prosthesis system, Sylmar. http://2-sight.eu/en/product-en. Last day accessed 4 Aug 2014

Shults MC, Rhodes RK, Updike SJ, Gilligan BJ, Reining WN (1994) A telemetry-instrumentation system for monitoring multiple subcutaneously implanted glucose sensors. IEEE Trans Biomed Eng 41(10):937–942

Skrivervik AK, Merli F (2011) Design strategies for implantable antennas. In: Antennas and propagation conference, Loughborough, Nov 2011

Soontornpipit P, Furse CM, Chung YC (2004) Design of implantable microstrip antenna for communication with medical implants. IEEE Trans Microw Theory Tech 52(8):1944–1951

Soontornpipit P, Furse CM, Chung YC (2005) Miniaturized biocompatible microstrip antenna using genetic algorithm. IEEE Trans Antennas Propag 53(6):1939–1945

Sun W, Yuan YX (2006) Optimization theory and methods. Springer, New York

Tang Z, Smith B, Schild JH, Peckham PH (1995) Data transmission from an implantable biotelemeter by load-shift keying using circuit configuration modulator. IEEE Trans Biomed Eng 42(5):524–528

Valdastri P, Menciassi A, Arena A, Caccamo C, Dario P (2004) An implantable telemetry platform system for in vivo monitoring of physiological parameters. IEEE Trans Inf Technol Biomed 8(3):271–278

Vidal N, Curto S, Lopez Villegas JM, Sieiro J, Ramos FM (2012) Detuning study of implantable antennas inside the human body. Prog Electromagn Res 124:265–283

Vidal N, Lopez-Villegas JM, Curto S, Colomer J, Ahyoune S, Garcia A, Sieiro JJ, Ramos FM (2013) Design of an implantable broadband antenna for medical telemetry applications. In: 7th European conference on antennas and propagation

Virtanen H, Keshvari J, Lappalainen R (2006) Interaction of radio frequency electromagnetic fields and passive metallic implants-a brief review. Bioelectromagnetics 27:431–439

Warty R, Tofighi MR, Kawoos U, Rosen A (2008) Characterization of implantable antennas for intracranial pressure monitoring: reflection by and transmission through a scalp phantom. IEEE Trans Microw Theory Tech 56(10):2366–2376

Weiss MD, Smith JL, Bach J (2009) RF coupling in a 433 MHz biotelemetry system for an artificial hip. IEEE Antennas Wirel Propag Lett 8:916–919

DOI 10.1007/978-981-4560-75-7_94-1

Wessels D (2002) Implantable pacemakers and defibrillators: device overview and EMI considerations. In: Proceedings of the IEEE international symposium electromagnetic compatibility (EMC 2002)

Xia W, Saito K, Takahashi M, Ito K (2009) Performances of an implanted cavity slot antenna embedded in the human arm. IEEE Trans Antennas Propag 57(4):894–899

Zarlink (2006) Medical implantable RF transceiver ZL70101 datasheet, Zarlink Semiconductor, Ottawa

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_95-1

Antennas in Partial Discharge Sensing System

Guillermo Robles[a]*, Ricardo Albarracín[b] and José Luis Vázquez[c]
[a]Department of Electrical Engineering, University Carlos III of Madrid, Leganes, Madrid, Spain
[b]Design and Innovation of distribution networks Technological surveillance, Boslan S.A., Madrid, Spain
[c]Department of Signal Theory and Communications, University Carlos III of Madrid, Leganes, Spain

Abstract

The exponential growing demand of electricity has stimulated the manufacture of electric equipment with high rated powers withstanding tens and hundreds of kilovolts. These devices have to be insulated to ensure a safe and reliable service while their size and cost are reduced. Unfortunately, insulations deteriorate over time by being in operation under load and exposed to harsh environments that can degrade their behavior and lead to unexpected equipment outages and failures. The continuous monitoring of these assets is paramount in the operation of electric power systems, and one of the most popular methods to evaluate the ageing is the detection of partial discharges. Partial discharges are ionization processes that take place in voids filled with gas or oil inside the insulation, in dielectric surfaces, and in the proximity of sharp metallic objects. The chemical and physical structure of the insulation is changed, and eventually weakened, by the continuous action of the discharges. Then, their apparition can be directly a signal of problems in the insulation, but they can also be the consequence of other degradation processes. Partial discharges can be measured with a wide range of detectors including inductive, capacitive, acoustic, and light sensors. Because partial discharges occur in extremely short times, well below 1 ns, the radiofrequency measurement of the phenomenon in the HF, VHF, and UHF bands is also part of the unconventional methods used for their detection. EM sensors or antennas have the ability of performing a complete study on the measurement of partial discharges. They can detect pulses, localize the partial discharge site, and, to some extent, classify the type of partial discharge online. However, one of the most important challenges when using antennas in the diagnostic of insulations is the difficulty of relating the RF emissions to the severity of the PD. Another determent in the wide application of antennas as partial discharge detectors is the sample rate needed to obtain information from the signals in the time and frequency domains. The chapter also explores the most common configurations of antennas used in the detection of partial discharges as well as how they are installed and used in different electrical machines.

Keywords

Partial discharges; Ultrahigh-frequency detection techniques; Electrical insulation diagnosis; Sensors; Instrumentation; Electrical measurements; Ultrahigh-frequency measurements; Antenna parametrization; Electrical maintenance

*Email: grobles@ing.uc3m.es

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_95-1

Introduction

The Importance of Insulation Systems

The exponential growing demand of electricity all over the world and the traditional dispersion of power plants from the consumption areas require an expansion of the network capacities, an improvement of their availability, and a reduction of congestion and outages (IEC 2010). The demand has also stimulated the manufacture of electric equipment with higher rated powers, and, considering that voltages have to be elevated to reduce losses in the transmission networks, the consequences are apparatuses withstanding tens and hundreds of kilovolts that have to be separated from ground through relatively small distances. The role of insulation systems is to ensure that the electrical assets can work in a reliable and safe environment while their size and cost are reduced. Unfortunately, insulations deteriorate over time by being in operation under load and exposed to harsh environments, as well as to unpredictable mechanical, thermal, and electric stresses that can degrade their behavior and lead to unexpected equipment outages and failures.

Electric Maintenance

Knowing the condition of the insulation is paramount to the reliability of power systems. For this reason, maintenance strategies are scheduled to keep reliability within a margin of quality throughout the lifetime of assets reducing the costs of malfunction, failure of equipments, and blackouts. These corrective, preventive, predictive, or proactive strategies facilitate the adoption of decisions regarding the management of the assets of the power system through different procedures (Mobley 2011; Gill 2008):

Corrective (or reactive) maintenance is the traditional maintenance strategy based on the replacement or reparation of the asset after its breakdown. It is also called run-to-failure (RTF) maintenance or unplanned maintenance. Its application depends on how critical the asset is in the power system and whether the costs of the failure are affordable. It does not take into account the forecasted life before failure, and the availability of the assets depends on their reliability, so it is implemented for small items that are redundant or unlikely to fail.

In time-based (or preventive) maintenance, the inspections are predefined on a fixed period of calendar time, operating cycles, or operating hours. The substitution is performed after a predefined lifetime, allowing high availability of the asset although the cost of maintenance is also higher. This technique is commonly used in equipment subjected to wear, consumable replacements or when there is a pre-known failure pattern.

With the advancement of technology and the need of higher reliability in electric equipment, new maintenance strategies such as the condition-based maintenance (CBM) and the reliability-centered maintenance (RCM) have been introduced. These strategies are a step forward on preventive maintenance, and, nowadays, they are the leading trends in the maintenance of electrical assets. With these techniques, assets are inspected or replaced depending on their condition and failure history, effectively reducing maintenance costs, failures can be predicted, unscheduled downtime is lower, and machinery life is significantly extended (IEC 2010). Due to its importance, special attention has to be put to condition-based maintenance.

Condition-Based Maintenance

CBM aims to carry out maintenance operations only when there is evidence of abnormal behavior in a piece of equipment. Hence, a correct implementation of a CBM through an effective condition monitoring (CM) significantly reduces maintenance costs by reducing the number of unnecessary preventive maintenance interventions (Jardine et al. 2006). CM contributes to network optimization both in terms of efficiency and reliability, permitting reliable diagnostics using valuable information. In effect, CM reunites all monitored data from the operation of the asset and, with the analysis of deviations or trends in

DOI 10.1007/978-981-4560-75-7_95-1

the variables, estimates the status of the asset and predicts the need of maintenance before serious deterioration or breakdown occurs (Han and Song 2003). Hence, prognosis can predict how much time is left to failure or can predict the probability that an asset operates without faults until the next inspection, which is sometimes much useful because it permits to schedule successive interval inspections. In other situations, when a fault is detected, the evidences of the problem can be found in some of the parameters that are being monitored. In these cases, when the failure of the electrical asset is irremediable, it still is possible to use the processing of the data to make a diagnosis. Under these assumptions, a CBM program would require three main steps (Jardine et al. 2006):

1. Data acquisition or information collecting where the relevant data is recorded to determine the health of the physical asset. The data can be single values such as oil analysis data, temperature, pressure, moisture, humidity, and weather or environment data. In the analysis of tendencies, it is also necessary to acquire time series from CM such as vibration data, acoustic data, and partial discharge pulses during a specific time. Finally, multidimensional data such as image data, infrared thermographs, and X-ray can also be collected.
2. Data processing or information handling. In this step, the data acquired are manipulated, filtered, transformed, and sorted to be understood and interpreted in a better way.
3. Maintenance decision-making or decision-making. At this point, appropriate operations of maintenance are recommended by using statistical approaches, artificial intelligence, explicit mathematical models, and expert decisions to determine the remaining useful life, incorporate prognostics maintenance policies, and select a condition monitoring interval.

The strength of CBM relies on the fact that it is applied online. Off-line testing requires to put electrical assets out of service to perform the necessary tests to provide information about their health. This implies evident related costs, such as electricity production shortages, reduction of power quality supply, labor costs, temporary replacement equipments, etc., which mostly are unaffordable for an optimal system reliability (Klempner and Kerszenbaum 2004). Alternatively, online condition monitoring requires a massive deployment of sensors throughout the network and every single asset that is going to be monitored. In some cases, the expenses are justified such as in power plants with generators with 1000 MVA ratings and more, where the costs of forced outages are undoubtedly higher than the installation of additional instrumentation to do a CBM (Klempner and Kerszenbaum 2004). In other situations, the number of sensors has to be reduced or they have to monitor several assets. In any case, all information generated by the sensor and the instrumentation systems has to be gathered and processed to improve the efficiency of the power network.

Smart Grids

The term *smart grid* is used here in the sense given by (Eur 2010) and defined as an electricity network that can intelligently integrate the actions of all users connected to it in order to efficiently deliver sustainable, economic, and secure electricity supplies. This is a key aspect on the modernization of the networks pursued by all the actors involved in generation, consumption, and regularization of the electricity market. Though there still are many aspects that have to be clarified and negotiated, the roadmap to the final implementation of smart grids has been outlined in the past years by the most important standardization organisms and institutions (IEC 2010; CEN et al. 2011; NIST 2012).

In NIST (2012), the National Institute of Standards and Technology identifies eight priority areas that are considered critical in the ongoing and short-term deployment of smart grid technologies and services. One of the key functionalities in this group is the wide-area situational awareness which includes the monitoring of power-system assets to optimize their management, behavior, and performance as well as to

DOI 10.1007/978-981-4560-75-7_95-1

prevent electrical disruptions. In this regard, the International Electrotechnical Commission (IEC 2010) has also defined the asset management using condition-based monitoring as one of the key aspects in the implementation of smart grids due to the expected growing demand for electrical energy and the consequent congestion and outage problems in the network. These technical reports are addressing new requirements that represent opportunities to develop new measuring and testing procedures for all types of assets. Under these assumptions, the massive deployment of sensors is crucial to gather all useful data that can help in the maintenance of electrical power networks.

Main Causes of Failure

There are many parameters that give information about aging mechanisms that can be used in early-warning systems to prevent the breakdown of high-voltage insulations and, consequently, electrical disruptions (Gill 2008). All these parameters are associated to temperature changes, chemical by-products, vibrations, and abnormal values in electrical quantities such as voltage, current, and power. The continuing presence of any of these indicators can lead to the appearance of partial discharges (PD), or vice versa, the presence of partial discharges can reveal hidden ageing mechanisms.

Power transformers are the core of transmission and distribution networks with the key mission of raising voltage levels at generation stations to reduce losses in the transport of electricity. The insulation in large- and medium-size transformers is cellulose impregnated with oil which also acts as refrigerant. This insulation is subjected to cyclic changes of temperature in function of the demanded power. It can degrade due to chemical changes in the cellulose and to mechanical reasons for the efforts to which the coils are exposed when the transformer is operated under abnormal conditions. Nevertheless, the most important damaging ageing in these assets is water that is traded between cellulose and oil in function of the temperature. Water, as moisture, weakens sharply the dielectric constant of the cellulose–oil insulation (Kiameh 2003), reduces the partial discharge voltage inception, increases cavity sizes inside the insulation, and reduces the effective dielectric margin between high and low voltages. Hence, the monitoring of humidity in oil and the activity of partial discharges are clear indicators of insulation degradation. Other systems that can be degraded are on-load tap changers (OLTC) that are in charge of changing the voltage levels usually in the secondary of transformers (Han and Song 2003). They can induce mechanical and electrical faults and insulation problems that can be monitored with accelerometers, temperature and current sensors, and partial discharge detectors.

Generators are the main assets in the production of electricity. The insulation is manufactured mainly from solid resins and mica coatings. As in the case of transformers, the aging process is caused by thermal, mechanical, and dielectric stresses. Temperature delaminates the insulation, and wedge wearing leads to vibration and abrasion and a general reduction on the thickness of the insulation that increases the dielectric stress. All these processes can cause the appearance of partial discharges that contribute to the insulation damage rate (Kurtz et al. 1980).

Power cable systems transport and distribute electricity, so the reliability of the electrical supply relies on the conditions of these assets. They also represent a large capital investment for electrical utilities that, under some circumstances, have to cope with cables over 30 years old that may be operated, exceeding their ratings (Densley 2001). Most power cable insulation is solid, made from cross-linked polyethylene (XLPE) or ethylene propylene rubber (EPR). The ageing causes are similar to the rest of assets: temperature factors such as high and low temperatures and temperature cycling; electrical in the case of overrated voltages and currents; environmental as humidity, lubricants, corrosive gases and chemicals, and radiation; and mechanical, for example, bending, tension, compression, torsion, and vibration (Densley 2001). Basically, the most popular online monitoring method is the distributed temperature measurement with optical fibers, when available, and partial discharges. Both methods also locate the damaged cable section.

DOI 10.1007/978-981-4560-75-7_95-1

For all electrical assets, there exist common parameters that can be continuously monitored to identify the symptoms of an eventual equipment failure. One of the most popular methods to estimate the aging status of electrical machines and power lines is the measurement of partial discharges. According to the (IEC 2010), PD monitoring is essential in the evaluation of the aging conditions of power transformers, GIS (gas-insulated substations), circuit breakers, isolators and earthing switches, overhead lines, cables, and surge arresters. PD might exist even when the time to failure is still far, so it is necessary to monitor their activity continuously to detect changes in the number of PD per network cycle (50/60 Hz) or the peak of the charge released. The origin and nature of partial discharges is unveiled in the next section.

Types of PD

Partial discharges are ionization processes that take place in voids filled with gas or oil, in dielectric surfaces, and in the proximity of sharp metallic objects (Kreuger 1989; Boggs 1990). The process depends on the electric field intensity applied in the area and the nonuniformity of that field. Being a phenomenon that develops in short gaps, partial discharges may be perceived as rapid and slow-rise-time pulses that generate small electrical sparks or arcs, true pulseless glows, or pseudoglow discharges (Bartnikas 2002). Though all these three types of discharges can occur separately, pulseless glows and pseudoglow discharges are usually accompanied by pulse discharges. The immediate consequence of the discharge are current pulses that can be detected far from the ignition site. From a practical point of view, there are different types of PD attending the geometry and the characteristics of the site where the discharges occur.

Internal

Internal discharges appear in nonhomogeneous areas of a dielectric. Voids within epoxy mica dielectrics used in turn-to-turn and ground insulations in rotating machines may appear when operating at high temperatures. A similar process can occur in oil–paper insulation systems in power transformers with the formation of gas cavities inside the insulation layers (Stone 2005). In the case of polyethylene-insulated power cables (Eager and Bader 1967), the imperfections can appear in different locations during the manufacturing process at the interfaces between the shield, insulation, and conductor. Improper cable assembly can be a common cause of apparition of partial discharges at the connections and cable accessories. Finally, metallic particles erected by electrostatic forces or surface wearing or erosion in gas-insulated switchgear (Niemeyer 1995; Tanaka 1986) can create nonuniformities in the insulation and encourage the apparition of partial discharges.

Once the cavity has been formed, a slow erosion process starts, including one or some of the following phenomena (Tanaka 1986; Mason 1951): ion production inside the void followed by ion bombardment and the final detachment of insulation molecules; high temperature localized at PD sites that can cause a melting of materials and enhances the oxidation of the insulation; and chemical reactions favored by the rise in temperature that attack the insulation. The final effect is an effective increase in the cavity with the formation of channels or treeing and the consequent decrease in the insulation and the open possibility of a total breakdown. Gaseous by-products such as carbon monoxide and carbon dioxide and sometimes hydrogen and methane and liquid by-products such as moisture droplets are found inside dielectrics in the first stages of partial discharges activity (Morshuis 2005). After a continuous activity, hundreds of hours according to (Hudon et al. 1991), solid by-products such as hydrated oxalic acid ($C_2H_2O_6 \cdot 2H_2O$) in epoxy and other crystals in polyethylene are found when the former droplets crystallize under the action of partial discharges. All these chemical reactions modify the shape of the cavities, changing the characteristics of the discharge site and, in consequence, modifying the partial discharge behavior to become a nonstationary process.

The appearance of partial discharges depends on the electric stress in the cavity, the breakdown strength of the cavity (Kreuger 1989), and the residual charge left by the previous discharge activity

DOI 10.1007/978-981-4560-75-7_95-1

(Niemeyer 1995). The stress in the cavity depends on the geometry. Cavities perpendicular or parallel to the applied electric field or spherical cavities have different electric stresses. The breakdown stress depends on the cavity dimensions, the type of gas, and its pressure inside the void. In this case, the Paschen curve that defines the inception voltage in function of the thickness of the cavity and the dielectric constant of the gas plays an important role.

Once the conditions to the appearance of partial discharges have been fulfilled, there are several mechanisms at stake (Bartnikas 2002; Niemeyer 1993, 1995; Morshuis 1993). Weak discharges are characterized by very low and quasi-continuous currents practically undetectable by pulse detection methods. Low-amplitude and slow-rise-time pulses are typical of Townsend discharges, and high-amplitude and fast-rise-time pulses are characteristics of streamer-like discharges. Though most partial discharges lead to a relevant ageing of the dielectric, streamer-like discharges are one of the most harmful in the process of total failure.

All these issues can help to understand the difficulties of identifying and classifying partial discharge pulses to assess the aging status of a piece of electric equipment.

Surface

Surface discharges may also occur in inhomogeneous locations on dielectrics when there exists a tangential component of the electric field parallel to the dielectric surface (Kreuger 1989). The original Lichtenberg figures are surface discharges. The inception voltage is notably lower than in the case of internal discharges. This phenomenon is prone to happen in bushings, insulators, oil–paper interfaces, ends of cables, or when an external discharge touches the surface. The contamination of the surface plays such an important role that there are artificial pollution tests on dielectrics (IEC507 1991). Once the discharge has been triggered, the electric field is modified and the effect can be extended along large areas, provoking a flashover and a total discharge that short-circuits the insulation.

This type of discharges is especially problematic in coastal electrical infrastructures due to salt accumulation, in agricultural areas where there are pesticides and dust, and in highly contaminated urban areas. In the case of electrical machines, surface discharges can occur at the edges of electrodes or when there are small particles adhered to the surface. In oil insulations, the presence of moisture and other oil pollutants are the major cause of surface discharges.

Corona

Historically, the term corona had been used indistinctively to describe all types of discharges. The American Society for Testing and Materials (ASTM) defines corona or partial discharge as a type of localized discharge resulting from transient in an insulation system when the voltage stress exceeds a critical value (Bartnikas et al. 1979), which includes all types of discharges. The same society, when referred to the significance and use of the standard, states that *the presence of partial discharges (corona) in an apparently solid insulation is a potential indication of the existence of internal cavities. Partial discharge tests have been useful in the design and inspection of molded, laminated, and composite insulation, as well as specimens in the form of cables, capacitors, transformers, bushings, stator bars, and rotating machines* (ASTM1868 2013). However, the generalized trend is to refer corona discharges exclusively to those discharges that occur in gases or liquids (oil) under the effect of highly divergent electric fields in the proximities of sharp points (Kreuger 1989).

The inception voltage depends on the radius of the sharp electrode; for a radius of 10 μm, it is around 450 V. Hence, corona starts at relatively low voltages with the ionization of the surrounding air in the negative cycles of the ac (alternating current) voltage. If the voltage is further increased, positive corona appears due to the flow of positive ions away from the electrode.

DOI 10.1007/978-981-4560-75-7_95-1

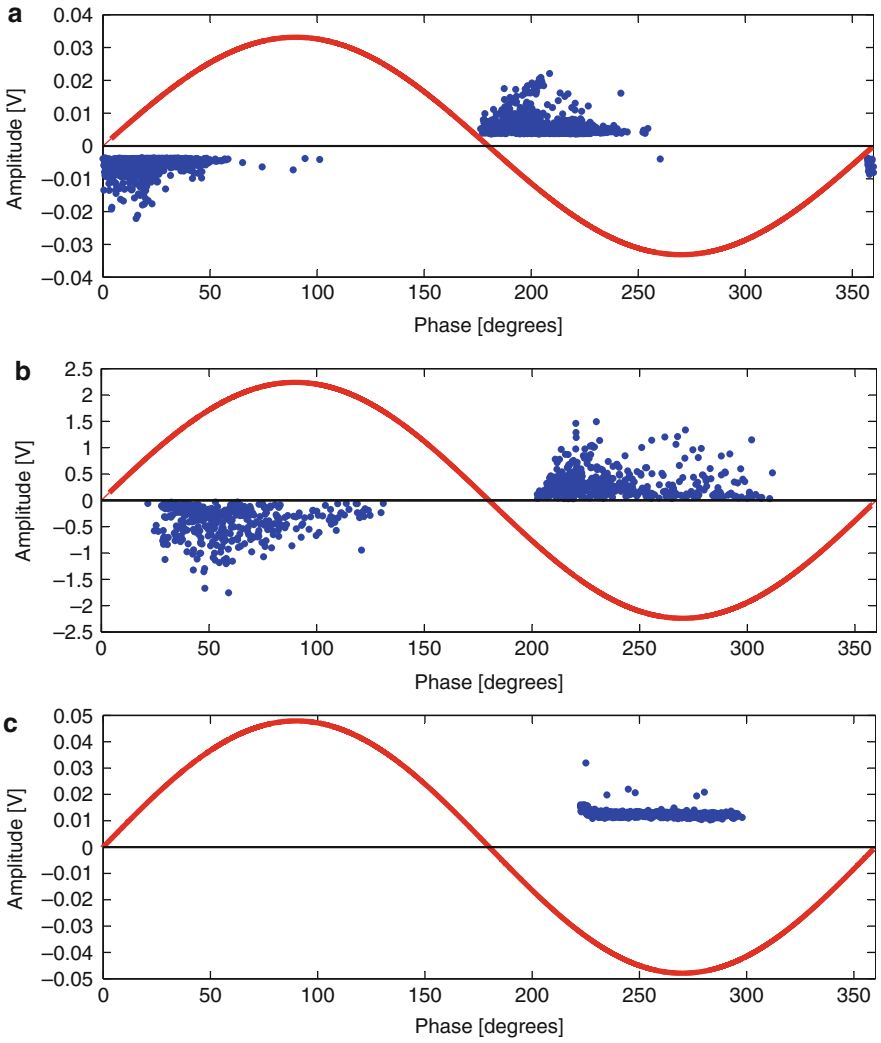


Fig. 1 PRPD patterns (**a**) internal, (**b**) surface, and (**c**) corona

The inherent problems of corona discharges are mainly related to active power losses, radio-electric disturbances, and their influence in the corrosion of conductors. They are not the cause of direct electrical degradation of insulators, but their presence can hide the activity of other harmful discharges such as internal.

Identification and Classification of Partial Discharges

Detecting partial discharges is the first stage in a condition-based monitoring system. The truly valid information should determine the nature and origin of the partial discharge in order to discriminate the insulation problem and set up an efficient maintenance. Many efforts have been directed to obtain fingerprints that are common to partial discharge from similar sources. One of the most popular and trusted methods among maintenance engineers is based on the collection of signals with the same phase referred to the grid frequency, and, hence, phase-resolved partial discharge (PRPD) patterns are analyzed. The shape and duration of the pulses are common characteristics for pulses of the same nature, so other methods are based on the extraction of the time and spectral characteristics of the signal such as time-resolved tools and time-frequency maps. Power spectral density and spectral power share in different

DOI 10.1007/978-981-4560-75-7_95-1

frequency bands have also been proved to be reliable techniques to extract characteristics of the nature of the discharge.

Phase-Resolved Partial Discharge Patterns

Phase-resolved partial discharge patterns are an electrical classification method used to identify noise sources and PD (Kreuger 1989). This method is based on the measurement of the amplitude or peak of the discharge, so it does not really require measurements in the high-frequency (HF) band; below 500 kHz is enough according to the standard IEC60270 (2000). The amplitudes of the partial discharges in pC or mV are plotted versus the phase of the grid ac voltage, Fig. 1. In this representation, all PD phases and amplitudes in a period of time are superimposed to one cycle of the ac voltage. The dot accumulation defines different patterns depending on the type of partial discharge, and by identifying this pattern, it is possible to find the origin of the fault. Internal discharge clusters span from zero crossings, 0° or 180°, of the sinusoidal waveform to 50–70° in the positive cycle and 230–250° in the negative cycle (plot a). Surface discharges are accumulated around the maximum and minimum of the ac signal (plot b). And, patterns for corona discharges only have pulses in one semi-cycle. Figure 1 shows the PRPD patterns for internal, surface, and corona PD measured with a monopole antenna synchronized with the grid ac voltage. Noise is not correlated with the phase of the sinusoidal signal, so in a PRPD plot, it is represented as a dispersed band of points at all phases.

Time-Frequency Maps

In this technique, the problem of PD classification is addressed, considering that the pulses from different sources have different waveforms and durations. Time-frequency (TF) maps are a classification tool commonly used in the field of signal theory and involve the separation of similar-shaped wave pulses into groups (Contin et al. 2002; Cavallini et al. 2003). The equivalent time duration and the equivalent bandwidth are calculated for all pulses. Background noise will have similar durations and spectra, so the signals will be grouped in the same cluster. Interferences are expected to have different parameters than background noise and will be assigned to other zones of the TF map. Partial discharges have also different waveforms and will occupy other areas. The process begins by normalizing all signals with time, $s(t)$:

$$\tilde{s}(t) = \frac{s(t)}{\sqrt{\int_0^T s(t)^2 dt}} \tag{1}$$

Subsequently, the standard deviation of the normalized signal is calculated in the time domain (Eq. 2) and in the frequency domain (Eq. 3).

$$\sigma_T = \sqrt{\int_0^T \left(t - t_g\right)^2 \tilde{s}(t)^2 dt} \tag{2}$$

$$\sigma_F = \sqrt{\int_0^\infty f^2 \left|\tilde{S}(f)\right|^2 df} \tag{3}$$

where f is the frequency, $\tilde{S}(f)$ is the Fourier transform of the pulse $\tilde{s}(t)$, and t_g is the time gravity center for the normalized signal, defined by:

DOI 10.1007/978-981-4560-75-7_95-1

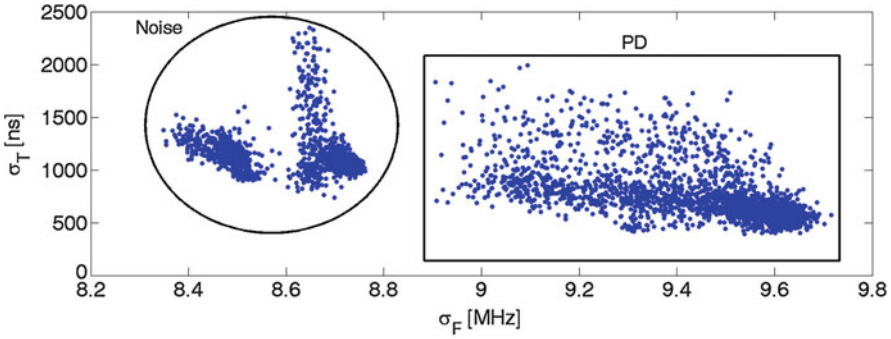


Fig. 2 TF map example for internal PD and electromagnetic noise

$$t_g = \int_0^T t\tilde{s}(t)^2 dt \tag{4}$$

In Fig. 2, a TF map for internal PD and electromagnetic noise measured with a monopole antenna is presented. In this figure, both main clusters of internal and noise sources are distinguished.

Spectral Methods

The characteristics in the pulse waveforms can also be analyzed through the study of their Fourier transform or their power spectral density. The spectrum can be divided in relevant frequencies where signals from different sources show differences. This information can be found with a support vector machine (SVM) by comparing two types of partial discharges or noise and partial discharges using training signals. The SVM selects a set of significant spectra capable of defining the two classes independently, so new pulses will be assigned to one of the classes depending on their similarity.

This method can be simplified when the relevant frequencies are congregated in a continuous band. These bands are selected as those intervals where there is higher dispersion in the accumulated power. The spectrum is previously normalized, and the power in every selected band is calculated as a percentage of the total. There can be as many frequency intervals as necessary; however, for the sake of clearness, only two are selected, an interval in a low-frequency band, PRL, and another in a high-frequency band, PRH. The percentages or power ratios are calculated with Eqs. 5 and 6, respectively, and plotted in a two-dimension map.

$$\%PRL = \frac{\sum_{f_{1L}}^{f_{2L}} \left|\tilde{S}(f)\right|^2}{\sum_0^{f_t} \left|\tilde{S}(f)\right|^2} \times 100 \tag{5}$$

$$\%PRH = \frac{\sum_{f_{1H}}^{f_{2H}} \left|\tilde{S}(f)\right|^2}{\sum_0^{f_t} \left|\tilde{S}(f)\right|^2} \times 100 \tag{6}$$

where $\left|\tilde{S}(f)\right|$ is the fast Fourier transform (FFT) magnitude calculated for the partial discharge pulse, $s(t)$; $[f_{1L}, f_{2L}]$ is the lower frequency band, $[f_{1H}, f_{2H}]$ is the higher frequency band, and f_t, is the highest frequency value to be analyzed.

In some cases, there can be sources with similar PRL and PRH that would give overlapped clusters in the map. Adding a third axis with the duration of the signal (t_{eff}) will simplify the separation of the pulses

DOI 10.1007/978-981-4560-75-7_95-1

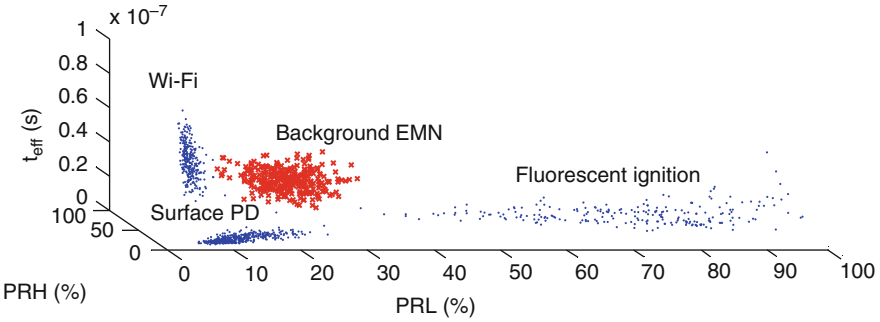


Fig. 3 Power ratio-time map for several electromagnetic sources and noise (*EMN*) measured with a Vivaldi antenna (Albarracin et al. 2015)

(Albarracin et al. 2015). Figure 3 shows the separation of several electromagnetic sources measured with an antenna.

Statistical Methods

Other methods select statistical parameters that define a certain type of partial discharge from the pulse count distribution $H_n(\phi)$ and from the mean pulse height distribution $H_{qn}(\phi)$ at a certain phase of the ac signal ϕ (Contin et al. 2000). The representative parameters for these two distributions are the third moment or skewness that describes the asymmetry compared to a normal distribution, Eq. 7, and the fourth moment or kurtosis that describes the sharpness, Eq. 8. These quantities are different in function of the type of partial discharge.

$$S_k = \sum_{i=1}^{N} \frac{(s_i - s_m)^2}{N} \tag{7}$$

$$K_u = \frac{1}{\sigma^4} \sum_{i=1}^{N} \frac{(s_i - s_m)^4}{N} \tag{8}$$

where s_i is the value of the pulse at sample i, s_m is the pulse height average, N is the number of samples, and σ is the pulse height standard deviation.

The pulse charge height distribution $F(q)$ follows a Weibull distribution, Eq. 9. Finding the scale σ and shape β parameters of the adjusted Weibull distribution will also give information about the type of partial discharge.

$$F(q) = 1 - e^{-\left(\frac{q}{\alpha}\right)^{\beta}} \tag{9}$$

Measuring Partial Discharges

Sensors

As stated in section "Main Causes of Failure," the ignition of partial discharges provokes changes in physical variables such as voltage, current, pressure, light, and electromagnetic radiation. All sensors applied to detect partial discharges have to be designed to measure any of these variables.

The standard IEC60270 recommends the use of a coupling capacitor, C_k, to provide a low impedance path for discharges in the device under test (IEC60270 2000). The capacitor is connected in parallel with

DOI 10.1007/978-981-4560-75-7_95-1

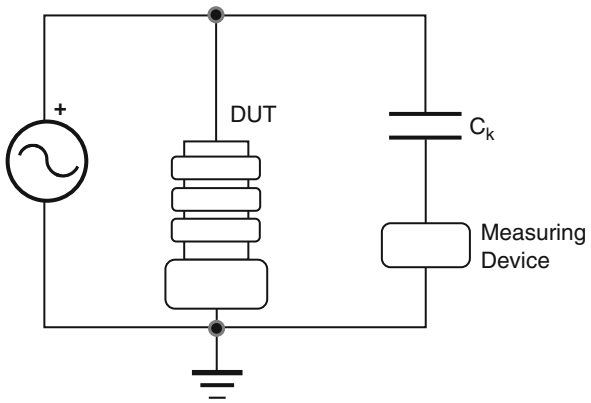


Fig. 4 Indirect measuring circuit as proposed in (IEC60270 2000). The coupling capacitor C_K is in series with a measuring device that can be an impedance or an inductive sensor

the test object directly to ground or through a measuring impedance or detector, Fig. 4. Capacitances are in the range of 80–1000 pF (Stone 2005), with a bandwidth limited below 10 MHz for large capacitances and that can be increased by reducing the capacitance value (Hauschild and Lemke 2014). Other capacitive couplers detect the electric field created by partial discharges in shielded power cables using a pair of electrodes separated by a known distance (Katsuta et al. 1992).

Inductive sensors are based on Faraday's law and measure the induced voltage in a coil through the magnetic field created by the partial discharge pulse. High-frequency current transformers (HFCT) are extensively used in measuring circuits described in the standard in series with measuring impedance, Fig. 4. They have split ferrite cores which improve the sensitivity but limit the bandwidth usually in the HF-VHF range. If current transformers are to be used directly on high-voltage leads, it is preferable to choose Rogowski coils because they have air cores and do not saturate with line currents at grid frequency (Argüeso et al. 2005). Inductive couplers which wound around power cables with helicoidal shield are also used to measure partial discharges (Dubickas and Edin 2004) and, most importantly, locate them along the cable.

Acoustic sensors can detect the pressure wave emitted by the discharge in a wide band of frequencies, including the ultrasonic range (beyond 20 kHz). With corona and some surface discharges, the pulse is clearly audible, so simple microphones could be used (Bartnikas et al. 1979). In the case of internal discharges, it is preferable to use piezoelectric transducers or narrowband microphones with central frequencies in the range from 40 to 150 kHz. Directional microphones are also used to identify the location of the source of surface or corona pulses. The pressure wave can also be detected with fiber optic interferometric sensors which present advantages when compared with electric sensors such as immunity to electromagnetic interferences, wide bandwidth, and chemical inert when installed in oil transformers and nonconducting (Zargari and Blackburn 1996). On the other hand, interferometer installation is expensive, difficult, and intrusive, and they are only valid in equipment with liquid insulation such as oil.

Light sensors have been used in studies of arcing in transformers to provide information about the electron density and temperature (Bårmann et al. 1996). They have to be focused to the partial discharge site so they are not very popular because they are difficult to integrate in real environments where partial discharges can occur anywhere and can even happen inside the insulation, so no light radiation would leak out. On the contrary, there are commercial sensors using light detectors in the ultraviolet range to measure corona in overhead lines.

The next section explains why electromagnetic (EM) sensors, dissected in section "Applications of Antennas in the Detection of Partial Discharges," are also suitable for the measurement of partial discharges.

DOI 10.1007/978-981-4560-75-7_95-1

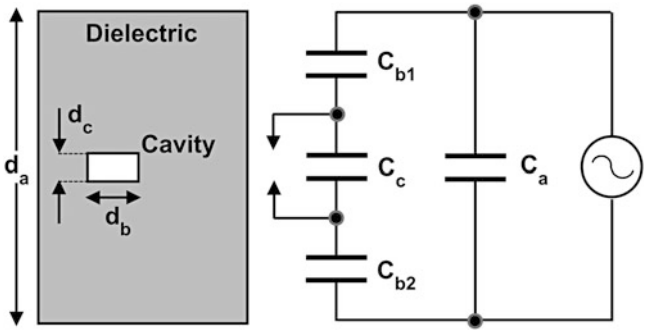


Fig. 5 Capacitive model in internal cavities

Partial Discharge Cavity Modeling

As explained in section "Types of PD," partial discharges involve the displacement of ions, both positive and negative, inside a cavity, along a surface, or in the proximity of areas with high divergent electric fields. These avalanches of particles occur in lapses as short as hundreds of picoseconds or nanoseconds which, in turn, create current pulses with very fast rise times. In general, it is not possible to measure this current at the ends of the PD site, so the actual shape of the pulses is unknown. The closest measurement can be done at the terminals of a test object; however, even under these circumstances, the signal would be deformed due to the attenuation and dispersion suffered inside the insulation from the PD site to the terminals (Hauschild and Lemke 2014). Moreover, if the test object is connected to a measuring system, the true pulse shape is further modified and the signal obtained will be very different from the original. It has been deeply investigated what is the relation between the original pulse and the signal measured at the ends of the test object. Assuming that the cavity has a regular geometric shape such as spherical, cylindrical, or elliptical, the study is greatly simplified and it is even possible to calculate the inception voltages.

Capacitive Model

The equivalent a-b-c circuit shown in Fig. 5 was introduced in the middle of the twentieth century to calculate the ratio between the actual PD charge and the measured charge in the case of internal discharges. C_a is the capacitance of the homogeneous dielectric; C_b is the accumulated capacitance of the dielectric between the electrodes and the cavity, though some authors prefer the use of C_{b1} and C_{b2} representing the dielectric over and below the cavity, respectively (Boggs 1990); and C_c is the capacity of the cavity which is shunted by a spark gap representing the eventual short circuit of the void. The ratio between the charge q_c involved in the discharge and the measured charge q_a is given by:

$$\frac{q_a}{q_c} \approx \epsilon_r \frac{d_c}{d_a} \tag{10}$$

being ϵ_r the relative permittivity of the dielectric and d_a and d_c the thickness of the homogeneous dielectric and the cavity, respectively. Recently, Eq. 10 has been modified under the assumption that during a discharge, the effective geometry of the equivalent circuit changes (Hauschild and Lemke 2014). Then, the ratio is governed by:

$$\frac{q_a}{q_c} \approx \epsilon_r \frac{d_c}{d_a} \left(\frac{d_b}{d_d}\right)^2 \tag{11}$$

DOI 10.1007/978-981-4560-75-7_95-1

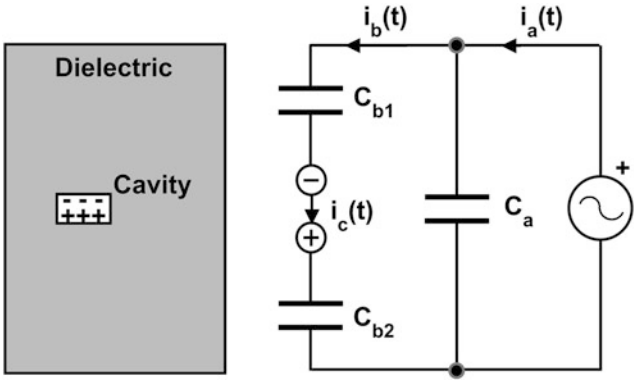


Fig. 6 Dipole model in internal cavities

with d_b the diameter of the equivalent cylinder containing the cavity and the sound dielectric and d_d the diameter of the discharge channel inside the cavity. If the measuring circuit is previously calibrated, injecting a known charge, it is possible to calculate the effective charge displacement during a discharge using inductive sensors connected to the recommended measuring circuits in IEC60270 (2000).

Dipole Model

Another widely accepted approach to the calculation of the charge involved in the ignition of partial discharges, which is alternative to the a-b-c model, is the use of a dipole model (Pedersen et al. 1991) (Fig. 6). The defenders of this model claim that the operation of an equivalent circuit may not correspond with the real physical processes that take place inside the dielectric (McAllister and Crichton 1998). In this case, the cavity turns into a dipole due to the polarization process under the application of an external electric field. The current $i_c(t)$ flowing inside the cavity when the dipole collapses continues as a displacement current $i_b(t)$ through the homogeneous dielectric that can be readily identified as capacitances C_{b1} and C_{b2}. In this case, the measurable charge q_a at the ends of the test object is equal to charge q_c moving inside the cavity (Hauschild and Lemke 2014).

Partial Discharge Pulse Modeling

Traditionally, the idealized partial discharge pulse has been simplified as a Gaussian pulse (Boggs and Stone 1982), but there are other valid approximations such as Wanninger's equation and the double exponential pulses (Reid et al. 2006).

Gaussian Pulses

A Gaussian pulse is shaped as a Gaussian function centered in $t = 0$, with a peak of I_0 amperes and a width at its half amplitude of T_h. From its definition, the proposed current pulse for partial discharges is:

$$i_c(t - t_d) = I_0 e^{-\left(\frac{t-t_d}{t_0}\right)^2}, \quad t_0 = \frac{T_h}{2\sqrt{\mathrm{Ln}2}} \tag{12}$$

where t_d is the time delay that shifts the pulse to the right, placing the peak at $t = t_d$, and t_0 is related to the pulse width and, hence, the standard deviation as $t_0 = \sigma\sqrt{2}$. Considering that the pulse is extinguished at 3σ, the time delay can be chosen as $t_d = 3t_0/\sqrt{2}$. In Fig. 7, a delayed and normalized ($I_0 = 1$ mA) Gaussian pulse is plotted. For this representation, $T_h = t_2 - t_1$ is assumed to be 1 ns, with $t_1 = t_d - t_0\sqrt{\mathrm{Ln}2}$ and $t_2 = t_d + t_0\sqrt{\mathrm{Ln}2}$.

DOI 10.1007/978-981-4560-75-7_95-1

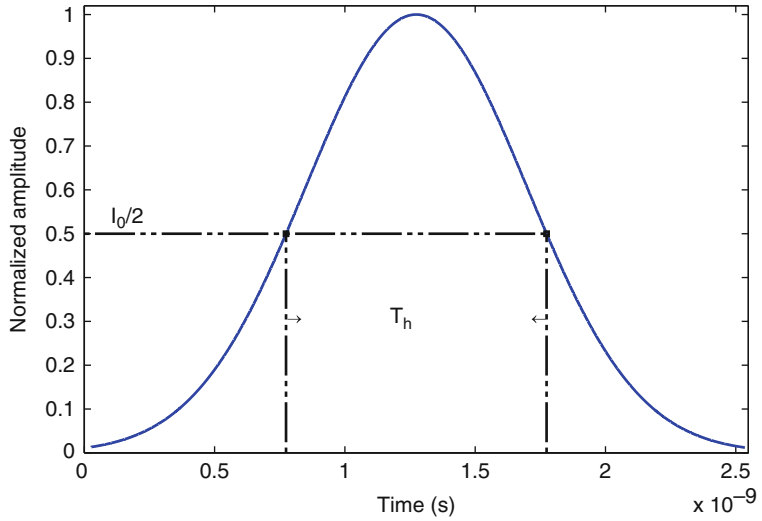


Fig. 7 Delayed and normalized Gaussian pulse with $I_0 = 1$ mA, $T_h = 1$ ns, and $t_d = 3\sigma$

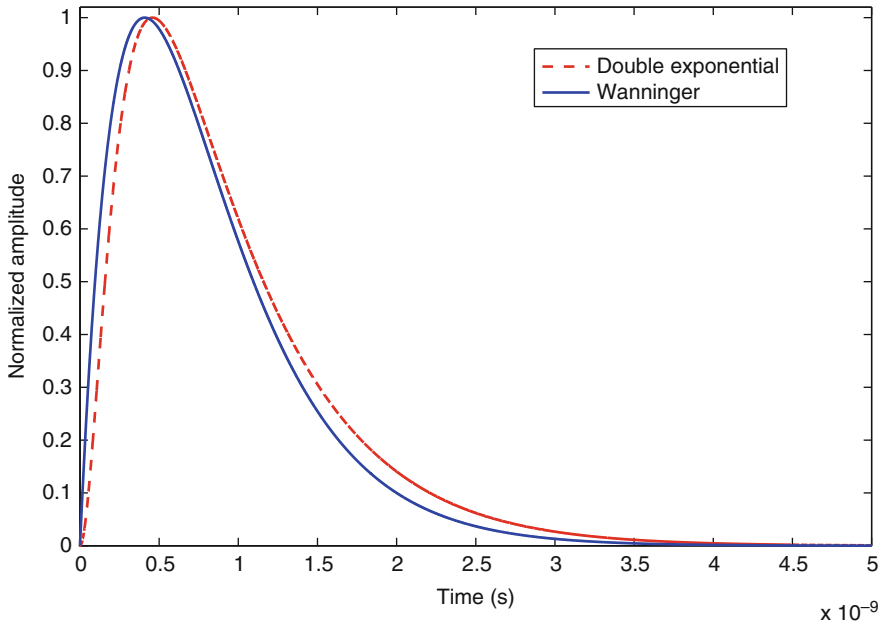


Fig. 8 Normalized Wanninger and double exponential pulses with 1 ns width at its half amplitude

Wanninger's Equation

The Wanninger's equation was used to model idealized discharge currents in (Wanninger 1997) using exponential functions in the time domain as follows:

$$i_c(t) = \frac{I_0}{\tau_1} t e^{(1-t/\tau_1)} \tag{13}$$

where I_0 is again the current amplitude in amperes and τ_1 is the rise time. From this equation, it is possible to represent discharge currents of free moving particles of single, double, and multiple pulses. This is made by combining several equations and varying τ_1 and the ratio of the amplitudes. In Fig. 8, the solid

DOI 10.1007/978-981-4560-75-7_95-1

line is an idealized pulse using the Wanninger's equation with a width at its half amplitude of 1 ns and $I_0 = 1$ mA.

Double Exponential Pulses

The double exponential was used in Ari and Blumer (1987) to represent transient electromagnetic fields due to switching operations and can also be applied to model partial discharges. The time function for the discharge current defined by a double exponential is:

$$i_c(t) = I_1\left[(1 + t/\tau_1)e^{-t/\tau_1} - (1 + t/\tau_2)e^{-t/\tau_2}\right] \tag{14}$$

where I_1 corresponds to the current amplitude in amperes. Noteworthy is that the maximum value of I_1 is 0.6517, while I_0 is 1 for both Gaussian and Wanninger pulses. For this reason, the double exponential function is divided into its maximum value to normalize the representation in Fig. 8. Parameters τ_1 and τ_2 are both time constants in seconds that define the pulse shape. The time at the peak value is calculated as:

$$t_{I_1} = 2\frac{\tau_1\ \tau_2}{\tau_2 - \tau_1}\mathrm{Ln}(\tau_2/\tau_1) \tag{15}$$

so, the condition $\tau_2 > \tau_1$ has to be held to obtain positive values.

Partial Discharge Radiation

Considering that the dipole created in the void explained in section "Partial Discharge Cavity Modelling" is an Hertzian dipole, it is possible to derive the equations of the electrical and magnetic fields radiated by the partial discharge. Let us assume that the dipole is along the z-axis and that the current $i_c(t) = \mathrm{dq_c/dt}$ or $\mathbf{I_c} = j\omega\mathbf{q_c}$ in the frequency domain for sinusoidal time-varying signals. The magnetic vector potential can be calculated as:

$$\mathbf{A_z} = \frac{\mu}{4\pi}\int_c \frac{\mathbf{I_c}}{r}e^{-j\beta r}dz \tag{16}$$

where μ is the permeability of the dielectric, r is the distance from the center of the dipole to a generic point P, and $\beta = \omega\sqrt{\mu\epsilon}$ is the wave number of the medium (Guru and Hiziroglu 2004). In small cavities, the current can be considered the same along the dipole, so Eq. 16 can be simplified as:

$$\mathbf{A_z} = \frac{\mu}{4\pi r}\mathbf{I_c}d_c e^{-j\beta r} \tag{17}$$

Changing into spherical coordinates to simplify the calculations, $\vec{\mathbf{A}}$ yields:

$$\vec{\mathbf{A}} = \frac{\mu}{4\pi r}\mathbf{I_c}d_c e^{-j\beta r}\left(\cos\theta\ \vec{\mathbf{a}}_r - \sin\ \vec{\mathbf{a}}_\theta\right) \tag{18}$$

with $\vec{\mathbf{a}}_r$ and $\vec{\mathbf{a}}_\theta$ defined as unitary vectors. Then, the magnetic field component can be calculated using:

$$\vec{\mathbf{H}}_\phi = \frac{1}{\mu}\left(\nabla\times\ \vec{\mathbf{A}}\right) = \frac{j\beta}{4\pi r}\mathbf{I_c}d_c\left(1 + \frac{1}{j\beta r}\right)\sin\theta e^{-j\beta r}\ \vec{\mathbf{a}}_\phi \tag{19}$$

DOI 10.1007/978-981-4560-75-7_95-1

with $\vec{\mathbf{H}}_r = 0$ and $\vec{\mathbf{H}}_\theta = 0$. The electric field component is obtained from the Maxwell–Ampère equation, assuming that the current density due to free charges is negligible outside the cavity and far from the discharge site:

$$\vec{\mathbf{E}} = \frac{1}{j\omega\epsilon}\left(\nabla\times\ \vec{\mathbf{H}}\right) \tag{20}$$

then,

$$\begin{aligned}
\vec{\mathbf{E}}_r &= \frac{\eta}{2\pi r^2}\mathbf{I_c}d_c\left(1+\frac{1}{j\beta r}\right)\cos\theta e^{-j\beta r}\ \ \vec{\mathbf{a}}_r \\
\vec{\mathbf{E}}_\theta &= \frac{j\beta\eta}{4\pi r^2}\mathbf{I_c}d_c\left(1+\frac{1}{j\beta r}-\frac{1}{\beta^2 r^2}\right)\sin\theta e^{-j\beta r}\ \ \vec{\mathbf{a}}_\theta \\
\vec{\mathbf{E}}_\phi &= 0
\end{aligned} \tag{21}$$

Equations 19 and 21 can be simplified considering far-field conditions at distances r, so $r \gg \lambda$ with $\lambda = 2\pi/\beta$ the wavelength of the electromagnetic field. Then:

$$\begin{aligned}
\vec{\mathbf{H}}_\phi &= \frac{j\beta}{4\pi r}\mathbf{I_c}d_c\sin\theta e^{-j\beta r}\,\vec{\mathbf{a}}_\phi \\
\vec{\mathbf{E}}_\theta &= \frac{j\beta\eta}{4\pi r}\mathbf{I_c}d_c\sin\theta e^{-j\beta r}\,\vec{\mathbf{a}}_\theta
\end{aligned} \tag{22}$$

eliminating the terms in $1/r^2$ and $1/r^3$. Equation 22 is, by definition, a transverse electromagnetic (TEM) wave propagating in the radial direction, so the far fields represent a spherical wave (Guru and Hiziroglu 2004). Note that $\vec{\mathbf{E}}_\theta/\vec{\mathbf{H}}_\phi = \eta$ with $\eta = \sqrt{\mu/\epsilon}$, the intrinsic impedance of the dielectric. Under these assumptions, partial discharges radiate TEM waves that can be captured with antennas in far-field regions. The next step should be to determine what ranges of frequencies have the radiation based on the type of pulse and its time characteristics such as rise time and width.

Electromagnetic Radiation of a Gaussian Pulse

The radiation of a dipole has just been calculated for sinusoidal signals, but it is easy to obtain the electric and magnetic fields at a point P when the dipole is excited with a Gaussian pulse. The magnetic vector potential in spherical coordinates has the following equation in the time domain when the current pulse is Gaussian:

$$\vec{\mathbf{A}} = \frac{\mu I_0 l}{4\pi r}e^{-\left(\frac{t-r/v}{t_0}\right)^2}\left(\cos\theta\ \ \vec{\mathbf{a}}_r - \sin\theta\ \ \vec{\mathbf{a}}_\theta\right) \tag{23}$$

where $v = 1/\sqrt{\mu\epsilon}$ is the speed of propagation in the media. Calculating the magnetic field using the first term in Eq. 19, the result is:

$$\vec{\mathbf{H}}_\phi = \frac{1}{\mu}\left(\nabla\times\ \vec{\mathbf{A}}\right) = \frac{I_0 l e^{-\left(\frac{t-r/v}{t_0}\right)^2}\sin\theta}{4\pi r^2}\left(1-\frac{2r(t-r/v)}{vt_0^2}\right)\vec{\mathbf{a}}_0 \tag{24}$$

DOI 10.1007/978-981-4560-75-7_95-1

Then, using Eq. 20 in the time domain:

$$\begin{aligned}
\vec{\mathbf{E}}_r &= \frac{\eta I_0 l e^{-\left(\frac{t-r/v}{t_0}\right)^2} \cos\theta}{2\pi r^2}\left(1+\frac{t_0\sqrt{\pi}}{2r}\right)\vec{\mathbf{a}}_r \\
\vec{\mathbf{E}}_\theta &= \frac{\eta I_0 l \sin\theta}{4\pi r^2}\left(e^{-\left(\frac{t-r/v}{t_0}\right)^2} - \frac{2r(t-r/v)e^{-\left(\frac{t-r/v}{t_0}\right)^2}}{vt_0^2} + \frac{v\sqrt{\pi}}{2r}\int_0^t e^{-\left(\frac{t-r/v}{t_0}\right)^2} dt\right)\vec{\mathbf{a}}_\theta \\
\vec{\mathbf{E}}_\phi &= 0
\end{aligned} \tag{25}$$

If the measuring point is far from the partial discharge site, the term $1/r$ dominates and the equations can be reduced to:

$$\begin{aligned}
\vec{\mathbf{H}}_\phi &= -\frac{I_0 l e^{-\left(\frac{t-r/v}{t_0}\right)^2} \sin\theta(t-r/v)}{2\pi r v t_0^2}\vec{\mathbf{a}}_\phi \\
\vec{\mathbf{E}}_\theta &= -\frac{\eta I_0 l e^{-\left(\frac{t-r/v}{t_0}\right)^2} \sin\theta(t-r/v)}{2\pi r v t_0^2}\vec{\mathbf{a}}_\theta
\end{aligned} \tag{26}$$

These equations are inversely proportional to t_0^2 which is related to the pulse width. So the narrower the pulse is, the larger the radiation of the partial discharge would be.

Considering that,

$$\frac{\partial i_c(t-r/v)}{\partial t} = -\frac{2I_0(t-r/v)}{t_0^2} e^{-\left(\frac{t-r/v}{t_0}\right)^2} \tag{27}$$

The electric and magnetic fields can be rewritten using the derivative of the delayed Gaussian pulse with $t_d = r/v$. In the case of the electric field is:

$$\vec{\mathbf{E}}_\theta = \frac{\eta I_0 l \sin\theta}{4\pi r v}\ \frac{\partial i_c(t-r/v)}{\partial t}\vec{\mathbf{a}}_\theta \tag{28}$$

Frequency Emission of a Gaussian Pulse

The Fourier transform of the electromagnetic field created by a Gaussian pulse gives the frequency band at which the discharge emits. The following calculations are done for the electric field, but the results are analogous for the magnetic field. The Fourier transform, $\mathcal{F}(\cdot)$, of a derivative function is:

$$\mathcal{F}\left(\frac{\partial i_c(t-r/v)}{\partial t}\right) = \int_{-\infty}^{\infty} \frac{\partial I_c(t-r/v)}{\partial t} e^{-j\omega t} dt \tag{29}$$

From Eq. 27, the derivative depends on a squared exponential function and a time delay. The Fourier transform of a squared exponential function is:

DOI 10.1007/978-981-4560-75-7_95-1

$$\mathcal{F}\left(e^{-t^2}\right) = \sqrt{\pi t_0^2} e^{\frac{-\omega^2 t_0^2}{4}} \tag{30}$$

Besides, the differentiation and time shift properties of the Fourier transform, Eqs. 31 and 32, respectively, must also be applied to solve Eq. 29:

$$\mathcal{F}\left(\frac{d^n i_c(t)}{dt^n}\right) = (j\omega)^n I_c(\omega) \tag{31}$$

$$\mathcal{F}(t - t_d) = \mathcal{F}(\omega) e^{-j\omega t_d} \tag{32}$$

Then, the Fourier transform of the amplitude of the electric field radiated by a Gaussian pulse can be written as follows:

$$\mathcal{F}(E_\theta) = \omega\sqrt{\pi} t_0 I_0 e^{-\frac{\omega^2 t_0^2}{4}} e^{j\left(\frac{\pi}{2} - \omega\frac{r}{v}\right)} \tag{33}$$

Considering only the module of the Fourier transform and substituting $\omega = 2\pi f$, the equation that represents the spectrum amplitude for a Gaussian pulse is:

$$|\mathcal{F}(E_\theta)| = 2\pi f\sqrt{\pi} t_0 I_0 e^{-(\pi f t_0)^2} \tag{34}$$

For a Gaussian pulse defined in terms of its width T_h as in Eq. 12, the Fourier transform is:

$$|\mathcal{F}(E_\theta)| = \pi f\sqrt{\pi}\frac{T_h}{\sqrt{\mathrm{Ln}2}} I_0 e^{-\left(\pi f\frac{T_h}{2\sqrt{\mathrm{Ln}2}}\right)^2} \tag{35}$$

The influence of the width of the pulses in the spectrum is represented in Fig. 9 and has been calculated with varying T_h in Eq. 35. This Figure and Table 1 help to determine the frequency band where partial discharges emit electromagnetic energy in function of their width. The band is $f_{\max} - f_{\min}$ where $f_{\max}$ and $f_{\min}$ are the cutoff frequencies for a drop of 3 dB in the amplitude of the electric field.

In (Shibuya et al. 2010), a compilation of pulse widths for partial discharges in small gaps and conductive particles is reported. The band in which the pulses emit electromagnetic energy from the data in (Shibuya et al. 2010) are calculated and represented in Table 2, sorted by decreasing pulse widths. This table summarizes different phenomena in a wide range of dielectric types. For short duration idealized Gaussian pulses approximately less than 2.5 ns, partial discharges emit energy in UHF. For larger time widths, antennas in the VHF range should be recommended. In any case, the casuistic of partial discharges is so vast that generalizing is sometimes a futile task. The important fact is that PD can be measured with antennas due to their short duration and fast rising front wave.

Applications of Antennas in the Detection of Partial Discharges

Measuring Partial Discharges in Electric Equipment with Antennas

Radiofrequency measurements in the HF, VHF, and UHF bands are part of the unconventional methods used for partial discharge detection (IEC62478 yet). As mentioned in section "Types of PD," partial discharges occur on a limited spot which comes with the property of extremely short breakdown time.

DOI 10.1007/978-981-4560-75-7_95-1

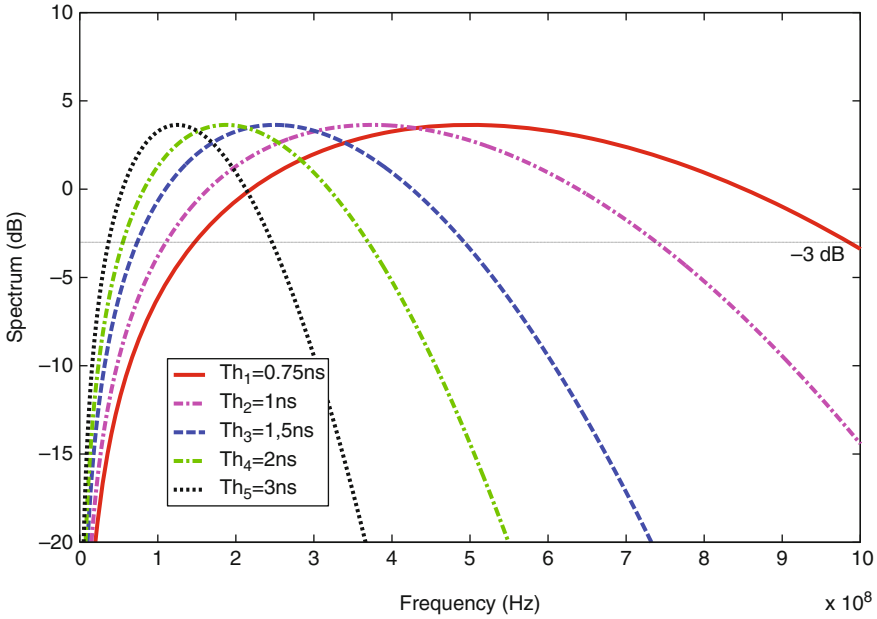


Fig. 9 Influence of varying T_h for the spectrum of Gaussian pulses (Albarracín-Sánchez 2014)

Table 1 Frequency band for PD in function of T_h after (Albarracín-Sánchez 2014)

Th (ns)	f^{min} (MHz)	f^{max} (MHz)	Band (MHz)
0.75	147	985	838
1	111	739	628
1.5	74	492	418
2	55	369	314
3	37	246	209

Table 2 Partial discharge currents for several gap conditions: (a) void discharge, (b) point-plane gap, (c) discharge involving conducting particles, and (d) short-gap discharge and electrostatic discharge. The acronyms stand for polyethylene (*PE*), polytetrafluoroethylene (*PTFE*), polymethylmethacrylate (*PMMA*), and sulfur hexafluoride (SF_6)

PD type	Gap conditions	T_h (ns)	f^{min} (MHz)	f^{max} (MHz)	Band (MHz)
(d)	Electrodes in air	0.047	2352	15720	13368
(c)	2 mm diameter Al particle in 0.38 MPa SF6	0.07	1579	10560	8981
(b)	0.1 MPa SF6	0.27	409	2737	2328
(a)	0.05 mm in PTFE	0.37	299	1997	1698
(b)	0.1 MPa SF6	0.49	226	1508	1282
(b)	0.4 MPa SF6	0.8	138	924	786
(a)	0.1 mm in PE	0.87	127	849	722
(a)	2 mm in PMMA	1.4	79	528	449
(d)	0.66 mm in air	2	55	369	314
(a)	In epoxy resin	2.4	46	308	262
(a)	0.5 mm in PTFE	2.4	46	308	262
(a)	0.64 mm in PE	5.3	21	139	118
(a)	0.64 mm in PE	10	11	74	63

DOI 10.1007/978-981-4560-75-7_95-1

Thus, the electromagnetic radiated pulse contains the combination of different frequency components reaching a few GHz falling within the range of VHF and UHF bands, in the low microwave band.

The detection of PD with EM sensors can be used in any type of insulating material and for multiple electrical assets such as gas-insulated substations, air-insulated substations (AIS), power transformers, insulators in aerial distribution cables, joints and splices of insulated cables, and rotating machinery. The main advantage of detecting partial discharge pulses in RF is the possibility of monitoring these assets online. Most diagnostic techniques such as tangent delta and insulation resistance measurements, frequency response analysis (FRA), surge tests, and hipot tests are carried out off-line or require periodic interventions of operators to take samples such as in dissolved gas analysis (DGA), degree of polymerization, and moisture analysis in power transformers.

EM sensors can detect partial discharge pulses, localize the partial discharge site, and, to some extent, classify the type of partial discharge online. In the detection and localization of PD, there is never a direct connection to the electric equipment. However, to analyze the type of PD, it is necessary to have the reference of the grid frequency to plot conventional PRPD patterns; see section "Phase-Resolved Partial Discharge Patterns." In this case, a voltage divider has to be installed and connected to the acquisition system to record both PD pulses and time (or phase) simultaneously. In this sense, one of the most important challenges when using antennas in the diagnostic of insulations is the difficulty of relating the RF emissions to the severity of the PD (Reid et al. 2006). Currently, it only can be said that there is a direct relation between the energy emitted by the partial discharge and the rate of change of current in the pulse as demonstrated in Eq. 28. Moreover, the sample rate to obtain information from the signals in the time and frequency domains is sometimes prohibitively high, incurring in great expenses, so it is applied only when it is strictly necessary.

On the other side, RF interferences from power lines can propagate into transformers and other electrical assets reducing the PD measurement accuracy when traditional detection, between tens of KHz and several MHz, is used. The UHF PD detection thus takes advantages of strong anti-interference ability over the approach defined by (IEC60270 2000). However, other sources of interference such as FM radio, TV broadcasting, mobile telecommunications, Wi-Fi, etc., could difficult the PD detection as well.

Ultimately, the measurement of PD in the RF range offers new insight into the diagnostic of electrical machines with the possibility of measuring under conditions where other methods may fail or are impossible to apply.

Nowadays, the implementation of UHF sensors in electric equipments is still not widespread. Most of the testing solutions in this field are using adapted spectrum analyzers with a specialized software to process and analyze the data in the time and frequency domains, and they are restricted to conventional air-insulated substations. There are several manufacturers that give partial or global solutions for PD monitoring in UHF in GIS using internal or external sensors. These measuring devices execute automated continuous online monitoring and periodic measurements of PD.

Gas- and Air-Insulated Substations

Partial discharge measurement with antennas took its first steps in the 1960s (Helstrom 1961) and became more relevant at the end of the 1980s, with the detection of PD created by free metallic particles in gas-insulated substations (Hampton and Meats 1988). This reference can be considered as one of the pioneer studies in the measurement of partial discharges in the UHF range. It is also one of the first reports of pulse propagation and location in electric equipment with UHF couplers. The importance of gas-insulated substations relies on the use of sulfur hexafluoride (SF_6) gas as electrical insulation for their high-voltage maneuvering devices and transformers. The SF_6 dielectric strength is about 2.5 times larger than that of air, so the size of a conventional air-insulated substation can be reduced in a 70 % by enclosing high-voltage equipment inside structures with this gas. Its construction requires a sealed

DOI 10.1007/978-981-4560-75-7_95-1

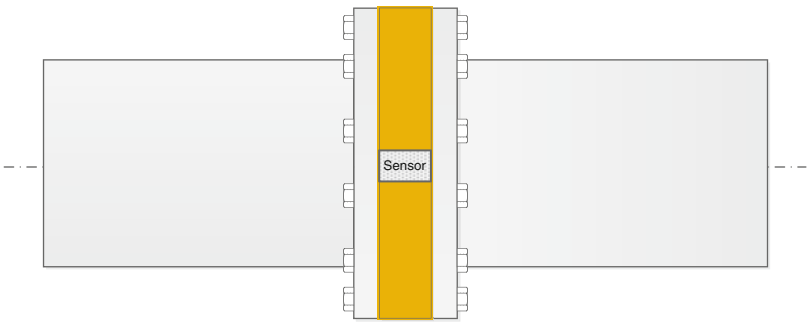


Fig. 10 GIS duct with a joint union, their bolts, and a flange-type sensor

shielding to confine all electrical elements and gas, so, once the substation is finished, introducing probes or installing sensors is practically impossible. The chambers behave as resonant cavities, so their size and shape set their resonance frequencies that can be measured in the range of UHF (Balanis 2012; Pearson et al. 1991). Then, whenever a pulse occurs inside, only a small fraction of the energy escapes the enclosure. For this reason, the most common sensor used to measure PD in UHF range is a capacitive disk coupler mounted on hatch cover plates in GIS substations (Hampton and Meats 1988; Pearson et al. 1991; Judd et al. 1996). Exceptionally, horn, biconical–log-periodic, loop, and monopole antennas have also been used for externally measuring partial discharges whose radiation leaks out through the flange connecting ducts (Kaneko et al. 2009).

The internal sensors are manufactured with metal shields, embedded electrodes, or field probes usually installed on a spare flange of the GIS or on dielectric windows (Judd et al. 2001). Apart from the capacitive disk couplers, some manufacturers also use conical antennas in new GIS substations. In (Chen et al. 2011), a square microstrip antenna and a broadband amplifier are designed, and the instrument is externally displayed for partial discharge measurement in a gas-insulated switch. The installation of internal sensors that have not been previously mounted requires opening the GIS with the subsequent risk of particle contamination. Therefore, this should only be carried out for renovation of aged substations or when the external sensors are not a practicable solution. Aside from avoiding opening the sealed chambers, external sensors are also used when the internal sensors are located too far apart and do not cover sectors where detection is needed. In these cases, the sensors are mounted in unshielded isolated spacers such as joints between conduits. Through these connections, radiation can escape out from within the GIS and can be measured mainly in the UHF range. However, measuring under these circumstances is influenced by several factors. One of them is the arrangement and shape of the resonant cavity that forms the structure of the GIS and defines its resonance frequencies when partial discharges are active. Another factor are the spacers between tanks which are attached with bolts disposed along the circumference as in Fig. 10. The sectors without bolts form small slots that behave as resonant dipole antennas. The greater the number of bolts is, the smaller the slots are and, then, the greater their resonance frequencies are (Kaneko et al. 2009). Conical, ring antennas or dipoles with frequencies bands tuned to the frequency defined by the slot can be placed at a few centimeters to measure the small fraction of the radiation that escapes from the cavities. Alternatively, in substations where there are inspection windows available, external sensors such as flange type are installed, Fig. 10.

Some manufacturers have developed portable spectrum analyzers using directional or resonant antennas to identify radiation sources originating from outside the electric equipment. These analyzers are used in periodic measurements, mainly in open-air substations upon the request of the owner utility. When UHF measurements are carried out in AIS substations, several electromagnetic noise sources and impulsive noise can perturb both the acquisition of PD and their locations (Shan et al. 2011). Conversely, partial discharges as noise can perturb Wi-Fi and ZigBee communications in substations. Nevertheless,

DOI 10.1007/978-981-4560-75-7_95-1

most of the efforts have been directed at localizing the partial discharge site in high-voltage plants (Moore et al. 2005; Portugués et al. 2009).

Transformers

As mentioned before, power transformers are strategic assets in the electrical grid. The oil–paper insulation permittivity directly influences the speed of the electromagnetic waves created by the partial discharges. For example, light mineral oil has a relative permittivity value of 2.9, and the signals travel at a speed of $2.03 \cdot 10^8$ m/s, while oil-impregnated pressboard has a relative permittivity of 4.52 and the speed is $1.46 \cdot 10^8$ m/s (Tang et al. 2006). Additionally, the permittivity does not vary significantly with temperature during transformer operation.

The measurement of PD in power transformers in the UHF range was developed in the 1990s in the last century after the first tests in GIS using also pre-installed hatch sensors or disk couplers. These sensors are mounted in power transformers bushing taps or in windows in the shielding tank. Other sensors, such as monopoles and conical antennas, have been also employed, inserted inside the tank through the drain oil valve of the power transformer (López-Roldán et al. 2008; Tenbohlen et al. 2008), previously emptying the tank. Evidently, there are other severe disadvantages in mounting sensors that have not been previously installed. The power transformer must be decommissioned, incurring in electricity production shortages unless a second transformer is available, which is, in most cases, prohibitively expensive. Furthermore, as in the case of GIS, during the installation, particle contamination in the tank can happen, inducing PD during normal operation once commissioned. A defective installation of a sensor in the shielding tank can cause problems of poor sealing and insulation.

The recommendations for condition monitoring and condition assessment facilities for transformers are reviewed in (A2.27 2008). This brochure takes into account the benefits of a possible standardization of the software for monitoring and identifying locations for mounting PD sensors in power transformers such as drain valve UHF probes.

PD measurement in power transformers with external antennas is an open-research topic (Robles et al. 2013a), wherein the electromagnetic emission of PD is difficult to obtain and only can be measured through dielectric windows of the tank and transformer bushes.

Additionally, there are many studies and applications devoted to the location of the partial discharge inside transformers with antennas. The calculus can be influenced by obstacles such as insulating paper, iron core, columns, and winding that delay the arrival time of the PD to the sensors. Section "Location of Partial Discharges" explains these techniques more thoroughly.

Rotating Machines

Online partial discharge measurement in rotating machines can detect degradation mechanisms that would otherwise have been missed in off-line measurements. The force of the magnetic fields when the machine is in service is the cause of the movement of the stator coils within the slot, provoking vibrations and abrasion and the subsequent deterioration of the insulation (Stone et al. 1992). Apart from loose windings in slots, partial discharge online measurements are able to detect thermal deterioration, load cycling deterioration, semiconductor and grading coating problems, and contaminated windings. Unfortunately, asynchronous and random noise is very present when the machine is rotating, and sophisticated techniques are needed to detect partial discharges and reject all unwanted disturbances. These techniques are different according to the type of rotating machines, hydraulic generators, or large turbine generators. There also exists some punctual research in inverted-fed motors using small monopole antennas inserted directly inside the motor terminal box (Fabiani et al. 2008).

DOI 10.1007/978-981-4560-75-7_95-1

Hydraulic Generators

The detectors are not antennas but coupling capacitors connected to the high-voltage bus and separated at a fixed and known distance. They create a low-impedance path for signals above several megahertz that can arrive from the stator of the machine (partial discharges) or from the line (noise). The pulses will be separated based on their origin using time differences of arrival to the detectors (Stone et al. 1992).

Turbine Generators

Stator slot couplers (SSC), basically two-port stripline antennas or capacitive–inductive couplers, can be installed under the stator wedges or between top and bottom bars in a new or rewound machine (Sedding et al. 1991). They are typically located in the high-voltage end slots of a stator winding where voltage and likely partial discharges are highest. The SSC have a frequency response from 30 MHz to over 1 GHz and are preferentially sensitive to the electromagnetic emission from PD pulses occurring in the proximities. This wide bandwidth helps in the discrimination of partial discharges from electrical noise based on the pulse shapes of both signals and the coupler location very close to the stator winding. Typically, arcing from bushes and partial discharges occurring outside the machine are attenuated by the inductive characteristic of the winding. They reach the coupler as low-frequency signals, while partial discharges in the stator have sharp rise times and high-frequency components. Unfortunately, pulses occurring a slot or two further down in the stator winding are heavily attenuated in the frequency detection range of these couplers, so they can easily be missed.

Other Equipment

The most common discharge in transmission lines and the less dangerous in terms of insulation damage is corona and was firstly measured in (Helstrom 1961). Nevertheless, the detection of PD in these systems is of importance, considering that they can also be surface discharges in the insulators of aerial distribution cables. These discharges can be critical to avoid faults in electrical lines installed in high-polluted or maritime environments with high salinity and humidity. In (Moore et al. 2004), insulation defects in 132 kV overhead lines were detected and localized using an array of four antennas and an acquisition system with a bandwidth of 1 GHz. In (Fernando et al. 2012), a horn antenna is located 1 m away to measure surface discharges from an 11 kV three-phase uninsulated overhead distribution cable system with the aim of designing an RF detection and localization system.

The measurement of PD emission in insulated cables can only be carried out by measuring their radiation through their joints and splices. In (Tenbohlen et al. 2008), on-site and online EM PD acquisitions are made using clamped monopole antennas around the cable terminations in the manhole of a GIS. These so-called antennas are really inductive and capacitive couplers that have also been used by many other authors. There are not evidences of far-field antennas in the measurement of partial discharges in insulated cables.

It is important to bear in mind that the expenses of high-frequency acquisition equipment are an important drawback in the detection of partial discharges in overhead lines because they can be hundreds of kilometers long, and it is impossible to cover all the system with few antennas. As for insulated cables, most of them are actually buried, so the use of antennas is out of the question except when they are placed close to joints. Even in these cases, other alternative methods, such as the use of high-frequency current transformers, are widely preferred.

Other authors have installed microstrip (Yao et al. 2013) or horn (Cai et al. 2014) antennas inside switchgear to monitor partial discharges. In this case, the size of the antenna is one of the most important design constraints due to the lack of space inside the switchgear which is packed with partition boards, circuit breakers, bushings, insulators, bus bars, and other measuring devices such as current transformers. The sensor should be placed far from the current transformers and bus bars because they attenuate the

DOI 10.1007/978-981-4560-75-7_95-1

electromagnetic emission of the partial discharges; insulators would distort the waveform so they should also be avoided; on the contrary, the circuit breakers do not affect RF, so they can be in the path of the partial discharge emission.

Types of Antennas

The different types of EM sensors mentioned in the last section can be identified depending on whether the operation is to couple (capture signal in a cavity) or to pick up signals from air (free space), working as a conventional antenna in a communication system. Hence, depending on the position of the sensor, we can distinguish between:

- Internal sensors: electromagnetic probes coupled to high-voltage equipment where partial discharges take place (Pinpart and Judd 2009; Reid et al. 2011).
- External sensors: antennas working in free space at close range to the device under test. In that case, it is the leaked signal that is picked up and processed (Chen et al. 2011). The antennas operate in near- or far-field conditions depending on the case.

In both cases, the signals can be measured from one or more points if the position of the discharge site is to be determined, section "Location of Partial Discharges."

Internal Sensors

When dealing with internal sensors, safety is always a concern because they are always placed close to hazardous zones. High-voltage equipment such as power transformers, rotating machines, and GIS are usually shielded so they can be considered as lossy inhomogeneous cavities. Hence, when the antenna is used internally, it is typically cavity backed (shielded), screened against electromagnetic interference from external sources. Since the wavelength is relatively small, multiple resonating modes could be simultaneously present when the discharge occurs. Moreover, the antenna response for internal sensors could be highly affected by elements in the vicinity which must be taken into account in the design and integration of the measuring system. Recently, some research has been published in the literature where a transformer is electromagnetically simulated by means of a full-wave analysis computer-aided design (CAD) (Zheng and Bojovschi 2012).

In this type of sensors, we can further distinguish between:

- Dielectric window sensors. The high-voltage equipment includes small dielectric windows in some predefined positions which are manufactured as a part of the tank or encasement so the sensor can be easily connected. Obviously, their sizes should be as small as possible so as not to have any influence in the safety of the equipment in operation. For other dielectric windows, not usually provided by the high-voltage equipment manufacturer, the typical diameter of the sensor is 15 cm (Judd et al. 2005).
- Sensors placed in valves (oil or gas drain valves): When the antennas are placed in valves, they can have a certain maximum width and depth. The typical size of the drain valve is 5 cm in diameter and 10 cm in length.

External Sensors

When signals are picked up externally, the antennas will be located close to monitored equipment in the range from centimeters to a few meters. In these cases, the signal strength will be smaller than the obtained with internal sensors. Interference coming from nearby external sources, mainly from communications systems located in the same frequency bands, will also degrade the measurement. For external antennas

DOI 10.1007/978-981-4560-75-7_95-1

with no volume constraints, dipole or monopole antennas are used in their wired or printed versions, with the usual modifications to widen the matching band or to achieve compact designs.

Antenna Selection

The performance of UHF antennas determines, to a large extent, the quality of the received signal. It is important to note that the antennas for partial discharge detection work only in receive mode and are connected to the receiver device by a feeding line that must be matched in impedance, typically $Z_o = 50\ \Omega$. Oscilloscopes, spectrum analyzers, amplifiers, or analogue to digital converter inputs are set to low impedance.

In what follows, the description of the operation mode of the considered antennas will rely mainly on the radiation pattern (directivity) and the impedance matching. It will also be assumed that there is no dominant polarization for the electromagnetic waves radiated by the discharges.

The power frequency content of the signals from partial discharges can vary considerably, depending on the location of the discharge and the propagation path through the asset and the surrounding high-voltage equipment. Besides, the antenna has a response in terms of amplitude and phase which will affect the received pulse and will give rise to a time domain signal whose complexity will depend on the discharge type, the environment, and the antenna itself. So, the quality of the received signal will be largely affected by these circumstances, resulting in a distorted pulse, hindering its detection and subsequent processing. Thus, depending on whether we want to *detect* the occurrence of the discharge, to *classify* the type of discharge, or to *locate* the position of the point where the PD occurred, the role of the antenna will be more or less important.

Since partial discharges emit radiation over a wide frequency range, it is necessary to use antennas with bandwidths wide enough to optimize the measurement. This implies that the typical parameters of an antenna, such as the voltage standing wave ratio (VSWR) or the S_{11} parameter, and the radiation patterns must be stable throughout the operating band (Choi et al. 2004). Besides, a minimum change in the phase center position is required if the partial discharge site is to be determined using any of the techniques described in section "Location of Partial Discharges."

It is well known that the measurement of the signals emitted by the PD in the wideband mode has advantages over the use of narrow bandwidth (Stone et al. 1992). Partial discharge signal-to-noise ratio shows improvements with the increase of bandwidth and provides more accurate results (Boggs and Stone 1982). Due to this reason, wideband and ultra-wideband (UBW) antennas are of particular interest in partial discharge classification applications, and maximizing the bandwidth while maintaining compactness will be one of the design goals of this type of antennas (Robles et al. 2013c). Log-periodic and fractal schemes can be used to improve the bandwidth although their size can increase substantially when they are used in the low frequency range of the UHF band.

Resonant antennas have multiple resonant modes within the band where the signal is emitted. Furthermore, each resonance frequency corresponds to a different radiation pattern, which can strongly distort the received signal so they can be used in the detection or localization of partial discharges. Some well-known modifications in the geometry of these antennas can be introduced in order to enhance the bandwidth of the fundamental mode (broadband–wideband antennas).

Depending on the exigences of the measurement, the antennas must be balanced or can be unbalanced (grounded). The balanced antennas, such as dipole antennas, require a balun to feed the balanced structure from an unbalanced transmission line such as coaxial cables. If no balun is used, this directly affects both to the impedance matching of the antenna to the feeding transmission line and to the corresponding radiation pattern. The balun can also perform the function of impedance matching, which may be interesting when seeking for compact broadband antennas where the input impedance is high.

DOI 10.1007/978-981-4560-75-7_95-1

In this section, several antenna shapes that have been used in the recent years for PD detection will be simulated in their simplest version for the sake of their comparison. The antennas will be classified in three large groups: wire antennas, printed planar antennas, and aperture antennas. Bibliographic references of each type will also be provided. Among the sensors are short monopoles, monopole-conical shape, top-loaded monopoles, disk antennas, and spiral, fractal, and Vivaldi antennas. The balanced cases have been simulated in their planar versions. When a grounded antenna is simulated, a circular ground plane has been used to emulate the valve cover that closes electromagnetically the high-voltage equipment, with a radius of 5 cm and a maximum height of 10 cm. A FR-4 substrate ($\epsilon_r = 4.3$) with thickness $h = 1.5$ mm has been used in all cases as the antenna base unless otherwise stated.

All prototypes have been simulated in the frequency range 0.3–3 GHz, assuming perfect electric conductors and lossless materials. Ideal $Z_s = 50\ \Omega$ point generators have been used to drive the antennas for both unbalanced-grounded or balanced antennas. For all cases, a sketch of the antenna, its return loss (RL) for a reference impedance of $Z_0 = 50\ \Omega$ versus frequency, will be provided.

Wire Antennas

The simplest wire antennas are dipoles, monopoles, and loops. A complete review of these antennas for PD application can be found in (López-Roldán et al. 2008), where different implementations are compared, including straight, trapezoidal, helix, disk, and conical monopoles and also simple loop and spiral antennas. Table 3 includes the most common cases.

Dipole antennas are widely used for many different communication applications. Such resonant antennas consist of two radiating elements, the arms of the antenna, fed in opposite phases. The radiation pattern has a null in the antenna axis, being omnidirectional in the plane perpendicular to that axis.

Monopole antennas consist of only one resonant element which is fed by a transmission line through a ground plane. Hence, a balun is not required. Besides, with an appropriate design procedure, some monopole antennas can provide a wider bandwidth than dipole antennas. Table 3 shows the reference result for an ideal monopole tuned to a frequency around 750 MHz. The second resonance can be easily recognized on the right.

Many different implementations of monopole antennas which are variations of the simplest case can be found in the literature. The modifications are usually aimed at widening the operating frequency band or achieving more compact antennas, especially in those cases where the antenna must be installed in a transformer valve. In all cases, the central wire is replaced by cylindrical elements (wires or disks), conical elements, or combinations thereof (top-loaded monopoles). Different implementations are reviewed below.

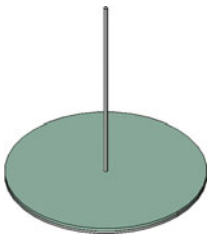
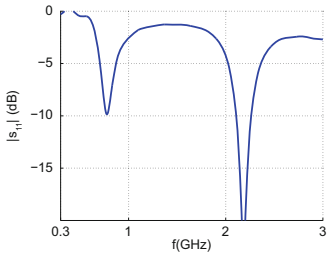
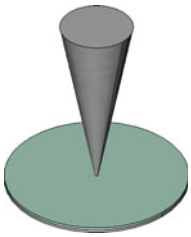
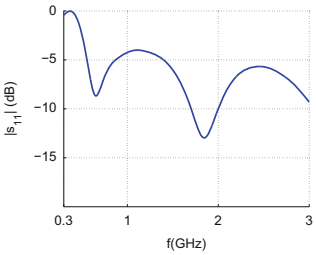
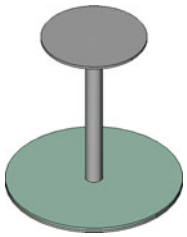
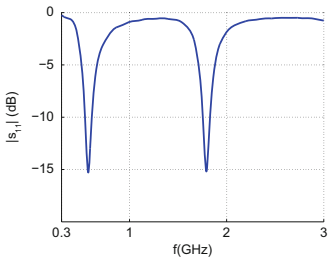
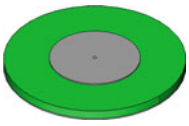
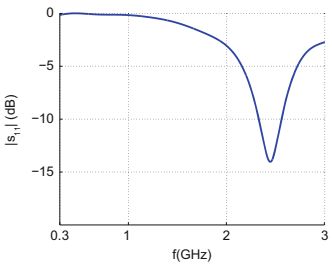
The first modification is to replace the central wire by an inverted cone of about the same height (conical skirt monopole). It is clear that the differences with the response for the conventional monopole are small. An increment in the bandwidth for each of the resonances as well as a slight displacement of the resonance frequencies towards lower frequencies is observed. The radiation patterns do not show significant changes.

In a second variation, a metallic disk is added to configure a top-loaded monopole antenna that typically exhibits an improvement in the matching level, maintaining the same resonance frequencies. Once again, radiation patterns do not exhibit significant changes.

The former elements can be combined in different ways. In patent (Raja 2007), an inner sensor is proposed using a dipolar-like structure where one of the electrodes is a disk and the other an inverted cone. The sensor operates in a frequency range between 0.2 and 1 GHz. In patent (Kato and Oshita 2000), another inner sensor is described, comprising a small dipole whose branches are two semicircular metallic disks.

DOI 10.1007/978-981-4560-75-7_95-1

Table 3 Monopolar antennas

Type		Matching
Monopole (grounded)		
Conical monopole		
Top-loaded monopole		
Disk (centered probe)		

When the antenna is used in a valve as in the latter cases, the metal walls have a significant effect on its electromagnetic response. In López-Roldán et al. (2008), a study is performed about the effect on the sensitivity of the depth of insertion of the cone into the pipe where the conical antenna is located, being the most favorable case when the antenna penetrates further in the transformer.

To show this effect, Fig. 11 represents the result of including an outer metallic cylinder of 8 cm in height in a top-loaded monopole antenna. It is observed how the first resonance frequency around 500 MHz has almost disappeared.

The upper part of the inverted cone can be modified to optimize the performance of the antenna. According to this idea, in Cleary and Judd (2006), a metallic semisphere is used above the base of the cone to achieve an improved response covering a good portion of the UHF band. Going further, the vertical

DOI 10.1007/978-981-4560-75-7_95-1

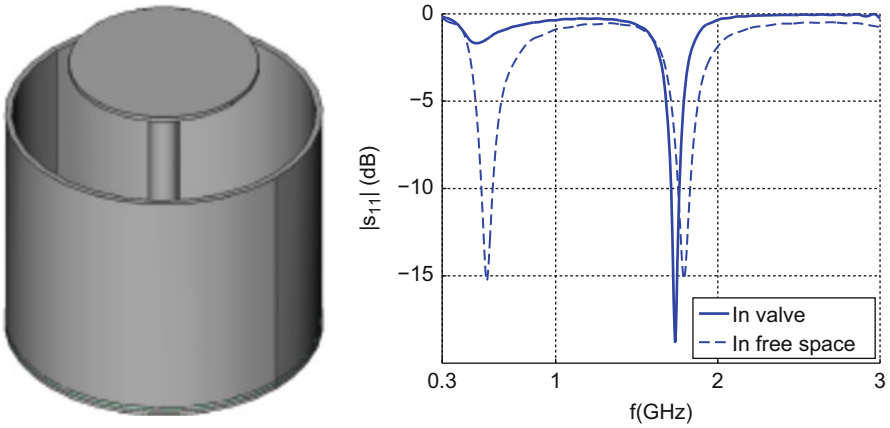


Fig. 11 Top-loaded monopole antenna in a simplified drain valve

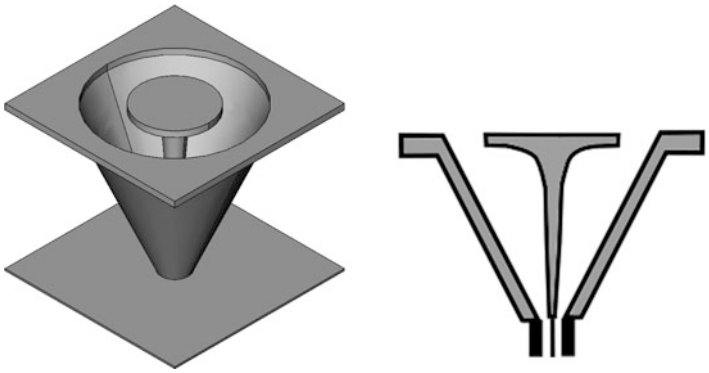


Fig. 12 Optimized top-loaded monopole antenna

monopole wire can be replaced by a different geometric element like in (Petchphung et al. 2007), where a metallic disk with a radius of 14 mm is used instead, achieving ultra-wideband response by means of a very compact sensor in a typical configuration for a UWB communication antenna.

In all the described cases, an optimized profile can be tailored in the central conductor to compensate for the effect of the outer metallic cylinder when the antenna is used as an internal sensor like in Fig. 12. For instance, patents (Reinhardt 1985) and de (Kock and Herbst 1998) are examples of such an approach. The combined effect of the monopole and the outer shielding element could also be seen as a tapered coaxial transmission line.

In a more complex configuration, patent (Fukasawa et al. 2014) describe different top-loaded monopoles shunted with a parallel stub that are used to improve the impedance matching in inner sensors. The stub is shorted to ground by connecting it to the outer metallic shell. Some dielectrics are employed to provide isolation and sealing.

The top-loaded monopole can be modified by increasing the size of the disk and reducing the height of the central post such as described in (Judd et al. 1995). In addition to these elements, some resistors are used in this antenna to earth the disk to the ground plane. In (Judd et al. 2000), a similar antenna that uses a composite material for insulation is presented. This antenna is compared to a single monopole in Ishak et al. (2012) using a finite difference time domain (FDTD) approach.

The disk in Table 3 represents a simplified model of this type of sensor. A PTFE substrate ($\epsilon_r = 2.2$) is used in this case. Note that the antenna radius has been increased to 10 cm. Substrates with higher

DOI 10.1007/978-981-4560-75-7_95-1

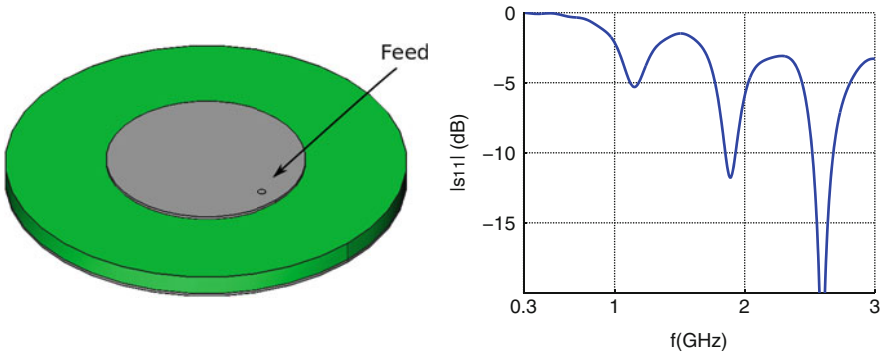


Fig. 13 Disk antenna with offset coaxial probe

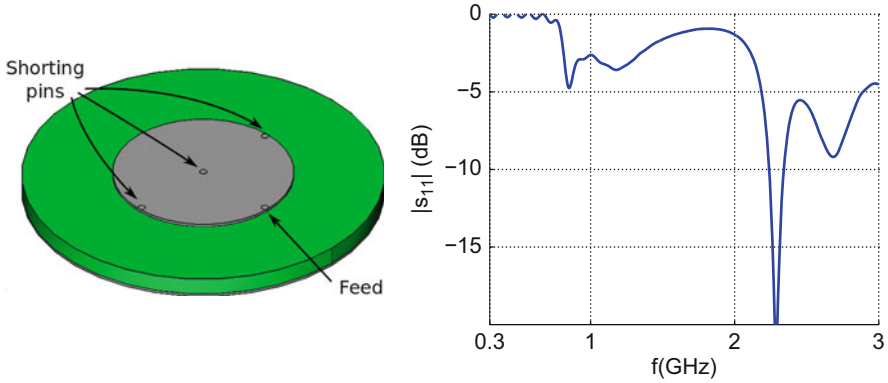


Fig. 14 Disk antenna with offset coaxial probe plus shorting central and lateral pins

permittivities can be used to reduce the antenna size, although this will have an impact on the matching level.

When the disk antenna is excited by a coaxial probe located at a point off-center closer to its edge, a good impedance response with essentially broadside radiation patterns can be achieved. Strictly speaking, this antenna is not a wire type but a circular patch-type antenna though it is shown in this section for simplicity. Figure 13 shows the response for this case. Multiple resonances are present within the band of interest, each having a different broadside radiation pattern depending on the field distribution of each mode.

This antenna supports a simple modification to achieve a compact and well-matched configuration. To this end, shorting pins are used in the center of the antenna and in different optimized positions as indicated in Fig. 14. Note that when a central shorting pin is used, the corresponding resonant mode for the antenna, according to its simplified cavity model, is rotationally symmetric (TM_{01} mode) (González-Posadas et al. 2006), which has a zero at the axis of radiation like a monopole. More pins can be added to improve the antenna response (Vázquez-Roy et al. 2010). Those pins also short to ground the central disk, protecting equipment and personnel of the risk of high-voltage contact.

Loop antennas are simple sensors with low efficiency when their size is not large. For instance, in Jin et al. (2008), small loops are used with relatively good results for detecting discharges in air cavities and surface and corona discharges in oil. Also, in (López-Roldán et al. 2008), a comparison is done with different monopoles with not-so-good results for PD detection in oil-filled transformers.

Loop antennas can be loaded with ferrite cores to improve their response (Ozaki et al. 1992). The radiation patterns are, essentially, the same as the dipole ones. If the direction of arrival of the wave front

DOI 10.1007/978-981-4560-75-7_95-1

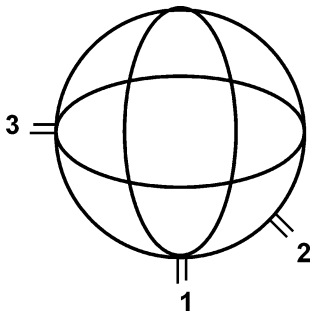


Fig. 15 Three-axis loop antenna for the detection of the angle of arrival of the wave front

in free space is to be estimated, three different loops oriented like in Fig. 15 can be used by properly combining the output of the antennas (Matsumoto et al. 2008).

Printed Antennas

Printed antennas have many interesting features including flexibility, compactness, low profile, and low cost. Many of the different possibilities are described in detail in (Reig and Ávila Navarro 2014). In the most common implementations, the planar structure can be located either perpendicular or parallel to the main radiation axis, in a broadside or end-fire radiation operation. Table 4 includes the most common cases.

Patch Antennas

Patch antennas, also known as microstrip antennas, are one of the most used printed antennas in a great variety of application fields. Some typical patch antenna geometries were patented for the measurement of PD in the 1990s (e.g., see Herbst and De 1996). However, in their basic configuration, they are resonant narrowband antennas, and, due to their big size in terms of wavelength, they have been used for PD detection to a lesser extent compared to other types. An analysis of the response of a rectangular patch antenna can be found in (Shibuya et al. 2010).

Wire Antennas

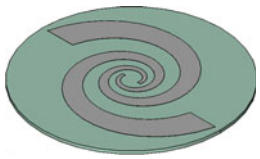
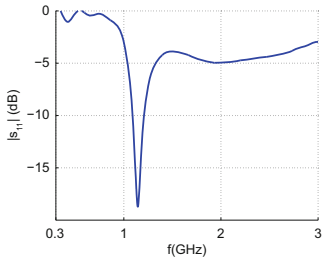
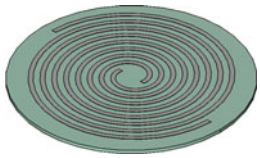
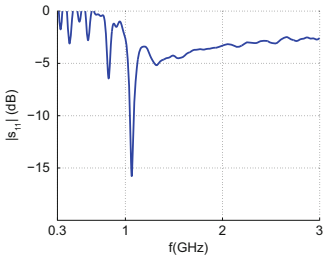
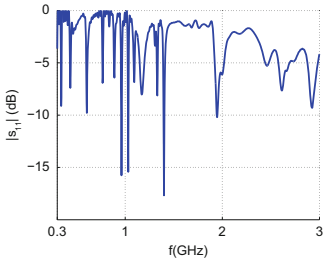
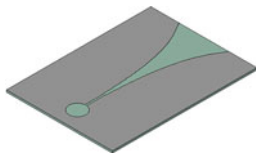
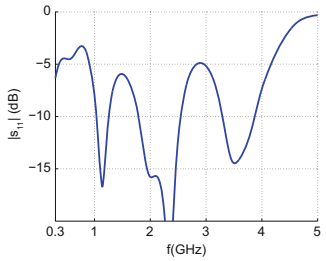
Many wire antennas can be implemented directly on a printed version. In that case, a low-loss substrate is used as support to print the metallizations on the same side of the substrate or on different sides (coplanar or antipodal antennas). In the same way, the wire in a monopole can be substituted by its planar printed version, and the geometry can be modified in order to optimize the antenna response. When the antenna is intended to work in the low-frequency band, a meandered line can be of use to achieve a compact antenna (Ayob and Ramli 2014; Li et al. 2014).

Logarithm Spiral Antennas

The logarithm spiral printed antenna is a differential wideband sensor commonly used for PD detection. In (Reid et al. 2009), this antenna is compared to a monopole and log-helical antenna using a FDTD algorithm. In Table 4, a result is included for such an antenna in free space. As with wire antennas, if the logarithm spiral is to be used inside a valve, it will operate inside a metallic cylinder which affects the antenna response considerably. Moreover, the radiation pattern curves show an important component of back radiation that could be reduced by placing a ground plane at half a wavelength for one specific frequency within the working band. To mitigate the unavoidable bandwidth reduction, the cavity which closes the antenna in its lower part can be filled up with a wave absorption material. It is worth noting that

DOI 10.1007/978-981-4560-75-7_95-1

Table 4 Printed antennas

Type		Matching
Logarithm spiral (balanced)		
Archimedean spiral (balanced)		
Hilbert (grounded)		
Vivaldi		

this antenna has a null radiation in the plane in which it is contained. With respect to the matching bandwidth, it comprises a very wide margin, with a clearly marked resonance at low frequencies.

The printed branches can be terminated in different ways besides the straight cut. In Zhang et al. (2014), printed butterfly elements are added to improve the bandwidth in a GIS sensor. These elements represent only a small penalty in size, and the manufactured antenna compares quite well with the horn antenna depicted in Fig. 16.

Not only logarithm spirals are used for PD detection. In Wang et al. (2001), an Archimedean spiral like the one shown in Table 4 is used. The response, in terms of matching and radiation pattern, is quite the same as that of the log-spiral antenna.

DOI 10.1007/978-981-4560-75-7_95-1

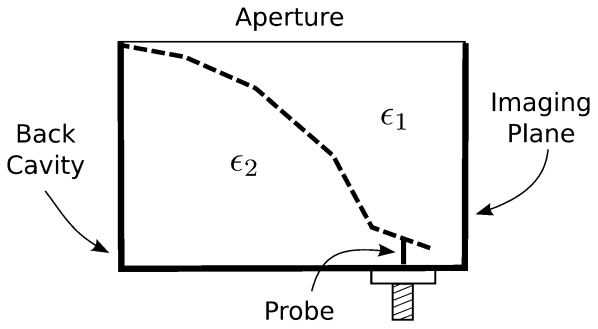


Fig. 16 Half TEM horn

Fractal Antennas

Fractal antennas are composed of long printed lines which are packed within a relatively small area to achieve compactness in a self-similarity structure. Thus, a motif is repeated over different scales providing the different resonances over an operating frequency band. One of the most used schemes is the Hilbert fractal depicted in Table 4. This antenna is analyzed in (Jin et al. 2008) and compared with a small-loop antenna. In (Li et al. 2012), a $N = 4$ order Hilbert antenna for PD measurements is studied, demonstrating how multiple resonances are present in the frequency band between 0.3 and 1 GHz. In that case, the fractal line is printed on a FR-4 substrate in a monopolar grounded configuration, being 1.5 mm the substrate thickness. The feeding point is selected to improve matching, and the sensor is combined with a wideband amplifier. Care must be taken when using this configuration, given the fact that the fractal line is printed on a relatively high permittivity substrate and the radiation efficiency might be very low. Moreover, some of the resonances within the frequency band could be related to dissipation and not to radiation.

To avoid this potential problem, the ground plane of the antenna and the substrate where it is printed can be spaced by air. Then, the coaxial probe must be extended through this space. In Yao et al. (2013), a fractal antenna which uses a Koch motif is designed according to this idea. Note that in the Hilbert antenna included in Table 4, the substrate was raised 3 cm in air. Many different resonances are present between 0.4 and 3 GHz.

Vivaldi Antennas

Wideband detection improves the measurement of the time of arrival when it comes to PD location (Tang et al. 2006). So, ultra-wideband antennas such as the Vivaldi antenna could be of interest to the PD detection antenna designer. Vivaldi antennas are low-profile planar antennas which include a tapered slot and an appropriate transition from a feeding transmission line. Since these antennas are not resonant in nature, bandwidths in excess of 3:1 in return loss can be obtained by optimizing the profile of the slot and the feeding network. Besides, a relatively directive pattern in the end-fire direction with linear polarization is achieved. In its simplest version, a microstrip line is used to feed the balanced tapered slotline and two radial and circular open stubs are included as a part of the feeding network in order to improve the matching. Table 4 includes the result for this type of antenna in an extended frequency range. An analysis of the response of a Vivaldi antenna placed outside the high-voltage equipment is performed in (Robles et al. 2013b). The measured return loss for the built prototype is below 10 dB in the range between 1.5 and 3.7 GHz with a total size of 9×12 cm. The sensitivity at low frequencies is poor, so FM radio bands and low TV broadcasting band are eliminated, obtaining good signals in the time domain.

Aperture Antennas

Aperture antennas are not so common in PD applications and only a few of them can be identified in the literature. For example, in xing (Zhang et al. 2013), a half TEM horn antenna is specifically designed for

DOI 10.1007/978-981-4560-75-7_95-1

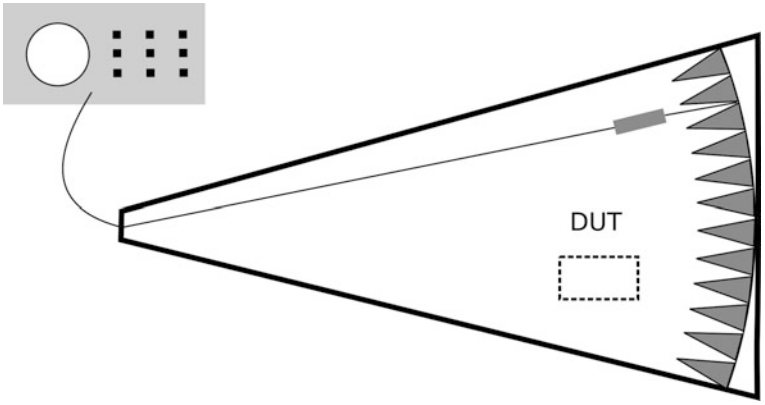


Fig. 17 GTEM cell

PD detection after (Shan et al. 2008). The antenna has the semi-closed cavity-backed structure depicted in Fig. 16. The horn is partially filled with dielectric to enable the reception of EM waves of shorter wavelengths. The simulated prototype is centered around 900 MHz and combines several different resonances to operate up to 3 GHz.

In a more conventional way, the design of a double ridge horn for PD detection in the range from 0.7 to 1.4 GHz is described in (Cai et al. 2014). The horn is fed using a coaxial probe and the set up was optimized for dual working frequency bands, namely, 0.7–0.9 GHz and 1.25–1.4 GHz. The dimensions of the aperture are 28 × 35 cm. This antenna was integrated in a medium-voltage switchgear system and was used to sense spatially distributed PD in AIS. The full transformer is simulated using a finite element method to determine the optimum locations for the sensing antennas.

Antenna Calibration

Due to the casuistic and the procedure typically followed in the detection of PD, a time domain analysis to obtain the involved signals is also of interest. The time domain signal at the output of the receiving antenna includes the nature of the partial discharge, the propagation through shields, metallic bulks and different media, and finally, the response of the antenna.

TEM cells are devices used for establishing a determined electromagnetic field in a shielded environment. They are commonly used for electromagnetic compatibility measurements, but they are also of interest in the context of PD detection since they provide clear reference signals which can be used as calibration references (Karst et al. 2002).

The GTEM cell (Fig. 17) is a frequency extended variant of the conventional TEM cell which can cover frequencies up to some GHz (Judd and Farish 1998). A slightly spherical wave propagates along the structure excited by a coaxial probe to simulate a free space planar wave. RF absorbers are used as a termination to avoid a standing wave buildup.

Several time-domain analyses have been performed to obtain the received signal in some relevant antennas to compare the simplified designs that have been presented in the previous sections. Instead of using TEM cells, a PD has been simulated in free space at a distance of 0.5 m of each antenna. That position has been set to be the optimum one for each antenna regarding radiation pattern with polarization matching. A time domain full-wave analysis algorithm has been used with this purpose like in (Ishak et al. 2012). An oriented point source has been used to drive the simulations. Free space conditions have been considered for the propagation between the source, which is the simulated PD, and the antenna under test. A Gaussian pulse with spectral content in the range of 0.5–3 GHz has been used as excitation signal,

DOI 10.1007/978-981-4560-75-7_95-1

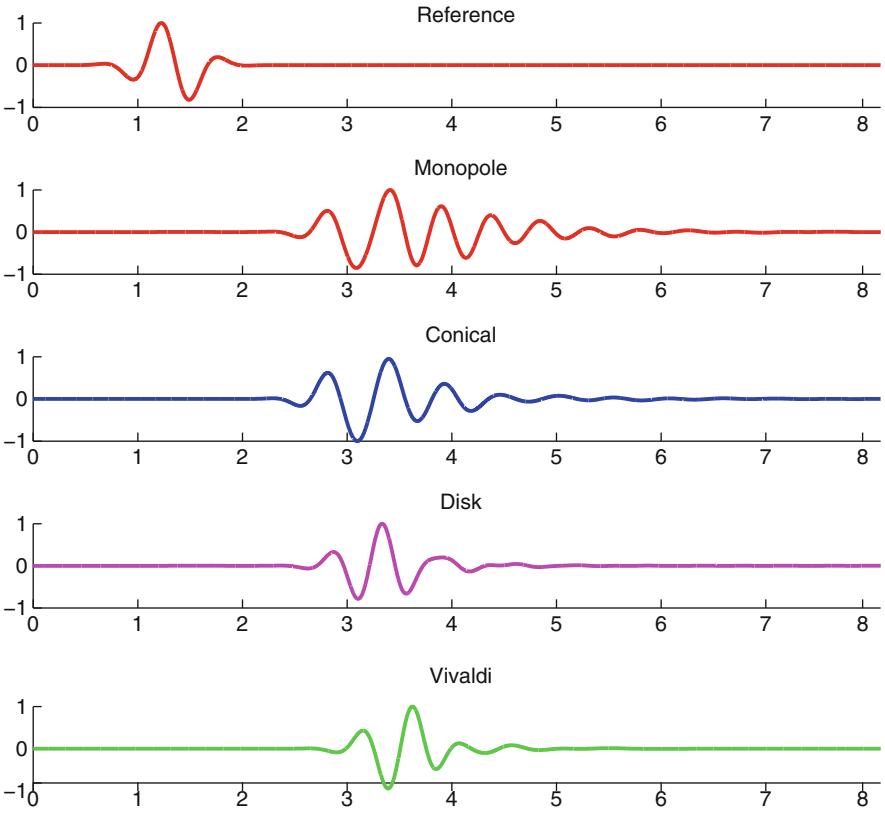


Fig. 18 Received signals in time domain for some of the simulated antennas

simulating, thus, a partial discharge in the order of ns. The results can be interpreted as the best signals that can be obtained with each type of antenna. The disk antenna displayed in Fig. 14 will be used in this case.

Figure 18 summarizes the obtained results where the source signal has also been included as a reference. All signals have been normalized to their maximum value. It can be observed how resonant antennas, such as conventional and conical monopoles, lead to more distorted signals with higher degrees of energy spread along time.

Finally, Fig. 19 displays the comparison of the envelopes of some of the received signals along with the envelope of the reference signal. By plotting all signals together, the degree of distortion introduced by the different antennas becomes more apparent. The signals have been normalized to the maximum of the Vivaldi antenna (best case). The Hilbert antenna of Table 4 is also included. It can be observed that the time delay is the same for all cases, while the different slopes of the leading edges are directly related to the antenna bandwidth. Moreover, the effect of the radiation pattern of the antenna can be deduced from the amplitudes of the pulses: the greater the directivity is, the higher the received signal level will be. This is significantly observed in the case of the Vivaldi antenna and the disk antennas.

As expected, the bandwidth of the antenna affects the received signal. Higher bandwidths will result in clearer time signals, facilitating the detection of the pulse and the subsequent signal processing.

Location of Partial Discharges

Location in Linear Structures

Partial discharges occurring in power cables, gas-insulated switchgear, and windings will travel along the waveguide or conductor and will be reflected at the ends (a bushing or terminal). Inductive sensors or

DOI 10.1007/978-981-4560-75-7_95-1

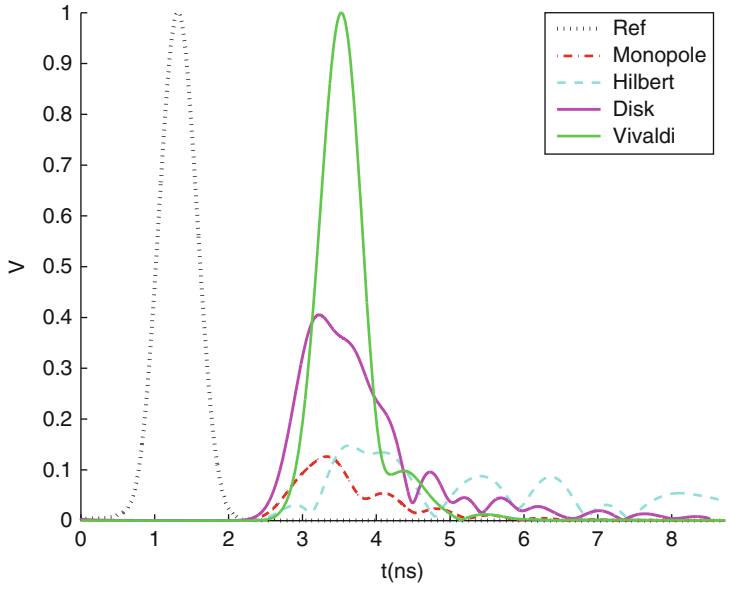


Fig. 19 Signal envelope comparison for some of the simulated antennas normalized to the maximum of the Vivaldi antenna signal

capacitive sensors are used in cables to detect the partial discharge, whereas electromagnetic sensors are used in GIS. Let us consider that the sensor is attached at $x = 0$ and the PD site is at a linear distance in $x = x_{PD}$ (Steiner et al. 1992). The first arriving time will be:

$$t_1 = \frac{x_{PD}}{v} \tag{36}$$

where $v = 1/\sqrt{LC}$, with L and C the inductance and capacitance per unit length of the cable, respectively; or $v = 1/\sqrt{\mu\epsilon}$ with μ and ϵ, the permeability and permittivity constants of the gas, respectively.

The front wave will be reflected at $x = L$, being L the total length of the cable, the GIS section, or the distance from the origin to a junction, terminal, or bushing. Then, the second time of arrival will be:

$$t_2 = \frac{2L - x_{PD}}{v} \tag{37}$$

Measuring $\Delta t = t_2 - t_1$ and knowing L, it is possible to calculate the distance from the PD site to the closest end of the device under test.

$$x_{PD} = L - \frac{1}{2} v \Delta t \tag{38}$$

Errors in the location of PD are caused by different sources (Kreuger et al. 1993). Focusing on the case of GIS, the pulse propagated through the encased switchgear can suffer attenuation and losses and provoke undetermined back reflections that may confuse the instrumentation and assign false arriving times. There has to be a common time base for all detectors to have accurate time differences; notice that an error of 1 ns for pulses travelling at the speed of light means an uncertainty of 0.3 m in the location of the source. Impulsive external noises or high noise levels affect the accuracy of the measurement. And, finally, the total length L of the resonant cavity has to be exactly measured to ensure that the distance to the PD site is accurate.

DOI 10.1007/978-981-4560-75-7_95-1

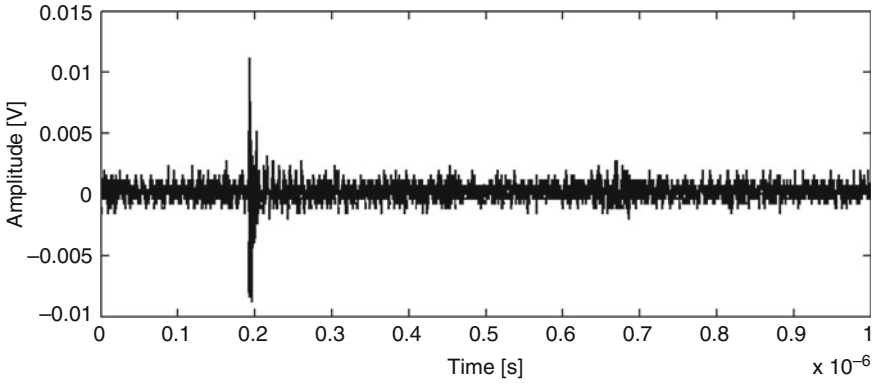


Fig. 20 Internal partial discharge obtained with a 15 cm monopole

Another application of this principle can be done off-line by injecting a known pulse and calculating the time difference of arrival of the pulses reflected at the PD site and at the end of the device under test. This approach is based on the fact that the dielectric properties change when the dielectric degrades. Therefore, when a pulse reaches these inhomogeneous zones, part of it is reflected and part of it is transmitted and then reflected at the open end of the cable. This technique can be applied to cables but not to GIS segments where partial discharges occur due to small particles that will not modify substantially the dielectric properties of the gas. Moreover, the defect can be located in new cables even before there is a continuous partial discharge activity and without energizing the cable.

Location in Three Dimensions

Partial discharges in transformers or open substations can be located using radiometric methods based on the time differences of arrival (TDOA) of the pulses to pairs of antennas.

Calculation of the Time Difference of Arrival

The incident signal emitted from the partial discharge site consists of a direct wave propagation and a long tail of oscillatory cycles due to multipath propagation and scattering. Additionally, the signal contains electromagnetic interferences from radio and TV broadcasting, mobile phones, and Wi-Fi among others, generating a noise floor that hides the start of the first pulse, Fig. 20. The first recommended step is to filter those interferences with any de-noising technique. Once the signal is cleaner, one of these algorithms can be applied to determine the time differences:

Cumulative energy. The energy of the acquisition is accumulated in time (Judd et al. 2005).

$$W_n = \sum_{k=0}^{n} s_k^2 \tag{39}$$

where W_n is the energy accumulation up to sample n and s_k is the signal sample k. First, there is an increment with a constant slope due to the noise acquired in the pre-trigger time span, and then, there is a dramatic change of the slope when the pulse starts. This instant is marked for all four antennas to calculate the time differences, Fig. 21.

Alternative cumulative energy. In this case, the cumulative energy is constantly reduced until the signal starts (Markalous et al. 2008). The minimum of the function is the starting time; see Fig. 22.

DOI 10.1007/978-981-4560-75-7_95-1

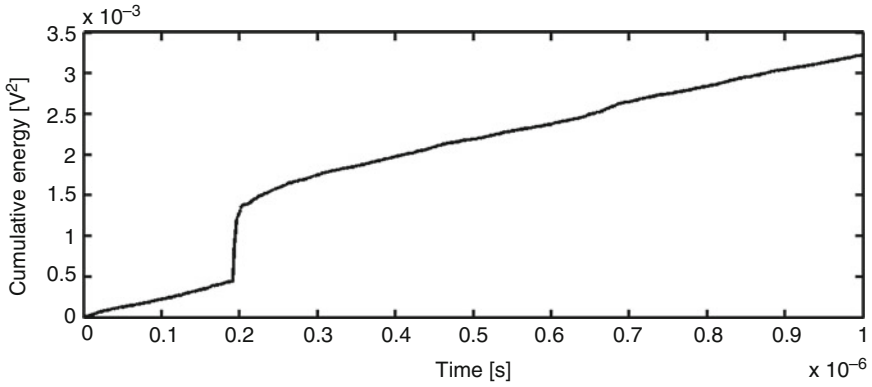


Fig. 21 Energy accumulation for the signal shown in Fig. 20

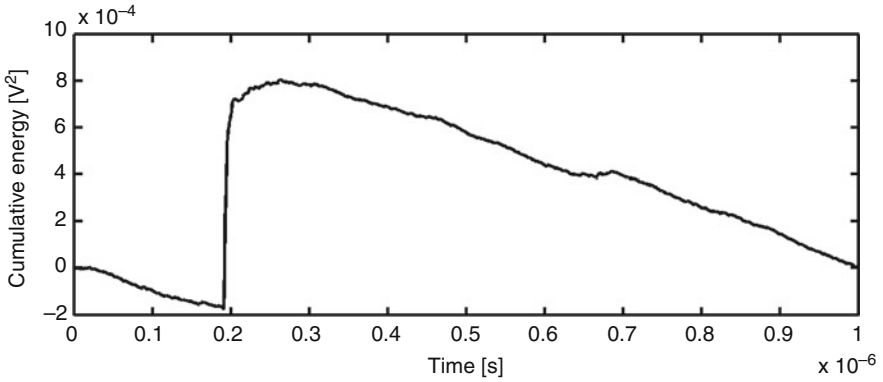


Fig. 22 Alternative cumulative energy. Notice that the minimum of this function coincides with the change of slope in Fig. 21

$$W_n = \sum_{k=0}^{n} \left(s_k^2 - n \frac{W_N}{N} \right) \tag{40}$$

with N the total number of samples and W_N the total energy of the signal.

Cross-correlation. The cross-correlation of the signals from two antennas has a maximum in the delay sequence that gives the time difference. The equation for the calculation of the cross-correlation for the finite sequences s_{k1} and s_{k2} from antennas 1 and 2, respectively, is:

$$r_{l,12} = \sum_{k=l}^{N-|n|-1} s_{k,1} s_{k-l,2} \tag{41}$$

with $n = 0$ for $l \geq 0$ and $n = l$ for $l < 0$.

The generalized correlation method uses weighting functions or pre-filters to enhance useful characteristics of the signals that will improve the peak detection of the standard cross-correlation. These filters are applied in time and modify the cross power spectrum with their frequency response:

$$R_{12}(t) = \int_{-\infty}^{+\infty} \psi(f) G_{12}(f) e^{j2\pi f t} df \tag{42}$$

DOI 10.1007/978-981-4560-75-7_95-1

where G_{12} is the cross power spectral density of the pre-filtered signals and $\psi(f)$ is a weighting function with the idea of accentuating the signals passed to the correlator at those frequencies at which the signal-to-noise ratio is highest. This would have the desired effect of obtaining a large sharp peak in R_{12} rather than a broad one to ensure good time-delay resolution (Knapp and Carter 1976).

The smoothed coherence transform (SCOT) uses the prewhitening filter shown in Eq. 43 to reduce the cross-correlation when the signal-to-noise ratio is low, assuming that the noise components affecting the two signals are uncorrelated and cannot be filtered since the partial discharge pulses may have broadband and narrowband components.

$$\psi(f) = \frac{1}{\sqrt{G_{11}(f)G_{22}(f)}} \tag{43}$$

The phase transform (PHAT) uses the weighting in Eq. 44 with the aim of avoiding the spreading of the cross-correlation function. Without noise components, the output of Eq. 41 with this filter is a delta function centered at the correct delay.

$$\psi(f) = \frac{1}{|G_{12}(f)|} \tag{44}$$

The maximum likelihood (ML) or Hannan and Thomson (HT) Eq. 45 processor is chosen to improve the accuracy of the estimated delay by attenuating the pulses in the spectral region where the signal-to-noise ratio is poorer.

$$\psi(f) = \frac{1}{|G_{12}(f)|} \quad \frac{|\gamma_{12}(f)|^2}{|\gamma_{12}(f)|^2} \tag{45}$$

with

$$|\gamma_{12}(f)|^2 = \frac{|G_{12}(f)|^2}{|G_{11}(f)| \cdot |G_{22}(f)|} \tag{46}$$

In any case, the accuracy of the TDOA depends ultimately on the sampling frequency. It is expected that the exact time lies between two samples of the cross-correlation or the cumulative energy sequence. An approximation using interpolation can improve the solution, but, nevertheless, it is imperative to have sampling times below 1 ns so the determination of the starting time or the cross-correlation peaks is clear and unequivocal.

Calculation of the Partial Discharge Site

For three-dimensional location, four antennas are necessary. Let us assume that the partial discharge site located in rectangular coordinates in (x_s, y_s, z_s) triggers a pulse in an unknown time t_0. Therefore, if the pulse wave front propagates spherically at velocity v, it eventually reaches antenna i located in (x_i, y_i, z_i) in time t_i. Times of flight t_i have to be calculated as accurately as possible to minimize the error in the location of the partial discharge site.

Then, the distance between the site and the antenna is $v(t_i - t_0)$ which is the radius of a sphere:

DOI 10.1007/978-981-4560-75-7_95-1

$$\begin{aligned} v^2(t_i - t_0)^2 &= (x_s - x_i)^2 + (y_s - y_i)^2 + (z_s - z_i)^2 \quad i = 1 \ldots 4 \\ R_i = v(t_i - t_0) &= \sqrt{(x_s - x_i)^2 + (y_s - y_i)^2 + (z_s - z_i)^2} \end{aligned} \tag{47}$$

The time t_0 and the position of the site are unknown; however, t_0 can be eliminated if the four implicit equations in Eq. 47 are subtracted in pairs. For the first two antennas:

$$v(t_1 - t_0) - v(t_2 - t_0) = v(t_1 - t_2) = vt_{12} \tag{48}$$

Then, $R_1 - R_2 = vt_{12}$. If this is repeated for all three pairs:

$$\begin{aligned} R_1 - R_2 - vt_{12} &= 0 \\ R_1 - R_3 - vt_{13} &= 0 \\ R_1 - R_4 - vt_{14} &= 0 \end{aligned} \tag{49}$$

Equation 49 represents a nonlinear non-convex problem with multiple local minima, depending on the positions on the source and the antennas. Recently, in mobile communications and wireless sensor networks (Xu et al. 2011), these equations have been treated as an optimization problem where the estimation of the source point gives a nonzero solution that has to be minimized, $R_i - R_j - vt_{ik} = \epsilon_e$. The technique to solve the equations converts a non-convex problem into a convex one through semi-definite relaxation so the location of the global minimum is easier.

Other methods that have been already used and proved to obtain (x_s, y_s, z_s) in partial discharge location are Newton–Raphson and, more recently, global search methods that include two types of evolutionary computation algorithms – stochastic optimization methods inspired in biologic processes, natural phenomena, or group behavior – such as genetic algorithms (GA) and particle swarm optimization (PSO). Though classical genetic algorithms are very efficient in searching the entire space and can rapidly locate the region in which the global optimum exists, they are relatively poor and take long times in finding the precise local optimal solution in the region in which the algorithm converges. This is why PSO has a lead over GA in partial discharge location and it is detailed below.

Newton–Raphson The Newton–Raphson method is a popular iterative technique to solve nonlinear equations in the form $f(\xi) = 0$ or systems in the form $\mathbf{F}(\boldsymbol{\xi}) = 0$ with $\boldsymbol{\xi}$ a column matrix. The idea is that $\xi_1 = \xi_0 - f(\xi_0)/f'(\xi_0)$ is a better solution than ξ_0 using the first derivative term of a Taylor's series.

The iterative form is $\xi_{k+1} = \xi_k - f(\xi_k)/f'(\xi_k)$ or $\boldsymbol{\xi}_{\mathbf{k}+1} = \boldsymbol{\xi}_{\mathbf{k}} - \mathbf{J}^{-1}\mathbf{F}(\boldsymbol{\xi}_{\mathbf{k}})$ with:

$$\boldsymbol{\xi}_{\mathbf{k}} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_N \end{pmatrix}_k \quad \mathbf{F} = \begin{pmatrix} f_1(\xi_1, \xi_2, \ldots, \xi_N) \\ f_2(\xi_1, \xi_2, \ldots, \xi_N) \\ \vdots \\ f_N(\xi_1, \xi_2, \ldots, \xi_N) \end{pmatrix}_k \tag{50}$$

and the Jacobian matrix **J** defined as:

DOI 10.1007/978-981-4560-75-7_95-1

$$\mathbf{J} = \frac{\partial(f_1, f_2, \ldots, f_N)}{\partial f_1(\xi_1, \xi_2, \ldots, \xi_N)} = \begin{pmatrix} \frac{\partial f_1}{\partial \xi_1} & \frac{\partial f_1}{\partial \xi_2} & \cdots & \frac{\partial f_1}{\partial \xi_N} \\ \frac{\partial f_2}{\partial \xi_1} & \frac{\partial f_2}{\partial \xi_2} & \cdots & \frac{\partial f_2}{\partial \xi_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial \xi_1} & \frac{\partial f_N}{\partial \xi_2} & \cdots & \frac{\partial f_N}{\partial \xi_N} \end{pmatrix} \tag{51}$$

The changes in the solution are given by $\Delta = \mathbf{J}^{-1}\mathbf{F}(\boldsymbol{\xi}_\mathbf{k})$ so the iteration can stop when $\max_k|\Delta| \leq \epsilon_e$ being ϵ_e a convergence error.

Equation 49 can be rewritten as:

$$\mathbf{F} = \begin{pmatrix} f_1(x_s, y_s, z_s) \\ f_2(x_s, y_s, z_s) \\ f_3(x_s, y_s, z_s) \end{pmatrix} = \begin{pmatrix} R_1 - R_2 - vt_{12} \\ R_1 - R_3 - vt_{13} \\ R_1 - R_4 - vt_{14} \end{pmatrix} = 0 \tag{52}$$

so,

$$\mathbf{J} = \begin{pmatrix} \frac{x_{s,k} - x_1}{g_1} - \frac{x_{s,k} - x_2}{g_2} & \frac{y_{s,k} - x_1}{g_1} - \frac{y_{s,k} - x_2}{g_2} & \frac{z_{s,k} - x_1}{g_1} - \frac{z_{s,k} - x_2}{g_2} \\ \frac{x_{s,k} - x_1}{g_1} - \frac{x_{s,k} - x_3}{g_3} & \frac{y_{s,k} - x_1}{g_1} - \frac{y_{s,k} - x_3}{g_3} & \frac{z_{s,k} - x_1}{g_1} - \frac{z_{s,k} - x_3}{g_3} \\ \frac{x_{s,k} - x_1}{g_1} - \frac{x_{s,k} - x_4}{g_4} & \frac{y_{s,k} - x_1}{g_1} - \frac{y_{s,k} - x_4}{g_4} & \frac{z_{s,k} - x_1}{g_1} - \frac{z_{s,k} - x_4}{g_4} \end{pmatrix} \tag{53}$$

and then:

$$\begin{pmatrix} x_{s,k+1} \\ y_{s,k+1} \\ z_{s,k+1} \end{pmatrix} = \begin{pmatrix} x_{s,k} - \Delta_x \\ y_{s,k} - \Delta_y \\ z_{s,k} - \Delta_z \end{pmatrix} \tag{54}$$

A single initial point has to be chosen to start the iteration. If it is close to the solution, the convergence will be fast and accurate; otherwise, the algorithm follows simple paths and it is prone to reach local minima, inducing great errors in the location. Under other circumstances, it may not be convergent at all. In these cases, changing the positions of the antennas can help to find a solution.

Particle swarm optimization The idea behind the algorithm is to exploit analogues of social interaction rather than individual cognitive abilities (Poli et al. 2007). In PSO, a certain number of individuals or particles are set to find the best fit to a problem using their own experience and the experience of one or more members of the community. Their position and velocity are changed according to their relative position to their individual best fit $\mathbf{p_i}$ and the group best fit $\mathbf{p_g}$. Then, every individual stochastically oscillates around $\mathbf{p_i}$ and $\mathbf{p_g}$ locations that are updated in every iteration. The movement of particles resembles the behavior of a flock of birds or fish searching for food in which every individual has information about their position to the target and information about the swarm movement.

The algorithm starts with a random deployment of particle positions and velocities in the space. The position of particle j in iteration k is $\mathbf{p_{k,j}} = \left(x_{s,k}, y_{s,k}, z_{s,k}\right)_j$. Then, every particle evaluates the fitness function shown in Eq. 52. If the particle current position gives a result closer to 0 than its individual best, $\mathbf{p_i}$ is updated with the new position, and if the current swarm best is closer to 0 than the previous best, $\mathbf{p_g}$ is

DOI 10.1007/978-981-4560-75-7_95-1

also updated. Then the velocity and positions of all particles are changed according to the following equations:

$$\begin{aligned} \mathbf{v_{k+1,j}} &= \mathbf{v_{k,j}} + \mathbf{U}(\mathbf{0}, \phi_1) \ \left(\mathbf{p_i} - \mathbf{p_{k,j}}\right) + \mathbf{U}(\mathbf{0}, \phi_2) \ \left(\mathbf{p_g} - \mathbf{p_{k,j}}\right) \\ \mathbf{p_{k+1,j}} &= \mathbf{p_{k,j}} + \mathbf{v_{k+1,j}} \end{aligned} \tag{55}$$

where $\mathbf{U}(\mathbf{0}, \phi)$ is a three-dimension vector of random numbers uniformly distributed in $[0, \phi]$ that is generated in every iteration to change the velocity of the particle in all directions (Poli et al. 2007). Sometimes an inertia factor ω is added, multiplying the velocity $\mathbf{v_{k,j}}$ in Eq. 55.

Summary

The great advantage of electromagnetic sensors in the measurement of partial discharges is the capability of detecting the pulses without galvanic connections to high-voltage equipment. All their applications are derived from this characteristic. While other conventional methods need the disconnection of the asset, partial discharge measurements with antennas can be done online, reducing considerably maintenance times and costs.

The first time an antenna was used to measure partial discharges was in gas-insulated substations. In these substations, all high-voltage equipment is insulated with sulfur hexafluoride, so all partial discharges occurring inside the containers would only be detected in the output cables where inductive sensors can be placed. However, connecting antennas to dielectric windows in several sections of the substation can locate the site with insulation problems.

Moreover, partial discharges inside oil-insulated power transformers can only be measured with sensors inside the tank because the inductive nature of the coils prevents the pulses from being detected with systems connected to the bushings. In these cases, antennas connected to dielectric windows are the best option compared to other methods such as acoustic detection. In this configuration, not only are the discharges detected; eventually they can also be localized inside the tank and point out the winding where they are occurring. In this case, several antennas should be deployed inside the transformer and connected to a high-speed acquisition system to carefully obtain the time differences of arrival of the pulses to the detectors. However, this method is still far from being massively applied due to the difficulties of dealing with reflections inside the tank and windings and the attenuation of the pulses when they pass through windings and core.

Currently there are several manufacturers dedicated to the measurement of partial discharges in air-insulated substations in two areas: detection and localization. The detection is basically done by walking around the substation with a spectrum analyzer and detecting anomalies in determined frequency bands where the discharges are emitting. The localization is done with several antennas mounted on vehicle and in the substation that are moved around, obtaining time signals to apply any localization method.

Unfortunately, the use of antennas has also important drawbacks. Though the sensors are inexpensive, the cost of the acquisition system can be sometimes prohibitively high if a detailed signal in the time or frequency domains is necessary such as in the case of localization or detection with spectrum analyzers, respectively. On the contrary, if the user only wants to detect pulses, a simple ultrahigh-frequency pulse detector can be used.

Finally, nowadays nobody has claimed that the identification of the type of partial discharge with antennas is possible. The only way to map the pulses in a phase-resolved partial discharge pattern is with a

DOI 10.1007/978-981-4560-75-7_95-1

synchronizing signal to the network frequency, so it is necessary to disconnect the equipment and connect a voltage divider, loosing the main advantage of antennas, namely, the galvanic insulation. The frequency content of the signals acquired with antennas seems to be related with the type of partial discharge, but nobody can claim that it is possible to do an effective identification in real environments.

The standard IEC62478 *High-voltage test techniques – measurement of partial discharges by electromagnetic and acoustic methods* is expected to see the light of day in the next years. In the meantime, manufacturers and utilities are working together to outline the specifications of detection systems based on antennas that can perform best in gas-insulated substations and transformers.

Cross-References

- ▶ Antenna Measurement/Setups-Introduction
- ▶ Antennas in High-power High-voltage Systems & Electromagnetic Pulse Systems
- ▶ Broadband and Multiband Planar Antennas
- ▶ EMI/EMC Chamber Design, Measurement and Instrument
- ▶ Loop Antennas
- ▶ Low-Profile Antennas
- ▶ Microstrip Patch Antennas
- ▶ Near-Field Antenna Measurement Techniques
- ▶ Omni-Directional Antennas
- ▶ Small Antennas (PIFA/PILA/Loading Antenna/etc)
- ▶ Spiral, Helical and Rod Antennas
- ▶ Taped Slot Antennas
- ▶ Ultra-Wideband Antennas

References

A2.27, Cigre W (2008) Technical Brochure 343

Albarracin R, Robles G, Martinez-Tarifa J, Ardila-Rey J (2015) Separation of sources in radiofrequency measurements of partial discharges using time-power ratio maps. ISA Trans http://dx.doi.org/10.1016/j.isatra.2015.04.006.

Albarracín-Sánchez R (2014) Measurement of partial discharges in radiofrequency (Spanish). Doctoral thesis, Universidad Carlos III de Madrid, Spain

Argüeso M, Robles G, Sanz J (2005) Implementation of a Rogowski coil for the measurement of partial discharges. Rev Sci Instrum 76(6):65107–65113

Ari N, Blumer W (1987) Transient electromagnetic fields due to switching operations in electric power systems. IEEE Trans Electromagn Compat EMC-29(3):233–237

ASTM1868 (2013) Standard test method for detection and measurement of partial discharge (Corona) pulses in evaluation of insulation systems ASTM International, West Conshohocken, PA http://www.astm.org/Standards/D1868.htm

Ayob MF, Ramli NH (2014) Meandered with parasite patch antenna for partial discharge detection applications. Int J Electron Comput Commun Technol 4(2):18–21

Balanis CA (2012) Advanced engineering electromagnetics, 2nd edn. Wiley, New York

Bårmann P, Kröll S, Sunesson A (1996) Spectroscopic measurements of streamer filaments in electric breakdown in a dielectric liquid. J Phys D Appl Phys 29(5):1188

DOI 10.1007/978-981-4560-75-7_95-1

Bartnikas R (2002) Partial discharges. Their mechanism, detection and measurement. IEEE Trans Dielectr Electr Insul 9(5):763–808

Bartnikas R, McMahon E, ASTM (1979) Corona measurement and interpretation. ASTM special technical publication. ASTM Committee D-9 on Electrical Insulating Materials

Boggs S (1990) Partial discharge. III. Cavity-induced PD in solid dielectrics. IEEE Electr Insul Mag 6(6):11–16

Boggs S, Stone G (1982) Fundamental limitations in the measurement of corona and partial discharge. IEEE Trans Electr Insul EI-17(2):143–150

Cai Z, Wilkinson RH, Bojovschi A (2014) Partial discharge detection approach for a medium voltage switchgear system. Prog Electromagn Res 144:221–228

Cavallini A, Montanari G, Contin A, Pulletti F (2003) A new approach to the diagnosis of solid insulation systems based on PD signal inference. IEEE Electr Insul Mag 19(2):23–30

CEN, CENELEC, ETSI (2011) Standards for smart grids. Final report. Technical report, Joint Working Group on Standards for Smart Grids

Chen H-C, Chen B-Y, Kuo C-C, Chao M-H (2011) UHF micro-strip antenna design for partial discharge detection of gas insulated switch. In: Lightning (APL), 2011 7th Asia-Pacific international conference on, Chengdu, pp 590–594

Choi J-G, Yi S-H, Kim K-H (2004) Development of a novel tapered balun for the UWB UHF coupler. In: Power modulator symposium, 2004 and 2004 high-voltage workshop. Conference record of the twenty-sixth international, San Francisco, CA, USA, pp 493–496

Cleary GP, Judd M (2006) UHF and current pulse measurements of partial discharge activity in mineral oil. IEE Proc Sci Meas Technol 153(2):47–54

Contin A, Montanari G, Ferraro C (2000) PD source recognition by weibull processing of pulse height distributions. IEEE Trans Dielectr Electr Insul 7(1):48–58

Contin A, Cavallini A, Montanari G, Pasini G, Puletti F (2002) Digital detection and fuzzy classification of partial discharge signals. IEEE Trans Dielectr Electr Insul 9(3):335–348

de Kock N, Herbst I (1998) Partial-discharge measuring device. US Patent 5,804,972

Densley J (2001) Ageing mechanisms and diagnostics for power cables. An overview. IEEE Electr Insul Mag 17(1):14–22

Dubickas V, Edin H (2004) Couplers for online time domain reflectometry diagnostics of power cables. In: Electrical insulation and dielectric phenomena, 2004. CEIDP'04. 2004 annual report conference on, Boulder, CO, USA, pp 210–214

Eager G, Bader G (1967) Discharge detection in extruded polyethylene insulated power cables. IEEE Trans Power Apparatus Syst PAS-86(1):10–34

Eur (2010) Smartgrids. Strategic deployment document for europeans electricity networks of the future. Technical report final, European Technology Platform

Fabiani D, Cavallini A, Montanari G (2008) A UHF technique for advanced PD measurements on inverter-fed motors. IEEE Trans Power Electron 23(5):2546–2556

Fernando S, Rowe W, Wong KL (2012) Long wire antenna-like behavior of uninsulated overhead distribution cables. IEEE Trans Power Deliv 27(3):1116–1123

Fukasawa T, Miyashita H, Itoh T (2014) Partial discharge sensor. US Patent App. 13/881,551

Gill P (2008) Electrical power equipment maintenance and testing, Power Engineering (Willis), 2nd edn. Taylor & Francis

González-Posadas V, Segovia-Vargas D, Rajo-Iglesias E, Vázquez-Roy J, Martín-Pascual C (2006) Approximate analysis of short circuited ring patch antenna working at tm01 mode. IEEE Trans Antennas Propag 54(6):1875–1879

DOI 10.1007/978-981-4560-75-7_95-1

Guru B, Hiziroglu H (2004) Electromagnetic field theory fundamentals. Cambridge University Press, Cambridge UK

Hampton B, Meats RJ (1988) Diagnostic measurements at UHF in gas insulated substations. IEE Proc C Gener Transm Distrib 135(2):137–145

Han Y, Song YH (2003) Condition monitoring techniques for electrical equipment. A literature survey. IEEE Trans Power Deliv 18(1):4–13

Hauschild W, Lemke E (2014) High-voltage test and measuring techniques. Springer, Heidelberg

Helstrom C (1961) The spectrum of corona noise near a power transmission line. Power apparatus and systems, Part III. Trans Am Inst Electr Eng 80(3):831–835

Herbst I, De K (1996) A partial discharge sensing device for a gas insulated apparatus. EP Patent App. EP19,940,113,728

Hudon C, Bartnikas R, Wertheimer M (1991) Analysis of degradation products on epoxy surfaces subjected to pulse and glow type discharges. In: Electrical insulation and dielectric phenomena, 1991. CEIDP. 1991 annual report. Conference on, pp 237–243

IEC (2010) Smart grid standardization roadmap, edition 1.0. Technical report, Internation Electrotechnical Commission. SMB Smart Grid Strategic Group (SG3)

International Electrotechnical Commission Geneva Switzerland IEC 60507:2013 Artificial pollution tests on high-voltage ceramic and glass insulators to be used on a.c. systems

IEC60270 (2000) High-voltage test techniques – partial discharge measurements, 3a edn

IEC62478 (Not published yet) High-voltage test techniques – measurement of partial discharges by electromagnetic and acoustic methods. Proposed horizontal standard, 1 edn, Draft standard, not published yet.

Ishak A, Judd M, Siew W, Baker PC (2012) Evaluation of FDTD modelling as a tool for predicting the response of UHF partial discharge sensors. In: Electrical insulation (ISEI), conference record of the 2012 I.E. international symposium on, San Juan, PR, pp 502–506

Jardine AK, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Process 20(7):1483–1510

Jin Z, Sun C, Cheng C, Li J (2008) Two types of compact UHF antennas for partial discharge measurement. In: High voltage engineering and application, 2008. ICHVE 2008. International conference on, Chongqing, pp 616–620

Judd M, Farish O (1998) A pulsed GTEM system for UHF sensor calibration. IEEE Trans Instrum Meas 47(4):875–880

Judd M, Farish O, Hampton B (1995) Broadband couplers for UHF detection of partial discharge in gas-insulated substations. IEE Proc Sci Meas Technol 142(3):237–243

Judd M, Farish O, Hampton B (1996) The excitation of UHF signals by partial discharges in GIS. IEEE Trans Dielectr Electr Insul 3(2):213–228

Judd M, Farish O, Pearson J, Breckenridge T, Pryor BM (2000) Power transformer monitoring using UHF sensors: installation and testing. In: Electrical insulation, 2000. Conference record of the 2000 I.-E. international symposium on, Anaheim, CA, USA, pp 373–376

Judd M, Farish O, Pearson J, Hampton B (2001) Dielectric windows for UHF partial discharge detection. IEEE Trans Dielectr Electr Insul 8(6):953–958

Judd M, Yang L, Hunter I (2005) Partial discharge monitoring of power transformers using UHF sensors. Part I: sensors and signal interpretation. IEEE Electr Insul Mag 21(2):5–14

Kaneko S, Okabe S, Yoshimura M, Muto H, Nishida C, Kamei M (2009) Detecting characteristics of various type antennas on partial discharge electromagnetic wave radiating through insulating spacer in gas insulated switchgear. IEEE Trans Dielectr Electr Insul 16(5):1462–1472

Karst J, Groh C, Garbe H (2002) Calculable field generation using TEM cells applied to the calibration of a novel E-field probe. IEEE Trans Electromagn Compat 44(1):59–71

Kato T, Oshita Y (2000) Partial discharge detector of gas-insulated apparatus. EP Patent App. EP19,970,922,123

Katsuta G, Toya A, Muraoka K, Endoh T, Sekii Y, Ikeda C (1992) Development of a method of partial discharge detection in extra-high voltage cross-linked polyethylene insulated cable lines. IEEE Trans Power Deliv 7(3):1068–1079

Kiameh P (2003) Electrical equipment handbook: troubleshooting and maintenance. McGraw-Hill, New York

Klempner G, Kerszenbaum I (2004) Operation and maintenance of large turbo-generators, IEEE Press series on power engineering. Wiley, New York

Knapp C, Carter GC (1976) The generalized correlation method for estimation of time delay. IEEE Trans Acoust Speech Signal Process 24(4):320–327

Kreuger F (1989) Partial discharge detection in high-voltage equipment. Butterworths, London

Kreuger F, Wezelenburg M, Wiemer AG, Sonneveld WA (1993) Partial discharge. XVIII. Errors in the location of partial discharges in high voltage solid dielectric cables. IEEE Electr Insul Mag 9(6):15–22

Kurtz M, Stone G, Freeman D, Mulhall V, Lonseth P (1980) Diagnostic testing of generator insulation without service interruption. CIGRE, Paris

Li J, Jiang T, Wang C, Cheng C (2012) Optimization of UHF Hilbert antenna for partial discharge detection of transformers. IEEE Trans Antennas Propag 60(5):2536–2540

Li M, Guo C, Peng Z (2014) Design of meander antenna for UHF partial discharge detection of transformers. Sens Transducers 171(5):232–238

López-Roldán J, Tang T, Gaskin M (2008) Optimisation of a sensor for onsite detection of partial discharges in power transformers by the UHF method. IEEE Trans Dielectr Electr Insul 15(6):1634–1639

Markalous S, Tenbohlen S, Feser K (2008) Detection and location of partial discharges in power transformers using acoustic and electromagnetic signals. IEEE Trans Dielectr Electr Insul 15(6):1576–1583

Mason J (1951) The deterioration and breakdown of dielectrics resulting from internal discharges. Proc IEE Part I General 98(109):44–59

Matsumoto S, Shiroi N, Suzuki I, Akiba T, Sobataka M, Kasajima K, Shibuya Y, Murooka Y, Kawamura T (2008) Three-axis loop antenna for the detection of partial discharge signal. In: Electrical insulating materials, 2008. (ISEIM 2008). International symposium on, pp 28–31

McAllister I, Crichton G (1998) Partial discharge transients: the field theoretical approach. In: Christophorou L, Olthoff J (eds) Gaseous dielectrics VIII, Springer, Springer US. New York. pp 239–252

Mobley R (2011) Maintenance fundamentals. Plant engineering. Elsevier Science, Oxford UK

Moore P, Portugues I, Glover I (2004) Remote diagnosis of overhead line insulation defects. In: Power Engineering Society General Meeting, 2004. IEEE, vol 2, Denver CO, USA. pp 1831–1835

Moore P, Portugués I, Glover I (2005) Radiometric location of partial discharge sources on energized high-voltage plant. IEEE Trans Power Deliv 20(3):2264–2272

Morshuis P (1993) Partial discharge mechanisms. Delft University Press, Netherlands

Morshuis P (2005) Degradation of solid dielectrics due to internal partial discharge: some thoughts on progress made and where to go now. IEEE Trans Dielectr Electr Insul 12(5):905–913

Niemeyer L (1993) The physics of partial discharges. In: Partial discharge, 1993., International conference on, Canterbury UK, pp 1–4

Niemeyer L (1995) A generalized approach to partial discharge modeling. IEEE Trans Dielectr Electr Insul 2(4):510–528

DOI 10.1007/978-981-4560-75-7_95-1

NIST (2012) Frame and roadmap for smart grid interoperability standards, release 2.0. Technical report 1108R2, National Institute of Standards and Technology

Ozaki T, Abe K, Umemura T (1992) Partial discharge detection using ferrite antenna. In: Electrical insulation, conference record of the 1992 I.E. international symposium on, Baltimore, MD USA pp 371–374

Pearson J, Hampton B, Sellars AG (1991) A continuous UHF monitor for gas-insulated substations. IEEE Trans Electr Insul 26(3):469–478

Pedersen A, Crichton GC, McAllister IW (1991) The theory and measurement of partial discharge transients. IEEE Trans Electr Insul 26(3):487–497

Petchphung P, Leelajindakrairerk M, Pattanadech N, Yutthagowith P, Aunchaleevarapan K (2007) The study of disc monopole antenna for partial discharge measurement. In: Power engineering conference, 2007. IPEC 2007. International, Singapore, pp 1292–1296

Pinpart T, Judd M (2009) Experimental comparison of UHF sensor types for PD location applications. In: Electrical insulation conference, 2009. EIC 2009. IEEE, Montreal, QC, pp 26–30

Poli R, Kennedy J, Blackwell T (2007) Particle swarm optimization. Swarm Intell 1(1):33–57

Portugués I, Moore P, Glover I, Johnstone C, McKosky R, Goff M, van der Zel L (2009) RF-Based partial discharge early warning system for air-insulated substations. IEEE Trans Power Deliv 24(1):20–29

Raja K (2007) Antenna for detection of partial discharges in a chamber of an electrical instrument. US Patent 7,208,958

Reid AJ, Judd MD, Stewart BG, Fouracre RA (2006) Partial discharge current pulses in SF6 and the effect of superposition of their radiometric measurement. J Phys D Appl Phys 39(19):4167

Reid A, Stewart M, Judd M (2009) FDTD modeling of UHF partial discharge sensor response. In: Sustainable power generation and supply, 2009. SUPERGEN '09. International conference on, Nanjing pp 1–4

Reid A, Judd M, Fouracre R, Stewart BG, Hepburn DM (2011) Simultaneous measurement of partial discharges using IEC60270 and radio-frequency techniques. IEEE Trans Dielectr Electr Insul 18(2):444–455

Reig C, Ávila Navarro E (2014) Printed antennas for sensor applications: a review. IEEE Sens J 14(8):2406–2418

Reinhardt H (1985) Measuring arrangement for detecting partial discharge within metal-shielded, pressure gas-isolated high-voltage commutator arrangements. EP Patent App. EP19,840,730,078

Robles G, Albarracín R, Martínez-Tarifa J (2013a) Shielding effect of power transformers tanks in the ultra-high-frequency detection of partial discharges. IEEE Trans Dielectr Electr Insul 20(2):678–684

Robles G, Albarracín R, Vázquez-Roy J, Rajo-Iglesias E, Martínez-Tarifa J, Rojas-Moreno M, Sánchez-Fernández M, Ardila-Rey J (2013b) On the use of Vivaldi antennas in the detection of partial discharges. In: Solid dielectrics (ICSD), 2013 I.E. international conference on, Bologna Italy, pp 302–305

Robles G, Sanchez-Fernandez M, Albarracin Sanchez R, Rojas-Moreno M, Rajo-Iglesias E, Martinez-Tarifa J (2013c) Antenna parametrization for the detection of partial discharges. IEEE Trans Instrum Meas 62(5):932–941

Sedding H, Stone G, Campbell S (1991) Measurement of partial discharges on operating turbine generators with a novel wideband coupler. In: Properties and applications of dielectric materials, 1991., proceedings of the 3rd international conference on, vol 2, Tokyo, pp 773–776

Shan Q, Glover I, Moore P, Portugués I, Judd M, Rutherford R, Watson R (2008) TEM horn antenna for detection of impulsive noise. In: Electromagnetic compatibility – EMC Europe, 2008 international symposium on, Hamburg, pp 1–6

Shan Q, Glover I, Atkinson R, Bhatti S, Portugués I, Moore P, Rutherford R, de Fátima Queiroz Vieira M, Lima A, de Souza B (2011) Estimation of impulsive noise in an electricity substation. IEEE Trans Electromagn Compat 53(3):653–663

Shibuya Y, Matsumoto S, Tanaka M, Muto H, Kaneda Y (2010) Electromagnetic waves from partial discharges and their detection using patch antenna. IEEE Trans Dielectr Electr Insul 17(3):862–871

Steiner JP, Reynolds P, Weeks W (1992) Estimating the location of partial discharges in cables. IEEE Trans Electr Insul 27(1):44–59

Stone G (2005) Partial discharge diagnostics and electrical equipment insulation condition assessment. IEEE Trans Dielectr Electr Insul 12(5):891–904

Stone G, Sedding H, Fujimoto N, Braun J-M (1992) Practical implementation of ultrawideband partial discharge detectors. IEEE Trans Electr Insul 27(1):70–81

Tanaka T (1986) Internal partial discharge and material degradation. IEEE Trans Electr Insul EI-21(6):899–905

Tang Z, Li C, Cheng X, Wang W, Li J, Li J (2006) Partial discharge location in power transformers using wideband RF detection. IEEE Trans Dielectr Electr Insul 13(6):1193–1199

Tenbohlen S, Denissov D, Hoek S, Markalous S (2008) Partial discharge measurement in the ultra high frequency (UHF) range. IEEE Trans Dielectr Electr Insul 15(6):1544–1552

Vázquez-Roy J, Inclan-Sánchez L, Quevedo-Teruel O, Rajo-Iglesias E (2010) Wideband zero-mode circular patch antenna. In: Antennas and propagation (EuCAP), 2010 proceedings of the fourth European conference on, Barcelona, Spain, pp 1–3

Wang G, Hao Y, Li J (2001) Study on the ultrahigh-frequency sensors for PD detection in power transformer. In: Electrical insulating materials, 200L (ISEIM 2001). Proceedings of 2001 international symposium on, Himeji, pp 793–796

Wanninger G (1997) Apparent charge measurement in GIS by modern diagnostic methods. Eur Trans Electr Power 7(4):251–255

Xing Zhang X, Chen Y, Tang J-Z (2013) A complex structural horn antenna for partial discharge detection. Przeglad Elektrotechniczny 2013(1b):51–55

Xu E, Ding Z, Dasgupta S (2011) Source localization in wireless sensor networks from signal time-of-arrival measurements. IEEE Trans Signal Process 59(6):2887–2897

Yao C, Chen P, Huang C, Chen Y, Qiao P (2013) Study on the application of an ultra-high frequency fractal antenna to partial discharge detection in switchgears. Sensors 13(12):17362–17378

Zargari A, Blackburn TR (1996) Modified optical fibre sensor for PD detection in high-voltage power equipment. In: Electrical insulation, 1996., Conference record of the 1996 I.E. international symposium on, vol 1, Montreal, Que, pp 424–427

Zhang X, Han Y, Li W, Duan X (2014) A rectangular planar spiral antenna for GIS partial discharge detection. Int J Antennas Propag 44:7

Zheng B, Bojovschi A (2012) Electromagnetic sensing of partial discharge in air-insulated medium voltage switchgear. In: Progress in electromagnetics research symposium proceedings. Moscow, Russia

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_96-1

Antennas in Automobile Radar

Wolfgang Menzel*
Institute of Microwave Techniques, University of Ulm, Ulm, Germany

Abstract

Automobile radars are under investigation since the 1960s. The first operational systems are on the market since 1992 for buses and trucks and 1999 for passenger cars, both in the frequency range around 24 as well as 76.5 GHz; a new frequency band for medium- and short-range sensors from 77 to 81 GHz has been allocated recently in Europe. Requirements for the sensor antennas are high gain and low loss combined with small size and depth for vehicle integration. Great challenges are due to the millimeter-wave frequency range, and a great cost pressure for this commercial application determines design and fabrication. Consequently, planar antennas are dominating in the lower frequency range, while lens and reflector antennas had been the first choice at 76.5 GHz, partly in folded configurations. With increasing requirements toward a much more detailed view on the scenery in front or around the vehicle, multi-beam antennas or scanning antennas have been designed. For actual systems, digital beamforming with a number of integrated antennas is in use or under development, and also MIMO concepts will be exploited. With such development, antennas for automotive radar no longer can be considered as stand-alone devices, but will be part of an "imaging" system including multiple transmit/receive units and digital signal processing.

General antenna concepts, partly including system aspects, as well as several realized antenna and sensor configurations will be described in detail in this chapter.

Keywords

Automotive radar; Millimeter-wave radar; Radar antennas; Microstrip antenna; Grid antenna; Reflector antennas; Lens antennas; Slotted waveguide array; Dielectric antenna

Introduction

With increasing traffic density, an increasing interest has emerged for additional electronic vision to improve comfort and safety, including video cameras, infrared sensors, laser, or radar. Although radar does not provide optical resolution, it has the great advantage of being mostly insensitive to fog or rain, it works independently on ambient light, and it directly provides speed based on the Doppler effect. Consequently, radar was considered as automotive sensor already back to 1959, when Cadillac presented two radar sensors in its experimental car "Cadillac Cyclone" (Cadillac 2013). No detailed information has been published, but the ogive-shaped radomes indicate military-type radars. In the 1970s, a number of companies again investigated automotive radars, starting at 10, 16, and later on 35 GHz (Meinel and Dickmann 2013) (Fig. 1). Later on, other systems were considered at 60 GHz (mostly in Japan, using the license-free ISM band) or 94 GHz based on military

*Email: wolfgang.menzel@uni-ulm.de

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_96-1

Fig. 1 Vehicle with a first 35 GHz radar (separate parabolic reflector antennas for transmit and receive, photograph TELEFUNKEN)

Fig. 2 VORAD radar in front of a bus (own photograph)

developments. The first commercial radar for collision avoidance was introduced in 1992 in the USA by VORAD for buses and trucks, working at 24 GHz (Woll 1995) (Fig. 2). It helped in reducing crashes, but was not really accepted by the drivers who felt too much controlled by this system. In 1998/1999, the first commercial radar in a passenger car was introduced by Mercedes in Germany, working in the 76–77 GHz frequency band, followed by other car manufacturers. These first radar sensors were used mostly in comfort functions, supporting an automatic speed control by keeping a safe distance to the preceding vehicle (adaptive cruise control, ACC).

The aim for further developments during the last years has been to make radar available for safety functions like forward collision alert, rear traffic crossing alert, blind spot detection, or side-impact warning; the latest developments include emergency breaking and autonomous stop-and-go driving (Mercedes E- and S-class). According to such different functions, sensors have to look into all directions around the vehicle with differently defined scanning and distance relations. Rear traffic crossing alert and blind spot detection, for example, operate in the near range and simply give information whether one or more targets are present in a certain area, whereas ACC and forward collision alert have to scan the entire traffic scene not only in the near range but also – dependent on the maximum speed of the object – in the mid- and far range. Typical range designations are low-range (LRR, up to a few 10 m), medium-range (MRR, about 40–100 m), and long-range (LRR, up to 200 m or even 250 m) radars. As a consequence, a much more flexible and wider field of view is required compared to the first applications, resulting in some kind of "imaging" of the scenery in

DOI 10.1007/978-981-4560-75-7_96-1

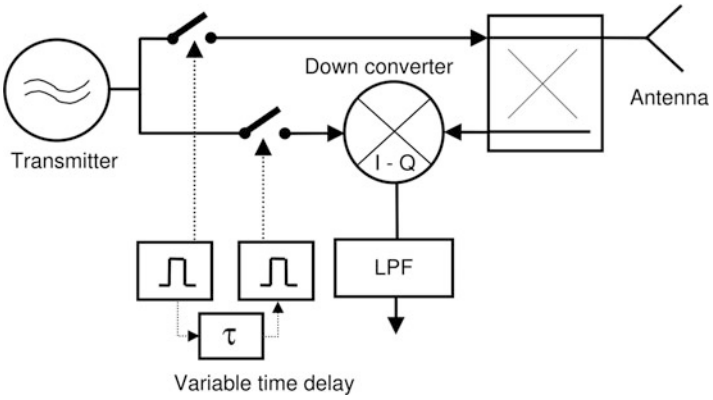


Fig. 3 Basic block diagram of an automotive pulse radar

front of the sensor. This, of course, has to be mapped to the antenna concepts, partly in combination with modified overall sensor arrangements, including multi-beam antennas, scanning antennas, switched antenna concepts, and beamforming approaches with multiple transmit and receive antennas. Recent trends show a change from the scanning or multi-beam antenna principle to the digital beamforming with increasing functional transfer to digital signal processing. As a consequence, automotive antenna concepts must not be considered independently from the overall system concept, but increasingly, the overall range and way of "viewing" the interesting range of angles is determined by an optimal combination of antenna elements, RF front end, and signal processing. The actual field of view of the sensor therefore is determined by the antenna or antenna system diagram but also by the required angle-dependent range (with possibly lower gain in directions away from the main lobe), the minimum radar cross section, and the required minimum signal-to-noise ratio.

Furthermore, when designing automotive systems including antennas, the engineer faces the task that all components must be suitable for mass production, operate in a temperature range from −40 °C to 85 °C or even 105 °C, and be shockproof, and the entire arrangement has to be compact in size. The implications of all these on the choice of materials and even the performance should not be underestimated.

Present frequencies allocations for short range are in the 24.05–24.25 GHz frequency range (ISM band), a 4 GHz wide band around 24 GHz, but with different allocations and restrictions in different countries, and the 76–77 GHz range for the far range. Recently, in the EU the band from 77 to 81 GHz has been allocated for short range and mid-range; efforts are undertaken to make this frequency range available for automotive radar worldwide. In the near future, almost all long- or mid-range antennas will operate in the 76–81 GHz range.

Key Properties of Automotive Radar

Automotive radar sensors provide information about other vehicles, pedestrians, and the road environment by measurement of the flight time of the electromagnetic wave from the radar to a target and back; relative speed is determined, extracting the Doppler frequency shift, and angular information is given by the antenna characteristics. Automotive sensors typically use either a pulse modulation (Gresham et al. 2004) or frequency-modulated constant wave (FMCW) signals with sawtooth or triangular frequency modulation (Hymans and Lait 1960). Most of the automotive pulse sensors operate in a sampling mode with some similarity to a sampling oscilloscope (Fig. 3). The received

DOI 10.1007/978-981-4560-75-7_96-1

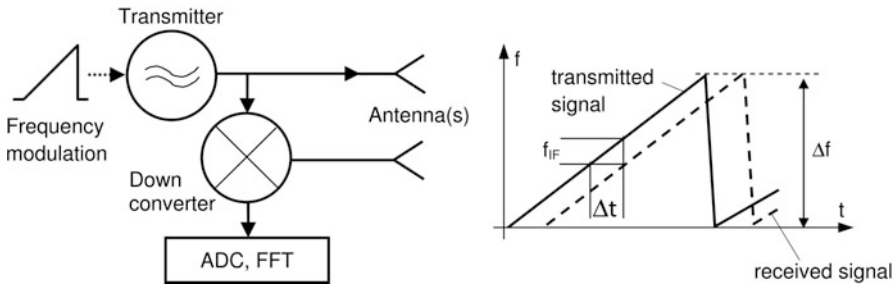


Fig. 4 Basic block diagram and time-frequency diagram of an FMCW radar

pulses are down converted with a pulse delayed by a time τ; an intermediate-frequency (IF) signal occurs only if the time of flight of the pulse is identical to the delay. Together with a low-pass filter (LPF) for the IF signal, bandwidth at the receiver output can be considerably reduced, and a further digital processing with moderate speed is possible. For the FMCW sensor, a signal with a linear frequency ramp is transmitted and then received with the respective delay time (Fig. 4). Down conversion with the actual transmitter signal then results in an IF signal, the frequency of which is proportional to the target distance. A typical further processing is done via analog-digital converter (ADC) and fast Fourier transform (FFT). As transmit/receive diplexer, mostly 3 dB couplers are used to avoid complicated and expensive circulators (as indicated in Fig. 3), or even individual antenna elements are used for transmitting and receiving (as indicated in Fig. 4).

In general, signal bandwidth, either via pulse width or FMCW frequency deviation, determines the minimum range resolution ΔR (separation between two targets):

$$\Delta R = c_0/(2\Delta f) \tag{1}$$

with c_0 as speed of light and Δf the modulation bandwidth. Typical values of ΔR range from about 1 m for early systems to some cm for newer short-range broadband systems (up to some GHz of bandwidth).

While for the first radar sensors, the RF circuits of the front ends were built based on discrete devices like Gunn elements and Schottky diodes, monolithic microwave integrated circuits (MMIC) based on GaAs (Camiade et al. 2000) and silicon-germanium (SiGe) are available today. SiGe MMICs even provide several RF front ends and some basic digital circuits on a single chip (Winkler et al. 2008).

The angular or cross-range observation depends on the sensor application and is determined by its antenna arrangement. In an early stage, when automotive sensors were used solely in comfort functions supporting the automatic speed control by keeping a safe distance to the preceding vehicle, a few (switched) narrow antenna beams were sufficient to monitor the own and adjacent lanes on a highway. Today, for safety purposes both for highway and for dense urban traffic scenarios, much more sophisticated antenna and overall system concepts are required.

Types of Antennas for Automotive Radar

General Requirements and Properties of Automotive Antennas

In general, the beamwidth of antennas is related directly to the physical size of an antenna. A rule of thumb estimates the 3 dB beamwidth ΔΦ by

DOI 10.1007/978-981-4560-75-7_96-1

$$\Delta\Phi = 60^\circ \ \lambda/D \tag{2}$$

where λ is the free space wavelength and D the overall antenna dimension in the respective plane. Requirements for a long-range sensor to restrict the 3 dB beamwidth to a 3.5 m-wide highway lane at a distance of about 50 m result in a beamwidth of 4° in azimuth and an antenna width of 15 wavelengths. This immediately shows that millimeter-wave frequencies are required to be able to integrate a radar sensor into the front of a car. At 77 GHz, this results in an antenna width of 60–80 mm. In elevation, a radar sensor should avoid strongly illuminating any bridge or traffic signs above the street (typically in a height of at least 5 m); this results in similar beamwidth requirements. Antenna gain then is in the range of 30 dB. To also observe neighboring lanes or operating in curves, several beams or some moderate beam scanning is required. At 24 GHz, antenna size and beamwidth typically are larger. Mid- and near-range systems have wider beamwidths and lower gain but require a wider angle of view by scanning, multiple beams, or beamforming.

Antenna losses should be low to maintain transmit gain as high as possible and not to deteriorate the receiver noise figure too much. Sidelobe levels need to be low enough to avoid false detection of targets at larger angles.

Compared to typical targets, wavelengths of about 12 and 4 mm at 24 and 77 GHz, respectively, are generally small, so polarization does hardly have any influence on target reflections. Some additional information may be gained if also orthogonal polarization was received, but this would add the implications of dual-polarization antennas and doubling the receiver part of the radar. In some sensors, polarization is adjusted at 45°; in this case, direct radiation from a radar into that of an oncoming vehicle is reduced. Ground reflections, however, will mostly foil the effects of this measure.

The antenna depth is of concern in integrating the sensor into the vehicle. Finally, antenna fabrication and integration into the sensor should be easy, reproducible, and cost-effective. The overall sensor including the antenna has to withstand shock, vibration, and the required automotive temperature range as mentioned above.

Reflector and Lens Antennas

In first test as well as in the first series versions of automotive radars, standard types of antennas were used, including horn antennas and parabolic reflectors. For sensors used in a car, antenna depth is of great concern for integration into the vehicle surface as well. A prominent example is a lens antenna with short focal length (Binzer et al. 2007), included in the commercial LRR3 sensor by Bosch. Another way to reduce the antenna depth is by "folding" the ray path of reflector or lens antennas as presented in Fig. 5 (Gresham et al. 2001; Millitech 1994). In these antennas, a feed initially radiates against a polarizing filter, and the wave is reflected back toward a planar structure where the wave is reflected again, but with a polarization twisted by 90°.

Such polarization twisting can be achieved by structures as shown in Fig. 6. The incident wave is polarized at 45° with respect to the axes of the reflecting structure (grid or rectangular patches). For the analysis of the reflected wave, the field is decomposed into components parallel to the two axes. In the case of the grid, the component parallel to the grid wires is reflected directly, while the orthogonal component passes the grid and is reflected at the ground plane placed a quarter of a wavelength behind the grid, resulting in an overall detour of half a wavelength or 180°. Combining the outgoing two components again, a wave results with a polarization twisted by 90°. For the structure with rectangular patches, the dimensions of the patches are selected in such a way that the reflection phase angles for the two components parallel to the patch axes differ by 180°, resulting in

DOI 10.1007/978-981-4560-75-7_96-1

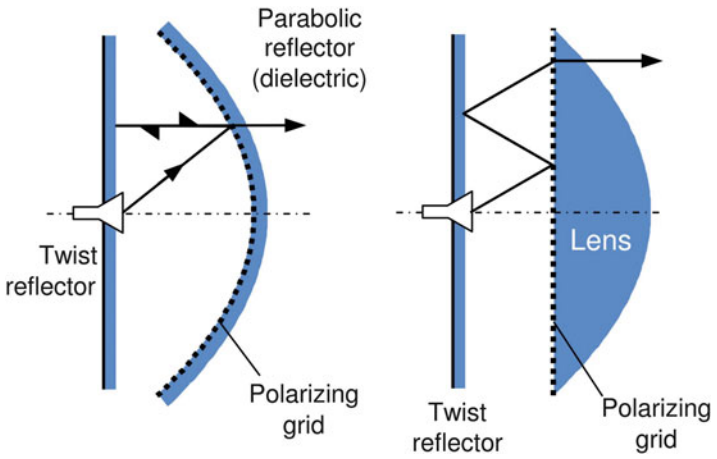


Fig. 5 Folded reflector (*left*) and lens antennas (*right*)

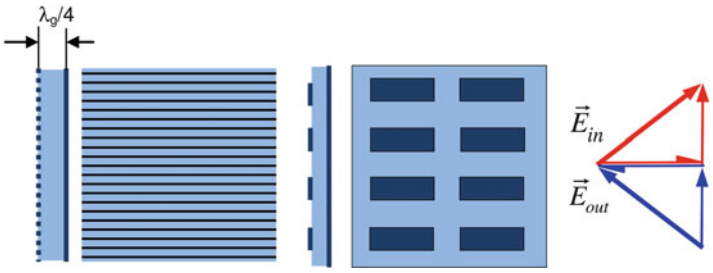


Fig. 6 Principles for polarization twisting (λ_g, wavelength in the dielectric)

the same 90° twisting of the polarization. In this case, the thickness of the substrate does not need to be a quarter wavelength thick.

In Menzel et al. (2002), the complex lens or curved reflector has been replaced by a planar printed structure (reflectarray) which, at the same time, can perform focusing of an incident spherical wave and twist the polarization (Fig. 7, top). The wave from a feed horn is incident on the polarizing grid (with its polarization parallel to the grid lines) and is reflected back to the planar array. Typical reflecting elements of such a reflectarray are rectangular patches which are oriented at an angle of 45° with respect to the incident polarization. Ninety degree twisting of the wave polarization is achieved in the same way as described above. Furthermore, combinations of the length and width of the patches can be selected to additionally give the required absolute phase shift to form an outgoing plane wave. Such an antenna with three beams (generated by three different feed locations) has been implemented into the second generation of Mercedes radars (Fig. 7, bottom), and this principle is also included in the third-generation sensor ARS 300 delivered by Continental (see the second application example in this chapter).

In general, lens and reflector antennas exhibit very low loss, advantageous in combination with millimeter-wave semiconductor elements for power generation of quite moderate level, but they require a non-negligible depth of a few cm; thus, purely planar antennas are gaining more and more interest.

DOI 10.1007/978-981-4560-75-7_96-1

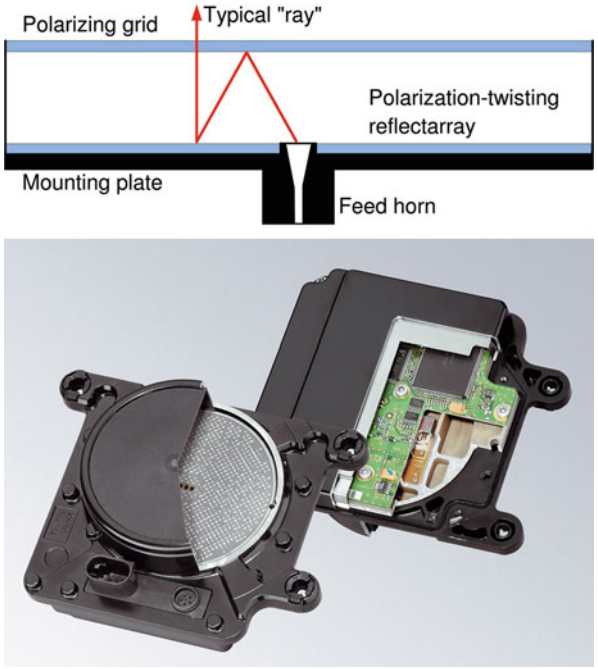


Fig. 7 Cross section and photograph (press photograph Continental ADC) of a folded reflectarray antenna (frequency 76.5 GHz, antenna diameter 90 mm, depth 23 mm)

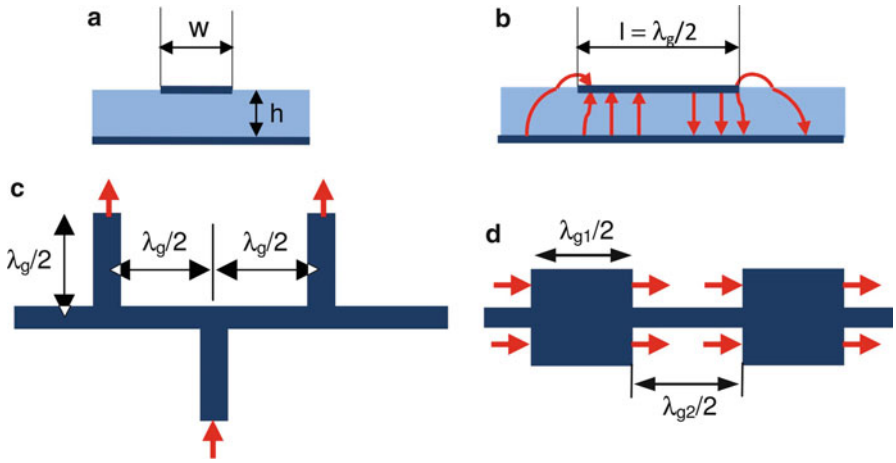


Fig. 8 Microstrip antennas: (**a**) microstrip cross section, (**b**) longitudinal cross section with an electric field of a half-wavelength resonator (patch antenna), (**c**) chain of radiating open-ended microstrip stubs, (**d**) chain of microstrip patches connected by half-wavelength lines

Planar Antennas

Planar antennas are found in various forms; the most common types are microstrip antennas (Carver and Mink 1981; James et al. 1981); for details see also the respective chapter in this book. Most often, half-wavelength resonators or dipoles (patch antennas) or open ends of microstrip stubs (Fig. 8) are employed as single antenna elements. Figure 8b depicts the longitudinal cross section of a radiating patch including the principal electric field distribution. In a very simplified model, the stray fields at the ends are the radiating sources – with this half-wavelength patch, the horizontal components of the stray field have the same direction and combine in phase for a broadside radiation

DOI 10.1007/978-981-4560-75-7_96-1

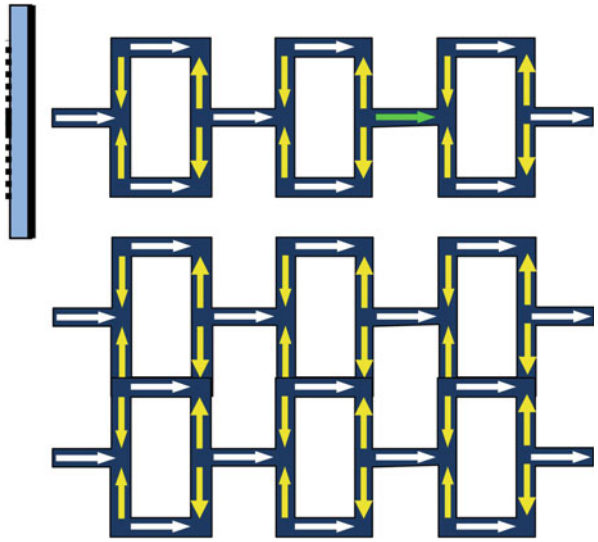


Fig. 9 Single- and double-row grid antenna

(the respective stray fields at the patch sides are in antiphase and cancel to a great extent). For narrower beams and higher gain, such elements can be combined in series and/or parallel arrangements (Fig. 8c, d) to form antenna arrays with the required overall antenna diagram. In Fig. 8c and d, the red arrows show the horizontal stray fields at the radiating edges and indicate the in-phase behavior, resulting in a broadside radiation.

Microstrip patches are resonators and typically exhibit bandwidths in the range of only a few percent with limited antenna efficiency, improving with lower dielectric constant of the substrate and increasing substrate thickness. The latter, however, leads to a higher degree of radiation by the feeding network as well as to the excitation of surface waves. Feed radiation may be mitigated by feeding the patch from the backside via a slot in the ground plane, but surface waves are more difficult to handle.

Single-patch antennas can be employed as feeding elements for a lens antenna (Binzer et al. (2007); see the first application example in this chapter); arrays of microstrip patches can be used directly as automotive antennas (Gresham et al. 2004; Winkler et al. 2008; Russell et al. 1997; Tokoro et al. 2003). With increasing geometrical dimensions of patch antenna arrays, feed network losses, especially for arrays with corporate feed networks, may pose a severe limit for antenna size. Series feeding of patches or open-ended stubs (Fig. 8) is based on connecting line segments of fixed length to ensure proper phase angle adjustment of each element; this however is valid only at a single frequency. Varying frequency, a progressive phase variation occurs from element to element, resulting in some degree of beam scan. For small array with relatively large beamwidth, this generally is of minor importance, but with an increasing length of a series-fed chain, beam scanning is no longer negligible.

Being predominantly resonating structures, patch antennas and radiating stubs exhibit limited bandwidth and increased losses; these effects pose a severe limitation for microstrip substrates with increased dielectric constant, e.g., for highly integrated circuits on low (or high)-temperature cofired ceramic (LTCC, HTCC) substrates. Already in the 1960s, a type of wire antenna was proposed, called grid antenna (Kraus 1964). Conti et al. (1981) proposed a microstrip version of such a grid array antenna. This array consists of a double meander microstrip structure as depicted in Fig. 9, top. All segments are about half a guided wavelength long; the currents (plotted white in Fig. 9) in the longitudinal segments have the same direction and contribute in phase to radiation. As there are no

DOI 10.1007/978-981-4560-75-7_96-1

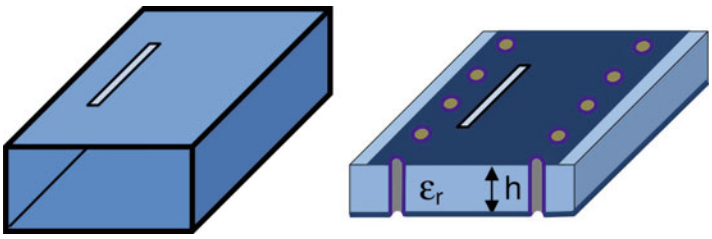


Fig. 10 Metal waveguide and substrate integrated waveguide (*SIW*) with radiating slots

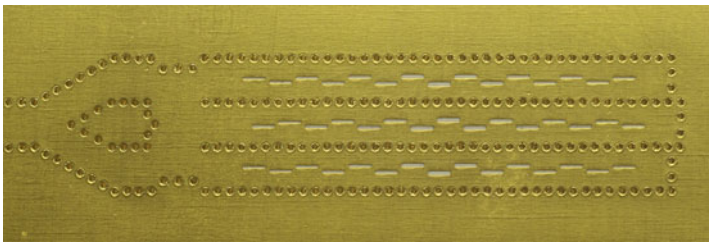


Fig. 11 Photograph of an SIW slotted array example (Massen et al. 2013, own photograph)

strong resonances, this type of antenna is more broadband, even for high-dielectric-constant substrates, and less sensitive to tolerances. Several segments can be aligned together in series, and even two or more chain can be combined in parallel (Fig. 9, bottom). Feeding can be done at one end or at suitable positions within the chain by coaxial or probe-type feeds from the substrate backside (Bauer and Menzel 2011, Bauer et al. 2013), via slot coupling through the ground plane, or even with a symmetric stripline from the side (opening one of the loops of the antenna (Frei et al. 2011)). Feeding a grid antenna chain from one end results in an antenna with slight scanning properties; feeding it in the center results in two shorter sections with reduced scanning into opposite directions and thus in a more broadband design.

The grid antenna has been implemented in different types of application, e.g., 60 GHz communication (Zhang and Zhang 2012), or in an experimental LTCC front end for a short-range automotive sensor (Bauer et al. 2013); this will be described in more detail later on in this chapter.

A well-known type of antenna is the waveguide slotted array (Fig. 10). It consists of a metal waveguide with longitudinal or transversal slots. Once again, radiation from a number of slots superimpose in phase to form the desired radiation pattern. One approach to make this antenna suitable for automotive radar is to machine a waveguide pattern into a metal plate and to glue a metal foil with the slots on top of this (Sakakibara et al. 2006). For an easy and low-cost fabrication, however, a synthetic or substrate integrated waveguide (SIW) can be realized within a dielectric substrate where the sidewalls are formed by rows of vias (Hirokawa and Ando 2000; Xu et al. 2009; Massen et al. 2013). Slotted arrays based on SIW can even be realized with LTCC substrates. As waveguide width is reduced due to the dielectric filling of this waveguide, only longitudinal slots are used. A critical item is tolerances for the dimensions of the slots due to the high frequency and the (high) dielectric constant; with a novel shape of the slots including bends, this problem can be overcome to some extent (Bauer and Menzel 2013a). In some applications, a resonant cavity is placed above each slot, either by an additional metal plate including the cavities (Sakakibara et al. 2006) or integrated within a (LTCC) multilayer substrate (Shino et al. 2005; Bauer and Menzel 2013b), giving a more broadband and better radiation performance.

DOI 10.1007/978-981-4560-75-7_96-1

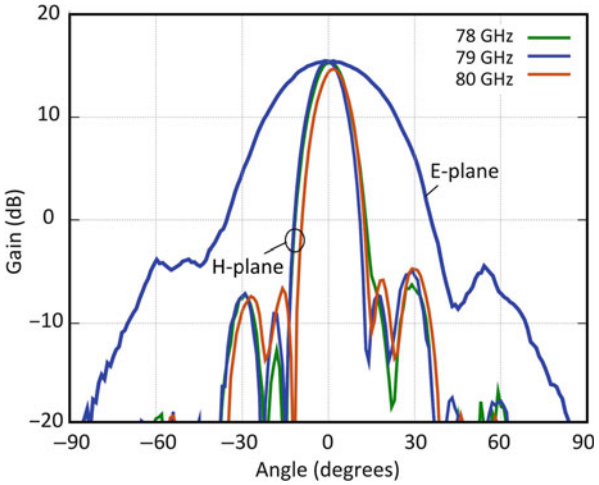


Fig. 12 Measured E- and H-plane radiation diagrams of the array shown in Fig.11

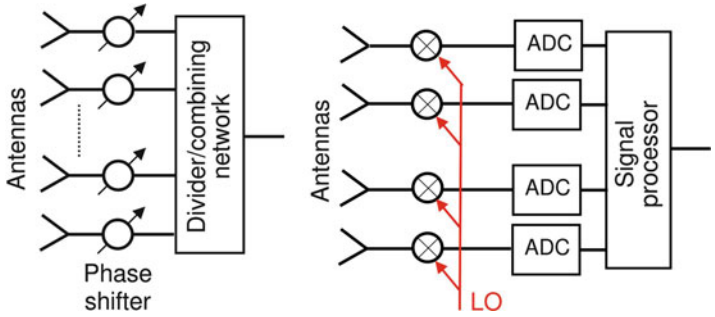


Fig. 13 Basic structures of phased array (*left*) and beamforming antenna (*right*)

As an example, an SIW slotted array is shown in Fig. 11, realized on a Rogers RO3003 substrate. It consists of three rows of SIW with 15 radiating slots in each row, together with a 1:3 power divider in the same technique. The longitudinal distance between the slots is half a wavelength with the SIW; the power distribution along the rows is controlled by the slot widths. An amplitude taper both within the rows and between the rows (unequal power divider) is selected to achieve a −20 dB sidelobe level. Details of the design procedure are given in Massen et al. (2013). H-plane radiation diagrams (perpendicular to the slots) at 78, 79, and 80 GHz are plotted in Fig. 12, together with the E-plane diagram at 79 GHz (the E-plane diagram does not change in the frequency range of interest). Beamwidths of 11° and 34° in H- and E-plane can be found. In both planes, the required sidelobe level is nearly met. A slight beam scan can be observed in the H-plane due to the series-fed slots. This array is one of six subarrays finally integrated on one substrate to form an antenna for digital beamforming (Massen et al. 2013).

Digital Beamforming

Antenna elements of arrays or subarrays as described in the previous subsection often are connected by transmission lines to (larger) arrays and provide fixed beams. Antenna elements or subarrays can alternatively be connected to phase shifters (and attenuators), allowing electronic beam scanning (Fig. 13, left side); this typically results in relatively complex, lossy, and costly arrangements at mm-wave frequencies. An alternative approach is to use several antenna elements or subarrays and

DOI 10.1007/978-981-4560-75-7_96-1

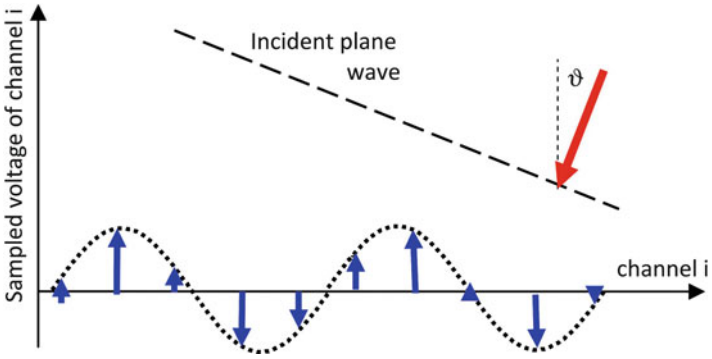


Fig. 14 Real part of IF samples for a slanted incidence of a wave

either switch these successively to a receiver or transmitter or connect them to multiple transmit/receive circuits, easily done with modern MMICs like in Winkler et al. (2008). Figure 13, right side, shows the example block diagram of a receiver with digital beamforming. The scene of interest is illuminated by a separate transmit antenna, or a combined transmit/receive circuit may be connected to one or (successively) several of the antenna elements. The receiver channels down convert and amplify the respective signals. The IF signals are sampled and converted to a sequence of digital values which can be processed further. After range processing for each channel, typically using an FFT for FMCW sensors, amplitude and phase angle are available for each range cell and receive channel. Multiplication with a complex number is equivalent to modify the amplitude and phase angle for the analog signal. A more efficient approach is based on the relations indicated in Fig. 14. An incident plane wave from a given slant angle is received by n receivers; according to the different path lengths, the signal at each receiver suffers a progressive delay (phase angle), resulting, after down conversion, in signals proportional to $\cos(i\, k_0\, d \sin\vartheta)$ with i, the number of the element; d, the distance between the elements; and k_0, the free space propagation constant. Accordingly, the spatial frequency of the sampled values indicated by the dotted line in Fig. 14 increases with increasing angle ϑ of incidence. As a consequence, applying another Fourier transform in cross range, different directions of wave incidence can be detected simultaneously, thus being able to measure the scenery in front of a car with high precision (Gresham et al. 2004; Asano et al. 2001; Stelzer et al. 2010). In comparison to analog beamforming with a phased array, this approach is referred to as digital beamforming.

The beamwidth of the subarray(s) in such an approach must be sufficient to cover the angle range of interest for the respective application. At least two of the subarrays should be placed at a distance as close as possible to half a wavelength to ensure an unambiguous cross-range resolution, but further subarrays may be arranged at larger but different distances between each other, so a high lateral resolution can be achieved, while the different distances can be optimized to remove ambiguities (grating lobes) (Moffet 1968; Md Tan et al. 2008; Feger et al. 2009). Combining this beamforming with configurations as described in the following, even an effective aperture width larger than the physical one can be achieved.

MIMO Concepts for Automotive Radar

In Kees et al. (1995), a concept was proposed to improve the angular resolution of an imaging radar system by transmitter location multiplexing. Today, this concept partly is called MIMO radar (multiple in, multiple out), although in most cases transmitters or receivers (Mayer et al. 2006) are operated successively. Just recently, a concept has been presented to simultaneously transmit via

DOI 10.1007/978-981-4560-75-7_96-1

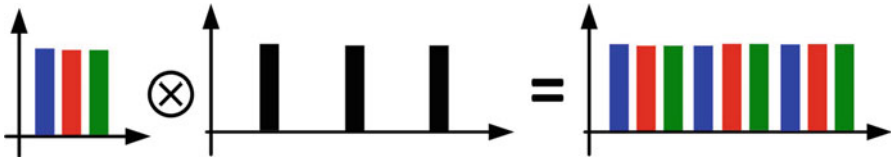


Fig. 15 Convolution principle for MIMO radar

several channels in frequency domain multiplex and to receive in parallel by a respective number of receivers (Pfeffer et al. 2013).

According to antenna theory, the angular distribution of the radiated fields (radiation diagram) of an antenna is proportional to the Fourier transform of its aperture distribution in the respective plane. On the other hand, the signal transmitted from one antenna and received by another one is proportional to the multiplication of the radiation diagrams of these two. Consequently, this product can also be derived from an aperture distribution formed by the convolution of the two individual aperture distributions. Superposition of the measurements from n transmits and m receive subarrays then results in the convolution of the respective excitation distributions (Fig. 15). Accordingly, the configuration as given in the example in Fig. 15 results in a synthetic antenna array with densely arranged subarrays. Other combinations of transmit and receive antennas even result in an increased aperture (Mayer et al. 2006). Weighting the individual measurements with complex numbers (representing phase and amplitude adjustments), different beam directions or modifications of the radiation diagram can be adjusted and evaluated simultaneously.

Fabrication, Packaging, and Mounting Considerations

As already mentioned in the introduction, automotive radars have to operate in a temperature range from −40 °C up to 85 °C or even higher, they have to withstand shock and vibrations, and they have to operate under conditions of rain, snow, or ice. Materials, fabrication processes, packaging as well as mounting of the sensor therefore have to be selected carefully.

Substrate materials need to be suitable for the mm-wave frequency, their physical and electrical properties should be reasonably constant over the temperature range, and they should not absorb moisture. On the other hand, standard microwave substrates may be rather expensive, so a compromise has to be made in this respect. Etching of the planar structures should, as far as possible, be done using standard printed circuit board (PCB) processes but may require accuracies down to 20 µm for the 77 GHz frequency range – this also poses a great challenge toward tolerance-optimized antenna design. Metal surfaces, in addition, need to be protected against corrosion by suitable plating.

Antennas need to be connected to the millimeter-wave circuits and/or MMICs. While bonding at 24 GHz is still not too critical, this provides major efforts and therefore higher cost at 77 GHz. Typically, MMICs have to be placed in some cavity in the circuit board, and bonding is done in the same level to keep parasitics as low as possible (see also the first application example in this chapter). One way out of this is to provide a suitable package for the MMIC as described in Fig. 16 (Wojnowski et al. (2011)), known as eWLB packaging. The MMIC is embedded into a plastic mold, further metallization levels connect the MMIC to lines on the mold, and a ball grid array then allows an easy interconnect to the circuit board, possibly directly to the antenna substrate. In this way, the assembling of antenna and MMICs is feasible without specialized and costly equipment.

The antenna and the radar sensor as a whole, finally, have to be protected by a suitable package. This typically is done by an overall housing, as can be seen also at the examples given in the following section. The cover in front of the antenna (radome) must be transparent to the

DOI 10.1007/978-981-4560-75-7_96-1

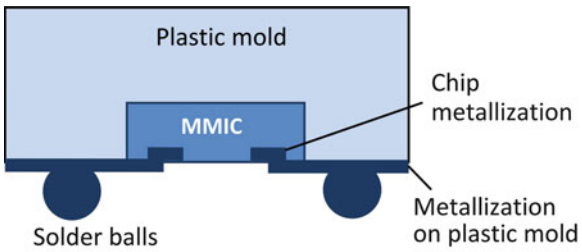


Fig. 16 eWLB package

electromagnetic wave in the respective frequency range, leading to an optimized thickness for the used plastic materials in the order of multiples of half a guide wavelength within the material. This optimized thickness only is perfect for one angle of radiation. For all other angles, there is performance degradation due to reflection. Major efforts are necessary to keep the influence of such a reflection below a given threshold. If the cover is close to the antenna, even a codesign of antenna and cover has to be done.

Mounting of the radar sensor at the car requires a position from which the radar can cover the relevant angle of view, but space limitations and overall car design concepts require some compromise; so typical mounting positions are behind the radiator grille and the bumper. Today, the bumper often is painted, and efforts have to be done to ensure a sufficient transparency of the bumper, even with paintings containing metallic particles. This can be done by optimizing the thickness of the bumper material or by adding some matching layer in front of the radar (Fitzek et al. 2010).

Furthermore, during operation, the radome must not be covered by layers of water, ice, or snow; this would severely deteriorate its performance. To avoid this, a protected mounting position, e.g., behind the bumper, a water-repellent radome surface, or a suitable form or orientation of the radome to blow away water or snow just by the natural airstream of the car, can be selected.

Application Examples for Automotive Antennas

Bosch Sensor LRR3

The LRR3 automotive radar sensor of Bosch came onto the market in 2009. It is based on the FMCW principle, and the mm-wave transmit and receive circuits are integrated on a four-channel SiGe monolithic chip (Winkler et al. 2008) connected to four single microstrip patch antennas combined with parasitic elements to adjust the bandwidth and beamwidth. In the original version, the MMIC is mounted into a cavity in the printed circuit board and bonded to the antenna feed lines. In an advanced version, eWLB packaging as described above is used for the MMIC. The antenna elements are tilted by 45° to reduce interference from oncoming cars. The antenna elements serve as feeds for a lens antenna (Binzer et al. 2007), resulting in four narrow beams. An exploded view of the complete sensor and the planar antenna section is shown in Fig. 17. With overlapping radiation diagrams of the beams (Fig. 18), the angular detection of targets is done by monopulse principles between two neighboring beams, resulting in a resolution considerably better than the beamwidth, provided that only one target is in the respective range cell.

DOI 10.1007/978-981-4560-75-7_96-1

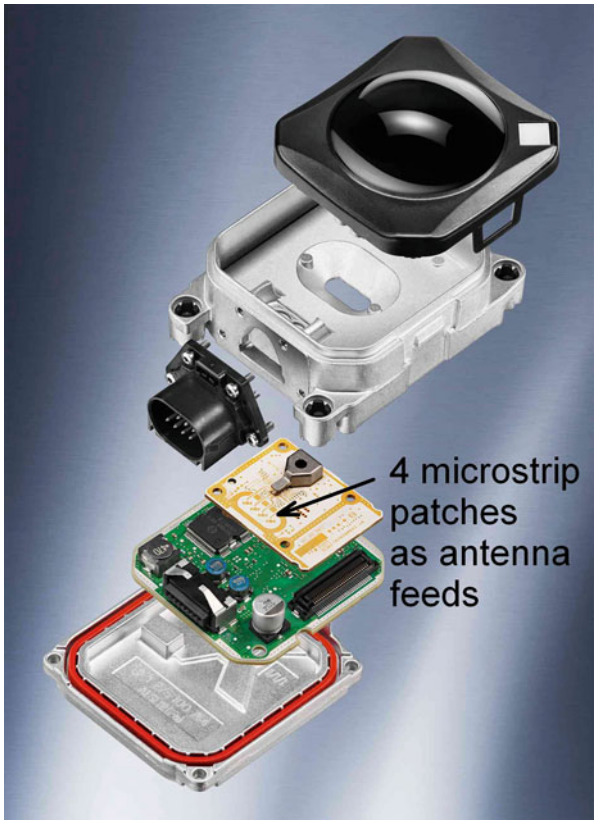


Fig. 17 Exploded view on the Bosch LRR3 sensor and details of the planar antenna elements feeding the lens (Sensor photograph provided by Bosch)

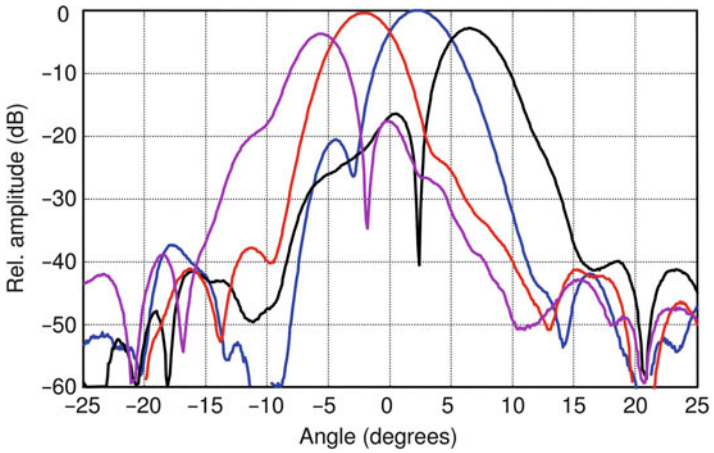


Fig. 18 Azimuth radiation diagrams of the four beams of the LRR3 radar (Fig. 16)

Continental Sensor ARS300

Also in 2009, Continental ADC introduced its third generation of automotive radar with completely new features and a novel approach for the radar antenna. The antenna of the ARS300 scans with a narrow beam over a given azimuth range; it performs auto alignment and is of compact size.

As can be seen in Fig. 19, the entire arrangement consists of a dielectric waveguide in the vicinity of a grooved rotating drum and a folded space consisting of a polarizing grid, also serving as

DOI 10.1007/978-981-4560-75-7_96-1

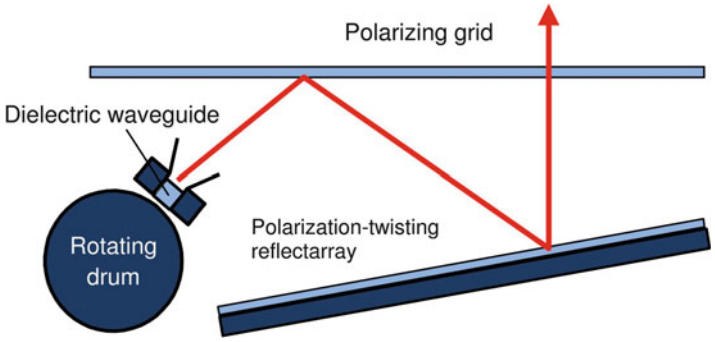


Fig. 19 Principal cross section of the Continental ARS 300 antenna

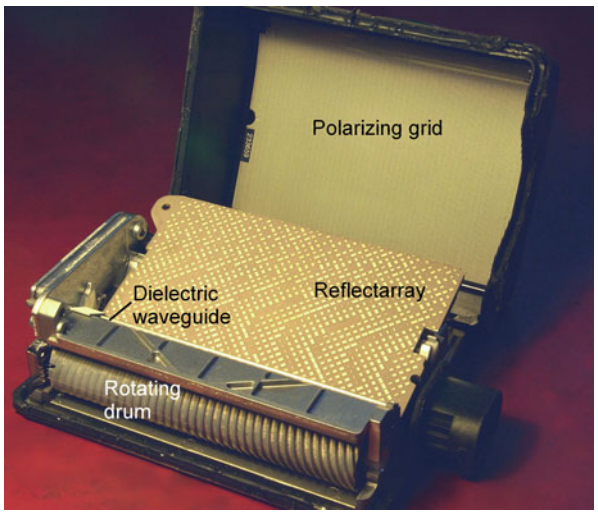


Fig. 20 Photograph of the Continental ARS 300 sensor with opened cover (own photograph)

mechanical protection and radome, and a focusing reflectarray. A photograph of the opened sensor is given in Fig. 20. The basic radiation structure is a dielectric guide with propagation constant k_d. The drum is placed close to the dielectric waveguide, and the metal ridges between the grooves of the drum form periodic perturbations at which the dielectric guide radiates (Fig. 19); the radiated partial fields superpose to form the overall far field radiation. For a given direction of radiation, these partial fields must superimpose in phase. According to Fig. 21, the following relation must hold

$$k_0 \Delta l \ \sin(\delta) + 2\pi = k_d \Delta l \tag{3}$$

where k_0 and k_d are the phase constants of free space and the dielectric guide, respectively, and Δl is the distance between the metal ridges (Fig. 21).

To scan the radiated beam over a given angle, the period of the grooves has to be modified – the groove period depends on the rotational angle, and the actual groove period in the direct vicinity of the guide determines the direction of the radiated beam. Rotating the drum then leads to a scanning of the beam (Manasson et al. 1996). A great effort has been spent in designing the drum in a way to find the optimum shape and size of the grooves such that at each azimuth angle a beam of a given scan angle with sidelobes below a prescribed threshold is obtained. A unique feature of this waveguide drum unit is that it contains different angular sections in which both the direction of

DOI 10.1007/978-981-4560-75-7_96-1

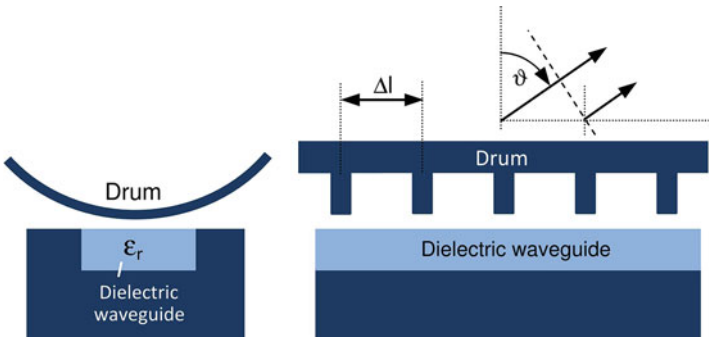


Fig. 21 Details of radiating dielectric guide

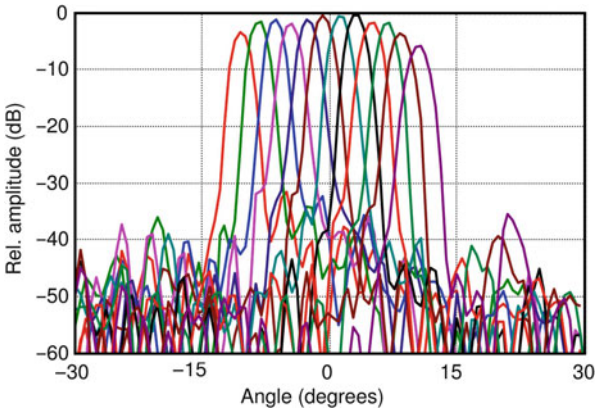


Fig. 22 Selected two-way azimuth radiation diagrams of the Continental ARS 300 sensor

radiation and the beamwidth can be altered such that different sensor modes from short to long range (up to 200 m) can be covered by a single setup.

Additionally, the radiated wave has to be formed to provide a narrow diagram in elevation as well. This is done by a folded reflectarray configuration as described earlier (see also Fig. 7). The required polarizer is integrated into the cover of the sensor such that the beam radiated from the dielectric waveguide/drum unit is reflected toward a reflectarray consisting of a dielectric layer glued onto a conducting plate. Phase angles are adjusted such that a narrow beam results in elevation, together with the necessary polarization twisting. Both the polarizing grid printed to the sensor cover and the reflectarray can clearly be seen in Fig. 19. Measurements of the radiation diagrams have been done using the complete radar, resulting in two-way diagrams. A selection of far field radiation diagrams for different drum positions can be seen in Fig. 22. Two-way beamwidths are in the range of 2.5°.

In addition, this antenna exhibits another unique feature. While the headlights of a car can be adjusted in elevation by any skilled person, this is different for a radar sensor with its non-visible radiation. With this antenna, the reflectarray can be tilted to some extent by a motor unit which performs an auto alignment in elevation.

Experimental 79 GHz LTCC-Based Sensor with Grid Antennas

For high-density 3D integration, low- and high-temperature cofired ceramics (LTCC, HTCC) are gaining increasing interest. The starting point for their fabrication is mixtures of ceramic and glass powder together with binding agent for LTCC and solely ceramic powder with binding agent for HTCC. These materials are pressed into thin sheets (one tenth or a few tenth of a millimeter thick)

DOI 10.1007/978-981-4560-75-7_96-1

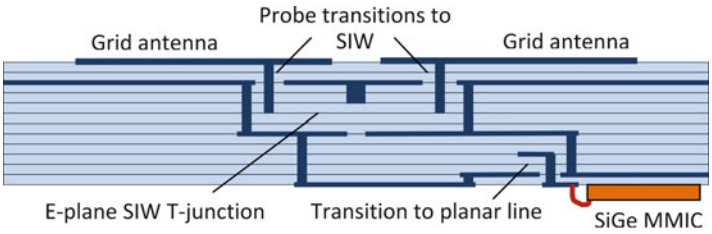


Fig. 23 Cross section of LTCC sensor

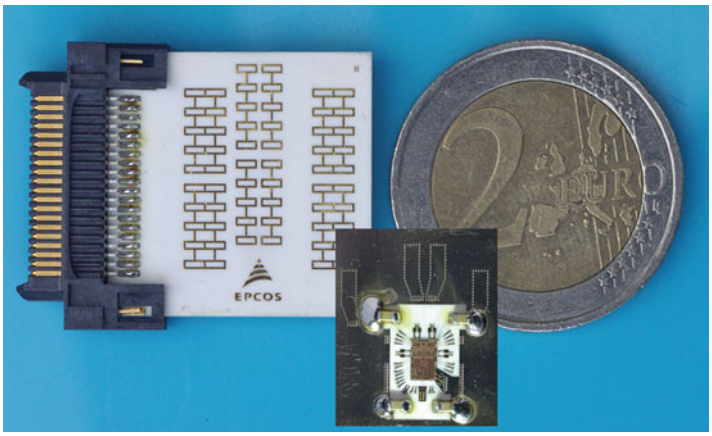


Fig. 24 Photograph of the 79 GHz LTCC sensor; antenna side and MMIC side in inset (own photograph)

and cut into suitable panels. In the first step, holes for vias are punched into the sheets and filled with metal paste. Following this, screen printing is applied to print metal patterns onto each sheet. Finally, a required number of sheets (even several tens of sheets) are laminated and pressed together, precut, and fired. For LTCC, only the glass powder needs to melt, so temperatures in the range of 900 °C are sufficient. For HTCC, the ceramic powder must be joined together, so temperatures of 1,600 °C are required. For very fine and accurate structures on the outer sides of the multilayer, even sputtered and electroplated metal with standard etching processes can be added (partly called “fine line” techniques (Brunner et al. 2012)). Including these processes, LTCC can be extended to the design of highly integrated millimeter-wave circuits.

This technique was applied to develop an experimental 79 GHz short-range sensor using grid antennas and a four-channel SiGe radar front end (Bauer et al. 2013). A principal cross section of the multilayer circuit is given in Fig. 23. The multilayer contains 12 layers of 0.118 mm thickness each. The dielectric constant for the material used here is 7.4. Vias can be fabricated down to a diameter of 0.1 mm, easily enabling SIW structures; the effective SIW width is about 1.2 mm. Two grid antennas with opposite orientations are located on the top surface, fed by probes to a first layer of SIW. An E-plane T-junction feeds the signals to a lower level SIW; from there, another probe-type transition provides a connection to a planar symmetric line at the backside where the MMIC is located and bonded to the planar lines.

A photograph of the antenna side of the sensor together with a section of the backside with the MMIC is depicted in Fig. 24. While the LTCC substrate has dimensions of 23 × 23 mm, the overall usable size for the antenna arrangement is 13 × 20 mm. This consists of four chains of grid antennas, each connected to one channel of the MMIC. As at least two channels need to be closely spaced to ensure unambiguity, two chains consist of single rows, while the other two are double chains to

DOI 10.1007/978-981-4560-75-7_96-1

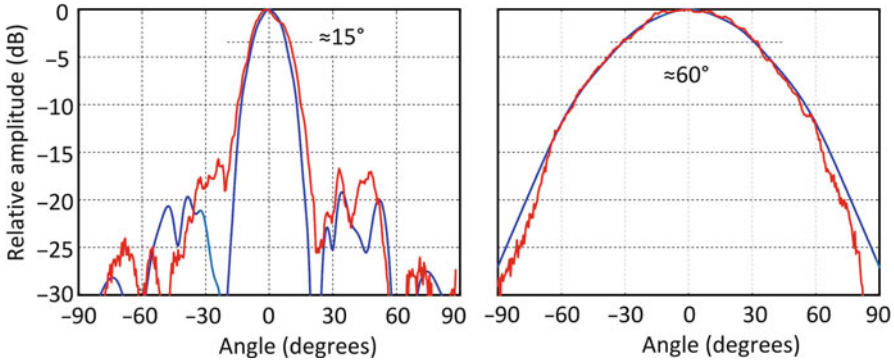


Fig. 25 Elevation (E-plane) and azimuth (H-plane) radiation diagrams of a single-row grid antenna. *Blue* simulation, *red* experiment

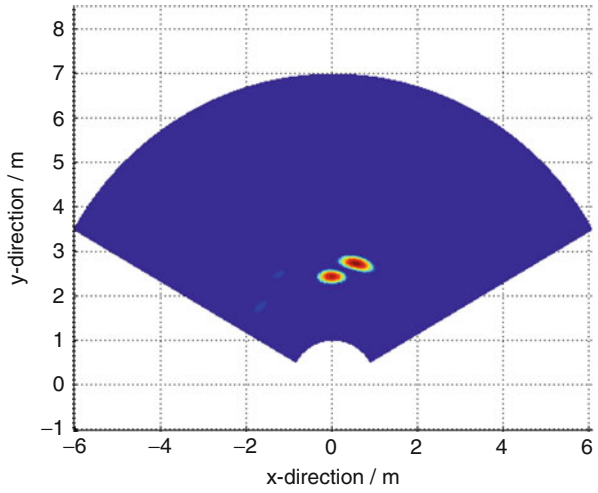


Fig. 26 Close-range image of two targets

increase the overall (processed) signal output (at the cost of a reduced angle of view) and placed at larger distances to the single chains. Mid-to-mid distances between the antenna rows are 3.6, 2.2, and 5.4 mm. Radiation diagrams of a single-row antenna are plotted in Fig. 25; beamwidths in elevation and azimuth are about 15° and 60°, respectively. The elevation diagram of the double-chain antenna is equal to that of the single-row one; the azimuth diagram of the double-row antenna has a 3 dB beamwidth of 40°.

Each radar channel is able to both transmit (FMCW) and receive; all channels are active in receiving whatever channel is transmitting. MIMO principles with optimized weighting functions are then applied for cross-range imaging according to Feger et al. (2009). According to the convolution principle, ten different synthetic antenna positions result with a nearly doubled antenna aperture. The response to a target then has a 3 dB width of about 10° at the broadside (Bauer et al. 2013). Together with the FMCW principle for range, a typical radar image for two targets is shown in Fig. 26. With this experimental sensor, a range of up to 20 m can be covered; employing a new LTCC material with lower losses and a further optimization of the backside transition to the MMIC will increase range further. Of course, also a larger antenna size will do the same.

DOI 10.1007/978-981-4560-75-7_96-1

Conclusion

After an introduction to automotive radar sensors and their principle architecture, this chapter provides a guide to different types of antennas for automotive radars. While in the early approaches more or less standard antenna configurations were dominating, different types of planar antennas and array antennas have been employed later on. As shown in the second system example, also different antenna types like a dielectric line-based radiator with a mechanical scanning unit and a printed and folded reflectarray antenna are combined for best performance. Presently, the original function of "stand-alone" antennas is being replaced by an overall arrangement of antenna elements and arrays, transmit/receive configuration, and digital signal processing, as it is demonstrated by digital beamforming together with (quasi-) MIMO concepts. The latter allows a reduction of transmit/receive modules and antenna elements, and it even allows to generate synthetic apertures larger than the physical ones.

Signal processing is also a means to further increase resolution in all radar relevant domains, i.e., range, cross range, and Doppler frequency. In the present commercial sensors, real-time signal processing is still limited to the basic functions, but soon high-resolution methods like MUSIC (Schmidt 1986), ESPRIT (Roy et al. 1986), or autoregressive (AR) signal estimation (Proakis and Manolakis 1996) will be considered. In a number of studies, AR signal estimation has proven to be quite robust in automotive applications, e.g., Mayer et al. (2006). In cross range – where the antenna configuration is relevant – beamforming leads to a number of sampled values from which the lateral target distribution is calculated using an FFT. Explained in a very simple way, such a series of sampled values can be modeled by a kind of digital filter, and with this, new artificial sampling values can be generated on either side of the original series, typically increasing the number of samples by 2 or 3. An FFT over the enlarged series then results in an improved cross-range resolution, equivalent to an antenna aperture of twice or three times as wide.

With more stringent requirements for safety application, also an improved resolution with respect to elevation gets more and more important – is there any target which the vehicle can run over or pass below, or is it a relevant obstacle? As a consequence, systems – and therefore the antenna configurations – need to be extended, either to switch or scan in elevation as well, or an extra antenna element may be added to observe higher elevation angles.

As discussed in the second chapter, operation frequency of automotive sensors mostly is shifting from the 24 GHz frequency range to the 76/77 and 77/81 GHz range; the driving motivation is toward smaller sensors; basically, the effective size is determined by the antenna. Based on this, also higher frequency ranges are under first investigations (Köhler et al. 2013). Frequencies may be in the 120 or 150 GHz range. For the same antenna aperture as with lower frequencies, gain is increased, or for the same gain, antenna dimensions get smaller. One has to be aware, however, that with smaller antenna aperture size, the received power decreases as well. From the antenna design point of view, a simple scaling of the structure seems to be possible; care has to be taken for increased losses and more critical tolerance requirements. Semiconductor devices show increasing cutoff frequencies (Pfeiffer 2012), but output power will be lower and receiver noise figure higher. Nevertheless, estimations in Köhler et al. (2013) show that MRR and SRR sensors should be feasible. Another challenge with higher frequencies will be interconnects, but efforts are going on to integrate antenna elements directly on a MMIC or providing specific packaging technologies with potentially integrated antenna elements provided by the MMIC manufacturer, e.g., Wojnowski et al. (2011) and Hasch et al. (2012).

DOI 10.1007/978-981-4560-75-7_96-1

Acknowledgments

Thanks are due to Bosch and Continental ADC for providing part of the material for this chapter. Other parts of the results shown here have been achieved by projects funded via the "RoCC project (project number 13 N9824) of the German Federal Ministry of Education and Research (BMBF)" and the "Austrian BMVIT and the Austrian Research Promotion Agency (FFG) within the co-funded project InRaS in the strategic objective FIT-IT Systems on Chip."

Cross-References

▶ Beam-Scanning Leaky-Wave Antennas
▶ Grid Antenna Arrays
▶ Microstrip Patch Antennas
▶ Millimeter-Wave Antennas and Arrays
▶ Reflectarray Antennas
▶ Substrate Integrated Waveguide Antennas
▶ Waveguide Slot Antennas and Arrays

References

Asano Y, Ohshima S, Harada T, Ogawa M, Nishikawa K (2001) Proposal of millimeter-wave holographic radar with antenna switching. IEEE Inter Microw Symp 2:1111–1114

Bauer F, Menzel W (2011) A 79 GHz microstrip grid array antenna using a laminated waveguide feed in LTCC. In: IEEE AP-S/URSI symposium 2011, Spokane, pp 2067–2070

Bauer F, Wang X, Menzel W, Stelzer A (2013) A 79-GHz radar sensor in LTCC technology using grid array antennas. IEEE Trans Microw Theory Tech 61:2514–2521

Bauer F, Menzel W (2013a) A 79-GHz resonant laminated waveguide slotted array antenna using novel shaped slots in LTCC. IEEE Antennas Wirel Propag Lett 12:296–299

Bauer F, Menzel W (2013b) A 79-GHz planar antenna array using ceramic filled cavity resonators in LTCC. IEEE Antennas Wirel Propag Lett 12:910–913

Binzer T, Klar M, Groß V (2007) Development of 77 GHz radar lens antennas for automotive applications based on given requirements. In: 2nd international ITG conference on antennas (INICA '07 Munich), pp 205–209

Brunner S, Stadler M, Wang X, Bauer F, Aichholzer K (2012) Advanced high frequency LTCC technology for applications beyond 60 GHz. In: Proceedings of the 8th international conference on Ceramic Interconnect and Ceramic Microsystems Technologies, Erfurt, pp 77–81

Camiade M, Domnesque D, Ouarch Z, Sion A (2000) Fully MMIC-based front end for FMCW automotive radar at 77 GHz. In: Proceedings of the 30th European Microwave conference Paris, pp 1–4

Carver KR, Mink JW (1981) Microstrip antenna technology. IEEE Trans Antennas Propag 9:2–24

Conti R, Toth J, Dowling T, Weiss J (1981) The wire grid microstrip antenna. IEEE Trans Antennas Propag 29:157–166

Feger R, Wagner C, Schuster S, Scheiblhofer S, Jäger H, Stelzer A (2009) A 77-GHz FMCW MIMO radar based on a SiGe single-chip transceiver. IEEE Trans Microw Theory Tech 57:1020–1035

DOI 10.1007/978-981-4560-75-7_96-1

Fitzek F, Rasshofer RH, Biebl EM (2010) Metamaterial matching of high-permittivity coatings for 79 GHz radar sensors. In: European Microwave conference, Paris, pp 1401–1404

Frei M, Bauer M, Menzel W, Stelzer M (2011) A 79 GHz differentially fed grid array antenna. In: European Microwave conference, Manchester, pp 1320–1323

Gresham I, Jain N, Budka T, Alexanian A, Kinayman N, Ziegner B, Brown S, Staecker P (2001) A compact manufacturable 76–77-GHz radar module for commercial ACC applications. IEEE Trans Microw Theory Tech 49:44–58

Gresham I, Jenkins A, Egri R, Eswarappa C, Kinayman N, Jain N, Anderson R, Kolak F, Wohlert R, Bennett J, Lanteri J-P (2004) Ultra-wideband radar sensors for short-range vehicular applications. IEEE Trans Microw Theory Tech 52:2105–2120

Hasch J, Topak E, Schnabel R, Zwick T, Weigel R, Waldschmidt C (2012) Millimeter-wave technology for automotive radar sensors in the 77 GHz frequency band. IEEE Trans Microw Theory Tech 60:845–860

Hirokawa J, Ando M (2000) 76 GHz post-wall waveguide fed parallel plate slot arrays for car-radar applications. IEEE Int Symp Antennas Propag 1:98–101

Hymans AJ, Lait J (1960) Analysis of a frequency-modulated continuous-wave ranging system. Proc IEE Part B Electron Commun Eng 107:365–372

James JR, Hall PS, Wood C (1981) Microstrip antenna theory and design. Peregrinus, London

Kees N, Schmidhammer E, Detlefsen J (1995) Improvement of angular resolution of a millimeterwave imaging system by transmitter location multiplexing. IEEE Int Microw Symp 2:969–972

Köhler M, Hasch J, Blöcher HL, Schmidt L-P (2013) Feasibility of automotive radar at frequencies beyond 100 GHz. Int J Microw Wirel Technol 5:49–54

Kraus J (1964) A backward angle-fire array antenna. IEEE Trans Antennas Propag 12:48–50

Massen J, Frei M, Menzel W, Möller U (2013) A 79 GHz SiGe short-range radar sensor for automotive applications. Int J Microw Wirel Technol 5:5–14

Mayer W, Gronau A, Menzel W, Leier H (2006) A compact 24 GHz sensor for beam-forming and imaging. In: 9th international conference on control, automation, robotics and vision (ICARV 2006 Singapore), pp 153–158

Md Tan MN, Rahim SKA, Ali MT, Rahman TA (2008) Smart antenna: weight calculation and side-lobe reduction by unequal spacing technique. In: IEEE international RF and microwave conference, Kuala Lumpur, pp 441–445

Manasson V, Sadovnik L, Mino R (1996) MMW scanning antenna. IEEE Aerosp Electron Syst Mag 11:29–33

Meinel HH, Dickmann J (2013) Automotive radar: from its origin to future directions. Microw J 56:24–40

Menzel W, Pilz D, Al-Tikriti M (2002) MM-wave folded reflector antennas with high gain, low loss, and low profile. IEEE Antennas Propag Mag 44:24–29

Millitech Corporation (1994) Crash avoidance FLR sensors. Microw J 37:122–126

Moffet A (1968) Minimum-redundancy linear arrays. IEEE Trans Antennas Propag 16:172–175

Pfeffer C, Feger R, Wagner C, Stelzer A (2013) FMCW MIMO radar system for frequency-division multiple TX-beamforming. IEEE Trans Microw Theory Tech 61:4262–4274

Pfeiffer UR (2012) Silicon CMOS/SiGe transceiver circuits for THz applications. In: IEEE 12th topical meeting on silicon monolithic integrated circuits in RF systems (SiRF), Santa Clara, pp 159–162

Proakis JG, Manolakis DG (1996) Digital signal processing: principles, algorithms, and applications. Prentice-Hall International, Upper Saddle River

DOI 10.1007/978-981-4560-75-7_96-1

Roy R, Paulraj A, Kailath T (1986) ESPRIT – a subspace rotation approach to estimation of parameters of cisoids in noise. IEEE Trans Acoust Speech Signal Process 34:1340–1342

Russell ME, Crain A, Curran A, Campbell RA, Drubin CA, Miccioli WF (1997) Millimeter-wave radar sensor for automotive intelligent cruise control (ICC). IEEE Trans Microw Theory Tech 45:2444–2453

Sakakibara K, Mizutani A, Kikuma N, Hirayama K (2006) Design of narrow-wall slotted hollow waveguide array for arbitrarily linear polarization in the millimeter-wave band. In: IEEE international symposium on antennas propagation, Albuquerque, 3141–3144

Schmidt R (1986) Multiple emitter location and signal parameter estimation. IEEE Trans Antennas Propag 34:276–280

Shino N, Uchimura H, Miyazato K (2005) 77 GHz band antenna array substrate for short range car radar. In: IEEE MTT-S international microwave symposium Long Beach, pp 2095 – 2098

Stelzer A, Feger R, Jahn M (2010) Highly-integrated multi-channel radar sensors in SiGe technology for automotive frequencies and beyond. In: ICECom conference, Dubrovnik, pp 1–11

The (new) Cadillac Database©Dream Cars on Cadillac Chassis (2013) http://www.cadillacdatabase.org/Dbas_txt/Drm_cycl.htm. Last updated 23 May 2013

Tokoro S, Kuroda K, Kawakubo A, Fujita K, Fujinami H (2003) Electronically scanned millimeter-wave radar for pre-crash safety and adaptive cruise control system. In: Proceedings of the IEEE intelligent vehicles symposium Columbus, pp 304–309

Winkler V, Feger R, Maurer L (2008) 79 GHz automotive short range radar sensor based on single-chip SiGe-transceivers. In: European Microwave conference, Amsterdam, pp 1616–1619

Wojnowski M, Lachner R, Böck J, Wagner C, Starzer F, Sommer G, Pressel K, Weigel R (2011) Embedded wafer level ball grid array (eWLB) technology for millimeter-wave applications. In: IEEE 13th electronics packaging technology conference (EPTC), Singapore, pp 423–429

Woll JD (1995) VORAD collision warning radar. In: IEEE international radar conference, Alexandria, pp 369–372

Xu JF, Hong W, Chen P, Ke W (2009) Design and implementation of low sidelobe substrate integrated waveguide longitudinal slot array antennas. IET Microw Antennas Propag 4:790–797

Zhang B, Zhang YP (2012) Grid array antennas with subarrays and multiple feeds for 60-GHz radios. IEEE Trans Antennas Propag 60:2270–2275

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_97-1

Applications of Phased Array Feeders in Reflector Antennas

S. G. Hay[a]* and T. S. Bird[b]
[a]CSIRO Digital Productivity, Epping, NSW, Australia
[b]Antengenuity, Eastwood, NSW, Australia

Abstract

An array feeder for a reflector or a lens provides considerable flexibility in the beams that can be produced. A single beam could be obtained by combining all elements; multiple beams can be created with clusters of subarrays or the elements combined in such a way that a beam is steered by adjusting the phase and amplitude of the elements. A variety of applications of array feeders are possible from satellite communications, radar, and radio astronomy. The latter application is described here through fixed arrays, and individual horns of the cluster are excited step-by-step to produce multiple beams at radio observatories such as Parkes, Australia, and Arecibo, Puerto Rico. A full phased array solution is described for the Australian Square Kilometre Array Pathfinder (ASKAP). For fixed arrays, an optimum excitation is described. It is shown that without coupling effects, the excitation should be the complex conjugate of the focal field. The size of an array is defined and the sensitivity of a multibeam feed is shown to be proportional to $\sqrt{N}$ where N is the number of elements of the array. The theory relating to phased array feeding of reflectors is developed, and the definition of sensitivity is extended to phased arrays. The survey speed of a phased array feed interferometer is defined, and conditions are developed for survey speed as a function of the number and spacing of the beams. The application of fixed arrays as multibeam feeds for radio telescopes is described for the 13-element multibeam feed for the Parkes radio telescope and for the phased array checkerboard feed for ASKAP in Western Australia.

Keywords

Arrays; Phased; Feeds; Beams; Fixed beam; Steering; Mutual coupling; Sensitivity; Radio telescope; Parkes; ASKAP; Survey speed; Multibeam

Introduction

There are an increasing number of applications with reflectors that benefit from the flexibility and performance improvement from more sophisticated feeds than a single element. Arrays have been used as feeds sometimes combined and sometimes singly. A phased array is a collection of antennas combined together through the excitation to perform as a single antenna. The beam produced is able to be steered or moved by adjusting the amplitude and/or the phases of the signals rather than by physically moving the entire antenna structure. A phased array feeder is the means of achieving this adaptability with another larger antenna, usually a reflector or a lens as shown in Fig. 1, which is efficient at forming a focused beam. The absence of moving parts enables the beams to be steered rapidly, making the phased array feeder useful in applications requiring one or multiple scanned or adaptable beams such as in radar and radio astronomy. For a description of the latter systems, refer to the chapter "► Antennas in Radio

*Email: stuart.hay@csiro.au

DOI 10.1007/978-981-4560-75-7_97-1

Telescope Systems." The beam of a spherical reflector is steered simply by rotating the feed at the focus. However, spherical reflectors tend to have lower gain and higher sidelobes than a paraboloid. For a parabolic reflector, beam steering is more complicated and can be achieved by a variety of methods. It can be steered over a small range by displacing a single feed transversely from the reflector vertex. Unless the aperture distribution is highly tapered, there is a large reduction in gain. Phase correction can be introduced, but this restricts the beam scanning to within a few beamwidths of the boresight direction before gain loss becomes too large. Lateral displacement of the single feed by a distance d in the focal plane results in a scanned beam at an angle θ_b, which is less than the angle from the focus to the vertex $\theta_v = \text{atan}(d/f)$. This difference is called the beam deviation factor (BDF). Thus

$$\theta_b = \theta_v \text{BDF} \tag{1}$$

where BDF typically has values $0.7 - 0.9$. A useful approximation when the f/D is not small is

$$\text{BDF} \approx \left[1 + \frac{1}{32(f/D)^2}\right]^{-1}. \tag{2}$$

Feeds that are displaced are unable to provide low sidelobes at wide scan angles due to aberrations as a result of not matching the fields in the focal region.

An improved method of achieving a scanned beam is to use information derived from the focal plane field and provide a suitable conjugate match. This means the feed has to be sufficiently flexible to provide the appropriate excitation to generate the required field. This is possible with an array of feeds that contains a sufficient number of elements. In another approach, Rudge et al. (1986) use signal processing to provide a second Fourier transform, as well as the one provided by the reflector, to generate the scanned beam. The transform is of the intercepted electric fields of the focal plane on some defined locus where only phase shifters are needed to correct for aberrations.

Phased arrays have been used as feeds in applications involving rapid beam steering. The complexity of the feeding network, cost, and weight have often in the past been a deterrent to their use. However, in recent years, the flexibility that is possible as well as new methods of feeding has stimulated a reassessment of phased arrays as feeds for reflectors. Some of the advances made in the area will be reviewed in this chapter.

Array Feeds for Reflectors

A reflector antenna radiates a pencil beam when excited from its focal region by a feed antenna as shown in Fig. 1. In most applications, a single feeder is used, which may consist of a dipole, horn, or waveguide. Efficient feeding antennas and associated techniques have been developed for reflectors as described in the chapter on "▶ Reflector Antennas." The reflector could be a single reflector such as a symmetrical paraboloid or a dual-reflector in a Cassegrain configuration as shown in Fig. 1. In order to efficiently illuminate a reflector to produce a beam on boresight or at some angle away from it as in steering, the aperture field of the feed should replicate the focal region field corresponding to that beam in some way. This can be done by operating the reflector in a receive mode with the beam incident from desired directions or in transmit by adjusting the phases of the elements of the array in a controlled manner.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_97-1

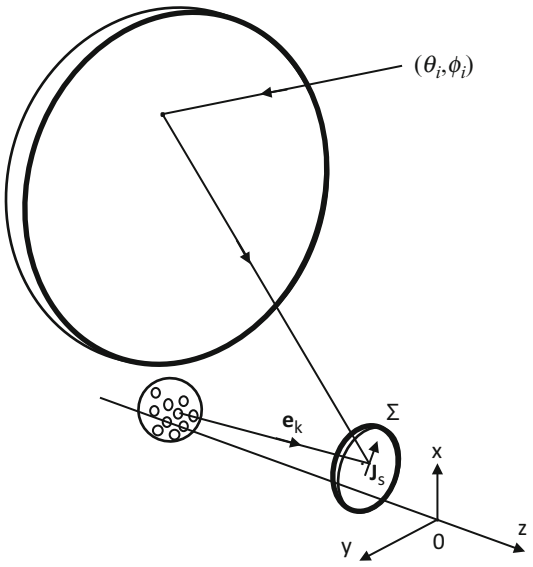


Fig. 1 Reflector antenna with feed array in receive mode with a beam from the direction (θ_i, ϕ_i)

A Dipole Array-Fed Reflector

In an earlier chapter in this book entitled "▶ Reflector Antennas," the field radiated by a single dipole at the focus of a paraboloid has been derived. This analysis is extended here as an initial example to a rectangular array of half-wave dipoles. The array has an odd number of elements in the x-direction N_x with a spacing of d_x and, similarly, N_y elements with spacing d_y in the y-direction. The dipoles are polarized in the x-direction. It is assumed that the array is in the far-field of the reflector and mutual coupling between the dipoles is neglected. Therefore, from Chapter 30 Eqns. 65 and pattern multiplication of the array and element functions, the fields radiated by the array are

$$\mathbf{E}_f = E_o \frac{e^{-jkr}}{r} A(\theta,\phi)\left[\hat{\theta}\,\cos\theta\cos\phi - \hat{\phi}\,\sin\phi\right] \tag{3}$$

$$\mathbf{H}_f = \frac{1}{\eta_o}\hat{r}\times\mathbf{E}_f \tag{4}$$

where

$$A(\theta,\phi) = \frac{\cos\left(\frac{\pi}{2}\sin\theta\right)}{\cos\theta}\sum_{m=-N_{x1}}^{N_{x1}}\sum_{n=-N_{y1}}^{N_{y1}} V_{mn}\exp\left[j\left(mT_x(\theta,\phi)+nT_y(\theta,\phi)\right)\right] \tag{5}$$

consists of the pattern function for a half-wave dipole times the array factor. E_o is a constant, η_o is the wave-impedance of free-space, and the steering phase factors in the x- and y-directions are

$$T_x(\theta,\phi) = kd_x\sin\theta\cos\phi$$

and $T_y(\theta,\phi) = kd_y\sin\theta\sin\phi$.

V_{mn} is the excitation of element mn, and $N_{x1,y1} = \left(N_{x,y}-1\right)/2$. The resulting far-fields of the reflector are given by Eqs. 4.6 and 4.7 of Chapter 30 and are expressed as follows:

DOI 10.1007/978-981-4560-75-7_97-1

$$E_\theta = -\frac{jkfE_o}{\pi}\ \frac{e^{-jkr}}{r}B(\theta,\phi) \tag{6}$$

and

$$E_\phi = \frac{jkfE_o}{\pi}\ \frac{e^{-jkr}}{r}C(\theta,\phi) \tag{7}$$

where

$$B(\theta,\phi) = \int_0^{2\pi} d\xi \cos\xi \int_0^{\psi_c} d\psi\, A(\psi,\xi)\exp[jk\rho(\sin\theta\sin\psi\cos(\phi-\xi) - (1+\cos\theta\cos\psi))] \times \left[\cos\theta(\cos\psi\cos\xi\cos(\phi-\xi) - \sin\xi\sin(\phi-\xi)) - \sin\theta\tan\frac{\psi}{2}\cos\phi\cos\xi\right]\tan\frac{\psi}{2} \tag{8}$$

$$C(\theta,\phi) = \int_0^{2\pi} d\xi \sin\xi \int_0^{\psi_c} d\psi\, A(\psi,\xi)\exp[jk\rho(\sin\theta\sin\psi\cos(\phi-\xi) - (1+\cos\theta\cos\psi))] \times [\cos\psi\cos\xi\sin(\phi-\xi) + \sin\xi\cos(\phi-\xi)]\tan\frac{\psi}{2} \tag{9}$$

where f is the focal length, D is the diameter of the reflector, $\rho = 2f/(1+\cos\theta)$ is the radial distance from the focus to the surface, and $\psi_c = 2\text{atan}[1/(4f/D]$ is the reflector's half-cone angle.

Radiation patterns computed from Eqns. 6 and 7 are plotted in Fig. 2 for a reflector with a diameter $D = 50\lambda$ and $f/D = 0.35$ that is fed by an array of 3 × 3 half-wave elements. The element spacing is $d_x = 0.55\,\lambda$ and $d_y = 0.5\lambda$. The excitation of the array has been obtained from the complex conjugate of the x-directed electric field in the focal region of a paraboloid with $f/D = 0.35$. A plot of the peak gain as

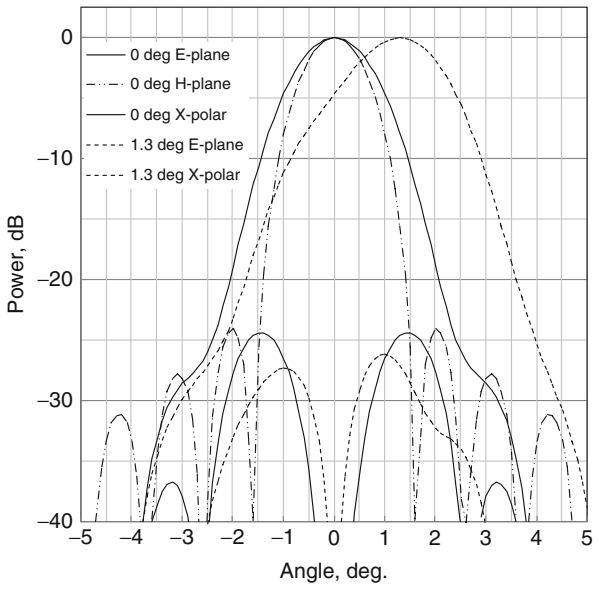


Fig. 2 Normalized radiation pattern of paraboloid reflector with 3 × 3 half-wave dipole feed array with excitation that is a conjugate match to the focal fields. Reflector diameter $D = 50\lambda$ and $f/D = 0.35$

DOI 10.1007/978-981-4560-75-7_97-1

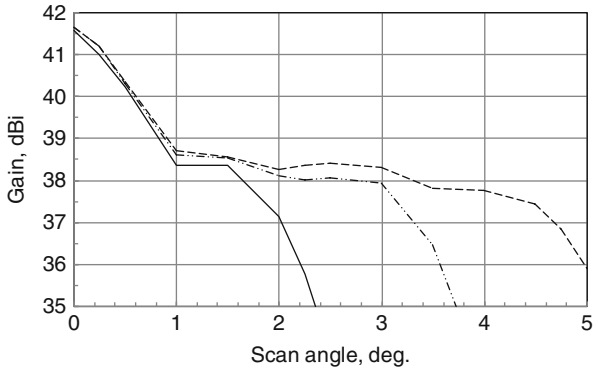


Fig. 3 Gain versus scan angle for paraboloid with half-wave dipole array feed. Reflector diameter $D = 50\lambda$ and $f/D = 0.35$. *Solid line*: 3 × 3 array; *long dashes*: 5 × 3 array; *short dashes*: 7 × 3 array

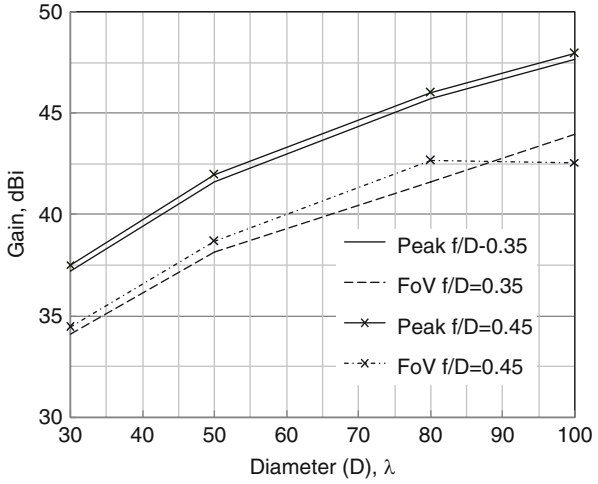


Fig. 4 Gain as a function of diameter D and f/D of a paraboloid with a 5 × 3 element dipole array feed that is matched to the focal fields corresponding to on-axis incidence (peak) and at a scan angle in the field of view at about two beamwidths (FoV)

the beam is steered is shown by the solid line in Fig. 3. It shows that for a 3 × 3 array, the gain decreases with scan angle until it reaches a critical point at about 2° scan angle or about $\sim 2\lambda/D$, after which the gain rapidly falls. This is because the small array, which is centered at the focus, is only able to match part of the focal region fields. As the beam is steered in the x-z plane, the focal region increases in extent in this plane. When the array size is increased further by adding more of elements in the x-direction, the scan in this direction is maintained over a wider scan range albeit with a gain loss of $3dB$. With 5 elements in the scan direction, a $\pm 3^\circ$ scan range is achieved, while with 7 elements, the scan range is over $\pm 4^\circ$. The scan range of $2\Delta\theta$ and hence the field of view (FoV) increase in proportion to the number of elements in the array in that direction. Thus 3 elements can match to a focal field and with conjugate matching can achieve a $FoV = 3^\circ$, while with 7 elements, the $FoV = 8^\circ$.

The scan range and the gain achieved across it are dependent on the reflector diameter and *f*/*D*. For example, Fig. 4 shows the gain achieved in the scan range as a function of paraboloid diameter for *f*/*D* values of 0.35 and 0.45 with the 5 × 3 element array. The peak gain on axis and a typical value of the field of view increase with the diameter as long as the array is large enough to encompass the significant parts of the focal region. If the array does not cover the important parts of the focal region or is incorrectly located, the gain begins to decrease as demonstrated for $D > 80\lambda$ when $f/D = 0.45$. The larger f/D case covers a

DOI 10.1007/978-981-4560-75-7_97-1

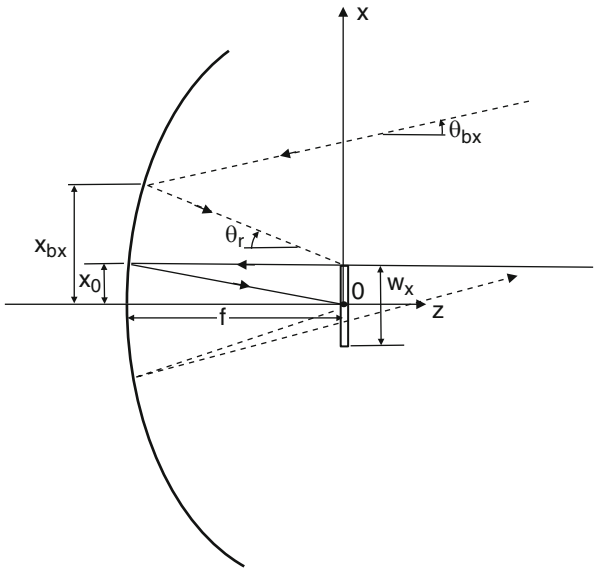


Fig. 5 Estimating array feed size for a symmetrical reflector

wider area of the focal region than for $f/D = 0.35$. It is emphasized that the results quoted above are only an indication of what happens as actual results vary with the feed element type and location and because mutual coupling between the array elements has been neglected.

When coupling is included, the scan performance compared with the uncoupled case can vary significantly and depends on the array geometry and particularly the proximity of the elements. In some configurations, the overall performance improves, while in others, the performance can be worse. However, to ensure the performance is close to the design prediction, it is important to include mutual coupling.

Size of the Feed Array

The size of the array that is required can be estimated by means of the geometric argument illustrated in Fig. 5. This shows a symmetrical reflector with an array feed of width w_x that is illuminated by a plane wave inclined at an angle θ_{bx} in the x-z plane. At this angle, the geometric ray through the antenna is incident on the outermost element of the array. Then, by means of Eq. 1, the array width in the x-direction is

$$w_x \approx 2f \tan\left(\theta_{bx}\left[1 + \frac{1}{32(f/D)^2}\right]\right). \tag{10}$$

Suppose the element spacing is chosen to be d_x. The number of elements required in the x-direction of the array is, therefore,

$$N_x = I\left[\frac{w_x}{d_x}\right] \tag{11}$$

where $I[x]$ is the integer value of the argument x. A similar argument applies in the y-z plane to obtain the maximum dimension and number of elements in the y-direction, N_y, of a rectangular array.

DOI 10.1007/978-981-4560-75-7_97-1

As an example, suppose a $\pm 2^\circ$ scan angle is required with minimum scan loss with a reflector having $D = 100\lambda$ and $f/D = 0.35$. What size of the feed array is required? Suppose the minimum spacing for elements in the feed array is $d_x = 0.55\lambda$. Equation 10 predicts an array width of $w_x \approx 3\lambda$, and as a result, Eq. 11 predicts the number of elements as $N_x = 5$. This compares with the results in Fig. 3 where for the 3×3 array, only about half this number of elements lie in the x-direction. When the number of array elements in the scan direction is increased to $N_x = 7$, which is predicted by Eq. 10 to achieve a scan range of about $\pm 3^\circ$, Fig. 9 shows that the scan angle in that plane has in fact increased to $> \pm 4^\circ$, which indicates Eq. 10 is a conservative estimate when the element spacing is close to 0.5λ. By judicious selection of the array element aperture field, however, the element spacing could be increased, and as a result, the number of array elements in the array feed should be able to be reduced from that predicted by Eq. 11.

Array Feed Excitation

When feeds are displaced from the central focal point, the resulting beam includes aberrations such as coma, and there is a loss of gain depending on the angle from boresight. This can be overcome to some extent by adjusting the array excitation to match the focal field, and with enough array elements, this can be done with several beams at one time or alternatively to create high performance steered beams. To obtain the excitation of a feed array more rigorously, consider an array of N apertures in the focal region of a reflector. A uniform plane wave from the direction $(\theta_\alpha, \phi_\alpha)$ $(\alpha = 1, \ldots, N_b)$, where N_b is the number of beams that are assumed incident on the reflector and hence to the array. By reciprocity, a beam in these directions produced when the array is excited appropriately. Suppose there are M modes in each aperture of the array of which there is access to only J of them through the beamforming network. Thus, there are $L = JN$ accessible ports of the MN available ports. The remaining modes will be excited indirectly through effects such as mutual coupling. As a consequence of the latter between the apertures of the array, the incident and reflected modes are related by a scattering matrix as follows:

$$\mathbf{b} = \mathbf{S}\,\mathbf{a}$$

where

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{MN} \end{bmatrix}$$

are all the inputs including those that are inaccessible such as for high-order modes.

The normalized power density radiated in a direction $(\theta_\alpha, \phi_\alpha)$ by the reflector antenna is given by

$$\eta(\theta_\alpha, \phi_\alpha, \mathbf{v}) = \frac{Q^* Q}{P_f} \tag{12}$$

where

$$Q = \mathbf{u}^T \mathbf{v},$$

DOI 10.1007/978-981-4560-75-7_97-1

$$\mathbf{v} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_L \end{bmatrix}$$

is a vector of the mode amplitudes at the L input ports and

$$P_f = \mathbf{v}^{\dagger} \mathbf{\Lambda} \mathbf{v}$$

is the total radiated power of the array where $\mathbf{\Lambda} = \mathbf{P} - \mathbf{S}^{\dagger}\mathbf{P}\mathbf{S}$,

$$\mathbf{P} = \begin{bmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{MN} \end{bmatrix}$$

in which p_i is the power contained in mode i. The T superscript indicates the matrix transpose, the asterisk * the complex conjugate, and † is the Hermitian conjugate. The row vector

$$\mathbf{u}^T = \mathbf{t}^T \mathbf{W} \tag{13}$$

includes the effect of mutual coupling as $\mathbf{W}$ is a matrix of dimension $MN\,x\,L$ between the L input ports and the MN output ports with elements given by

$$W_{ij} = \begin{cases} 1 + S_{ii} & ; \quad i = j = 1, 2, \ldots, JN \\ S_{ij} & ; \qquad \text{otherwise} \end{cases}$$

and $\mathbf{t}$ is a vector related to the fields radiated by every aperture in the far-field direction (θ_b, ϕ_b), which for the kth element in receive mode is given by (refer to Fig. 1)

$$t_k(\theta_\alpha, \phi_\alpha) = \frac{1}{4\sqrt{P_{inc}}} \iint_{\Sigma} dS' \;\; \mathbf{e}_k \cdot \mathbf{J}_s(\theta_\alpha, \phi_\alpha | \mathbf{x}') \tag{14}$$

where $\mathbf{e}_k$ is the electric field radiated by the kth output port when all other ports are removed; P_{inc} is the power incident in the plane wave from the direction (θ_α, ϕ_α); and $\mathbf{J}_s(\theta_\alpha, \phi_\alpha | \mathbf{x}')$ is the current induced at $\mathbf{x}'$ on the reflector that is nearest to the feed from an incident wave. An equivalent expression can also be obtained in the transmit mode from the far-field radiated by the reflector due to output port k.

An exact solution exists to Eq. 12 for the maximum gain in any of the beam directions (θ_α, ϕ_α) ($\alpha = 1, \ldots, N_b$). These beams can be sequential as part of a beam scan or randomly required in a search, for instance. In each case, there is a distinct excitation vector $\mathbf{v}_\alpha$ given by

$$\mathbf{v}_\alpha = \mathbf{\Lambda}^{-1} \mathbf{u}_\alpha{}^* \tag{15}$$

where $\mathbf{v}_\alpha$ is obtained from Eq. 15 for direction (θ_α, ϕ_α). If the mutual coupling is negligible or zero, $\mathbf{\Lambda} = \mathbf{P}$ and Eq. 15 reduces to $\mathbf{v}_\alpha = \mathbf{P}^{-1}\,\mathbf{t}_\alpha^*$, which is equivalent to saying that the input amplitudes are a conjugate match to the focal region fields. This is the approximation mentioned at the beginning of this section and for obtaining the results in Fig. 3.

DOI 10.1007/978-981-4560-75-7_97-1

The excitation could also be determined by numerical means such as through optimization of particular requirements on each beam other than for maximum gain, such as through envelope constraints on the sidelobes. In order to apply envelope constraints, define N_s specified points in the coverage region. At the jth point in the region corresponding to the direction (θ_i, ϕ_i) in the far-field, let

$c_{u,j}$ = upper beam level in direction j
$c_{L,j}$ = lower beam level in direction j.

Thus, it is required that

$$c_{L,j} \le \eta\left(\theta_j, \phi_j\right) \le c_{u,j}. \tag{16}$$

Many different methods exist for finding a solution to Eq. 16. One approach is the least pth technique (Bandler and Charalambous 1972). A least pth index is defined by

$$I = H\left[\sum_{f_i>0}\left(\frac{f_i}{H}\right)^q\right]^{1/q} \tag{17}$$

where $H = \max(f_i)$, $q = \mathrm{sgn}(H)p$, and κ is the set of specifications

$$\kappa = \begin{cases} \text{all} \quad f_i \;\; if\, H < 0 \\ \text{positive}\, f_i \quad if\, H > 0 \end{cases}.$$

Equation 17 has the advantage that it will be zero if all constraints are satisfied and then optimization may be terminated. Although, in principle, any p >1 is allowable, p = 2,4 and 10 have been found useful in shaped beam synthesis. If a solution is known approximately, a higher value of p can prove useful. Following an initial estimate of the array excitation coefficients either from Eq. 15 or an earlier search, I is further minimized by one of a number of search methods such as the genetic algorithm or gradient search such as the quasi-Newton method.

Another approach to determining the array excitation is also based on knowledge of the focal plane fields and is the technique initially described by Rudge and Withers (1971). This uses the Fourier transform relation between the aperture fields and the focal plane fields. In particular, a uniform aperture distribution results in a focal region distribution in a paraboloid of the form $J_1(u)/u$ where J_1 is the first-order Bessel function. Suppose the main reflector subtends a maximum half-cone angle ψ_c from the geometric focus. It can be shown that for off-axis incidence, as in scanning, the same relationship is maintained between the aperture and the focal point when the new focal point moves off-axis. The cone angle subtended by the reflector aperture as measured from the new focal point is constant with value given by $2\psi_c$. A new focal plane is defined to be normal to the line from the new focal point bisecting the angle $2\psi_c$. The path of free motion of the feed passes through the new focus as well as the edges of the aperture plane and follows a circle with a diameter

$$f\,\sec\psi_c\left[1+\left(\frac{D}{2f}\right)\right]^2. \tag{18}$$

Once a new focal plane has been obtained, the focal field is a Fourier transform of the aperture field whatever the angle of incidence. A second Fourier transform placed at the center of the new focal region will output a uniform amplitude and a phase that is the complex conjugate of the aperture field. Therefore,

DOI 10.1007/978-981-4560-75-7_97-1

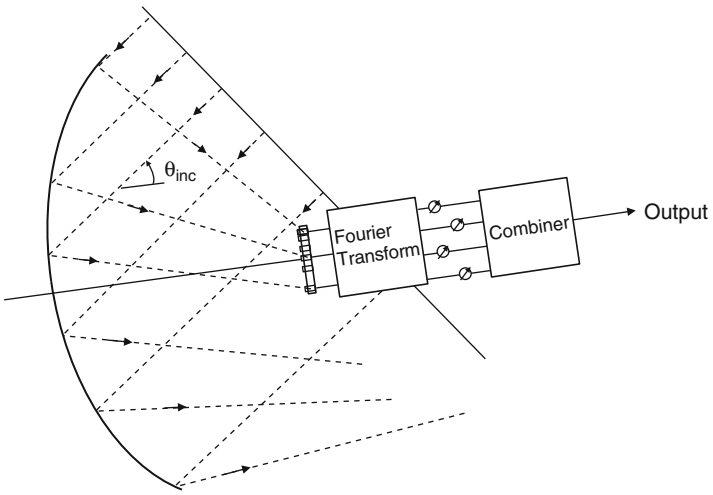


Fig. 6 Reflector with movable Fourier transform feed

the parabola performs as an ideal Fourier transformer of the off-axis incident wave for a feed at the new focal plane. Also, as the second transformer duplicates a uniform aperture field, only phase shifters are required to refocus the antenna. In practice, the second Fourier transformer is often implemented with a Butler matrix. The latter is a beamforming network that can steer the antenna beam in a specific direction in one plane depending on the input port. To create scanning in three dimensions, two Butler matrices can be combined. A single-plane scanner that utilizes a moving Fourier transform feed is illustrated in Fig. 6.

Sensitivity

Of particular importance in radio astronomy is the effective sensitivity of the instrument (Christiansen and Högbom 1969). For a single beam, the measure of sensitivity is the smallest detectable change in noise temperature, which is related to the loss in transmission line, η_t, and the overall change in the system noise temperature, $(\Delta T)_{\min}$, as follows:

$$\begin{aligned}(\Delta T_a)_{\min} &= \frac{1}{\eta_t}(\Delta T_a) \\ &= \alpha M \frac{T_{\text{sys}}}{\left(\sqrt{\Delta \nu \tau}\right)\eta_t}\end{aligned} \tag{19}$$

where T_{sys} is the system noise temperature, Δv is the bandwidth, τ is the integration time, and αM is a multiplier related to the type of detector (Christiansen and Högbom 1969) and selected from experience at a particular observatory. With a single beam telescope, only a fraction of the sky is occupied by the main beam. The remainder is called the stray factor. Therefore, over the main beam region

$$(\Delta T_b)_{\min} = \frac{1}{\eta_b}(\Delta T_a)$$

where η_b is the beam efficiency, which is the antenna efficiency for a single beam. Thus

$$(\Delta T_b)_{\min} = \alpha M \frac{T_{\text{sys}}}{\left(\sqrt{\Delta \nu \tau}\right)\eta_t \eta_b}. \tag{20}$$

DOI 10.1007/978-981-4560-75-7_97-1

The minimum detectable matched flux density is given by

$$k_B(\Delta T_b)_{\min} = A(\Delta S_m)_{\min}$$

where S_m in $Wm^{-2}Hz^{-1}$ is the flux density, ΔS_m is its change, A is the area of the aperture, and $k_B = 1.38.10^{-23}\ WHz^{-1}\text{degree}^{-1}$ is Boltzmann's constant. As a result,

$$(\Delta S_m)_{\min} = \alpha M \frac{k_B}{\left(\sqrt{\Delta v \tau}\right)\eta_t} \frac{T_{\text{sys}}}{A_{\text{eff}}} \tag{21}$$

where $A_{\text{eff}} = \eta_b A$ is the effective area. Equation 21 shows that the sensitivity is proportional $T_{\text{sys}}/A_{\text{eff}}$, which is dependent only on the antenna system.

For a multibeam telescope that has N beams, it is possible to extract N simultaneous pieces of information from a sky measurement. If the sky measurement area is Ω_{map} steradians for the multiple beams and Ω_b is the area mapped with a single beam, the survey speed time is $\tau = \left(\Omega_b/\Omega_{\text{map}}\right)\tau_s$ where τ_s is the total time taken to complete the survey of the sky with a single beam. Hence, the surveying sensitivity is defined as

$$(\Delta SS_m)_{\min} = \sqrt{\left(\Omega_b/\Omega_{\text{map}}\right)}\,(\Delta S_m)_{\min}.$$

If a map or image consists of N beams, then $\Omega_{\text{map}} \approx N\Omega_b$ and

$$(\Delta SS_m)_{\min} = \sqrt{1/N}\,(\Delta S_m)_{\min}. \tag{22}$$

Therefore, an N-beam image-forming reflector has $\sqrt{N}$ times the surveying sensitivity of a single beam instrument. However, the survey speed time is $\tau \approx (1/N)\tau_s$.

Single-Element Feeds and Associated Beamforming Networks

In some applications, clusters of feeds have proved useful for scanning over directions that can be established a priori. In that instance, feed elements of the array have their own feed network. Under computer or autonomous system control, the excitation for a cluster, or a single feed is loaded individually from one beam position to the next. One such application where one horn at a time of the array has been used very successfully is in radio astronomy. Another approach used in satellite communications is that the beamforming network has been segmented. Each segment feeds a sub-array that produces part of the shaped beam. By combining sub-arrays or switching between them, an overall shaped beam can be produced with some re-configurability that can be changed quite rapidly.

The single horn at a time in an array has been used in radio astronomy to avoid a lossy beamforming network and to maintain a low noise temperature. If the array spacing is chosen correctly, the center spacing of adjacent beams from the telescope is about a beamwidth apart. Moving the telescope through half a beamwidth and interleaving the field of view with beams fill in the coverage with relatively little loss of gain at the crossover. In the same way, several horns in the cluster could be combined, but again, this is not used because of losses and higher noise temperature. As will be seen in section "Phased Array Feeds," recent developments have enabled feeds to be developed where low-loss beamforming can be achieved.

DOI 10.1007/978-981-4560-75-7_97-1

In relation to multibeam feeds creating single beams, the Parkes Observatory in Australia, the Arecibo Observatory in Puerto Rico, and the Lovell radio telescope in the UK all have used such arrays in L- and C-bands. The feed horns and spacing of the elements in these arrays are different in each case due to different focal lengths of the reflectors and different frequency bands. However, where possible the horn input diameters have been chosen to be the same to enable the same array of cooled low-noise receivers to be shared. At the observatories mentioned, reflectors are prime focus fed, and therefore, for efficient illumination, wide beamwidth elements are required. Both the Parkes and Lovell telescopes use prime-focus feeds, although the latter dish is deeper (the f/D of Lovell is 0.3 while for Parkes $f/D = 0.41$). The elements of L- and C-feed arrays for Parkes are stepped circular horns. At C-band, the aperture diameter is about 1.05λ, and element spacing in the array is 1.16λ. They were based on the Arecibo multibeam L-band feed array, which was a refinement of the first Parkes L-band design (Bird 1994). The stepped circular horn is unsuitable for a deep dish such as the Lovell telescope, and therefore, coaxial horns, which have a wider beamwidth, were employed on Lovell (Bird 1997).

The main parameters in the design of a fixed multibeam feed are the input waveguide diameter, the aperture geometry to achieve the required edge illumination, the peak cross-polarization level, and a matching network to give a good input match to the receiver. The maximum dimension of elements of the array should be small enough so that there is only about 6 dB between successive beams, and loss of gain at the edge of the outermost feeds is not too significant. This limits the aperture diameter to be $0.8 - 1.2\lambda$. The close spacing between array elements eliminates from consideration many well-known techniques for improving pattern symmetry using the flange such as parasitic rings or back-set corrugated flanges. Nevertheless, from the properties of circular or coaxial apertures, the beam is quite axisymmetric, and there is low cross polarization at a frequency given by $kb \approx 1.25(a/b)^2 - 3.3(a/b) + 3.6$ where a and b are the inner and outer conductor radii. In addition, through the use of steps and irises inside the waveguide, successful feed elements have been produced.

The feed elements in satellite communications have ranged from uniform circular waveguides to tapered rectangular horns. The latter are often preferred in order to maximize the packing density. Also, the latter produces a uniform illumination in the E-plane, and with the excitation of high-order modes or dielectric loading, the aperture field in the H-plane can also be made almost uniform. The element spacing in shaped beams for satellites can be greater as the coverage of a field of view is of importance. The feed arrays can consist of more than 200 elements. The associated beamforming network needs to provide the correct amplitude and phase to each element, and this can be achieved with low-loss power dividers and lengths of transmission line.

To carry out the design of the shaped beam, the reflector, the array, and the beamforming network, a suite of computer programs should be available for the purpose. For instance, the software should be able to analyze arrays in a rigorous manner taking into account such effects as mutual coupling by numerical methods or mode-matching techniques. The latter are sometimes preferred for the feeder because they accurately model the internal geometry of circular and rectangular horns as a series of stepped waveguides and can also separately include the effects of mutual coupling between horns. To minimize extraneous signals reaching the feed and introducing noise, the array is often terminated in a large ground plane. The effect of a finite size array can be included in the radiation pattern by the method of physical theory of diffraction or by the method of moments.

The analysis of the reflector is usually based on the method of physical optics, and this is linked to the feed array software to determine the shaped beams. The beamforming network can be designed separately from the reflector and feed array using techniques based on numerical methods such as finite elements. The results obtained in this way are very accurate as they are able to include variations that are part of the fabrication such as surface imperfections, bends, and twists.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_97-1

Phased Array Feeds

Phased array feeds (PAF) have been used for feeds for rapid beam steering and in situations usually in reception where the system noise temperature has not been important. The feeding network used to combine the elements of the array in the appropriate amplitude and phase has a significant loss, and this contributes to the overall noise temperature of the antenna. Applications of array feeds have varied widely from the production of multiple beams by individual elements of the array, to a fixed shaped beam such as from a geostationary satellite, and to a fully steered beam with a phased array for rapid tracking such as for missile tracking. In all cases, the array elements must be excited in the correct amplitude and phase by a beamforming network. These can be constructed from microstrip line, coaxial components, or waveguide structures depending on the requirement of low loss.

PAFs have received increased attention over the past decade, particularly for high-sensitivity receiving applications to radio astronomy (Fisher and Bradley 2000). PAFs produce multiple simultaneous antenna beams throughout a contiguous field of view (FoV). They use a dense array of antenna elements, typically spaced by less than half the minimum wavelength. This spacing is important for reception of information in incident fields and control of beam radiation patterns. Various element types have been used, including dipole, tapered slot or Vivaldi, and checkerboard. The PAF beams are formed typically by digital beamforming. Thus, the individual array-element signals are digitized and combined by complex weighting and summing independently in multiple frequency channels. This process allows control of reflector illumination, spillover, and other beam properties such as spacing, sidelobes, and cross polarization. The close spacing of the array elements implies significant mutual coupling, and this requires co-design of the array elements, front-end amplifiers, and beamforming. Beams can be optimized to maximize sensitivity or signal-to-noise ratio and produce wide FoV. Some applications also require operation over large frequency ranges >2:1. A significant benefit for radio astronomy is increased survey speed. In this section, introductory formulations of PAF concepts are provided and also some illustrative examples of recent PAF systems.

Incident Field Representation

The design of PAFs for high-sensitivity receiving application requires use of models of the incident field. The PAF incident field contains both desired signal and noise components and depends on the geometry of the reflectors used to concentrate the signal energy.

The incident field is represented in terms of a spectrum of plane waves produced by an equivalent current in the far-zone of the antenna. Thus, the incident electric field is

$$\mathbf{E}(\mathbf{r}) = \frac{-jk\eta_o\, e^{-jk\,r'}}{4\pi\, r'} \int_{S'} dS' \;\; \left[J_p(\mathbf{r}')\hat{\mathbf{p}}(\mathbf{r}') + J_q(\mathbf{r}')\hat{\mathbf{q}}(\mathbf{r}')\right] e^{jk\mathbf{r}\cdot\hat{\mathbf{r}}'} \tag{23}$$

where $\mathbf{r}$ is the position vector of the observation point, the integral is over a spherical surface S' in the far-zone of the antenna, dS' is the element of area on S', $\mathbf{r}'$ is the position vector of the point on S', $\hat{\mathbf{r}}' = \mathbf{r}'/r', r' = |\mathbf{r}'|$, $\hat{\mathbf{p}}$ and $\hat{\mathbf{q}}$ are orthogonal unit vectors corresponding to co- and cross polarization, respectively, each tangential to S', and J_p and J_q are the corresponding current components. The current components are random variables with spatial independence, and they are represented in terms of temperatures according to

$$\mathrm{E}\begin{bmatrix} J_p(r_1')\overline{J}_p(r_2') & J_p(r_1')\overline{J}_q(r_2') \\ J_q(r_1')\overline{J}_p(r_2') & J_q(r_1')\overline{J}_q(r_2') \end{bmatrix} = \delta(r_1' - r_2')\,\frac{8k_B}{\eta_o}\begin{bmatrix} T_{pp}(r_1') & T_{pq}(r_1') \\ T_{qp}(r_1') & T_{qq}(r_1') \end{bmatrix} \tag{24}$$

DOI 10.1007/978-981-4560-75-7_97-1

where E(.) is the expectation function, r_1' and r_2' are any two points on S', δ is the delta function with respect to dS', and the terms in the matrix on the right are equivalent temperatures in degrees Kelvin.

PAF Sensitivity Formulation

Figure 7 shows a Norton-equivalent network representation of the PAF and first-stage low-noise amplifiers (LNAs). Mutual coupling between the array ports is represented by the admittance matrix Y_A. Fields incident on the antenna and noise associated with loss in the antenna are represented by equivalent sources I_i each equal to the current at the corresponding port when all ports are short circuited. Noise generated in the LNA is represented by an equivalent source comprising the correlation admittance Y_c and uncorrelated voltage and current sources with equivalent resistance R_n and conductance G_n, respectively. All LNAs in the array are of a common design. Subsequent stages of the receiver then produce beamformed outputs that take the form

$$V_{\text{beam}} = w^T V^{LNA} \tag{25}$$

where V^{LNA} and w are column vectors of LNA output voltage phasors and weighting coefficients, respectively.

A sensitivity measure that is related to Eq. 21 is expressed as

$$\frac{A_{\text{eff}}}{T_{\text{sys}}} = \frac{k_B}{S_m}\ \text{SNR} \tag{26}$$

where SNR is the signal-to-noise ratio of the beamformed response of the antenna given by

$$\text{SNR} = \frac{E\left(|V_{\text{beam}}^{\text{signal}}|^2\right)}{E\left(|V_{\text{beam}}^{\text{noise}}|^2\right)}, \tag{27}$$

where $E(.)$ is the expectation function. In Eq. 26 S_m is the power density of the plane-wave field incident on the antenna.

An expression for the array sensitivity Eq. 26 can be derived from the network of Fig. 7. The SNR is readily found as

$$\text{SNR} = \frac{1}{8k_B T_0} \frac{|\left(I^{\text{signal}}\right)^T \phi|^2}{\overline{\phi}^T G_{\text{tot}} \phi} \tag{28}$$

where T_0 is the thermodynamic temperature of the antenna and LNAs, I^{signal} is the vector of short-circuit currents representing the signal field incident on the antenna, the overbar denotes the complex conjugate, and the vector ϕ is related to the beamformer weight vector w through

$$\phi = P^T w \qquad P = Q(Y_A + Y_{in}U)^{-1} \tag{29}$$

where Y_{in} is the input admittance of the output-terminated LNA, U is the identity matrix, and Q is the matrix that maps the LNA input voltages into the LNA output voltages that are selected for beamforming. If all LNA output voltages are selected for beamforming, then Q is a diagonal matrix. Figure 21 (below) illustrates another important case where the LNAs are paired and the differential output voltages of the

DOI 10.1007/978-981-4560-75-7_97-1

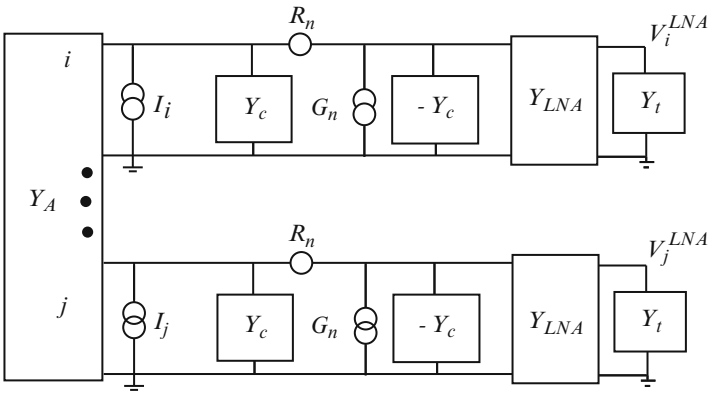


Fig. 7 Norton-equivalent network of the array and low-noise amplifiers

pairs are beamformed, in which case Q is an M/2 × M matrix where M/2 is the number of LNA pairs. This case is considered in the application section.

The matrix G_{tot} in Eq. 28 can be expressed as

$$G_{\text{tot}} = G_{\text{ext}} + G_{\text{loss}} + G_{\text{rec}} \tag{30}$$

where the contributions are due, respectively, to the external noise incident on the antenna, the noise associated with loss in the antenna, and the noise produced by the LNAs. These terms may be expressed as

$$(G_{\text{ext}})_{i,j} = \frac{1}{T_0\,\eta_0}\int_{S'} dS' \left(\overline{\mathbf{E}}_{i,p} \quad \overline{\mathbf{E}}_{i,q}\right) T \begin{pmatrix} \mathbf{E}_{j,p} \\ \mathbf{E}_{j,q} \end{pmatrix} \qquad T = \begin{bmatrix} T_{pp} & \overline{T}_{pq} \\ \overline{T}_{qp} & T_{qq} \end{bmatrix} \tag{31}$$

$$G_{\text{loss}} = G_A - G_{\text{ext}}\Big|_{T=\begin{bmatrix} T_0 & 0 \\ 0 & T_0 \end{bmatrix}} \tag{32}$$

$$G_{\text{rec}} = (F_{\min} - 1)G_A + \quad N\left(\overline{Y}_A^T - \overline{Y}_{\min}U\right)G_{\min}^{-1}(Y_A - Y_{\min}U) \tag{33}$$

where $\mathbf{E}_i$ is the electric field radiated by the array when 1 V exists at port i while all other ports are short circuited and $G_A = \Re Y_A$. Equation 31 follows from the incident field representation Eq. 23 and the reciprocity theorem. In Eq. 33, $F_{\min}$ is the minimum noise figure of the LNA, obtained with the LNA input connected to source of admittance $Y_{\min} = G_{\min} + jB_{\min}$ and $N = R_n G_{\min}$. The quantities $F_{\min}$ and N are Lange invariants, so named because they remain constant under the addition of any lossless impedance-transforming networks to the input or output circuits of the LNA. Thus the noise-match admittance $Y_{\min}$ is the main parameter that can be varied through design of the LNA. The Lange constants are related by

$$(F_{\min} - 1)/4 < N < (F_{\min} - 1)/2. \tag{34}$$

The left side of the inequality in Eq. 34 is a general result, whereas the right inequality has been derived for LNAs composed of field effect or bipolar junction transistors.

In Eq. 28, the vector I^{signal} of short-circuit currents representing the signal field incident on the antenna can be computed from the reciprocity relation

DOI 10.1007/978-981-4560-75-7_97-1

$$I_i^{\text{signal}} = \hat{\mathbf{p}} \cdot \mathbf{E}_i(\mathbf{r}') \sqrt{2\eta_0 S_m} \frac{4\pi |\mathbf{r}'| \exp(jk|\mathbf{r}'|)}{-jk\eta_0}. \tag{35}$$

Equation 35 applies to a plane wave incident from the direction of the far-field position vector $\mathbf{r}'$, with electric field polarized in the direction of the unit vector $\hat{\mathbf{p}}$ and power density S_m.

Array and Amplifier Matching

Of great interest in practice is the maximum SNR obtainable by optimizing the beamforming coefficients and the LNA noise-match admittance $Y_{\min}$. Equations for the optimum values can be found by equating to zero the derivatives of Eq. 26, where the SNR function is given by Eq. 28, with respect to the real and imaginary parts of $Y_{\min}$ and the elements of the beamforming weight vector w. This procedure results in the equations

$$B_{\min} = \frac{\overline{\phi}^T B_A \phi}{\overline{\phi}^T \phi} \quad G_{\min} = \sqrt{\frac{\overline{\phi}^T \overline{Y}_A^T Y_A \phi}{\overline{\phi}^T \phi} - B_{\min}^2} \tag{36}$$

and

$$w = \left(\overline{P} G_{\text{tot}} P^T\right)^{-1} \overline{P} \overline{I}^{\text{signal}} \tag{37}$$

where $B_A = \Im Y_A$. Equations 36 and 37 are readily solved in practice by a small number of iterations.

Spatial-Filtering Interpretation of PAF Beamforming

To show how high-sensitivity operation of the array can be obtained when the elements are closely spaced with large mutual coupling, the equations relating the currents and voltages at the array ports are recast as follows:

$$\begin{aligned} I_{m,n}^x &= \sum_{k,l} Y_{m-k,n-l}^{x,x} V_{k,l}^x + \sum_{p,q} Y_{m-p,n-q}^{x,y} V_{p,q}^y \\ I_{m,n}^y &= \sum_{k,l} Y_{m-k,n-l}^{y,x} V_{k,l}^x + \sum_{p,q} Y_{m-p,n-q}^{y,y} V_{p,q}^y. \end{aligned} \tag{38}$$

In Eq. 38, $I_{m,n}^x$ and $V_{m,n}^x$ are the current and voltage at the x polarized array port at the point $(x,y) = (m,n)\delta/2$, where δ is the spacing of the like-polarized ports and (m, n) are the even positive and negative integers or zero. For the y polarized ports, the summations in Eq. 38 run over the odd positive and negative integers. Equation 38 is possible since, in the regular periodic array, the mutual admittances are functions only of the vector separation of the array ports. Equation 38 becomes the algebraic equations

$$\begin{bmatrix} \hat{I}^x(u,v) \\ \hat{I}^y(u,v) \end{bmatrix} = \begin{bmatrix} \hat{Y}^{x,x}(u,v) & \hat{Y}^{x,y}(u,v) \\ \hat{Y}^{y,x}(u,v) & \hat{Y}^{y,y}(u,v) \end{bmatrix} \begin{bmatrix} \hat{V}^x(u,v) \\ \hat{V}^y(u,v) \end{bmatrix} \tag{39}$$

where the Fourier transforms are defined as

$$\hat{I}^x(u,v) = \sum_{m,n} I_{m,n}^x e^{-jk(mu+nv)\delta/2}. \tag{40}$$

DOI 10.1007/978-981-4560-75-7_97-1

The SNR defined by Eq. 27 can then be re-expressed as

$$\mathrm{SNR} = \frac{1}{8k_B T_0} \frac{\left| \iint dudv \ \left(\hat{I}^{\mathrm{signal}}(-u, -v) \right)^T \hat{\phi}(u, v) \right|^2}{\iint dudv \ \hat{\phi}^{\dagger}(u, v) \hat{G}_{\mathrm{tot}}(u, v) \hat{\phi}(u, v)} \tag{41}$$

and, analogous to Eq. 30,

$$\hat{G}_{\mathrm{tot}}(u, v) = \hat{G}_{\mathrm{ext}}(u, v) + \hat{G}_{\mathrm{loss}}(u, v) + \hat{G}_{\mathrm{rec}}(u, v) \tag{42}$$

and the LNA contribution is given by

$$\begin{aligned} \hat{G}_{\mathrm{rec}}(u, v) = (F_{\min} - 1) \left[\hat{Y}(u, v) + \hat{Y}^{\dagger}(u, v) \right] / 2 \\ + N \left[\hat{Y}^{\dagger}(u, v) - \overline{Y}_{\min} U \right] G_{\min}^{-1} \left[\hat{Y}(u, v) - Y_{\min} U \right] \end{aligned} \tag{43}$$

where $\hat{Y}(u, v)$ is a 2×2 matrix in Eq. 43 and U is the 2×2 identity matrix. With optimum beamforming for maximum SNR, the optimum value of the spectral function in Eq. 41 is readily obtained as

$$\hat{\phi}(u, v) = \hat{G}_{\mathrm{tot}}^{-1}(u, v) \overline{\hat{I}}^{\mathrm{signal}}(-u, -v) \tag{44}$$

with the maximum SNR given by

$$\mathrm{SNR} = \frac{1}{8k_B T_0} \iint dudv \ \hat{I}^{\mathrm{signal}}(u, v)^T \hat{G}_{\mathrm{tot}}^{-1}(u, v) \overline{\hat{I}}^{\mathrm{signal}}(u, v). \tag{45}$$

The admittance function and short-circuit current given in Eqs. 39 and 41 have been computed by numerically solving the integral equation for the current induced on the self-complementary array illustrated in Fig. 8. Figure 9 shows the minimum eigenvalue of the matrix $\hat{G}_{\mathrm{rec}}(u, v)$ and the corresponding integrand in Eq. 45. For comparison, Fig. 10 gives the corresponding results for an array with twice the element density. Increasing the density results in more LNA noise in the spatial frequency spectrum. However, the additional noise occurs outside the spatial frequency bandwidth of the signal, and therefore the resulting sensitivity is high in both cases, provided that accurate optimal beamforming is applied. A detailed design example of a connected array design that was developed for the Australian SKA Pathfinder (ASKAP) application is described in section "Australian SKA Pathfinder (ASKAP) PAF."

Survey Speed of a PAF Interferometer

The survey speed of next-generation radio telescopes such as the Square Kilometre Array (SKA) will be increased greatly by the use of phased array feeds (PAFs). In PAFs, the signals of a dense array of antenna elements are digitized and beamformed, producing multiple simultaneous antenna beams of high sensitivity throughout the field of view. In this section, an early analysis of a connected array PAF for ASKAP (Bunton and Hay 2010) is extended to include optimum linear combining of PAF beam correlations with

DOI 10.1007/978-981-4560-75-7_97-1

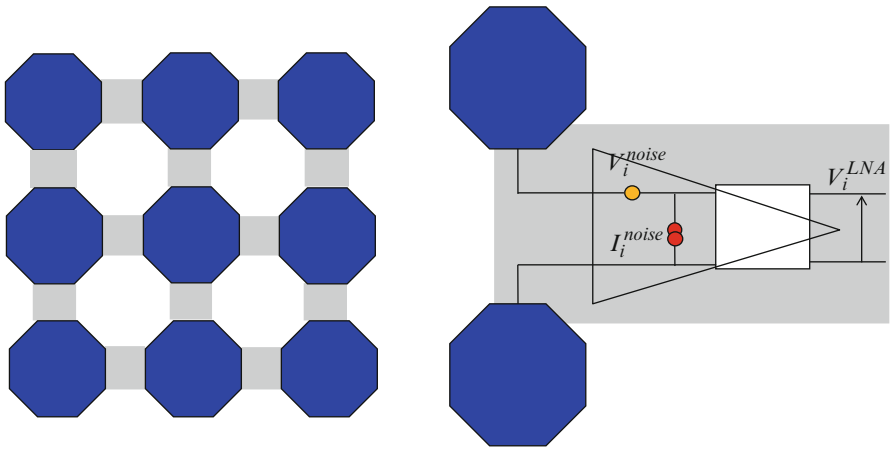


Fig. 8 Schematic of self-complementary patch array with adjacent patch corners connected by a low-noise amplifier (LNA). The LNA produces noise, represented by equivalent current and voltage sources at the input, and its effect on signals is characterized by an admittance matrix Y

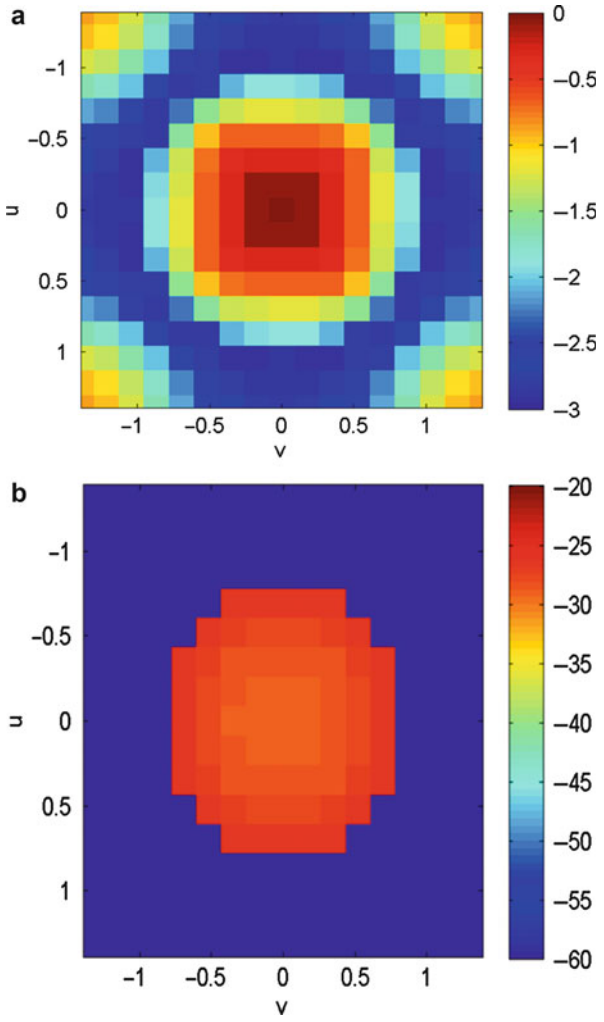


Fig. 9 Minimum eigenvalue of spectral-domain receiver noise matrix given in Eq. 43(**a**) and optimum spectral-domain sensitivity distribution given by the integrand in Eq. 45(**b**) for array with 90 mm element spacing at 1.2GHz. The quantities plotted are magnitudes in dB.

DOI 10.1007/978-981-4560-75-7_97-1

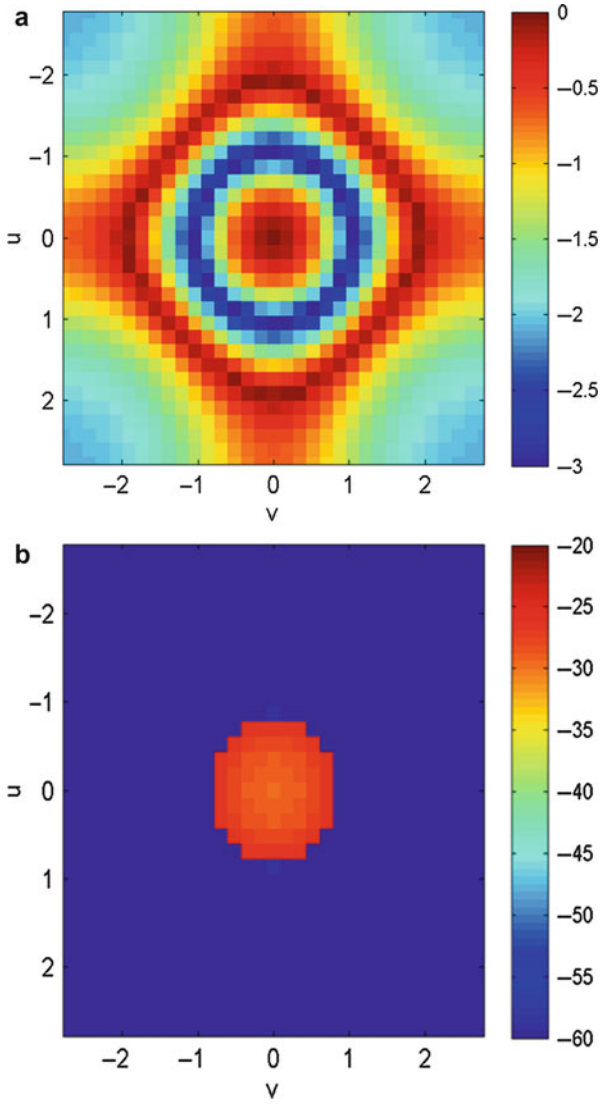


Fig. 10 Results corresponding to Fig. 9 with 45 mm element spacing at 1.2GHz

DOI 10.1007/978-981-4560-75-7_97-1

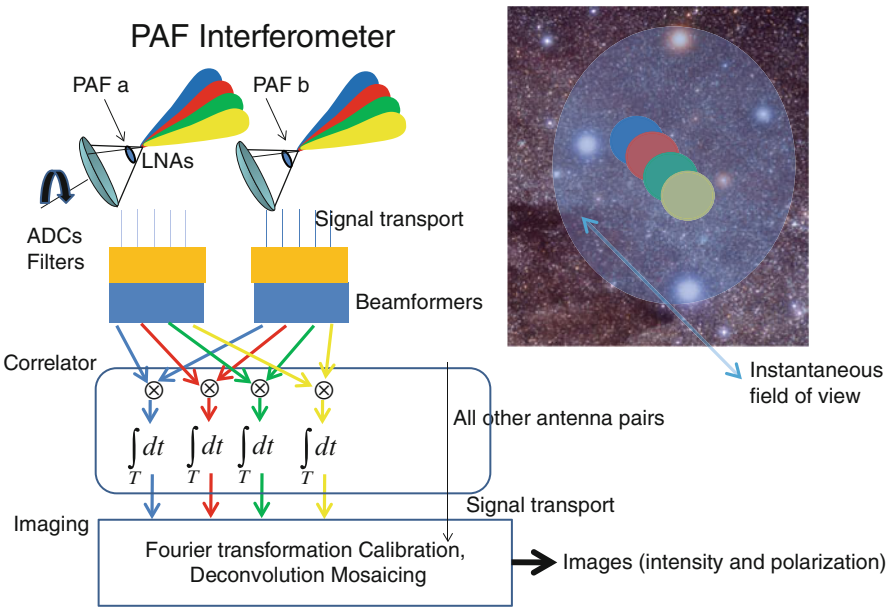


Fig. 11 Schematic of the PAF interferometer radio telescope

maximum selective sensitivity to all coherence parameters of the incident field. As illustrated in Fig. 11, the ASKAP telescope is an interferometer using PAFs in each antenna. The PAFs expand the field of view by creating multiple simultaneous antenna beams. The simultaneous PAF beams have correlated noise due to common low-noise amplifiers. A general treatment of this noise is provided in the next section.

Picture: courtsey of CSIRO Australia

Formulation

Problem Definition Let $\mathbf{V}^a$ and $\mathbf{V}^b$ be column vectors containing the element voltages at PAFs of antennas a and b, respectively. These vectors are regarded as random variables. Each PAF produces a set of beams in polarization pairs or

$$\tilde{\mathbf{V}}_i^a = \begin{pmatrix} V_{i,1}^a \\ V_{i,2}^a \end{pmatrix} = \left(\mathbf{w}_i^a\right)^T \mathbf{V}^a \quad \mathbf{w}_i^a = \begin{pmatrix} \mathbf{w}_{i,1}^a & \mathbf{w}_{i,2}^a \end{pmatrix} \tag{46}$$

where $\mathbf{w}_{i,1}^a$ are column vectors of complex weights and i is an index of beam directions. It is assumed that the correlator produces sample averages

$$\mathrm{A}\left[\tilde{\mathbf{V}}_i^a \left(\tilde{\mathbf{V}}_i^b\right)^\dagger\right] = \frac{1}{M}\sum_{m=1}^{M} \tilde{\mathbf{V}}_{i,m}^a \left(\tilde{\mathbf{V}}_{i,m}^b\right)^\dagger \tag{47}$$

of M time samples of the 2×2 cross products of antenna a and b beam voltages at each beam direction i, with the sample vectors $\tilde{\mathrm{V}}_{i,m}^a$ independent in the sample index m. The correlator outputs Eq. 47 have expected values

$$\hat{\mathrm{A}}\left(\tilde{\mathbf{V}}_i^a \left(\tilde{\mathbf{V}}_i^b\right)^\dagger\right) = E\left[\mathrm{A}\left(\tilde{\mathbf{V}}_i^a \left(\tilde{\mathbf{V}}_i^b\right)^\dagger\right)\right]$$

$$= \int d\Omega \;\; \mathbf{J}_i^a \mathbf{C} \left(\mathbf{J}_i^b\right)^\dagger \exp\left[jk\left(\mathbf{r}^a - \mathbf{r}^b\right) \cdot \left(\hat{\mathbf{s}} - \hat{\mathbf{s}}_0\right)\right] \tag{48}$$

where the integral is over all directions of propagation of plane-wave components of the incident field, $\hat{\mathrm{s}}$ is the unit vector in the direction from which the plane wave arrives,

$$\mathbf{C} = E\begin{bmatrix} e_1\overline{e}_1 & e_1\overline{e}_2 \\ e_2\overline{e}_1 & e_2\overline{e}_2 \end{bmatrix} \tag{49}$$

is the coherence matrix of the electric-field components e_1 and e_2 in orthogonal directions p and q, respectively, orthogonal to $\hat{\mathbf{s}}$, $\mathbf{r}^a$ is the position vector of antenna a, $\hat{\mathbf{s}}_0$ represents a reference direction in the field of view, and $d\Omega$ is the element of solid angle. In Eq. 48

$$\mathbf{J}_i^a = \begin{bmatrix} J_{i,1,1}^a & J_{i,1,2}^a \\ J_{i,2,1}^a & J_{i,2,2}^a \end{bmatrix} = \begin{pmatrix} \left(\mathbf{w}_{i,1}^a\right)^T \\ \left(\mathbf{w}_{i,2}^a\right)^T \end{pmatrix} \begin{pmatrix} \mathbf{V}_1^a & \mathbf{V}_2^a \end{pmatrix} \tag{50}$$

where $\mathbf{V}_1^a$ and $\mathbf{V}_2^a$ are column vectors of element voltages produced at the PAF of antenna a by reference waves arriving from the direction $\hat{\mathbf{s}}$ with electric fields of 1 V/m polarized in the directions p and q, respectively. The coherence matrix $\mathbf{C}$ in Eq. 49 satisfies

$$\mathbf{C} = \mathbf{U}^\dagger \mathbf{C}' \mathbf{U} \quad \mathbf{U} = \begin{bmatrix} U_{1,1} & U_{1,2} \\ U_{2,1} & U_{2,2} \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \tag{51}$$

where $\mathbf{C}'$ is the coherence matrix of the field components in any other orthogonal directions p' and q' obtained by rotating p and q by any angle ϕ.

The correlator outputs in Eq. 47 also contain noise. The main sources of such noise are the PAF front-end amplifiers, the ground around the antennas, and the noise associated with loss in the antennas. This noise is large compared to astronomical signals. As a first step in estimating the coherence matrices of the sources in the field of view, the aim is to find a linear transformation of the $A[\tilde{\mathrm{V}}_i^a(\tilde{\mathrm{V}}_i^b)^\dagger]$ that approximates

the coherence matrix $\mathbf{C}'$ of the point source with minimum error due to the noise. As a result, a 2×2 response matrix is given by

$$\mathbf{R}^{a,b} = \begin{bmatrix} R_{1,1}^{a,b} & R_{1,2}^{a,b} \\ R_{2,1}^{a,b} & R_{2,2}^{a,b} \end{bmatrix} \tag{52}$$

which satisfies the requirement for the expected value

$$\hat{\mathbf{R}}^{a,b} = \phi^{a,b}\mathbf{C}' \quad \phi^{a,b} = \exp\left(jk\left(\mathbf{r}^a - \mathbf{r}^b\right) \cdot \hat{\mathbf{s}}\right) \tag{53}$$

and with each element a linear combination of all elements of the $A[\tilde{\mathbf{V}}_i^a(\tilde{\mathbf{V}}_i^b)^\dagger]$. The general linear transformation can be expressed as

$$\mathbf{R}^{a,b} = \sum_i \begin{bmatrix} \overline{\mathbf{L}}_i^{1,1} \bullet A\left[\tilde{\mathbf{V}}_i^a\left(\tilde{\mathbf{V}}_i^b\right)^\dagger\right] & \overline{\mathbf{L}}_i^{1,2} \bullet A\left[\tilde{\mathbf{V}}_i^a\left(\tilde{\mathbf{V}}_i^b\right)^\dagger\right] \\ \overline{\mathbf{L}}_{i,}^{2,1} \bullet A\left[\tilde{\mathbf{V}}_i^a\left(\tilde{\mathbf{V}}_i^b\right)^\dagger\right] & \overline{\mathbf{L}}_i^{2,2} \bullet A\left[\tilde{\mathbf{V}}_i^a\left(\tilde{\mathbf{V}}_i^b\right)^\dagger\right] \end{bmatrix} \tag{54}$$

where

$$\mathbf{L}_i^{\lambda,\mu} = \begin{bmatrix} L_{i,1,1}^{\lambda,\mu} & L_{i,1,2}^{\lambda,\mu} \\ L_{i,2,1}^{\lambda,\mu} & L_{i,2,2}^{\lambda,\mu} \end{bmatrix} \tag{55}$$

are 2×2 matrices, the overbar denotes the complex conjugate, and the dot denotes summation of the products of all pairs of corresponding elements. The expected value of the elements of $\mathbf{R}^{a,b}$ are readily obtained as

$$\hat{R}_{\lambda,\mu}^{a,b} = \sum_i \overline{\mathbf{L}}_i^{\lambda,\mu} \bullet \hat{\mathrm{A}}\left[\tilde{\mathbf{V}}_i^a\left(\tilde{\mathbf{V}}_i^b\right)^\dagger\right]. \tag{56}$$

Using $\mathbf{X} \bullet \mathbf{Y} = \mathrm{Tr}(\mathbf{X}^t\mathbf{Y}) = \mathrm{Tr}(\mathbf{Y}\mathbf{X}^t)$, where $\mathrm{Tr}(x)$ is the trace of x and the identity is applicable to any matrices $\mathbf{X}$ and $\mathbf{Y}$ of equal size, the elements of $\hat{\mathbf{R}}^{a,b}$ can be recast as

$$\begin{aligned} \hat{R}_{\lambda,\mu}^{a,b} &= \left[\sum_i \left(\mathbf{J}_i^a\right)^T \overline{\mathbf{L}}_i^{\lambda,\mu} \overline{\mathbf{J}}_i^b\right] \bullet \mathbf{C} \\ &= \left[\sum_i \left(\mathbf{J}_i^a\right)^T \overline{\mathbf{L}}_i^{\lambda,\mu} \overline{\mathbf{J}}_i^b\right] \bullet \left(\mathbf{U}^\dagger \mathbf{C}' \mathbf{U}\right) \\ &= \left\{\overline{\mathbf{U}}\left[\sum_i \left(\mathbf{J}_i^a\right)^T \overline{\mathbf{L}}_i^{\lambda,\mu} \overline{\mathbf{J}}_i^b\right] \mathbf{U}^T\right\} \bullet \mathbf{C}' \end{aligned} \tag{57}$$

times $\phi^{a,b}$. The transformation requirement Eq. 53 can therefore be expressed by the equations

DOI 10.1007/978-981-4560-75-7_97-1

$$\begin{aligned}
\sum_i \left[\left(\mathbf{J}_i^a\right)^\dagger \mathbf{L}_i^{1,1} \mathbf{J}_i^b \right] &= \mathbf{U}^\dagger \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \mathbf{U} \\
\sum_i \left[\left(\mathbf{J}_i^a\right)^\dagger \mathbf{L}_i^{1,2} \mathbf{J}_i^b \right] &= \mathbf{U}^\dagger \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{U} \\
\sum_i \left[\left(\mathbf{J}_i^a\right)^\dagger \mathbf{L}_i^{2,1} \mathbf{J}_i^b \right] &= \mathbf{U}^\dagger \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \mathbf{U} \\
\sum_i \left[\left(\mathbf{J}_i^a\right)^\dagger \mathbf{L}_i^{2,2} \mathbf{J}_i^b \right] &= \mathbf{U}^\dagger \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{U}.
\end{aligned} \tag{58}$$

Equation 58 can be written compactly as

$$\sum_{j,\alpha,\beta} P_{(x,y),(j,\alpha,\beta)} L_{j,\alpha,\beta}^{\lambda,\mu} = Y_{(x,y)}^{\lambda,\mu} \tag{59}$$

where

$$P_{(x,y),(j,\alpha,\beta)} = \overline{J}_{j,\alpha,x}^a J_{j,\beta,y}^b Y_{(x,y)}^{\lambda,\mu} = \left(\boldsymbol{\delta}_x\right)^T \mathbf{U}^\dagger \boldsymbol{\delta}_\lambda \left(\boldsymbol{\delta}_\mu\right)^T \mathbf{U} \boldsymbol{\delta}_y \tag{60}$$

$$\boldsymbol{\delta}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \boldsymbol{\delta}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The aim is to solve Eq. 59 and minimize the variance due to noise in each element of $\mathbf{R}^{a,b}$.

Expression for Variance of the Coherence Matrix Estimate The variances are obtained as follows. From Eq. 57, it can be found that

$$R_{\lambda,\mu}^{a,b} - \hat{R}_{\lambda,\mu}^{a,b} = \sum_i \overline{\mathbf{L}}_i^{\lambda,\mu} \bullet \left\{ \mathrm{A}\left[\tilde{\mathbf{V}}_i^a \left(\tilde{\mathbf{V}}_i^b \right)^\dagger \right] - \hat{\mathrm{A}}\left[\tilde{\mathbf{V}}_i^a \left(\tilde{\mathbf{V}}_i^b \right)^\dagger \right] \right\}. \tag{61}$$

Since the samples at different times are uncorrelated,

$$\mathrm{E}|R_{\lambda,\mu}^{a,b} - \hat{R}_{\lambda,\mu}^{a,b}|^2 = \frac{1}{M} \mathrm{E}|\sum_i \overline{\mathbf{L}}_i^{\lambda,\mu} \bullet \left\{ \left[\tilde{\mathbf{V}}_i^a (\tilde{\mathbf{V}}_i^b)^\dagger \right] - \mathrm{E}\left[\tilde{\mathbf{V}}_i^a (\tilde{\mathbf{V}}_i^b)^\dagger \right] \right\}|. \tag{62}$$

Assuming Gaussian distributed random variables, Eq. 62 can be simplified by expanding into a double sum and applying a known result on the expectation of the product of four Gaussian random variables (Janssen and Stoica 1987). Thus, the variance of the elements of $\mathbf{R}^{a,b}$ can then be obtained as

$$\mathrm{E}|R_{\lambda,\mu}^{a,b} - \hat{R}_{\lambda,\mu}^{a,b}|^2 = \frac{1}{M} \sum_{i,j} \sum_{x,y} \sum_{\alpha,\beta} \left[L_{i,x,y}^{\lambda,\mu} E\left(\overline{\tilde{\mathbf{V}}}_{i,x}^a \tilde{\mathbf{V}}_{j,\alpha}^a \right) E\left(\tilde{\mathbf{V}}_{i,y}^b \overline{\tilde{\mathbf{V}}}_{j,\beta}^b \right) \overline{L}_{j,\alpha,\beta}^{\lambda,\mu} \right] \tag{63}$$

where

DOI 10.1007/978-981-4560-75-7_97-1

$$E\left(\overline{\tilde{\mathbf{V}}}^{a}_{i,x}\tilde{\mathbf{V}}^{a}_{j,\alpha}\right) = \left(\mathbf{w}^{a}_{i,x}\right)^{\dagger} E\left(\overline{\mathbf{V}}^{a}(\mathbf{V}^{a})^{T}\right)\mathbf{w}^{a}_{j,\alpha}$$
$$E\left(\overline{\mathbf{V}}^{a}(\mathbf{V}^{a})^{T}\right) = 8k_B T_0 \mathbf{G}^{a}_{tot} \tag{64}$$

in which $\mathbf{G}^{a}_{\text{tot}}$ is the mutual conductance matrix of the PAF of antenna a (Hay 2010) and T_0 is the thermodynamic temperature of the antennas. Equation 63 is compactly written as

$$\mathrm{E}|R^{a,b}_{\lambda,\mu} - \hat{R}^{a,b}_{\lambda,\mu}|^2 = \frac{1}{M}\sum_{i,x,y}\sum_{j,\alpha,\beta} L^{\lambda,\mu}_{i,x,y}\Lambda_{(i,x,y),(j,\alpha,\beta)}\overline{L}^{\lambda,\mu}_{j,\alpha,\beta} \tag{65}$$

where

$$\Lambda_{(i,x,y),(j,\alpha,\beta)} = (8k_B T_0)^2 \times \left(\mathbf{w}^{a}_{i,x}\right)^{\dagger}\mathbf{G}^{a}_{\text{tot}}\mathbf{w}^{a}_{j,\alpha} \times \overline{\left(\mathbf{w}^{b}_{i,y}\right)^{\dagger}\mathbf{G}^{b}_{\text{tot}}\mathbf{w}^{b}_{j,\beta}} \tag{66}$$

The mutual conductance matrix $\mathbf{G}^{a}_{\text{tot}}$ is computed as the sum of three terms

$$\mathbf{G}^{a}_{\text{tot}} = \mathbf{G}^{a}_{\text{external}} + \mathbf{G}^{a}_{\text{loss}} + \mathbf{G}^{a}_{\text{receiver}}$$

representing, respectively, the effects of the external noise field incident on the antenna, noise associated with loss in the antenna, and noise produced by the low-noise amplifiers in the PAF. The component matrices are obtained by detailed modeling of the antenna electromagnetic and front-end electronic system.

Minimizing the Variance of the Coherence Matrix Estimate The solution to the problem of finding the $L^{\lambda,\mu}_{j,\alpha,\beta}$ that minimize the variance Eq. 65 subject to the transformation requirement Eq. 53, expressed by the linear system Eq. 59, can be derived in the form

$$\overline{L}^{\lambda,\mu}_{j,\alpha,\beta} = \sum_{r,s} H_{(r,s),(j,\alpha,\beta)} c^{\lambda,\mu}_{(r,s)} H_{(r,s),(j,\alpha,\beta)} = \sum_{i,x,y} \Lambda^{-1}_{(j,\alpha,\beta),(i,x,y)} P_{(r,s),(i,x,y)} c^{\lambda,\mu}_{(r,s)} = \sum_{(p,q)} G^{-1}_{(r,s),(p,q)} Y^{\lambda,\mu}_{(p,q)} \tag{67}$$

where

$$G_{(r,s),(p,q)} = \sum_{(j,\alpha,\beta)(i,x,y)} \overline{P}_{(r,s),(j,\alpha,\beta)} \Lambda^{-1}_{(j,\alpha,\beta),(i,x,y)} P_{(p,q),(i,x,y)} \tag{68}$$

and the minimum variance is given by

$$\min\ \mathrm{E}|R^{a,b}_{\lambda,\mu} - \hat{R}^{a,b}_{\lambda,\mu}|^2 = \frac{1}{M}\sum_{(r,s),(p,q)} Y^{\lambda,\mu}_{(r,s)} G^{-1}_{(r,s),(p,q)} \overline{Y}^{\lambda,\mu}_{(p,q)}. \tag{69}$$

The representation in Eq. 69 of minimum variance is dependent on the orthogonal directions p' and q' in the definition of the source coherence matrix $\mathbf{C}'$ in Eq. 51. To quantitatively evaluate this dependence, the aim is to find directions p' and q' that minimize the root-mean-square (rms) standard deviation

$$\sigma = \sqrt{\frac{1}{4}\sum_{\lambda,\mu} \min\ \mathrm{E}|R^{a,b}_{\lambda,\mu} - \hat{R}^{a,b}_{\lambda,\mu}|^2} \tag{70}$$

of all elements of $\mathbf{R}^{a,b}$, as follows. From Eqs. 51 and 60

$$Y^{\lambda,\mu}_{(r,s)} = \frac{1}{2}\left[A^{\lambda,\mu}_{(r,s)} + \cos(2\phi)\ B^{\lambda,\mu}_{(r,s)} - \sin(2\phi)\ C^{\lambda,\mu}_{(r,s)}\right] \tag{71}$$

where ϕ is the angle of rotation of p' and q' from reference directions p and q, and the coefficients are given by

$$\begin{aligned}
A^{\lambda,\mu}_{x,y} &= \begin{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{(x,y)} & \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}_{(x,y)} \\ \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}_{(x,y)} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{(x,y)} \end{bmatrix}_{(\lambda,\mu)} \\
B^{\lambda,\mu}_{x,y} &= \begin{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}_{(x,y)} & \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_{(x,y)} \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_{(x,y)} & \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}_{(x,y)} \end{bmatrix}_{(\lambda,\mu)} \\
C^{\lambda,\mu}_{x,y} &= \begin{bmatrix} \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}_{(x,y)} & \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}_{(x,y)} \\ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}_{(x,y)} & \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_{(x,y)} \end{bmatrix}_{(\lambda,\mu)}.
\end{aligned} \tag{72}$$

Introducing Eqs. 71 and 69 into Eq. 70 gives

$$\begin{aligned}
\sigma^2 = \frac{1}{16M} \times \Bigg\{ & \sum_{(\lambda,\mu),(r,s),(p,q)} \left[A^{\lambda,\mu}_{r,s} G^{-1}_{(r,s),(p,q)} A^{\lambda,\mu}_{p,q} + \frac{1}{2} B^{\lambda,\mu}_{r,s} G^{-1}_{(r,s),(p,q)} B^{\lambda,\mu}_{p,q} + \frac{1}{2} C^{\lambda,\mu}_{r,s} G^{-1}_{(r,s),(p,q)} C^{\lambda,\mu}_{p,q}\right] \\
& + \cos(2\phi) \sum_{(\lambda,\mu),(r,s),(p,q)} \left[A^{\lambda,\mu}_{r,s} G^{-1}_{(r,s),(p,q)} B^{\lambda,\mu}_{p,q} + B^{\lambda,\mu}_{r,s} G^{-1}_{(r,s),(p,q)} A^{\lambda,\mu}_{p,q}\right] \\
& - \sin(2\phi) \sum_{(\lambda,\mu),(r,s),(p,q)} \left[A^{\lambda,\mu}_{r,s} G^{-1}_{(r,s),(p,q)} C^{\lambda,\mu}_{p,q} + C^{\lambda,\mu}_{r,s} G^{-1}_{(r,s),(p,q)} A^{\lambda,\mu}_{p,q}\right] \\
& + \frac{1}{2}\cos(4\phi) \sum_{(\lambda,\mu),(r,s),(p,q)} \left[B^{\lambda,\mu}_{r,s} G^{-1}_{(r,s),(p,q)} B^{\lambda,\mu}_{p,q} - C^{\lambda,\mu}_{r,s} G^{-1}_{(r,s),(p,q)} C^{\lambda,\mu}_{p,q}\right] \\
& - \frac{1}{2}\sin(4\phi) \sum_{(\lambda,\mu),(r,s),(p,q)} \left[B^{\lambda,\mu}_{r,s} G^{-1}_{(r,s),(p,q)} C^{\lambda,\mu}_{p,q} + C^{\lambda,\mu}_{r,s} G^{-1}_{(r,s),(p,q)} B^{\lambda,\mu}_{p,q}\right] \Bigg\}.
\end{aligned} \tag{73}$$

Using Eq. 72, all terms in Eq. 73 involving ϕ can be shown to vanish, giving the result

DOI 10.1007/978-981-4560-75-7_97-1

$$\sigma^2 = \frac{1}{8M} \times \sum_{(r,s),(p,q)} \left[I_{r,s} G^{-1}_{(r,s),(p,q)} I_{p,q} + Q_{r,s} G^{-1}_{(r,s),(p,q)} Q_{p,q} + U_{r,s} G^{-1}_{(r,s),(p,q)} U_{p,q} + V_{r,s} G^{-1}_{(r,s),(p,q)} V_{p,q} \right] \tag{74}$$

where

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad V = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad Q = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad U = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{75}$$

Equation 74 simplifies to

$$\sigma^2 = \frac{1}{4M} \sum_{(r,s)} G^{-1}_{(r,s),(r,s)}. \tag{76}$$

Sensitivity and Survey Speed The equivalent antenna sensitivity in any direction Ω is defined as

$$S(\Omega) = \frac{k_B}{P_{\text{ref}} \sigma} \tag{77}$$

where σ is given in Eq. 76, is the rms noise in the estimate of the coherence matrix, and $P_{ref} = 1/(2 \times \eta_o)$ is the power density of the reference incident fields used in setting up Eq. 48. In using Eq. 76 in Eq. 77, just one time sample $M = 1$ is used, to separate antenna and integration effects. The result Eq. 77 is independent of the choice of coordinate systems used to represent the antennas and coherence parameters of the astronomical sources.

Correlation data obtained at different times with different beams or antenna pointing directions may be linearly combined to achieve maximum signal-to-noise ratio. The maximum sensitivity squared then becomes

$$S^2(\Omega) = \sum_{i=1,2,3,\cdots N_p} S^2_{\text{i}}(\Omega) \tag{78}$$

where $S^2_i(\Omega)$ is the sensitivity squared for pointing i and N_p is the number of pointings. Equation 78 results since the noise at different times is uncorrelated. Dithering of the beam and antenna pointings is used to flatten the sensitivity across the field of view. The survey sensitivity is then approximately equal to the average

$$\begin{aligned} \left(S^2\right)_{\text{ave}} &= \frac{1}{A_{\text{survey}}} \int_{A_{\text{survey}}} d\Omega \sum_{i=1,2,3,\cdots N_p} S^2_{\text{i}}(\Omega) \\ &\approx \frac{N_p}{A_{\text{survey}}} \int_{A_{\text{survey}}} d\Omega \quad S^2_{i^*}(\Omega) \end{aligned} \tag{79}$$

where A_{survey} is the solid angle of the surveyed area of sky and i^* is any representative pointing. Using Eqs. 78, 77, and 76, the square of survey sensitivity given by Eq. 79 is seen to be proportional to the product $N_p \times M$ of the number of pointings N_p and the number of time samples M averaged in the correlations (Eq. 47) at each pointing. The product $N_p \times M$ can also be expressed as $N_p \times M = \tau \times B$

DOI 10.1007/978-981-4560-75-7_97-1

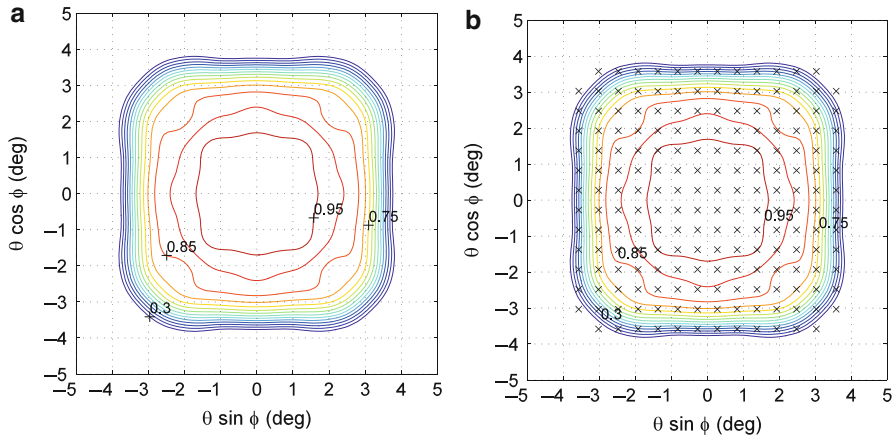


Fig. 12 Sensitivity distributions throughout the FoV of a PAF. Part (a) is the peak sensitivity of single dual-polarized beams at each point. Part (b) is the sensitivity obtained by optimally combining the correlations of a grid of dual-polarized beams in the directions marked with *crosses*. The contour labels refer to the sensitivity relative to the maximum sensitivity in the FoV

where τ is the total survey time and B is the bandwidth of the beamformed signal channels input to the correlator. The survey speed is proportional to

$$\text{SSFoM} = \left(S^2\right)_{\max} \times \text{SSFoV} \tag{80}$$

where

$$\left(S^2\right)_{\max} = \max_{\Omega} \; S^2_{i^*}(\Omega) \tag{81}$$

$$\text{SSFoV} = \frac{1}{\left(S^2\right)_{\max}} \int d\Omega \; S^2_{i^*}(\Omega) \tag{82}$$

is the survey speed field of view. Thus, the survey speed figure of merit is defined as

$$\text{SSFoM} = \int d\Omega \; S^2_{i^*}(\Omega). \tag{83}$$

In Eq. 80 to Eq. 83, $S^2_{i^*}(\Omega)$ is calculated from Eqs. 77 and 76 using just $M = 1$ time sample in order to separate antenna effects from the effects of integration time and channel bandwidth. Clearly, the survey speed is proportional to the square of the physical area of the telescope.

A general result that emerges from the survey speed analysis is that given sufficient number of beams, the sensitivity obtained at each point in the FoV by optimally combining the beam correlations is equal to that of a single dual-polarized beam at the same point. Figure 12 illustrates this in the case of the ASKAP PAF, which will be described in more detail in the following section. Part a of Fig. 12 shows the peak sensitivity of single dual-polarized beams formed at each point in the FoV. Part b of Fig. 12 shows the sensitivity obtained by optimally combining the correlations of a grid of dual-polarized beams, where the beam directions are indicated by the crosses. The beam spacing is $\lambda/(2D)$, where D is the aperture diameter of the reflector antenna. The spacing $\lambda/(2D)$ is sufficient to maximize the sensitivity obtained by optimally combining the beam correlations. This is a general result that can be proved using the bandlimited Fourier

DOI 10.1007/978-981-4560-75-7_97-1

transform relationship between the beam radiation patterns and the aperture fields of the reflector antennas.

Applications in Radio Astronomy

Radio telescopes have a number of special system requirements that many other applications do not have. For an overview of some international radio observatories, please refer to the chapter "► Antennas in Radio Telescope Systems." As well as antenna gain, input match, sensitivity, noise temperature, polarization, and the collecting area, other requirements include the survey speed and image discrimination. Through these largely scientific instruments, some new techniques and components in some unique frequency bands have been trialed in an operating system. In this section, two applications of arrays in radio telescopes are outlined, and unique design details are provided.

Design of a Multibeam Feed for the Parkes Radio Telescope

The 64 m radio telescope located at Parkes in the central west of New South Wales, Australia, was commissioned in 1964. It was involved in many early discoveries of quasars and pulsars, but by the mid-1990s, it needed enhancements. A multiple beam capability was investigated from the early 1990s for observations at a wavelength around 21 cm (actually, the frequency range 1.27 to 1.47 GHz) to undertake deep, large-area surveys of neutral hydrogen emission from external galaxies. In particular, a large-area survey of the southern hemisphere for sources of HI was required. With a single prime-focus feed, it was estimated that this survey would take about 10 years to complete. However, using a feed of N elements, each with two orthogonal linear polarizations, the observing period would be reduced by a factor of about 1/N. To establish a possible feed size, a first approximation is the sample spacing in the u-v plane that is needed to resolve images of extragalactic objects. This indicated that the desired beam spacings were approximately an HPBW (half-power beamwidth) apart, and for this to occur, feeds should be spaced to give beams an integer multiple of HPBW apart. Physically, this would be impossible without dielectric-loaded horns. However, dielectric loading is unattractive in radio astronomy because of losses and mismatch to free-space, both of which increase the noise temperature and significantly reduce sensitivity. In addition, the reflector scan-gain loss and the available number of separate receivers limit the size of the array. An alternative approach is to increase the spacing to about 2 HPBW and to fill in the sample space by moving the reflector as well to point at interstitial positions. As a result, conventional horns were able to be used and arranged in the focal region so that the beam spacing is approximately 2 HPBW. An hexagonal array of circular apertures was chosen because of its efficient packing density, and with limits on the number of receivers, the ultimate number of elements in the array was $N = 13$ after a study was performed with a smaller number of elements (Bird 1994). It was considered that the redundancy provided by the additional elements was worth the extra expense. A circular horn of about one wavelength diameter was selected as an element over other geometries because it has excellent pattern symmetry and cross-polar performance. Circular waveguide is preferred throughout, from the orthogonal mode transducers (OMT) to the horn input. The geometry of the final multibeam feed system is shown in Figs. 13 and 14.

The hexagonal cluster shown in Fig. 13 makes for a compact feed system because the OMTs (in the Dewar) can be attached directly to the rear of the horns to extract both linear polarizations.

Stepped Circular Horn Design

The geometry of the Parkes symmetrical parabolic reflector is summarized in Table 1. The approach taken to the design was to obtain the best possible performance with a single feed horn of the desired diameter

DOI 10.1007/978-981-4560-75-7_97-1

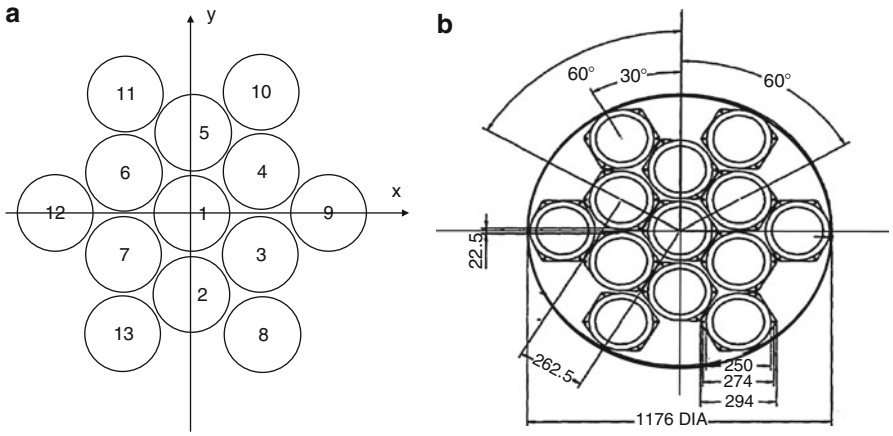


Fig. 13 Thirteen-element L-band multibeam feed for the Parkes radio telescope. (**a**) Layout and (**b**) mechanical drawing of ground planes, dimensions in mm

Table 1 Parkes paraboloid geometry

Quantity	Value
Diameter (D)	64 m
Focal length (f)	26.270 m
f/D	0.41047
Semi-angle	62.69°

and then to accommodate for any array effects. Initially, both circular and square horns were potential candidate elements as both yielded comparable gain with the reflector. However, circular horns have lower spillover loss, better pattern symmetry, and lower cross polarization than the square horn option. For any parabolic reflector fed at the focus by a circular waveguide, there is an optimum waveguide diameter for maximum gain. A circular waveguide of about one wavelength diameter maximizes the gain with the Parkes reflector and also has low cross polarization and spillover. During design, steps along the profile were introduced to achieve high beam efficiency, low input reflection coefficient, and low inter-element coupling.

There is, however, an upper limit to the feed diameter in the array. This is specified by the beam spacing for effective sampling in u-v space. Making use of the beam angle estimate in Introduction, $\theta_b \approx \mathrm{BDF}\,(s/\lambda)$ where s is the radius of the first ring of the hexagonal array and $2 \times$ HPBWs with a -12 dB edge illumination corresponds to $\theta_b \approx 2.4\lambda/D$. Combining these estimates gives an estimate of the radius of the first ring in the hexagonal array as

$$s \approx \frac{2.4\,(f/D)\,\lambda}{\mathrm{BDF}}\,. \tag{84}$$

For the Parkes radio telescope geometry (Table 1), Eq. 84 gives $s \approx 1.17\lambda$. To allow for an adequate wall thickness and other possible adjustments, the radius of the first ring of $s = 1.2\lambda$ was selected. From Eq. 84, it is seen that the spacing reduces with the reflector f/D. Thus, for the Lovell radio telescope at Jodrell Bank, UK, which has $f/D = 0.3$, based on the same beam spacing, the element spacing should be $s \approx 0.97\lambda$. With an adequate wall thickness, this limits the horn diameter to about 0.9λ, which limits the option for feeds and the use of choke rings in the flange to improve the efficiency and beam symmetry.

DOI 10.1007/978-981-4560-75-7_97-1

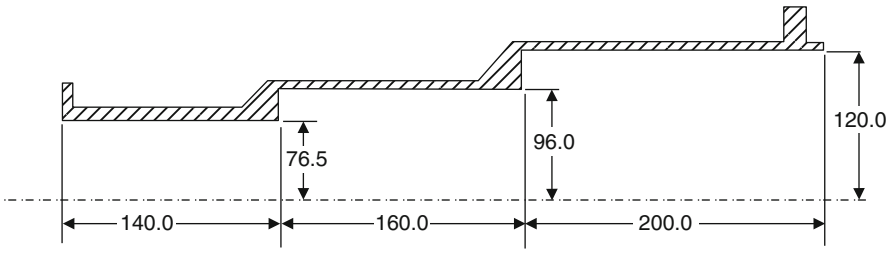


Fig. 14 L-band stepped horn for Parkes multibeam feed. Dimensions in mm

As a result, for the Lovell telescope multibeam feed design, a coaxial horn was ultimately chosen (Bird 1997).

Another important geometric parameter in the design is to establish the diameter of the input waveguide. To minimize losses and effects due to higher order modes, it was decided to have all modes other than the fundamental TE_{11} mode be cut off in the frequency range 1.27 to 1.47 GHz. A 153 mm diameter input pipe satisfies this requirement.

The circular horn with steps along its length were analyzed with mode-matching software CIRCAR to achieve the desired efficiency and reflection coefficient. The software can also be used to analyze mutual coupling between the horns (Bird 1996). This allowed the horns in a 13-element hexagonal array to be analyzed, and, as a result, further minor adjustments were made to the horn geometry while, at the same time, keeping all horns identical. After an investigation of several configurations, it was found that most objectives were attained with the two-step horn design shown in Fig. 14. Low cross polarization is easily achieved with the single isolated feed, but, in the array, significant levels of TM_{01} and TE_{21} modes can be generated. These modes can cause high cross-polar levels, which, in some instances, are as high as for the square horn. However, by adjusting the output aperture diameter and length of the circular waveguide steps, low cross-polar illumination (> -35 dB) of the reflector at the center frequency was ultimately achieved for each horn of the array. The calculated reflection coefficient and insertion loss over the frequency band of interest are shown in Fig. 15. Both mismatch and loss (including ohmic loss) are least near the center of the band and degrade at the band limits. Figure 16 shows the co- and cross-polarized radiation patterns of the central horn (element 1) in the array, and this has been degraded only slightly by the array environment, as are the horn mismatch and loss. Across the band, good beam symmetry is maintained, although the cross-polar level deteriorates to -27 dB at the lower band edge, which was still acceptable. It was found that satisfactory performance could be obtained, both separately and in the array, across an extended frequency range from 1.20 GHz to 1.55 GHz, which provided options for many different types of observations with the multibeam feed.

Multibeam Performance

CSIRO's array feed software was combined with the reflector analysis software to compute the radiation pattern of the overall antenna.

The radiation pattern of the reflector with the array was computed for each beam. It was assumed that the sources to be observed are mainly uncorrelated and, therefore, the signals radiated by the horns are not summed coherently. The computed co- and cross-polarized radiation patterns of one of the beams on the first ring at the band-center frequency of 1.37 GHz are shown in Fig. 17. The patterns were obtained by exciting horn 1 and include reflector aperture blockage, which is 5 m in diameter as a result of the feed cabin. The performance of the four independent beams is summarized in Table 2. The results for the other beams can be obtained from symmetry. At the extremes of the frequency band, the beam efficiency decreases. For example, at 1.27 GHz the efficiency of beam 5 is 69 % and at 1.47 GHz the value is 67 %.

DOI 10.1007/978-981-4560-75-7_97-1

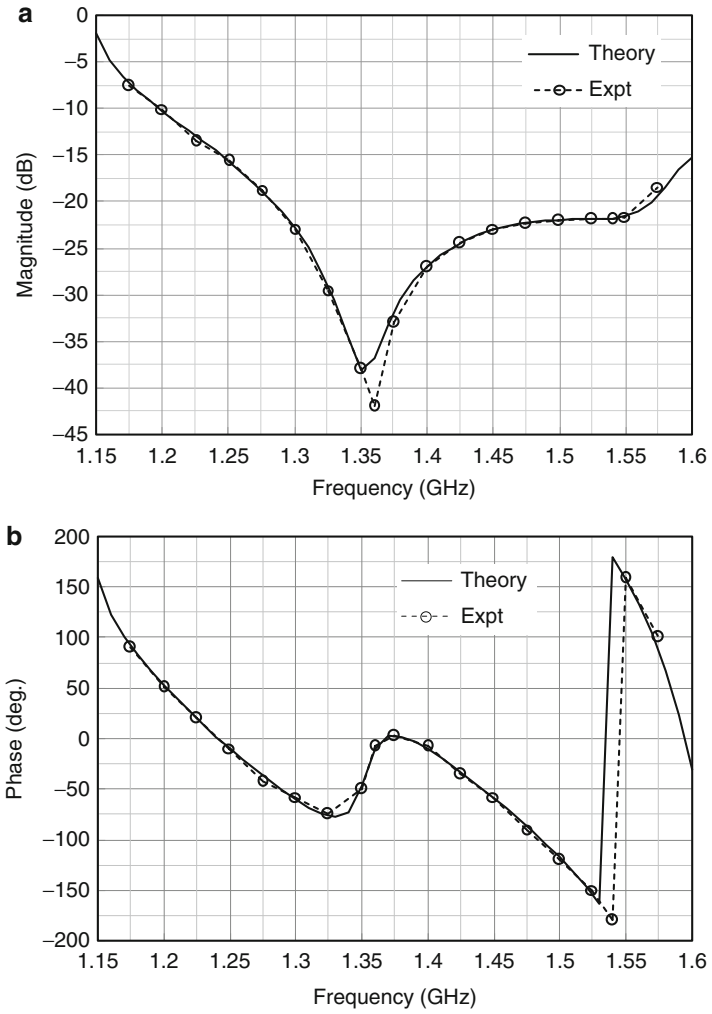


Fig. 15 Input reflection coefficient of isolated horn. *Solid line*: theory and *dashed line*: experiment. (**a**) Magnitude and (**b**) phase

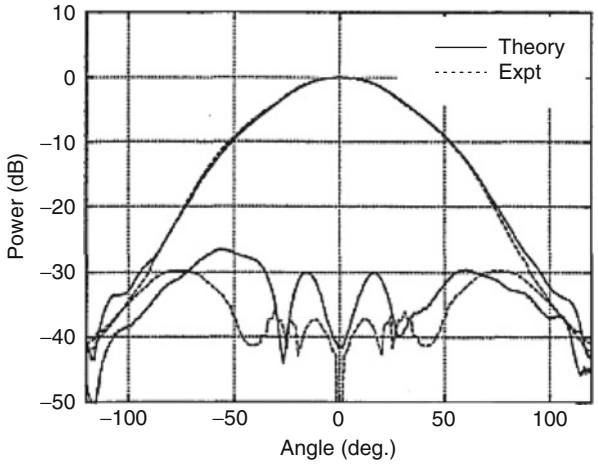


Fig. 16 Radiation patterns of the central horn, theory, and measurement, in the 45° plane at a frequency of 1.37 GHz

DOI 10.1007/978-981-4560-75-7_97-1

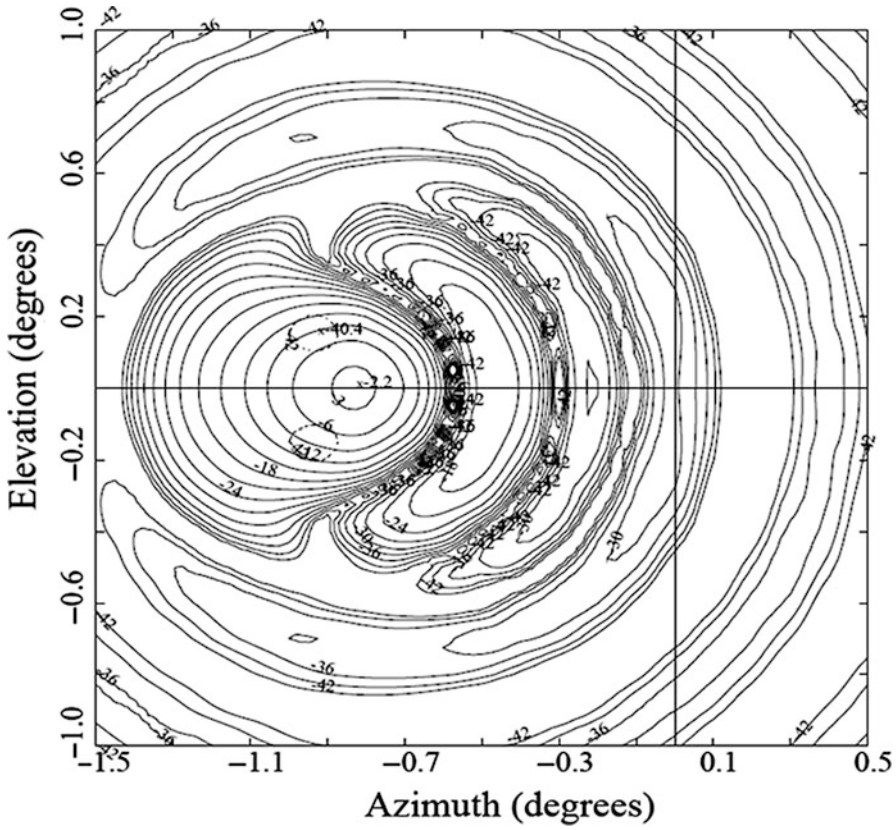


Fig. 17 Beam pattern for horn in second ring of hexagonal array (beam 9). Contours are at 3 dB intervals. The peak co-polar level is −2.23 dB corresponding to an efficiency of 60 %. Cross-polar lobes that are shown *dotted* near the center of the beam have a peak of −38.2 relative the peak co-polar level

Table 2 Summary of the beam performance at 1.37 GHz with a blockage diameter of 5 m

Beam no.	Efficiency %	HPBW deg	X-pol. peak dB	Spillover dB	Beam maximum u deg	v deg	Feed RL dB
1	70.9	0.24	−40.5	−0.172	0.000	0.000	31.5
2	66.7	0.24	−39.4	−0.180	0.000	−0.467	31.8
3	69.7	0.24	−40.0	−0.175	−0.428	−0.233	31.5
8	60.0	0.24	−37.3	−0.178	−0.428	−0.719	31.8
9	59.8	0.24	−48.2	−0.188	−0.836	0.000	32.2

Spillover efficiency rises with frequency, due to narrowing of the feed horn pattern, from 94 % at the lower to 97 % at the higher frequency.

The radiation pattern of the radio telescope for the central beam both computed and measured results are shown in Fig. 18. A 5 m central blockage that is due to a focal cabin from which feeds are rotated into the focus was assumed in the calculations. The sidelobes of the measured pattern are filled in and are consistently of the same form for most beams, which indicates they may be due to effects not connected directly with the multibeam feed. Figure 19 shows the completed feed as it was lifted into position. Table 2 shows that the efficiency of the central beam is predicted to be 70.9 % relative to a uniformly illuminated aperture without including reflector surface errors. This efficiency decreases at the band edges to 68 % at 1.27 GHz and 66 % at 1.47 GHz. The efficiency of the most displaced feed (element 9) is 59.8 %, and the latter's coma sidelobe is −14 dB below the peak value. The spillover efficiency was estimated to

DOI 10.1007/978-981-4560-75-7_97-1

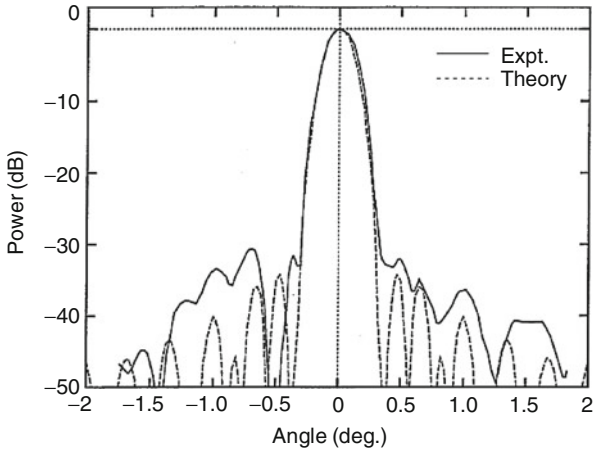


Fig. 18 Measured and computed radiation pattern of the Parkes radio telescope for the central horn of the multibeam array at a frequency of 1.37 GHz

Fig. 19 L-band multibeam feed lifted into position on Parkes radio telescope (Picture: courtesy of CSIRO Australia)

be $>$ 96 % for all beams, and this decreases slightly at lower frequencies due to broadening of the illumination. Overall, the cross coupling between beams is at a satisfactory level and has a maximum of -21 dB due to diffraction and mutual coupling in the array, but typically, it is < -35 dB while cross-polar isolation is $>$40 dB.

Australian SKA Pathfinder (ASKAP) PAF

The Australian Square Kilometre Array Pathfinder (ASKAP) is a 36-antenna interferometer radio telescope that has been developed and is currently being deployed by CSIRO at the Murchison Observatory in Western Australia. The specifications of the feed array for the ASKAP PAF application are listed in Table 3. Figure 20 shows the resulting connected array design that was developed to achieve these specifications.

As illustrated in Fig. 21, the PAF has conducting patches of similar form to the self-complementary array but also a parallel ground plane, used to give the array directivity and allow convenient placement of

DOI 10.1007/978-981-4560-75-7_97-1

Table 3 Specifications of the ASKAP telescope

Number of antennas	36
Reflector diameter	12 m
Maximum baseline	6 km
Resolution	30”
Sensitivity	65 m^2/K
Survey speed	1.3 × 105 $m^4/K^4/deg^2$
System noise temperature	50 K
Aperture efficiency	80 %
Observing frequency	700–1,800 MHz
Field of view	30 deg^2
Processed bandwidth	300 MHz
Spectral channels	16,384
Phased array feed	188 elements

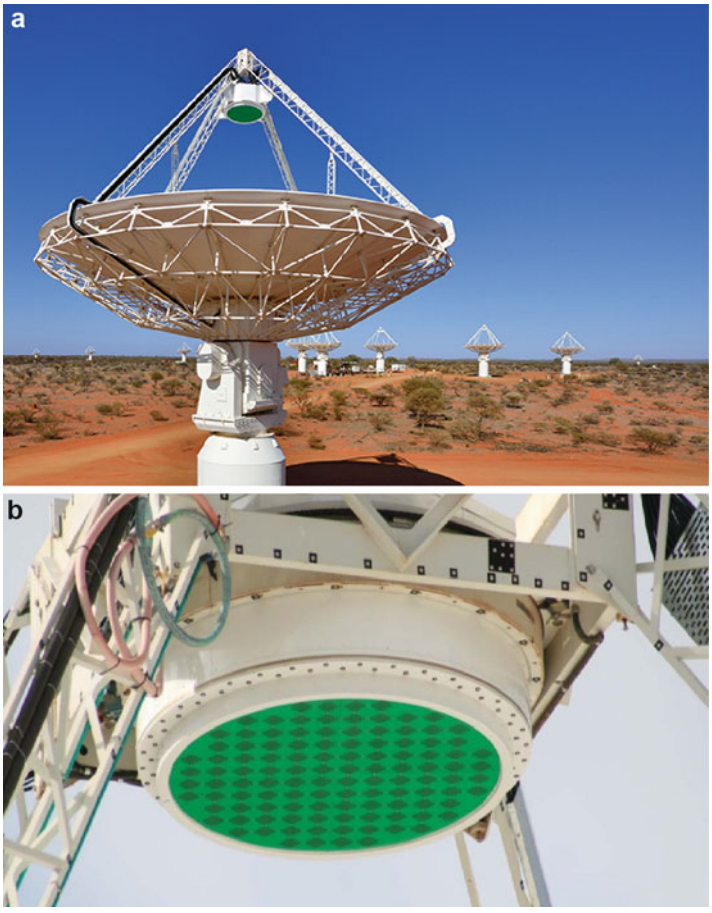


Fig. 20 ASKAP antennas. (**a**) Reflectors and (**b**) checkerboard PAF mounted at the focus (Picture: courtesy of CSIRO Australia)

LNAs and other circuitry on the other side to the patches. Two-wire transmission lines connect adjacent patch corners via the LNAs. A low-loss and low-permittivity material is used to support the patches, which are formed on a printed circuit board.

DOI 10.1007/978-981-4560-75-7_97-1

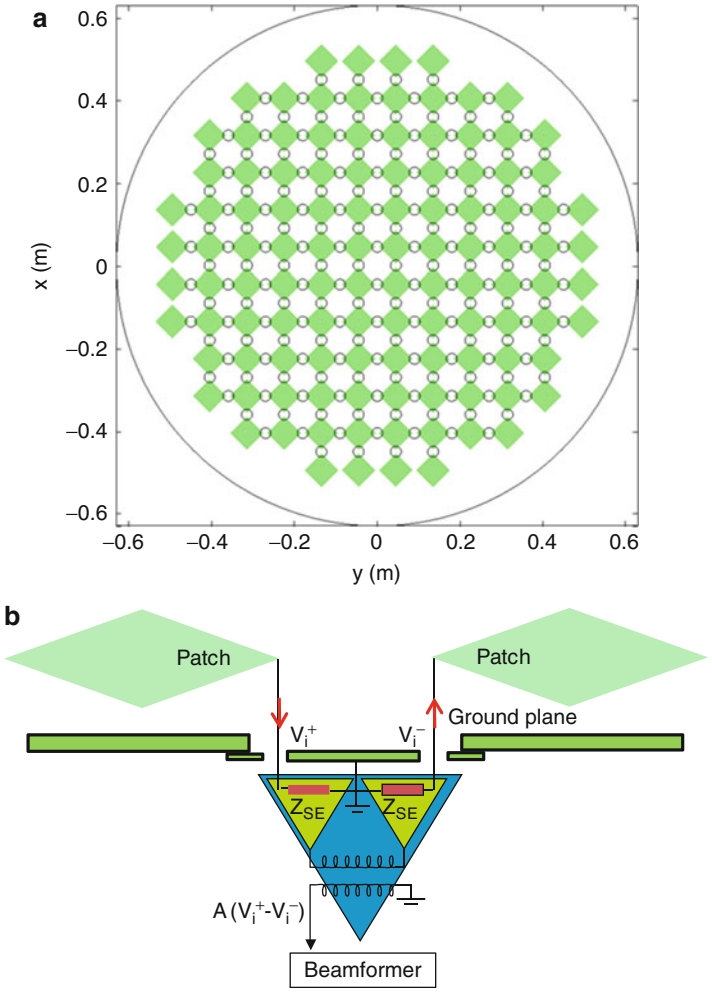


Fig. 21 One hundred and eighty eight-element connected array with ground plane (**a**) and schematic (**b**) of two-wire transmission lines diverting array currents to LNAs on the other side of the ground plane. The patches are 80 mm across the diagonal, and a 10 mm gap exists between neighboring patch corners, resulting in an array element spacing of 90 mm. The patches are 65 mm from the ground plane, which is 10 mm thick and has apertures 26 mm in diameter. The wires connecting the patch corners to the LNAs have a diameter of 0.86 mm

One key to solving the design challenges has been rigorous computational electromagnetic analysis of the array. This needs to account for the fine structure in the small array elements and also the mutual coupling across the entire array. An accurate and efficient technique was developed (Hay et al. 2011). This is an extension of the characteristic basis function method that accounts for the strong flow on conduction current between the connected array elements. The computational efficiency of the software allowed a study of the array effects, and, as a result, the new arrays were designed with improved performance including increased frequency range.

Another crucial development is a theory of the array operation in terms of common and differential modes of current on the two-wire transmission lines that divert the array currents to the LNAs behind the ground plane (Hay and O'Sullivan 2008). This has explained resonances that occur in the array with certain LNA configurations and led to an improved, resonance-free array/LNA design that is illustrated in Fig. 21.

A further important development has been the design and accurate experimental characterization of the LNAs (Shaw et al. 2012). Effective procedures have been developed for both the signal and noise

DOI 10.1007/978-981-4560-75-7_97-1

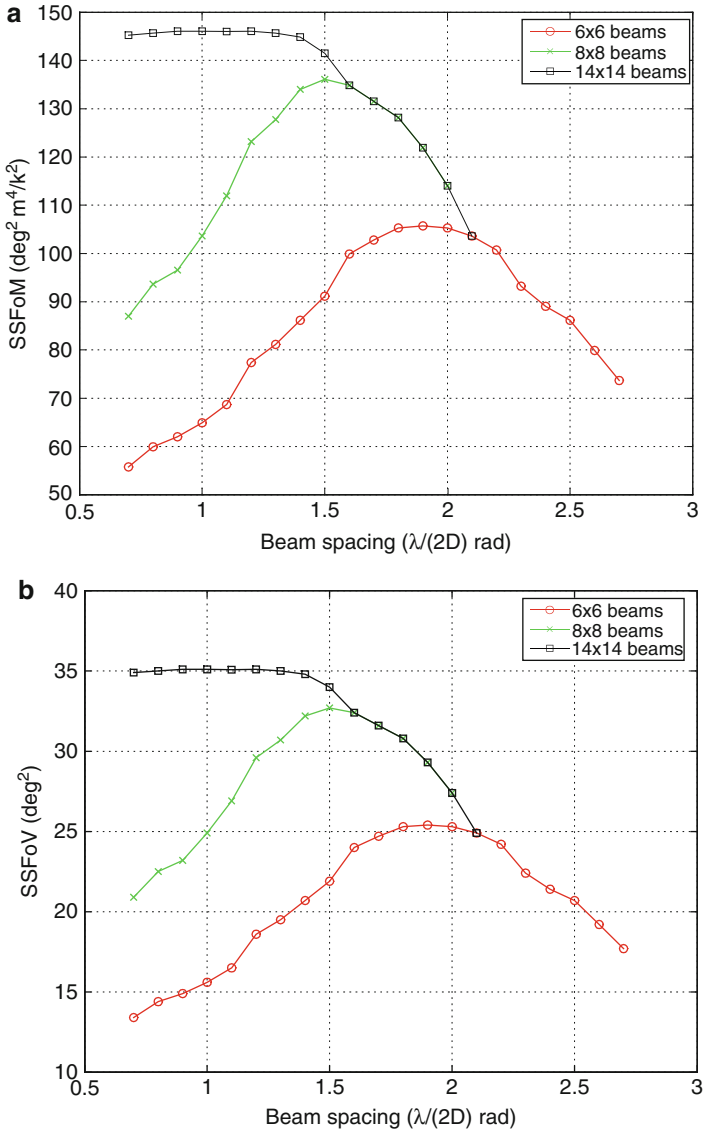


Fig. 22 Computed survey speed figure of merit (**a**, SSFoM, Eq. 83) and survey speed field of view (**b**, SSFoV, Eq. 82) for the 188-element ASKAP PAF at 1.3GHz as a function of the number and spacing of the beams. The LNA model is noise-matched to the array and has a minimum noise temperature Tmin of 30 K. The FoV is centered at zenith, and the assumed brightness temperature distribution is 5 K at zenith and transitions from this value to the ground temperature of 275 K between 80 and 120° from zenith

properties of the three-port LNAs, which have differential noise-match and input impedance of the order of 300 Ω.

An analysis of survey speed in section "Phased Array Feeds" provides a framework for optimizing the number and spacing of the PAF beams. Figure 22 shows the SSFoM and SSFoV at 1.3GHz for various beam numbers and spacings. Although performance is maximized with a beam spacing of λ/(2D), where D = 12 m is the aperture diameter of the ASKAP antennas, it is desirable in practice to use a larger beam spacing and smaller number of beams. This is because beamforming cost increases with the number of beams. For the ASKAP PAF, a 6 × 6 grid of beams was selected as a reasonable compromise between cost and performance. Figure 23 shows the computed SSFoM and SSFoV as a function of frequency. At each

DOI 10.1007/978-981-4560-75-7_97-1

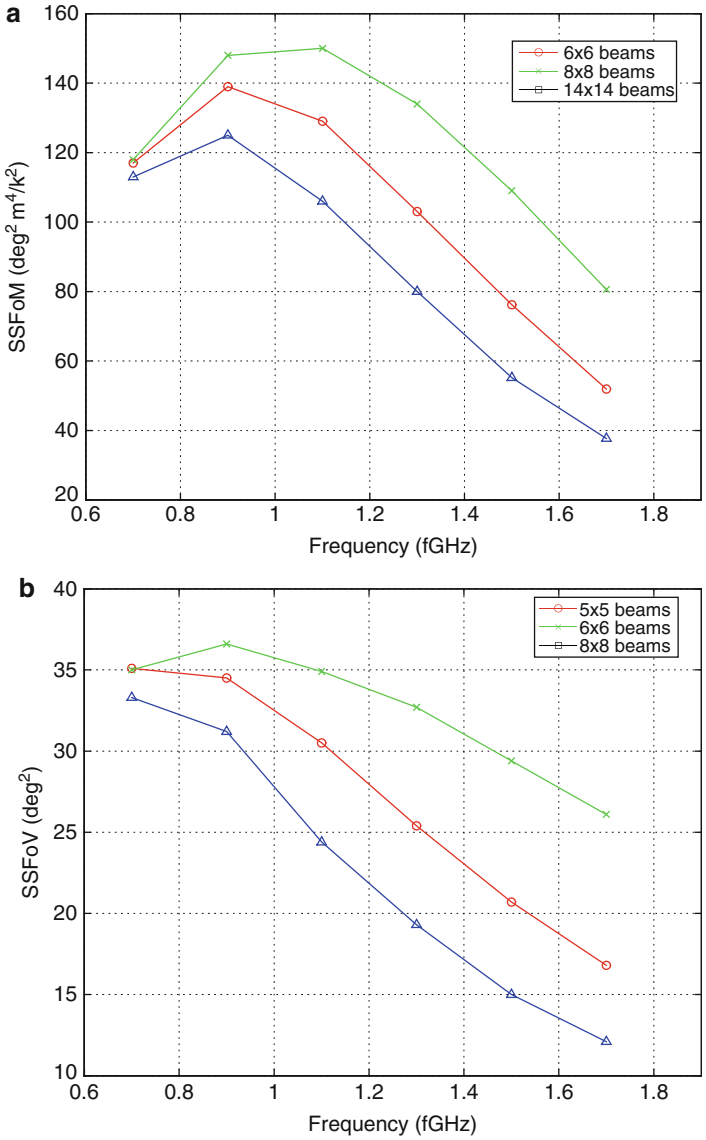


Fig. 23 Computed survey speed figure of merit (**a**, SSFoM, Eq. 83) and survey speed field of view (**b**, SSFoV, Eq. 82) for the 188-element ASKAP PAF as a function of frequency. The LNA model is noise-matched to the array and has a minimum noise temperature Tmin of 30 K. The FoV is centered at zenith, and the assumed brightness temperature distribution is 5 K at zenith and transitions from this value to the ground temperature of 275 K between 80 and 120° from zenith

frequency, the beam spacing has been optimized to maximize SSFoM. With a 6 × 6 beam, SSFoV is limited at low frequency mainly by the physical size of the PAF.

The development of the ASKAP PAFs has proceeded to a MKII design with a number of improvements including increased sensitivity. Figure 24 shows a 40-element prototype of the MKII PAFs. As illustrated in Fig. 25, the patch design is modified to allow improved matching of the array and LNAs. The overall geometry is similar to the original PAF, with the same element spacing of 90 mm and 10 mm of the wires in each of the two-wire transmission lines. The patches are reduced in size and are connected to the transmission lines via narrow strips in the same plane as the patches. This provides inductance that compensates for the ground plane at high frequency. Chip capacitors connect the strips to the patches, compensating also for the ground plane at low frequency. The diameter of the wires in the two-wire transmission lines is increased slightly to 1.6 mm.

DOI 10.1007/978-981-4560-75-7_97-1

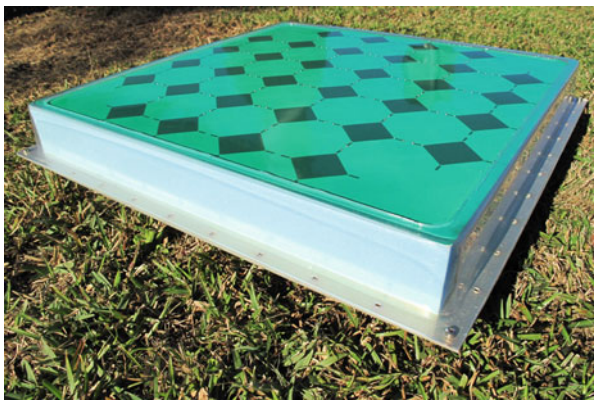


Fig. 24 Photograph of a 40-element prototype of the ASKAP MKII PAFs (Picture: courtesy of CSIRO Australia)

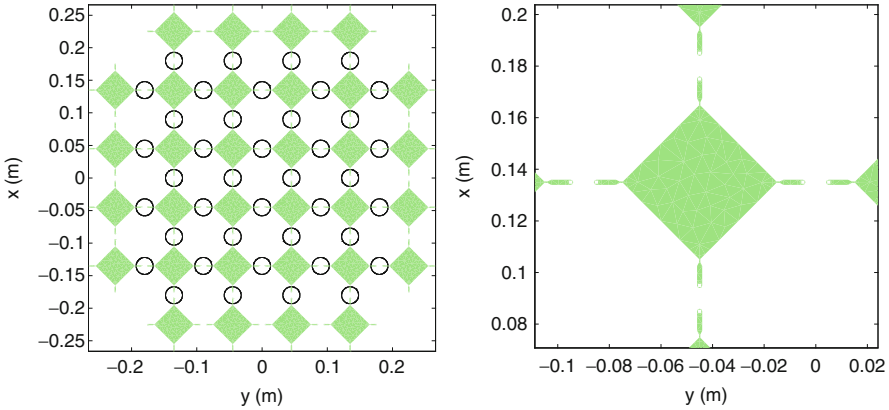


Fig. 25 Modified patch geometry of prototype MKII PAF

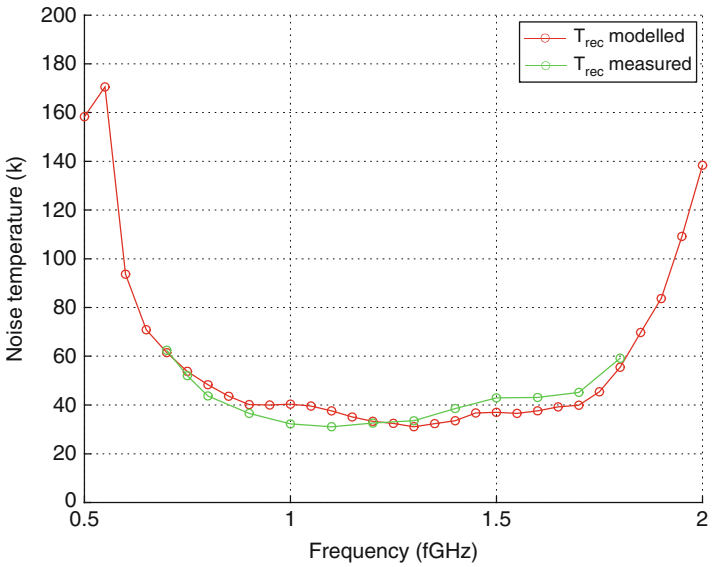


Fig. 26 Beamformed noise temperature of a 40-element prototype ASKAP MKII array when directly receiving a wave from the boresight direction

DOI 10.1007/978-981-4560-75-7_97-1

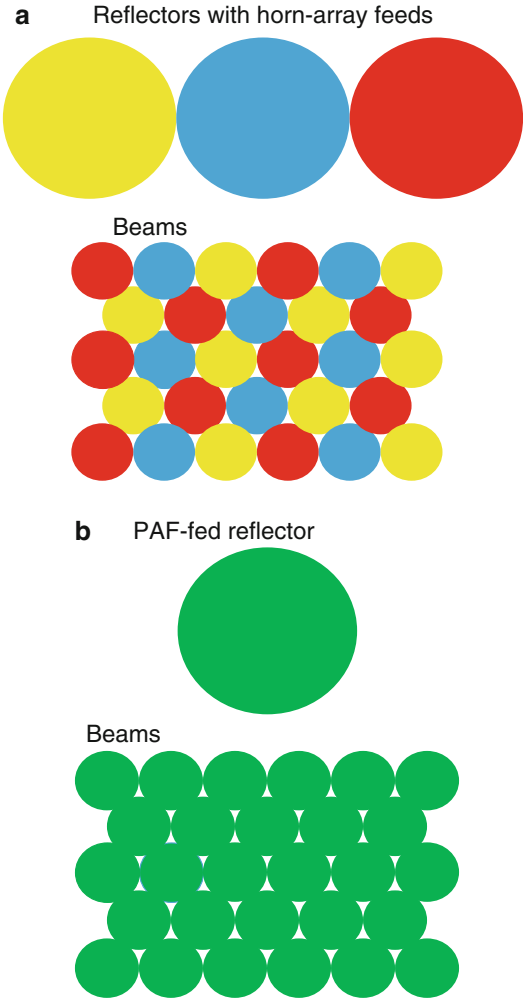


Fig. 27 Reflectors and beams of horn-fed (**a**) and PAF-fed (**b**) reflector antennas

Figure 26 compares measured and modeling results on a 5 × 4 prototype of the ASKAP MKII PAF. The results are for the beamformed receiver noise temperature

$$T_{\text{rec}} = T_0 \frac{\overline{\phi}^T G_{\text{rec}} \phi}{\overline{\phi}^T G_A \phi} \tag{85}$$

when tested as a directly receiving array pointed at zenith. The measured result is estimated from the measured Y-factor using the sky as the cold load and microwave absorber as the hot load (Chippendale et al. 2014). These results show good low noise performance across the 0.7–1.8 GHz frequency range required for ASKAP.

Conclusion

The area of phased array feeds has made significant progress through technology development and new applications. Advances are the large frequency range and digital beamforming, providing the ability to form multiple simultaneous antenna beams with highly controllable properties. The degree to which beam

DOI 10.1007/978-981-4560-75-7_97-1

properties can be controlled is increased over previous multibeam systems using horn feed arrays or arrays with analog beamforming networks. In both of these former systems, beam spacing is limited by the size of the feeds or by signal loss in analog beamforming networks. With digital beamforming, PAFs can produce beams with arbitrary spacing, without significant loss of signal. These features are of considerable interest for a range of applications in communications. These are:

1. Interference limited systems. The capacity of multiuser communication systems can be increased through frequency reuse using multiple simultaneous antenna beams at each frequency. The ability to closely control beam spacing, size, shape, sidelobes, and cross polarization should unlock the full potential of this approach. In addition, improved use of limited or expensive real estate will be possible. For example, as illustrated in Fig. 27, instead of multiple reflector apertures, each fed by horn-array feeds with limited spacing, a single reflector aperture with a PAF can produce a contiguous set of beams throughout the field of view.
2. Multiband systems. The large frequency range implies the ability to accommodate multiple frequency bands within the one receiving or transmitting aperture. For example, the satellite communication bands, or integration of the multiple bands used in wireless communications, in the base station antennas.
3. Cognitive systems. These systems are aware of the radio environment and adapt to make best use of the limited available resources of spectrum and power. This requires sensing the radio environment over a wide frequency range and then adapting antenna characteristics such as frequency, beam polarization, and radiation pattern.
4. Moving platforms. There is great interest in mobile platforms, including satellite systems for vehicles, ships, and aircraft. Also, in some systems, antennas are subject to perturbing motion of platforms, e.g., terrestrial point-to-point communication systems using towers that flex in the wind. The connected array offers correction for motion effects.
5. Spatial power combining. Transmitting systems can also be developed, allowing the power of multiple transmitters to be coherently combined. This is particularly attractive at high frequencies in mm-wave and THz frequency ranges, due to reduced power of electronic devices at these frequencies. Applications include very high throughput communications.

Radio astronomy has been a key driver including the requirement for high sensitivity obtained through integration of the array and low-noise amplifiers operating at ambient temperature. There exists good scope for further developments in this area. Wide FoV will continue to be a key area for development for radio astronomy. The benefits have been clearly illustrated by the various multibeam systems such as Parkes. Good progress has also been made in PAFs, e.g., ASKAP with further increased FoV and frequency range. There is significant interest in astronomy at higher frequencies including mm-wave and THz, evidenced by the recent development of the ALMA telescope, where wide FoV systems are a significant possibility and challenge for further development.

The ability to operate over a large frequency range with unprecedented beamforming accuracy and flexibility is likely to enable significant advances also in communications. Future developments will include extending the technology to higher frequency bands, where integration with electronic devices, signal transport, and array fabrication are key challenges.

DOI 10.1007/978-981-4560-75-7_97-1

Acknowledgement

We acknowledge the Wajarri Yamatji people as the traditional owners of the site of the ASKAP telescope.

Cross-References

► Antennas in Radio Telescope Systems
► Reflector Antennas

References

Bandler JW, Charalambous (1972) Practical least p-th optimization of networks. IEEE Trans. Microw. Theory, Vol. MTT-20, pp. 834–840.
Bird TS (1994) A multibeam feed for the Parkes radio-telescope. In: IEEE antennas & propagation society symposium, Seattle, pp 966–969
Bird TS (1996) Modelling arrays of circular horns with choke rings. In: IEEE Antennas & Propagation Society Symposium, Baltimore, USA, pp. 226–229.
Bird TS (1997) Coaxial feed array for a short focal-length reflector. In: IEEE antennas & propagation society symposium, Montréal, pp 1618–1621
Bunton JD, Hay SG (2010) Achievable field of view of checkerboard phased array feed. In: International conference on electromagnetics in advanced applications, Sydney, pp 728–730
Chippendale AP, Hayman DB, Hay SG (2014) Measuring noise temperatures of phased array antennas for radio astronomy at CSIRO. Publications of the Astronomical Society of Australia, 31, id.e01, April 1
Christiansen WN, Högbom JA (1969) Radiotelescopes. Cambridge University Press, London
Fisher JR, Bradley RF (2000) Full-sampling array feeds for radio telescopes. In: Proceedings of the SPIE, 4015, pp 308–319
Hay SG (2010) Maximum-sensitivity matching of connected-array antennas subject to Lange noise constants. Int J Microw Opt Technol 5(6):1889–1890
Hay SG, O'Sullivan JD (2008) Analysis of common-mode effects in a dual-polarized planar connected-array antenna. Radio Sci 43:RS4S06
Hay SG, O'Sullivan JD, Mittra R (2011) Connected patch array analysis using the characteristic basis function method. IEEE Trans Antennas Propag 59(3):1828–1837
Janssen PHM, Stoica P (1987) On the expectation of the product of four matrix-valued Gaussian random variables, Eindhoven University of Technology, EUT Report 87-E-178, ISBN 90-6144-176-1
Loux PC, Martin RW (1964) Efficient aberration correction with a transverse focal plane array technique. IEEE Int Conv Record (USA) 12, Part 2: 125–131
Rudge AW, Withers MJ (1971) New technique for beam steering with fixed parabolic reflectors. Proc IEE 118(7):857–863
Rudge AW, Milne K, Olver AD, Knight P (1986) The handbook of antenna design, vol 1 & 2. Peter Peregrinus, London
Shaw RD, Hay SG and Ranga Y (2012) Development of a low-noise active balun for a dual-polarized planar connected-array antenna for ASKAP. International Conference on Electromagnetics in Advanced Applications, Cape Town, pp. 438–441

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_99-1

MIMO Systems and Antennas for Terminals

Hui Li[a,b]* and Buon Kiong Lau[b]
[a]School of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China
[b]Department of Electrical and Information Technology, Lund University, Lund, Sweden

Abstract

MIMO technology has facilitated tremendous performance improvements in wireless communications, allowing the data rate to increase linearly with the number of antennas used, at no additional expense in transmit power or spectrum. However, the tremendous performance gain can only be achieved by multi-antenna designs that provide low coupling and correlation, as well as high total efficiency. Such design criteria are especially challenging for small terminal devices. The situation becomes even more complicated with increasing bandwidth requirements for terminals in existing and upcoming mobile communication standards. Beginning with the history of MIMO terminal antenna and its evaluation methods, this chapter is geared towards providing useful guidelines to researchers and practitioners alike on how to design efficient MIMO antennas for terminals. The focus is on decoupling and decorrelation techniques, including RF circuit level decoupling, antenna structure decoupling and characteristic mode based decoupling. Future directions in MIMO antenna design and some corresponding open problems are also described.

Keywords

MIMO systems; MIMO metrics; Channel capacity; Antenna correlation; Mutual coupling; Mobile communication; Characteristic mode; Parasitic element; Neutralization line; Defected ground plane

Introduction

The increasing popularity of smartphones, laptops and tablets is spurring the fast-growing demand for high speed mobile communications. This in turn leads to the widespread deployment of multiple-input multiple-output (MIMO) technology in wireless communication standards, where multiple antennas are employed at both base stations and terminals to enable linear increase in channel capacity (or ideal data transmission rate) with the number of antennas, without sacrificing additional frequency spectrum and transmitted power (Gesbert et al. 2003). However, the linear increase in channel capacity is the ideal case for MIMO. A lot of non-ideal factors, such as non-zero correlation between channels and antennas, imperfect antenna efficiency and finite power imbalance will deteriorate the link performance. To evaluate the MIMO performance of multiple antenna systems, figures of merits are introduced in this chapter and their effects on the data rate are illustrated.

From the antenna design perspective, MIMO technology greatly increases the complexity of antenna design in compact terminal devices, due to the requirement of implementing multi-antennas for single or multiple frequency bands within a limited space allocation. The closely spaced antennas can lead to high mutual coupling, which decreases the antenna efficiencies. At the same time, the close spacing can also increase the signal correlation between antennas, degrading the terminals' MIMO performances. Thus, it

*Email: hui.li@dlut.edu.cn

DOI 10.1007/978-981-4560-75-7_99-1

is important to decouple and de-correlate the antennas. The relationship between decoupling and decorrelation is discussed in the chapter. Generally, antenna design methods that are intended to achieve high isolation and low correlation are collectively called decoupling techniques. However, decoupling techniques are primarily aimed at increasing isolation and they do not ensure effective decorrelation in multi-antennas with moderate to low radiation efficiencies. The existing decoupling techniques based on various mechanisms are introduced in this chapter, including decoupling network, modified ground plane, neutralization/connecting line, characteristic modes, etc. Specifically, the characteristic mode based decoupling technique will be more systematically described, as it has been found to be effective for designing decoupled and decorrelated MIMO terminal antennas that offer good bandwidth performance.

In this chapter, the main metrics to characterize MIMO systems is first described. Then the discussions are focused on the design of multi-antennas for mobile terminals, since the same design principles herein apply to the larger form factors of tablet PCs, including phablets (phone-tablets) or mini-tablets. This is because the main challenges in multi-antenna design are related to the small physical sizes of the terminal in relation to the wavelength, which means that compact mobile terminals present the worst case. From another perspective, the design of multi-antenna elements for larger terminal devices than mobile handsets is equivalent to designing the antennas for mobile handsets for higher frequencies. For example, the electrical wavelength at the 700 MHz band is about three times longer than that at the 2 GHz band.

History of MIMO Terminal Antennas

MIMO technology is based on the use of multi-antennas. The term MIMO is commonly used in the context of spatial multiplexing, which is transferring multiple streams of data by applying multi-antennas at both the transmitter and receiver. However, the use of multi-antennas to improve wireless communications is not a new phenomenon. For example, multi-antennas were employed in the short-wave frequencies (3–30 MHz) to combat fading of radiotelegraph signals due to multipath propagation in the 1930s (Peterson et al. 1931). This fading mitigation approach is known as diversity reception, which relies on the received signals at different antennas having different signal strengths. Diversity reception has also been adopted in mobile communications when it is deployed in GSM base stations to improve signal quality. However, a common denominator of early adoptions of multi-antennas is that the multi-antennas are deployed at the base station rather than the user terminal, due to additional complexities required for processing the received signals as well as the cost of implementing the complexities. Moreover, in order for multi-antennas to yield signals that are sufficiently different (or decorrelated) from one another, one common approach is to implement antennas that are well separated in space, which is easier to achieve at the base station due to larger available space. The requirement of decorrelated antennas is critical for both diversity and spatial multiplexing techniques to work well, even though spatial multiplexing employs more advanced signal processing to enable transformation of the physical multipath propagation channel into independent transmission channels. The constraint of antenna spacing in the limited sizes of typical user terminals in relation to the signal wavelengths has made terminal multi-antenna design especially challenging.

The first attempt to deploy multi-antennas in mobile terminals was made for the Japanese Personal Digital Cellular (PDC) systems around the mid 1990s, based on antenna switched diversity (Kobayashi et al. 1994). A whip antenna (monopole) was used for both transmitting and receiving, whereas an internal planar inverted F-antenna (PIFA) was used for receiving. Multi-antenna diversity was also utilized to mitigate delay spread in the PDC systems, to reduce intersymbol interference (ISI) (Falconer et al. 1995). Notwithstanding, research and practical implementations of dual-antenna terminals predates the PDC systems, with the Japanese company NTT already proposing built-in (internal) two-PIFA based designs

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_99-1

Fig. 1 Samsung LTE USB dongle (Model name: GT-B3710) (Source: Samsung.com)

for both shoulder-carry and handheld mobile terminals as far back as 1986 (Taga and Tsunekawa 1986; Taga and Tsunekawa 1987).

During the late 1990s, the MIMO technology that was first introduced in (Winters 1987) came out of obscurity and began to attract a lot of attention in the research community, due to its potential to drastically improve data rate without additional spectrum use. At a time when acquiring spectrum license has become a major expense of mobile operators, the idea of making far better use of existing spectrum was indeed timely. Nevertheless, extensive efforts were needed in the research and development of the technology before it was ready for commercial deployment. The first introduction of MIMO in mobile communications was for the High Speed Packet Access Evolution (HSPA+) system in 2007, corresponding to the Third-Generation Partnership Project (3GPP) Release 7 (Rao et al. 2009). HSPA+ employs spatial multiplexing with two data streams on the downlink to enable a data rate of up to 28 Mbits per second (Mbps), which effectively doubles the maximum data rate of its predecessor. At around the same time, MIMO also appeared on the market in the form of wireless local area network (WLAN) terminal devices that complied with the IEEE802.11n draft 2.0 specifications (Perahia 2008). Though limited in range to 70 m indoor and 230 m outdoor, IEEE802.11n systems can provide data rates of up to 600 Mbps, using 4×4 MIMO to support up to 4 data streams.

However, despite the enthusiastic adoption of MIMO technology in HSPA+ and IEEE802.11n, the MIMO feature remains optional in these standards. Hence, it is not compulsory to implement multi-antennas in their terminals. This situation changed with the deployment of Long Term Evolution (LTE) networks corresponding to 3GPP Release 8, since LTE mandates the implementation of two or more antennas in the terminals. The first LTE terminals were USB dongles that enable wireless broadband access for laptop computers. One example of such terminals is the Samsung LTE dongle (see Fig. 1), which was used in the world's first commercial LTE network deployed by the Nordic mobile operator Teliasonera on 14 Dec 2009. The LTE standard specifies up to four antennas for Category 5 User Equipment (UE), with maximum downlink and uplink data rates of 300 and 75 Mbps, respectively.

Evaluation of MIMO Systems

Ideally, spatial multiplexing can increase the channel capacity of a MIMO system linearly with the number of transmit-receive antenna pairs used. However, its realization depends on a lot of practical factors in both the antenna design and the propagation environment. This section introduces figures of merits that can be used to characterize MIMO system performance, focusing on the antenna and

DOI 10.1007/978-981-4560-75-7_99-1

propagation aspects. Moreover, the influence of mean effective gain, branch power imbalance and correlation on capacity and multiplexing efficiency will be discussed.

Mean Effective Gain

Mean effective gain (MEG), introduced by Taga in 1990 (Taga 1990), is defined as the average power received by the antenna to the total mean incident power in the real propagation channel. In spherical coordinates with azimuth angle ϕ and elevation angle θ, MEG is given by:

$$\mathrm{MEG} = \int_0^{2\pi}\int_0^{\pi}\left(\frac{XPR}{XPR+1}G_\theta(\theta,\phi)P_\theta(\theta,\phi)+\frac{1}{XPR+1}G_\phi(\theta,\phi)P_\phi(\theta,\phi)\right)\sin\theta d\theta d\phi, \tag{1}$$

where XPR is the ratio of the average power in the θ-polarized component to that of the ϕ-polarized component. $G_\theta(\theta,\phi) = |E_\theta(\theta,\phi)|^2$ and $G_\phi(\theta,\phi) = \left|E_\phi(\theta,\phi)\right|^2$ are the antenna gain patterns in the θ and ϕ components, respectively, with $E_\theta(\theta,\ \phi)$ and $E_\phi(\theta,\ \phi)$ being the θ- and ϕ-polarized electric far field pattern of the antenna, respectively. $P_\theta(\theta,\ \phi)$ and $P_\phi(\theta,\ \phi)$ are the angular power density functions in the θ and ϕ components, respectively. Different from total radiated power (TRP) of an antenna, MEG takes into consideration the full interaction between the antenna and the propagation channel, and represents the performance of an individual antenna in non-ideal environments. A high MEG indicates a good match between the antenna and the channel, such that the antenna can receive more signal power in the channel. Good MIMO performance requires that the individual antennas have high MEGs. For an ideal channel with uniform three dimensional (3D) angular power spectrum (APS), $XPR = 1$, $P_\theta = P_\phi = \frac{1}{4\pi}$.

Branch Power Imbalance

Aside from MEG, in order to achieve good MIMO performance, it is also required that the power levels of the signals received by the antennas should not vary significantly from one another. This is described by branch power imbalance, which is the minimum ratio of MEG between two branches (1 and 2):

$$k = \min\left(\frac{\mathrm{MEG}_1}{\mathrm{MEG}_2}, \frac{\mathrm{MEG}_2}{\mathrm{MEG}_1}\right). \tag{2}$$

For an ideal channel with uniform 3D APS, k is equal to the ratio of antenna efficiency or TRP between the two branches.

Correlation

Signal correlation is another metric critical to MIMO systems. The complex correlation coefficient between antenna element i and antenna element j, based on the far field patterns, is expressed as:

$$\rho_{c,ij} = \frac{\int_0^{2\pi}\int_0^{\pi} A_{ij}(\theta,\phi)\sin\theta d\theta d\phi}{\sqrt{\int_0^{2\pi}\int_0^{\pi} A_{ii}(\theta,\phi)\sin\theta d\theta d\phi \int_0^{2\pi}\int_0^{\pi} A_{jj}(\theta,\phi)\sin\theta d\theta d\phi}} \tag{3}$$

DOI 10.1007/978-981-4560-75-7_99-1

where $A_{ij} = XPR\ E_{\theta,i}(\theta,\phi)E^*_{\theta,j}(\theta,\phi)P_\theta(\theta,\phi) + E_{\phi,i}(\theta,\phi)E^*_{\phi,j}(\theta,\phi)P_\phi(\theta,\phi)$, $E_{\theta/\phi,i/j}(\theta,\phi)$ is the θ (or ϕ)-polarized electric far-field pattern of the antenna i (or j) in spherical coordinates. The (normalized) complex correlation matrix at the receiver for M_r antennas can be expressed as:

$$\overline{\mathbf{R}}_r = \begin{bmatrix} 1 & \rho_{c,12} & \cdots & \rho_{c,1M_r} \\ \rho_{c,21} & 1 & \cdots & \rho_{c,2M_r} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{c,M_r1} & \rho_{c,M_r2} & \cdots & 1 \end{bmatrix}. \tag{4}$$

In wireless communications, envelope correlation coefficient is often of interest due to signal envelope (magnitude) being easier to measure than complex signal with both magnitude and phase. It has been found empirically that envelope correlation coefficient (of antenna i and j) can be approximated by the square of the magnitude of the complex correlation:

$$\rho_{e,ij} \approx \left|\rho_{c,ij}\right|^2. \tag{5}$$

When the correlation is high, the MIMO performance will deteriorate, especially when the envelope correlation exceeds 0.5. This is because MIMO systems rely on the independence among the received signals to provide diversity and/or independent channels in spatial multiplexing.

Evaluation of Correlation

The evaluation of correlation is important in determining MIMO system performance. According to Eq. 3, signal correlation between two antennas is calculated from the full spherical radiation patterns of the antennas, which are obtained with both phase and polarization information. However, complete pattern measurement is both time-consuming and expensive, due to the need for specialized equipment and facilities.

Consequently, it is of significant interest to the antenna community to avoid using antenna patterns in calculating correlation. In 1962, the relationship between antenna patterns and scattering (or S) parameters in closed form was derived by Stein in (Stein 1962), with special emphasis on lossless and reciprocal cases. Though the work of Stein contains all the necessary building blocks, the explicit expression of signal correlation in terms of S parameters is attributed to (Blanch et al. 2003), in which the envelope correlation for two antennas is given by:

$$\rho_e \approx |\rho_c|^2 = \frac{\left|S^*_{11}S_{12} + S^*_{21}S_{22}\right|^2}{\left(1 - |S_{11}|^2 - |S_{21}|^2\right)\left(1 - |S_{22}|^2 - |S_{12}|^2\right)}, \tag{6}$$

where uniform 3D APS and 100 % antenna radiation efficiency are assumed. This S-parameter based correlation expression can be easily generalized to M-port antenna systems (Thaysen and Jakobsen 2006). The expression has been widely used in the evaluation of MIMO terminal antenna design (see e.g., Su 2010; Sonkki et al. 2012; Xiong et al. 2012), since it is far more convenient to measure S parameters with vector network analyzer (VNA) than to measure radiation patterns. Moreover, S parameters are important multi-antenna parameters to measure in their own right, which means no additional effort is required to obtain them. However, the accuracy of the method relies heavily on the assumption of 100 % radiation efficiency, and significant error can occur even for relatively high radiation efficiencies (Li et al. 2013a).

DOI 10.1007/978-981-4560-75-7_99-1

Unfortunately, this aspect is often overlooked in some open literature on MIMO terminal antenna evaluation, which employs Eq. 6 despite significant dielectric and conductive losses.

The accuracy of using Eq. 6 to calculate correlation due to antenna losses was first investigated in (Hallbjörner 2005). An improved method was proposed to calculate the upper bound of correlation (i.e., the worst case) for dual-antenna systems, and the relationship between antenna radiation efficiency and correlation uncertainty was presented. The upper bound (or "guaranteed" upper limit) of correlation in terms of S parameters as well as radiation efficiency η_{rad} is given by:

$$|\rho_c|_{\text{guaranteed}} = \frac{\left|2\text{Re}\left(S_{11}S_{21}^*\right)\right|}{\left(1-|S_{11}|^2-|S_{21}|^2\right)\eta_{\text{rad}}} + \frac{1}{\eta_{\text{rad}}} - 1, \tag{7}$$

where the two antennas are assumed to be identical. Re(•) denotes the real part of (•). This estimation of correlation coefficient is quite good for antennas with high radiation efficiencies, but it can be very conservative for antennas with low efficiencies. For example, when the radiation efficiency is lower than 50 %, which is not uncommon for mobile terminal antennas (especially in frequency bands below 1 GHz), the calculated upper bound is larger than unity regardless of the S parameters (Li et al. 2013a). More intuitively, the guaranteed correlation coefficients largely depend on the radiation efficiency of the antenna when the efficiency is low; however, the true correlation coefficients do not vary much with radiation efficiencies when materials with higher losses are introduced, since this will only cause slight changes in the antenna patterns.

In (Li et al. 2013a), equivalent circuits were used to estimate the signal correlation, where antenna losses were modeled as series resistance or parallel conductance depending on the antenna behavior. Given radiation efficiencies and S parameters, the loss resistance (or conductance) was calculated in order to isolate the lossy component from the original lossy antennas. For the dual-antenna case, the original lossy antenna array was decomposed into three parts: two resistors (one for each antenna) and one lossless two-element array. Then, the correlation coefficient of the original dual-antenna system can be calculated from the S parameters of the lossless array through matrix operations. Significantly better accuracy had been achieved using this method, relative to other S-parameter based methods.

It is worth noting that in practice, the calibration of the phases of the S parameters is very important in the measurement, since they are utilized in all S-parameter based calculation methods. In contrast, the phase of the coupling (or transmission) coefficient S_{21} (or S_{12}) is unimportant for conventional multi-antenna evaluation. Other approaches to estimate signal correlation can also be found in (Stjernman 2005, Salonen et al. 2005).

Relationship Between Decoupling and Decorrelation

By definition, decoupling means increasing isolation or decreasing coupling among the antenna ports, which leads to lower mismatch loss and higher total efficiency. On the other hand, de-correlation means making the complex radiation patterns of the antennas more orthogonal to one another, which leads to lower signal correlation.

To examine the relationship between decoupling and decorrelation, it is assumed that the antennas in a dual-antenna system are identical (e.g., identical antenna elements symmetrically distributed on a ground plane) and have 100 % radiation efficiency. Hence, Eq. 6 reduces to (Dossche et al. 2004):

$$\rho_e \approx \frac{\left|2\text{Re}\left(S_{11}S_{21}^*\right)\right|^2}{\left(1-|S_{11}|^2-|S_{21}|^2\right)^2}. \tag{8}$$

DOI 10.1007/978-981-4560-75-7_99-1

According to Eq. 8, if the antennas are well matched (i.e., $|S_{11}|$ is very small), the correlation between antennas will be low when the mutual coupling ($|S_{21}|$) is very small. Therefore, the antennas are automatically decorrelated when they are decoupled. Other than minimizing $|S_{11}|$ and $|S_{21}|$, another method to decorrelate the antennas is by causing S_{11} to be in quadrature with S_{21}, which can be achieved by using a coupled matching network (Dossche et al. 2004). However, low mutual coupling is not required in this case.

Based on these observations, it can be concluded that decoupling between antennas with high efficiencies can lead to decorrelation, whereas decorrelation does not necessarily mean decoupling. For antennas with low radiation efficiencies (e.g., terminal handset antennas at frequencies below 1 GHz), there exists no simple relationship between decoupling and decorrelation. Therefore, both metrics of coupling and correlation need to be considered and evaluated when designing antennas.

Channel Capacity

In MIMO systems with M_t transmit antennas and M_r receive antennas, the transmitted signals **s** and the received signals **y** can be represented in the baseband as:

$$\mathbf{s} = [s_1, s_2, \ldots s_{M_t}]^T, \tag{9}$$

$$\text{and} \qquad \mathbf{y} = [y_1, y_2, \ldots y_{M_r}]^T. \tag{10}$$

The input–output signal model of such a system is described as:

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{11}$$

where **n** is the AWGN noise at the receive antennas and the channel matrix **H** is given by:

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1M_t} \\ h_{21} & h_{22} & \cdots & h_{2M_t} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M_r1} & h_{M_r2} & \cdots & h_{M_rM_t} \end{bmatrix}, \tag{12}$$

where h_{mn} is the complex transmission coefficient from the transmit antenna n to the receive antenna m. Then, the channel capacity (maximum achievable error-free data rate) for the MIMO system can be expressed as (Telatar 1999):

$$C = \max_{\mathrm{Tr}(\mathbf{R_{ss}})=M_t} \log_2 \det\left(\mathbf{I}_{M_r} + \frac{P_t}{M_t N_0}\mathbf{H}\mathbf{R_{ss}}\mathbf{H}^H\right), \tag{13}$$

where $\mathbf{I}_{M_r}$ is the $M_r \times M_r$ identity matrix, P_t is the transmit power and N_0 is the receiver noise power, and $(\bullet)^H$ denotes the Hermitian transpose operator. $\mathbf{R_{ss}}$ is the covariance matrix of **s** in Eq. 9 and must satisfy $\mathrm{Tr}(\mathbf{R_{ss}}) = M_t$ to constrain the total transmit power.

For antenna designers, the effect of non-ideal antennas on the channel capacity is an important design consideration. To study the impact of non-ideal behaviors, a simpler and more intuitive expression of Eq. 13 can be derived (Paulraj et al. 2003) using the Kronecker channel model (Kermoal et al. 2002), which has good accuracy for lower-order MIMO systems in non-line-of-sight scenarios with rich scattering and has been adopted in several IEEE standards

DOI 10.1007/978-981-4560-75-7_99-1

$$\mathbf{H} = \mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{R}_t^{1/2}, \tag{14}$$

where $\mathbf{R}_t \in C^{M_t \times M_t}$ and $\mathbf{R}_r \in C^{M_r \times M_r}$ are the transmit and receive antenna correlation matrices, respectively. $\mathbf{H}_w$ is the channel matrix for the independent and identically distributed (IID) Rayleigh channel. Substituting Eq. 14 into Eq. 13 for the case of equal power allocation (i.e., $\mathbf{R_{ss}} = \mathbf{I}_{M_t}$) gives

$$C = \log_2 \det\left(\mathbf{I}_{M_r} + \frac{P_t}{M_t N_0}\mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{R}_t\mathbf{H}_w^H\mathbf{R}_t^{H/2}\right). \tag{15}$$

Assuming $M_r = M_t = M$ and a high signal-to-noise ratio (SNR), the capacity of the correlated MIMO channel can be written as:

$$C \approx \log_2 \det\left(\frac{P_t}{M_t N_0}\mathbf{H}_w\mathbf{H}_w^H\right) + \log_2 \det(\mathbf{R}_r) + \log_2 \det(\mathbf{R}_t). \tag{16}$$

To focus on the antenna performance, the channel is assumed to be uniform 3D APS, then the correlation matrices can be expressed as $\mathbf{R}_t = \mathbf{\Lambda}^{1/2}\overline{\mathbf{R}}_t\mathbf{\Lambda}^{1/2}$ and $\mathbf{R}_r = \mathbf{\Lambda}^{1/2}\overline{\mathbf{R}}_r\mathbf{\Lambda}^{1/2}$ (Tian et al. 2011), where $\overline{\mathbf{R}}_t$ and $\overline{\mathbf{R}}_r$ are normalized correlation matrices similar to $\overline{\mathbf{R}}_r$ in Eq. 4, whereas $\mathbf{\Lambda}_t = \mathrm{diag}\left(\left[\eta_1, \ldots, \eta_{M_t}\right]\right)$ and $\mathbf{\Lambda}_r = \mathrm{diag}\left(\left[\eta_1, \ldots, \eta_{M_r}\right]\right)$ are diagonal matrices containing the total efficiencies of the antennas, with η_i being the total efficiency of the i^{th} antenna element.

In Eq. 16, $\log_2 \det(\mathbf{R}_r) \leq 0$ and $\log_2 \det(\mathbf{R}_t) \leq 0$, where the inequalities will occur for any non-ideal antenna behaviors in total efficiency, efficiency imbalance and correlation. This implies that non-ideal MIMO antennas will result in a capacity loss of $\log_2 \det(\mathbf{R}_r) + \log_2 \det(\mathbf{R}_t)$ bps/Hz. The effects of non-ideal behaviors on MIMO performance will be further explained in the next subsection using the metric of multiplexing efficiency.

Multiplexing Efficiency

Capacity is a very useful metric to describe MIMO performance at the system level, since it takes into account all antenna and propagation effects. However, expressing performance in data rates does not provide explicit information to antenna researchers and engineers regarding the impact of individual antenna design parameters. Moreover, antenna researchers and engineers are more accustomed to power based metrics such as total efficiency. Therefore, multiplexing efficiency was introduced as a new figure of merit to evaluate antenna performance in MIMO systems (Tian et al. 2011).

Multiplexing efficiency is defined as the power penalty of a non-ideal antenna system in achieving a given capacity, compared with an ideal antenna system with 100 % total antenna efficiencies and zero correlation among the antennas:

$$\eta_{\text{mux}} = P_0/P_t \leq 1. \tag{17}$$

In Eq. 17, P_0 and P_t are the transmit power (or SNR, if the same receiver noise is assumed) required by the ideal and the real antenna systems, respectively, to achieve the same capacity.

Following the procedure described in Tian et al. (2011), and assuming no correlation at the transmitter and the receive antennas being the MIMO antennas under test, the multiplexing efficiency for a $M \times M$ MIMO system can be derived as:

DOI 10.1007/978-981-4560-75-7_99-1

$$\eta_{\text{mux}} = \frac{\det(\mathbf{I}_M + \gamma_t \mathbf{R}_r)^{1/M} - 1}{\gamma_t} \quad [\text{dB}], \tag{18}$$

where $\mathbf{R}_r = \mathbf{\Lambda}^{1/2}\overline{\mathbf{R}}_r\mathbf{\Lambda}^{1/2}$ and $\gamma_t = P_T/N_0$. At high SNRs, the multiplexing efficiency can be well approximated by:

$$\eta_{\text{mux}} = \left(\prod_{i=1}^{M} \eta_i\right)^{\frac{1}{M}} \det\left(\overline{\mathbf{R}}_r\right)^{\frac{1}{M}}. \tag{19}$$

Equation 19 can be interpreted as a generalization of total efficiency of the single-antenna case to the multi-antenna case, where the extended form translates efficiency imbalance and correlation to an equivalent power loss in decibel (dB).

For the simple case with two antennas at both the transmitter and receiver, the multiplexing efficiency expression reduces to:

$$\eta_{\text{mux}} = \underbrace{\sqrt{\eta_1\eta_2}}_{\eta_g} \underbrace{\sqrt{\left(1 - |\rho_c|^2\right)}}_{\eta_c}. \tag{20}$$

In Eq. 20, η_g is the geometric mean (or arithmetic mean in dB) of the antenna efficiencies, which includes the effect of efficiency and efficiency imbalance, whereas η_c reveals the equivalent power loss caused by correlation. Equation 20 is useful for testing and comparing between different MIMO terminal antennas.

Decoupling Techniques for MIMO Terminals

As mentioned above, coupling is important in terms of its influence on total antenna efficiency, which together with correlation has significant impacts on MIMO capacity. In this section, existing decoupling techniques on both the circuit level and the antenna level for achieving low coupling and correlation are surveyed and discussed. Apart from the decoupling techniques covered in this section, it should be noted that polarization diversity can also be utilized to achieve low mutual coupling in multi-antennas (and multi-port antennas in general), with dual-polarized patch antenna being a classic example.

Decoupling Network

The purpose of a decoupling network is to reduce the mutual impedance or coupling coefficients among the antennas to zero, and at the same time retain the good impedance matching of each antenna element (Wallace and Jensen 2004). Therefore, it is also called decoupling and matching network (DMN) in the literature, (e.g., Volmer et al. 2008). Decoupling network is a circuit approach to achieve decoupled MIMO antennas.

One possible realization of decoupling network for dual-antennas is shown in Fig. 2 (Dossche et al. 2004). In this realization, the whole network consists of transmission lines, open-circuited stubs and a decoupling line. Firstly, the matching network is connected to the antennas to turn each reflection coefficient (S_{11} and S_{22}) into a real number and each coupling coefficient (S_{21} and S_{12}) into an imaginary number. According to Eq. 8, assuming the two antennas are identical, the correlation coefficient between the antennas is zero due to S_{11} being in quadrature with S_{21}. The imaginary S_{21} ensures that the mutual

DOI 10.1007/978-981-4560-75-7_99-1

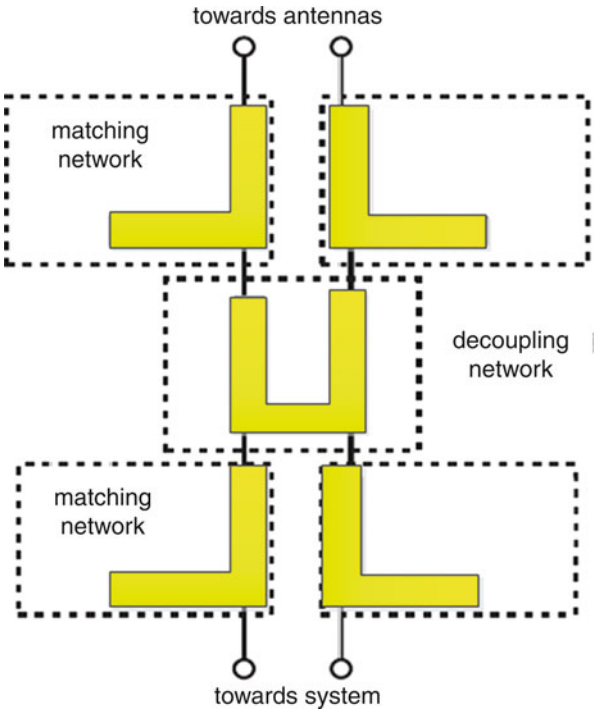


Fig. 2 Decoupling network for a two-port antenna system (Dossche et al. 2004)

impedance Z_{21} is imaginary, such that a lossless decoupling network can be employed to decouple the antennas by achieving $S_{21} = 0$. After that, a second matching network can be used to minimize S_{11} to ensure good impedance matching at the output ports. The dual-antenna system is then well matched and decoupled.

Other realizations of the decoupling network have been proposed based on a combination of lumped elements (including capacitors and inductors) (Chen et al. 2008), distributed elements (Dossche et al. 2005), the hybrid 180° coupler (Coetzee and Yu 2008b; Lin et al. 2012a; Volmer et al. 2008) or decoupler line (Dossche et al. 2005). Lumped elements are easier to tune and also conserve space on the PCB relative to distributed elements (Weber et al. 2006), yet at the same time when the number of lumped elements is increasing, the loss of the whole network increases, reducing the efficiency of the whole antenna system.

In order to enable multi-band and multi-mode operation for highly coupled terminal antennas, decoupling networks have also been designed to be tunable or work in multiple bands. One example of dual-band decoupling network has been illustrated in (Lin et al. 2012). A dual-mode 180° hybrid coupler is utilized to decouple the low band by creating orthogonal modes. In the high band, the shunt path between the upper and lower parts of the hybrid coupler is prohibited, and the coupler becomes only a pair of delay lines. Thus, the reactive elements, realized with standard matrix manipulation, take over to cancel the mutual coupling. The unusual dual mode 180° hybrid coupler, whose behavior differs at low band and high band, is realized by synthesized microstrip lines with additional LC tanks.

Modified Ground Plane

Modified ground plane is an antenna-level decoupling method, where the ground structure of the multi-antennas is modified (e.g., etched) in a certain way to change its electromagnetic properties, such as current distribution and surface wave. Due to the slow wave and stop-band characteristics of the Defected ground plane (DGS) (Ahn et al. 2001; Kin et al. 2002), it has been previously utilized to suppress

DOI 10.1007/978-981-4560-75-7_99-1

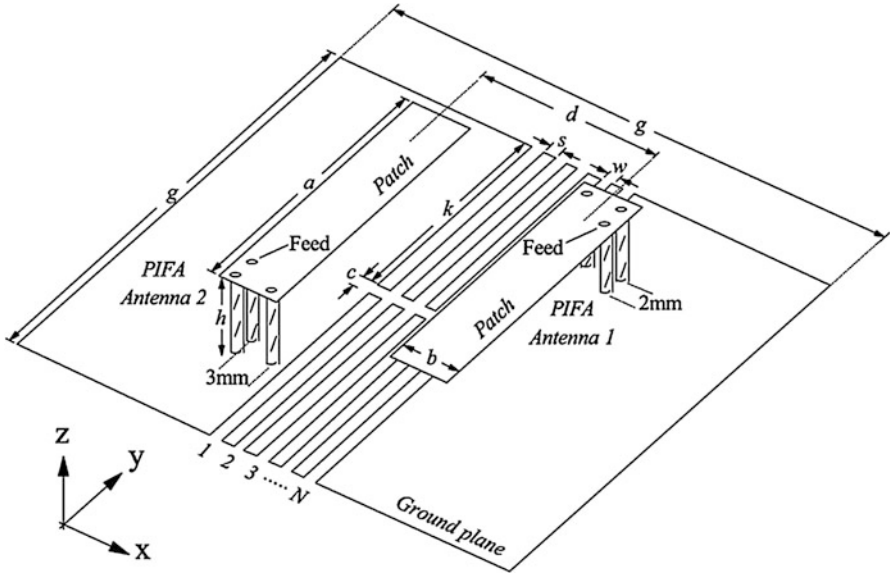


Fig. 3 Geometry of two closely-packed PIFAs with DGS (Chiu et al. 2007)

harmonics and cross polarization of a patch antenna in single-antenna design (Sung et al. 2003; Guha et al. 2005). For MIMO antennas, the effect of bandstop filter is employed to reduce the mutual coupling among the antenna elements. A common realization is to introduce multiple slits on the ground plane (Chiu et al. 2007; Li et al. 2009a; Huang and Lee 2009; Kokkinos et al. 2008; Zhu et al. 2009).

In the pioneering work of (Chiu et al. 2007), *N* pairs of slits were etched in the ground plane between two PIFA antennas, as presented in Fig. 3. The bandstop filter was built from slits interleaved with strips, where the capacitance comes from the strips, and the inductance is introduced along the central small connecting strip. The DGS part of the structure in Fig. 3 was extracted from the MIMO antennas and analyzed independently. Two transmission lines were connected to the DGS structure to test its reflection and coupling characteristics, which exhibit a stop-band from 2 to 3 GHz, i.e., most power is reflected back to the feeding port within the stop-band. Thus, when applied to MIMO antennas, most of electric current cannot flow from one antenna to the other, leading to improved isolation. The DGS approach was also successfully applied to other kind of antennas in (Chiu et al. 2007), including monopoles as non-planar radiating elements and patch antennas.

Following the same principles, different variations of DGS have been designed. Figure 4a shows an MIMO system with four quarter wavelength slots decoupled by a DGS between every two adjacent slot antennas (Li et al. 2010). Compared with the DGS in (Chiu et al. 2007), the required implementation space is halved and the terminal antennas can be more compact. In Fig. 4b, two coupled quarter wavelength slits with their openings on different edges of the ground plane are used (Kokkinos et al. 2008). Together with the ground plane, the two slots form an effective loop and resonate at a certain frequency. Thus, a transmission zero is introduced in the coupling path between the two antennas at the shorter edges, leading to high isolation of 50 dB. The shape of the DGS can also be different. For example, a simple ring shaped DGS was applied around a cylindrical dielectric resonator antenna (DRA) in (Guha et al. 2008), improving the isolation by 5 dB. The effect of different DGS structures, including dumb-bell shape, E shape and H shape were compared for closely packed microstrip patch antennas in (Elsheakh et al. 2010).

Apart from DGS, an electromagnetic band-gap (EBG) structure can also be used for decoupling. An EBG structure is constructed by arranging unit cells periodically on the ground plane (see Fig. 5). An EBG

DOI 10.1007/978-981-4560-75-7_99-1

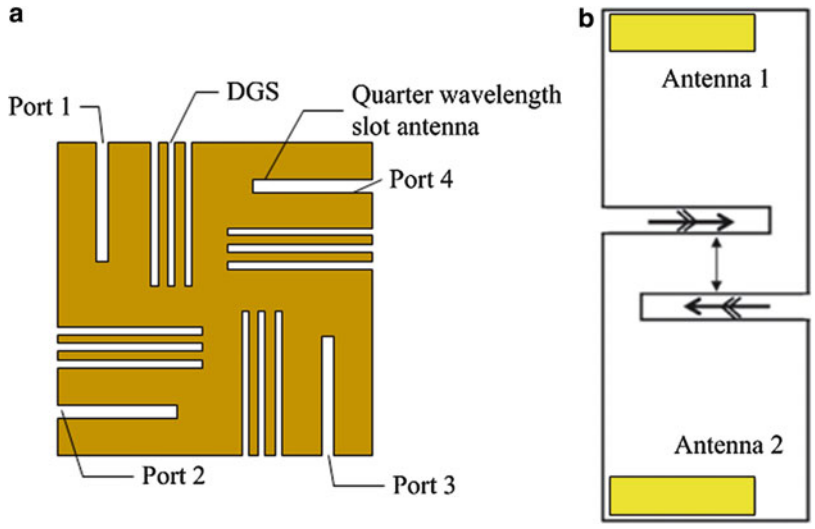


Fig. 4 (**a**) Orthogonally polarized slot antennas with DGS (Li et al. 2010); (**b**) DGS as a magnetic resonator (Kokkinos et al. 2008)

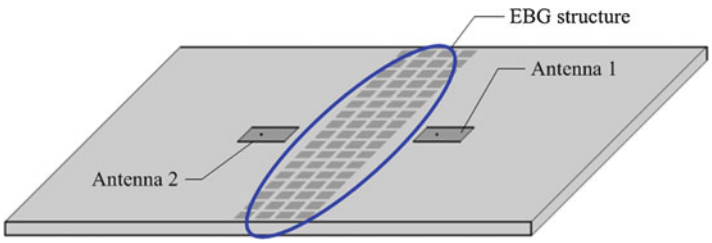


Fig. 5 The configuration of the EBG structure for decoupling (Yang and Rahmat-Samii 2003)

structrure behaves as a bandstop filter with a wide stop-band and high attenuation, which is well suited for decoupling multi-antennas (Yang and Rahmat-Samii 2003; Inclan-Sanchez et al. 2007). A typical EBG unit cell is a mushroom-like structure, consisting of a patch and a via connecting to the ground. By varying the dimensions of the unit cell and arranging them in two dimensions, the stop band can be achieved at different frequencies. In (Iglesias et al. 2008), using a 4-cell EBG structure, an isolation improvement of over 10 dB was achieved between two patch antennas at 2.8 GHz. However, restricted by the Bragg reflection condition, EBG structures are complicated and require a large implementation area, especially for low frequencies (Huang and Lee 2009).

Another similar decoupling technique is to use a soft surface in the modified ground plane, as shown in Fig. 6. Soft surface does the filtering in only one direction, so that it can have a smaller unit cell size when compared to EBG (Iglesias et al. 2009). With a soft surface, the isolation can normally be improved by around 10 dB. The use of more strips will result in higher isolation.

Neutralization/Connecting Line

Neutralization line approach is another antenna-level decoupling method. In general, a neutralization line is a conductive wire (or strip) that connects two antennas on a common ground plane to each other. Normally, the connection points are near the antenna feeds or shorting pins (if the antennas have shorting pins). In the beginning, the neutralization line is proposed to improve the isolation between two PIFAs whose resonating frequencies are very close to each other (Diallo et al. 2006). The technique was then

DOI 10.1007/978-981-4560-75-7_99-1

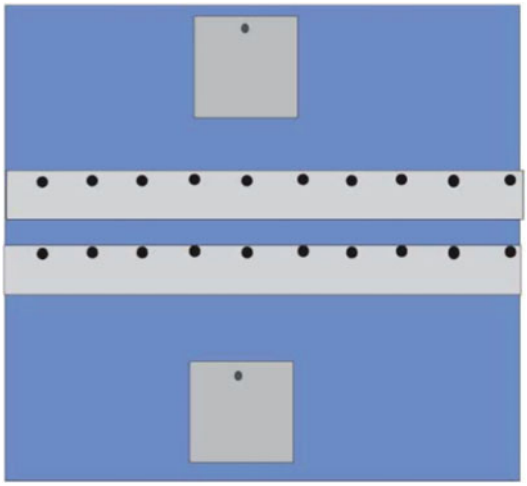


Fig. 6 Structure of soft surface to decouple two patch antennas (Iglesias et al. 2009)

applied to dual-antennas working at the same band, including different kinds of antennas, such as monopoles (Chung and Yoon 2007), patches (Ranvier et al. 2006) and PIFAs (Diallo et al. 2008). This technique is popular for designing MIMO antennas due to the simple design procedure and the relatively non-intrusive nature of the neutralization line structure.

The principle of the neutralization technique is to mitigate the existing electromagnetic coupling between the two antennas through introducing a direct connection (i.e., a new coupling path) via a suspended link. The design of PIFAs with the neutralization line between the feeding strips is shown in Fig. 7a. Strong current is observed on the connecting line at around 2 GHz, which cancels out the initial near field coupling and the coupling due to the shared ground plane. The effectiveness of the connecting line to achieve decoupling depends on the length and the width of the line, and a particular combination of the length and width of the line can lead to a better isolation level. Attributed to the connecting line, an isolation dip is achieved at around 2 GHz. The connecting line can also be placed between the shorting pins, which, unlike the connecting line between the feeding strips, can make the decoupling effect more frequency independent.

When the PIFAs are placed even closer to each other, as shown in Fig. 7b, there is insufficient space to accommodate the connecting line, which reduces its effectiveness. To decouple the antennas, one solution is to insert a folded line on the side of each PIFA, connecting the feeding strip to the ground plane (Chebihi et al. 2008). To obtain high isolation over multi-bands or a wide band, multiple neutralization lines of different lengths can be utilized (Wang and Du 2014), as shown in Fig. 8. Three neutralization lines operate at a center frequency of 1.81, 2.05 and 2.65 GHz, respectively, offering an isolation of over 15 dB from 1.7 to 2.7 GHz.

The neutralization line technique has also been analyzed from a generalized network perspective with a two-port model in (Shaker et al. 2008). Based on the model, the technique can be interpreted as a specific realization of the decoupling network in section "Decoupling Network", whereby the decoupling and matching network is built into the antenna by the use of neutralization line(s). Along the same line, another theoretical analysis of the neutralization lines was carried out in (Peng et al. 2013), where the neutralization lines were shown to be equivalent to 180° phase shifters and realized by feeding networks.

Parasitic Scatterer

In conventional single-antenna design, parasitic scatterers/elements are often used to create multi-resonances for bandwidth enhancement (Costa et al. 2007) or to obtain antenna pattern reconfigurability

DOI 10.1007/978-981-4560-75-7_99-1

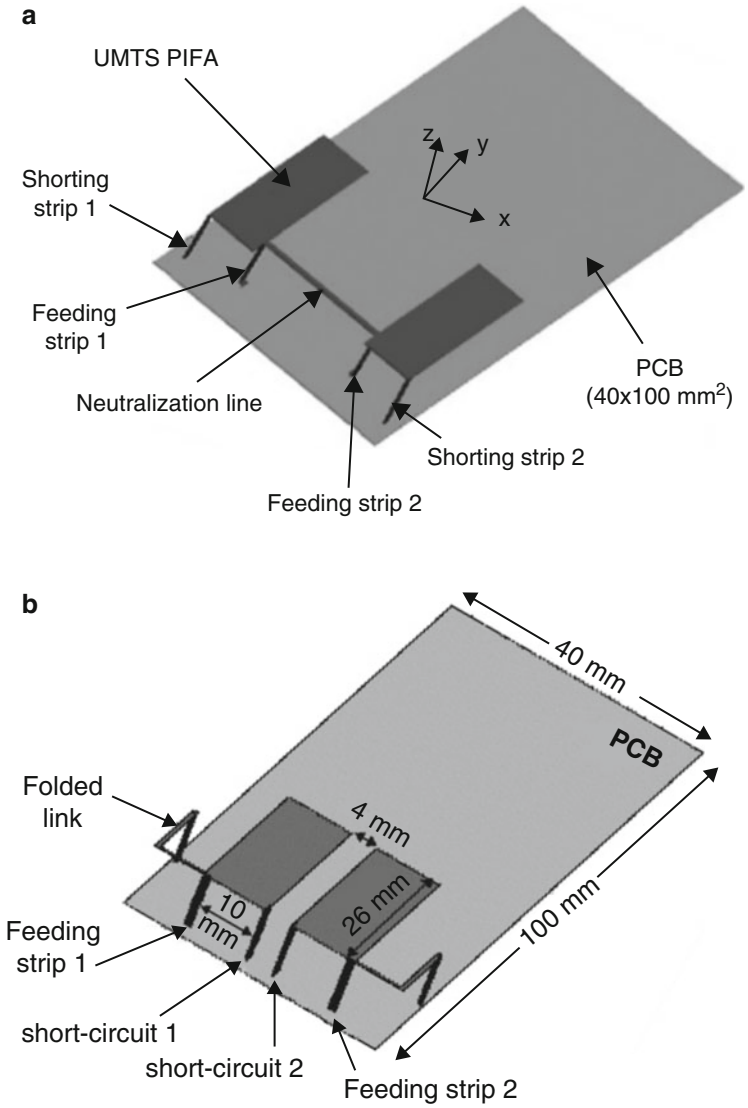


Fig. 7 (**a**) Two PIFA structure with the neutralization line between the feeding strips (Diallo et al. 2008); (**b**) two PIFA structure with the neutralization lines on the side (Chebihi et al. 2008)

(Lim and Ling 2010). In MIMO antenna design, it can also be added in between two radiating elements to reduce mutual coupling (Min et al. 2005; Lau and Andersen 2007; Mak et al. 2008; Minz and Garg 2010).

A parasitic scatterer artificially creates an additional coupling path between two antenna elements to reduce the coupling between them, which can be compared to a neutralization line adding a direct coupling path to achieve decoupling. By changing the configuration of the scatterer, the amplitude and phase of the coupling coefficient can be varied. A theoretical analysis of the parasitic scatterer can be found in (Lau and Andersen 2009, 2012). The principle of parasitic scatterer is illustrated with the setup in Fig. 9. In the "black box", a parasitic scatterer (port 2) is placed in between two antenna elements (ports 1 and 3). The parasitic scatterer is terminated by a load impedance Z_L, whereas the antenna elements are connected with matching circuits. A generalized three-port network can be used to analyze the setup, with which the voltage and current relationships between port 1 and port 3 can be obtained with the impedance of the antennas and the load Z_L. In order to decouple the antennas, it is required that the mutual impedance between port 1 and port 3 is zero. The load impedance of Z_L is then determined to fulfil the requirement. It

DOI 10.1007/978-981-4560-75-7_99-1

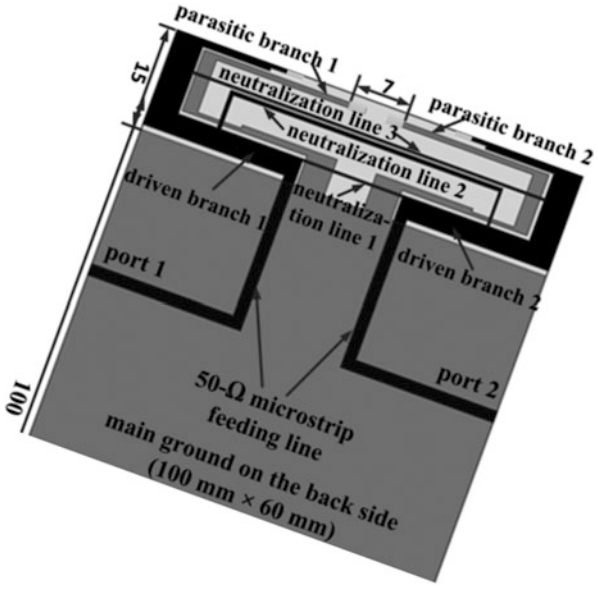


Fig. 8 A wideband dual-element MIMO antenna with good isolation achieved by using three neutralization lines (Wang and Du 2014)

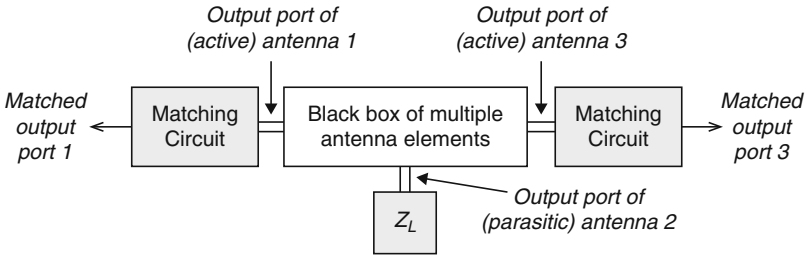


Fig. 9 Decoupling setup for a *black box* containing an arbitrary two-antenna structure (ports 1 and 3) and a parasitic scatterer (port 2) (Lau and Andersen 2012)

is worth noting that the closely coupled antenna and scatterer should be considered together as the radiating element. With the parasitic scatterer, a complete decoupling, i.e., zero coupling coefficient between antennas, can be achieved at the center frequency at the expense of bandwidth.

A more intuitive explanation of parasitic scatterer based on current distribution was provided in (Mak et al. 2008). Assuming a sinusoidal signal is fed into dipole antenna A, as seen in Fig. 10, a strong coupled current which is approximately 180° out of phase compared to the current in dipole A is then formed on dipole antenna B due to the close proximity of the two dipoles. When parasitic scatterer C, whose electrical length is around half wavelength, is inserted between the two dipoles, it creates another coupling path. The coupling starts from dipole A, goes through C and couples to dipole B. The current on Dipole B through this path is in phase with that on dipole A, because the scatterer C introduces an additional 180° phase. If coupling coefficient from A to B is denoted as α, and that from A to C is denoted as β, when β is properly optimized (e.g., using a capacitive load), the current on dipole B from the two coupling paths can cancel out each other, and decoupling is achieved.

Like the single-antenna counterpart, the parasitic scatterer has been successfully implemented on different dual-antenna structures. In Fig. 11a, a slotted meander line resonator is placed between two microstrip antennas (Alsath et al. 2013), improving the isolation by 10 dB. Though the parasitic scatterer

DOI 10.1007/978-981-4560-75-7_99-1

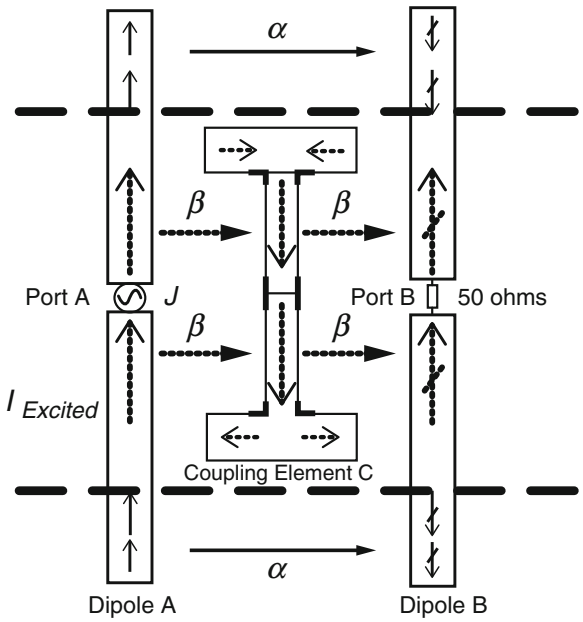


Fig. 10 Coupling between two dipoles with a parasitic scatterer (Mak et al. 2008)

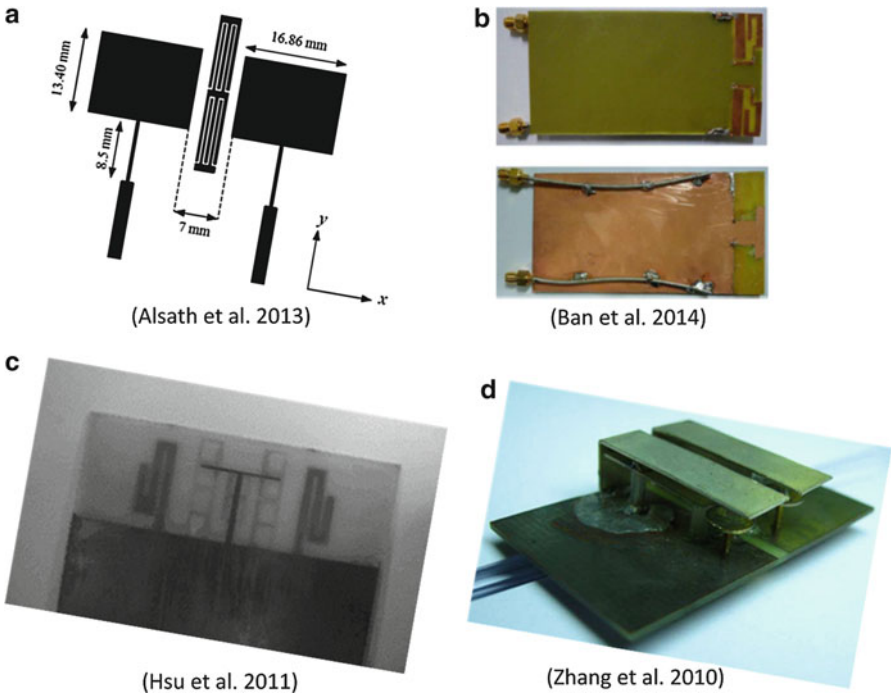


Fig. 11 Parasitic scatterers for isolation enhancement

in this case looks similar to the DGSs in section "Modified Ground Plane", the former DGS approach is based on the principle of stop-band filtering in choking out the coupling current, rather than introducing new coupling path(s). A popular implementation of parasitic scatterer is in the form of a grounded T-shaped structure (Ban et al. 2014; Lee et al. 2014; Hsu et al. 2011). The T structure as shown in Fig. 11b can also be utilized to accommodate some other electronic components, such as a USB port or a

DOI 10.1007/978-981-4560-75-7_99-1

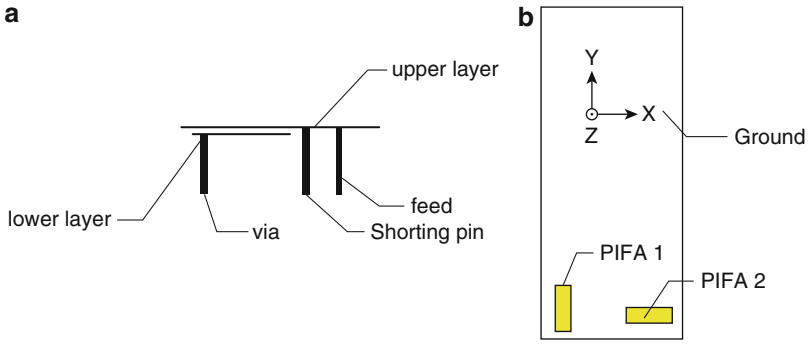


Fig. 12 Implementation of local ground plane to improve isolation (Li et al. 2009b)

loudspeaker. In Fig. 11c, the T-shaped branch is used together with an H-shaped metamaterial inspired insulator to achieve a 20 dB isolation bandwidth of about 12.8 %. Apart from adding a physical structure, the parasitic scatterer can also be formed opportunistically by placing two PIFAs side by side as illustrated in Fig. 11d, where the naturally formed slot in between the two PIFAs can be tuned to achieve decoupling (Zhang et al. 2010). This decoupling idea was further refined in (Zhang et al. 2012a) to allow for arbitrary ground plane size as well as dual-band operation.

Current Localization

As mentioned earlier, current flowing on the shared ground plane is a critical source of coupling for MIMO terminal antennas. The antenna-level decoupling techniques discussed so far in this chapter either create a new coupling path to cancel out the existed coupling or prevent the current flow by a bandstop filter. Different from these techniques, the technique of current localization aims to reduce the current on the shared ground plane through designing antennas that can better localize the current on the ground plane.

Adding a local ground plane is one way to realize current localization. As shown in Fig. 12 (Li et al. 2009b), a small local ground plane (lower layer) was inserted between PIFA and the main ground plane, so that the current is mostly confined to the lower patch underneath the PIFA. As a result, an isolation of over 20 dB was obtained. Another current localization method was proposed in (Gao et al. 2007).

Combination of Decoupling Techniques

Apart from the specific decoupling techniques described above, several decoupling techniques can be combined to further improve decoupling relative to the use of a single technique. In (Ling and Li 2011), good isolation is achieved for dual-band back-to-back monopoles by adding both a neutralization line and a parasitic scatterer between them, as shown in Fig. 13a. The neutralization line operates at the low band of 2.4 GHz, whereas the scatterer mainly takes effect at the high band of 5.8 GHz. Figure 13b shows a decoupling structure that consists of a grounded neutralization line and an ungrounded parasitic scatterer (Lee et al. 2009). The scatterer is a folded line with the electrical length of half wavelength. Another example of this decoupling approach is found in (Zhang et al. 2012b), where the parasitic scatterer that is formed from the gap between two ultra-wideband antennas further enhances the polarization diversity inherent to these antennas.

One common limitation for the decoupling techniques described in this section is that decoupling is normally achieved at the expense of bandwidth. This tradeoff can become a significant problem in frequencies below 1 GHz, since a given fractional bandwidth yields a lower absolute bandwidth for a lower center frequency, which can greatly complicate cellular band coverage. For example, 10 % bandwidth at 2.65 GHz (e.g., LTE Band 7) implies absolute bandwidth of 265 MHz, whereas 10 %

DOI 10.1007/978-981-4560-75-7_99-1

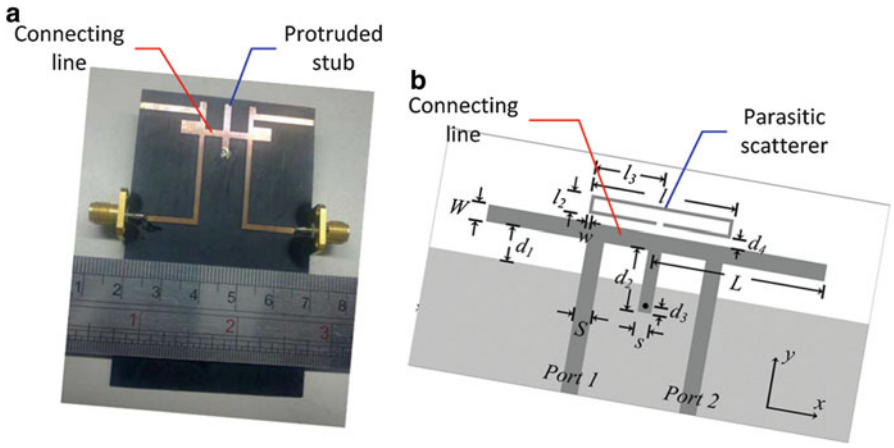


Fig. 13 (**a**) Combination method of connecting line and protruded stub (Ling and Li 2011) (**b**) Combination method of connecting line and parasitic scatterer (Lee et al. 2009)

bandwidth gives only 70 MHz absolute bandwidth at 700 MHz (e.g., LTE Band 13). Moreover, getting enough bandwidth is challenging at lower frequencies, even for single-antenna elements, due to decreasing electrical size of the mobile terminals. These problems can be addressed to some extent by the use of the Theory of Characteristic Modes (TCM) to design MIMO antennas, which is the subject of the following section.

Characteristic Mode Based Decoupling

In principle, the decoupling techniques in the previous section treat the ground plane of antennas as an electrically large structure, and hence the excitation of the ground plane (i.e., chassis) is not explicitly taken into consideration. However, for mobile terminal antennas, chassis radiation is critical for antenna design, especially for frequencies below 1 GHz. This is because at these frequencies, the largest dimension (i.e., length) of typical smartphone-type mobile terminals is around a third of a wavelength or less, with the antenna elements being even electrically smaller. Therefore, there is a large tendency for the chassis to be excited by the antenna elements to become the main radiator. Similar arguments apply for higher frequencies if terminals of smaller form factors are considered, such as USB dongles and smart watches. On the other hand, chassis radiation will become important for even lower frequencies if larger form factors are applied, including phablets and tablet PCs.

As will be further described in this section, chassis excitation complicates the design of decoupled and decorrelated MIMO antennas. Fortunately, recent results show that TCM can be used to significantly alleviate the design challenge. The main ideas of TCM are summarized below.

Theory of Characteristic Modes

Characteristic mode analysis is a method in electromagnetics, which gives insight into the potential resonant characteristics of a structure by finding and examining the inherent modes of the structure. Characteristic mode analysis is based on the Theory of Characteristic Modes (TCM), which was first introduced by Garbacz and Turpin (Garbacz 1968; Garbacz and Turpin 1971) and later refined by Harrington et al. (Harrington and Mautz 1971; Harrington et al. 1972).

Considering a conducting body with surface *S,* an incident or external (vector) electric field (or voltage) $\vec{E}^i$ can induce a surface current $\vec{J}$ on it. This surface current will further generate a scattered field $\vec{E}^s$.

DOI 10.1007/978-981-4560-75-7_99-1

According to electromagnetic boundary conditions, the tangential component on the surface of the conducting body satisfies the relation:

$$\left(\vec{E}^{i} + \vec{E}^{s}\right)_{\tan} = 0. \tag{21}$$

To relate the scattered electric field to the conductor surface current, an impedance operator of Z was defined, so that

$$Z\left(\vec{J}\right) = -\vec{E}^{s}_{\tan}, \tag{22}$$

where $Z\left(\vec{J}\right) = R\left(\vec{J}\right) + jX\left(\vec{J}\right)$, and R and X are the real and imaginary parts of Z, respectively.

The characteristic currents can be solved from the following eigenvalues equation:

$$X\left(\vec{J}_n\right) = \lambda_n R\left(\vec{J}_n\right), \tag{23}$$

where λ_n and $\vec{J}_n$ are the eigenvalues and eigenfunctions (eigencurrents), respectively, and both of them are real since R and X are real symmetric operators. Each eigenvalue represents the ratio of the modal stored reactive power (P_{stored}) to the radiated power (P_{rad}):

$$\lambda_n = \frac{P_{\text{stored}}}{P_{\text{rad}}} = \frac{\left\langle \vec{J}^{*}_n, R\vec{J}_n \right\rangle}{\left\langle \vec{J}^{*}_n, X\vec{J}_n \right\rangle}, \tag{24}$$

where $P_{\text{stored}} = \frac{1}{2}\left\langle \vec{J}^{*}_n, R\vec{J}_n \right\rangle, P_{\text{rad}} = \frac{1}{2}\left\langle \vec{J}^{*}_n, X\vec{J}_n \right\rangle$, and $\left\langle \vec{B}, \vec{C} \right\rangle = \oiint_S \vec{B} \cdot \vec{C}\ ds$.

Thus, the smaller the magnitude of λ_n, the more important the mode is for radiation, and $\lambda_n = 0$ corresponds to a resonant mode. One important property of the eigenvalue equation is that the characteristic currents and hence the modal electric far-fields are orthogonal, i.e.,

$$\left\langle \vec{J}^{*}_m, X\vec{J}_n \right\rangle = \frac{1}{\eta_0}\left\langle \vec{E}^{*}_m, \vec{E}_n \right\rangle = \delta_{mn}, \tag{25}$$

where η_0 is the free space wave impedance and δ_{mn} is a Kronecker delta function with $\delta_{mn} = 1$ for $m = n$, and $\delta_{mn} = 0$ for $m \neq n$. This property of orthogonal modes and patterns lends itself naturally to the design of multi-antennas for MIMO, which require low coupling and correlation between the antennas. With the above modal orthogonalities, any arbitrary surface current density can be expanded as a weighted sum of modal currents:

$$\vec{J} = \sum_n \alpha_n \vec{J}_n = \frac{\left\langle \vec{J}_n, \vec{E}^{i} \right\rangle}{1 + j\lambda_n} \vec{J}_n. \tag{26}$$

In this equation, λ_n and $\vec{J}_n$ depend solely on the structure and material of the conducting body, while α_n also depends on the excitation field $\vec{E}^{i}$, which is determined by the feeding method used. In antenna

DOI 10.1007/978-981-4560-75-7_99-1

design, α_n is decided by the number and location of the antenna ports, and also the amplitudes and phase of each excitation. In practice, one can excite mode n with a combination of inductive or capacitive feeds placed at appropriate location(s) on the conducting body (Martens et al. 2010).

Other than the conductor-only case, TCM was later extended to dielectric and magnetic bodies using a similar approach as for the conducting bodies, with the boundary condition in Eq. 21 replaced by the constitutive relationship (Harrington et al. 1972). For loss-free bodies, the current can still be expanded as a set of real characteristic functions, and the characteristic fields are orthogonal with each other. However for lossy materials, the characteristic far fields are no longer orthogonal. Most of the existing literature on TCM nowadays only treats the simple case of conducting bodies.

In the following, the TCM based MIMO antenna designs for frequencies above 1 GHz ("high frequencies") and below 1 GHz ("low frequencies") will be discussed separately. The two cases are treated differently because at high frequencies, the chassis readily offers multiple inherently orthogonal modes that can be excited to design decoupled and decorrelated MIMO antennas. However, at low frequencies, the chassis has only one resonant mode conventionally, and the single mode can become the main radiator that is shared by multi-antennas, which induces large coupling and correlation. Therefore, chassis excitation greatly complicates MIMO antenna design. To counteract this problem, different methods are used either to avoid the shared chassis radiation or to intentionally create more resonant modes for the same chassis form factor.

TCM Based MIMO Antenna Design for High Frequencies

For mobile terminals, the chassis plays a more important role than just being the ground plane of antenna elements. The impedance, bandwidth and radiation mode of an antenna element on a terminal ground plane is often defined by the location of the antenna and its feeding point on the ground plane (Best 2009). Efforts to apply TCM to terminal antenna design began around a decade ago and the topic has since been receiving increasing interest in the terminal antenna community. At the beginning, it was used to determine the optimum feeding arrangement to enhance the bandwidth of the antenna (Antonino-Daviu et al. 2004, 2005; Cabedo-Fabres et al. 2006; Schroeder et al. 2005, 2006). However, following the rise of MIMO technology, TCM is also employed to design MIMO terminal antennas. This is because TCM offers inherent orthogonal modes for a given structure, which is the ground plane in this case. The basic principle is that if the modes can be individually excited without affecting one other, ideal MIMO antennas with zero coupling and correlation can be obtained. In general, the orthogonal modes can be jointly excited to give multiple orthogonal antenna ports, as long as they are suitably combined in magnitude and phase to retain orthogonality at the antenna ports.

The eigenvalues of a candy-bar-type chassis with a size of 100 mm × 40 mm are shown in Fig. 14a. The figure shows four orthogonal modes which have closely spaced resonant frequencies at around 2.5 GHz. The current distributions of the four modes whose eigenvalues are close to zero are shown in Fig. 14b. The specific current distributions show sinusoidal behavior over the longer or shorter edges of the chassis. Each mode has its unique set of current maxima and minima at specific positions, which provide information on the location and type of feed to use. Through using four compact non-resonant capacitive couplers located at four vertices of the chassis and a feeding network composed of four 180° hybrid couplers, four orthogonal patterns are achieved with a small modification of the chassis. However, orthogonal patterns and good impedance matching cannot be achieved at the same time for all four antennas. It has been shown in (Ethier 2008) that for the 2.5 GHz band, the chassis of a typically mobile terminal is only capable of supporting at most three efficient radiating modes.

In addition to current distribution, near field distribution (including electric and magnetic fields) above the chassis is also important for mode excitation. The electric field and the current distribution will have a phase difference of 90°, i.e., the electric field distribution will have its minima at the maxima of the current

DOI 10.1007/978-981-4560-75-7_99-1

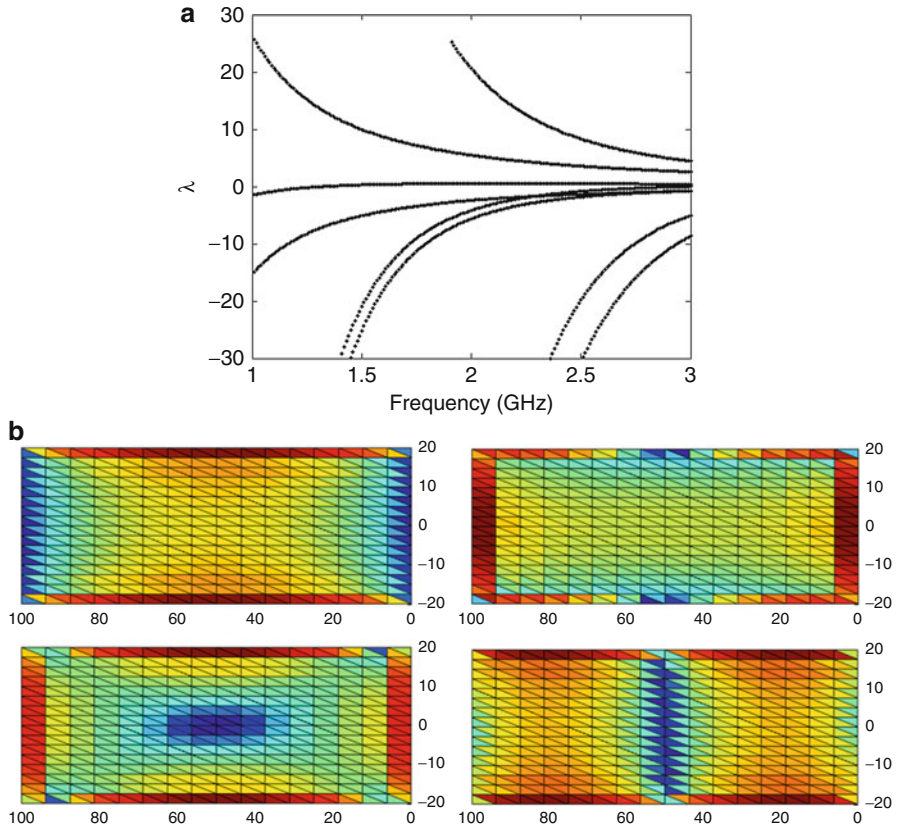


Fig. 14 (**a**) Characteristic eigenvalues over frequency for 100 × 40 mm chassis (**b**) Characteristic current distribution of the chassis. The chassis material is Perfect Electric Conductor (*PEC*)

distribution and vice versa. It means that if an inductive coupler is used, it should be located at one of the maxima in the current distribution. On the other hand, a capacitive coupler should be placed at one of the maxima in the electric field distribution.

In (Martens et al. 2010), inductive and capacitive feeds were used to selectively excite three modes of the chassis at 2.45 GHz. Each non-resonant inductive coupling elements (ICE) was created by a wire loop etched into one edge of the chassis, with its location corresponding to the location of one of the maxima in the current distribution of the desired mode, in order to excite a specific characteristic mode. The phase of each ICE was adjusted according to the requirement of the modes. For capacitive coupling, a small co-planar plate close to the main chassis was used as a non-resonant capacitive coupling element (CCE). Multiple CCEs were then located to excite the desire modes according to the locations of electric field maxima. In both cases of inductive and capacitive coupling, matching networks were needed to enhance the impedance bandwidth. A comparison between the inductive and capacitive coupling approaches is given in (Martens et al. 2010). For the inductive coupling approach, a three-port MIMO antenna was designed to achieve isolations of above 25 dB, and the correlations between all antenna pairs are lower than 0.02. However, one major drawback in this approach of using multiple non-resonant elements to excite any one of the desired modes is that the efficiency of the antennas is limited by the losses in the complicated matching network needed to properly excite the modes and to obtain good impedance bandwidth.

To improve coupling into the desired modes when fewer coupling elements are used, electrically larger coupling elements can be used, such as those shown in Fig. 15 (Kishor and Hum 2013). Different from using coupling elements as pure mode exciter, the complete antenna structure (ground plane and the exciters)

DOI 10.1007/978-981-4560-75-7_99-1

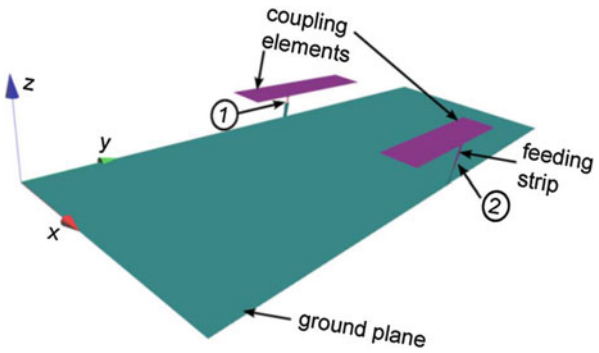


Fig. 15 Design of chassis mode based antenna with large exciters (Kishor and Hum 2013)

needs to be analyzed using TCM, because the eigenmodes and the radiation characteristics will change due to the increased electrical size of the coupling element. Moreover, unlike the MIMO antennas in (Martens et al. 2010), which rely on complicated matching networks with hybrid 180° couplers to excite non-overlapping sets of characteristic modes, only simple impedance matching is needed for the dual antenna system in Fig. 15 to excite two separate modes. Even though these two modes are simultaneously excited by both ports, the out-of-phase behavior between the ports ensures that pattern orthogonality is retained. As a result, the isolation of above 20 dB and the envelope correlation coefficient of below 0.1 are achieved.

In a separate study that accounts for the impact of larger coupling elements on characteristic modes, a two-port pattern reconfigurable MIMO antenna was designed using the method of port analysis (Kishor and Hum 2014). In this design, two coupling elements that were used as antenna ports were located at diagonally opposite corners of the chassis, where the electric field is at its maximum. The remaining two coupling elements were used as parasitic elements with reconfigurable impedance loads (i.e., Z_{L3}, Z_{L4}). Four orthogonal modes were selected and manipulated by TCM analysis to provide two sets of orthogonal patterns across the two ports. Each set of two orthogonal patterns, which represents one reconfigurable state, was obtained using different Z_{L3}, Z_{L4} selected based on port analysis. The design provides both low intra-state and inter-state correlations.

TCM Based MIMO Antenna Design for Low Frequencies

As explained earlier, a typical terminal chassis conventionally offers only one excitable mode at frequencies below 1 GHz. Moreover, multiple electrically small antenna elements tend to rely on this single mode as their dominant radiator. Hence, the problem of coupling is more severe at lower frequencies as compared to higher frequencies, and the isolation of MIMO terminal antennas with cellular band coverage is typically around 6 dB or less, e.g., (Plicanic et al. 2009).

To quantify this problem, the characteristic modes of a typical candy-bar-type chassis with a size of 100 mm × 40 mm are shown in Fig. 16. It is seen that there is only one mode that is excitable at around or less than 1 GHz. To further understand the single-mode problem, the normalized characteristic electric and magnetic fields on a plane 5 mm above the 100 mm × 40 mm chassis are shown for the single resonant mode at 1 GHz in Fig. 17a and b, respectively. The strong electric field at the chassis edges in Fig. 17a reveals that an electric antenna that is placed at either of the two shorter edges of the chassis can strongly excite the mode. Therefore, if two electric antennas are placed at the two chassis edges, both antennas will share the chassis as their main radiator, causing severe mutual coupling.

To avoid the excitation of the shared chassis by more than one antenna, the simplest method is to optimize the position of the second antenna to effectively reduce chassis excitation for that antenna (Li et al. 2012a). The dual-antenna design with PIFA and slot monopole elements as shown in Fig. 18a

DOI 10.1007/978-981-4560-75-7_99-1

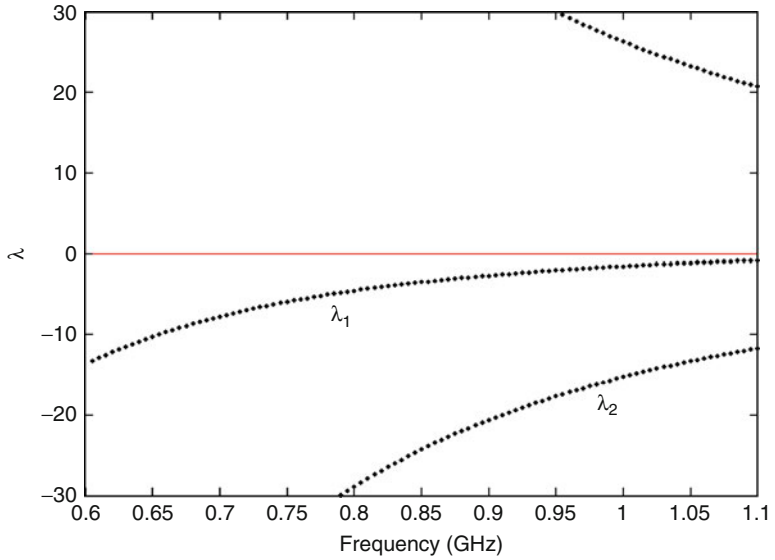


Fig. 16 Characteristic eigenvalues over frequency for 100 × 40 mm chassis (Li et al. 2012a). The chassis material is Perfect Electric Conductor (*PEC*)

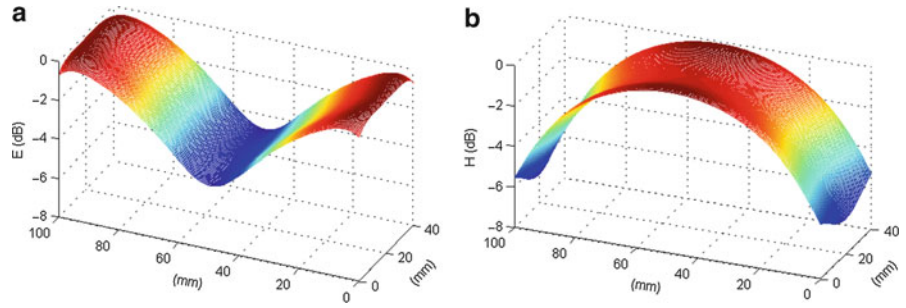


Fig. 17 The normalized magnitude of (**a**) the total electric field and (**b**) the total magnetic field for the first mode of the chassis (chassis size: 100 × 40 mm) (Li et al. 2012a)

was used to demonstrate this design principle (Li et al. 2012a). To optimize the antenna location, the first antenna, which is a slot monopole, was kept at one edge of the chassis (to excite the single chassis mode), whereas the second antenna, which is a PIFA, was moved from the edge of the chassis towards the center position (to avoid exciting the mode). As a result, the isolation was enhanced from below 4 to 10 dB and the capacity for equal power condition was increased from 5.5 to 7.5 bps/Hz when averaged over a bandwidth of 30 MHz. It is interesting to note that the significant isolation improvement occur even though the distance between antenna elements is halved. Therefore, this simple example also reveals that chassis coupling dominates over other coupling mechanisms, which should increase with the smaller inter-element distance.

Another solution to the chassis induced coupling problem is to utilize magnetic antenna instead of electric antenna for one of the two antennas (Li et al. 2012b). As opposed to an electric antenna, a magnetic antenna stores mainly magnetic energy (rather than electric energy) in the near field. According to Fig. 17b, the magnetic field of the chassis mode is quite weak at the shorter edges. Thus, if a magnetic antenna is placed at a shorter edge, the chassis will not be excited. This means that an electric antenna, if placed on the opposite shorter edge to excite the chassis, will be relatively unaffected by the magnetic antenna, resulting in high isolation. Following this principle, a coupled fed loop antenna that radiates like

DOI 10.1007/978-981-4560-75-7_99-1

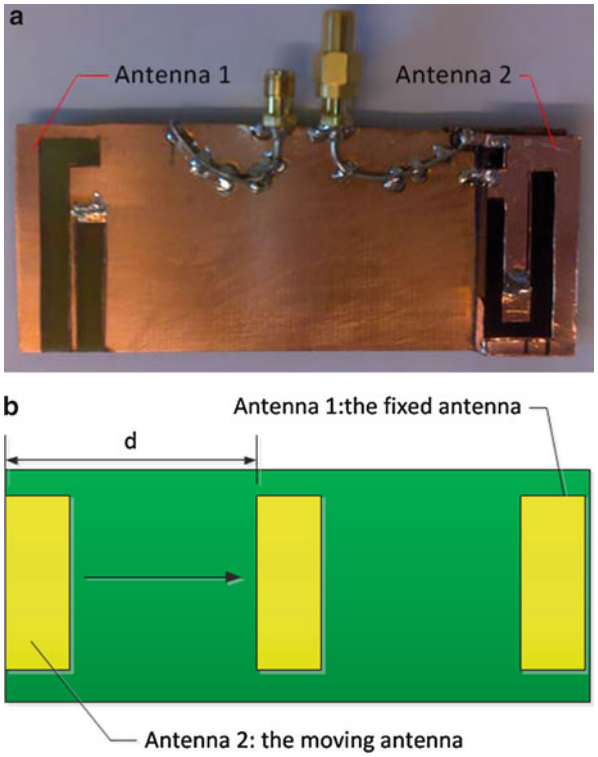


Fig. 18 (**a**) Prototype when the PIFA is at the short edge; (**b**) the study of MIMO performance for different positions of the PIFA (Li et al. 2012a)

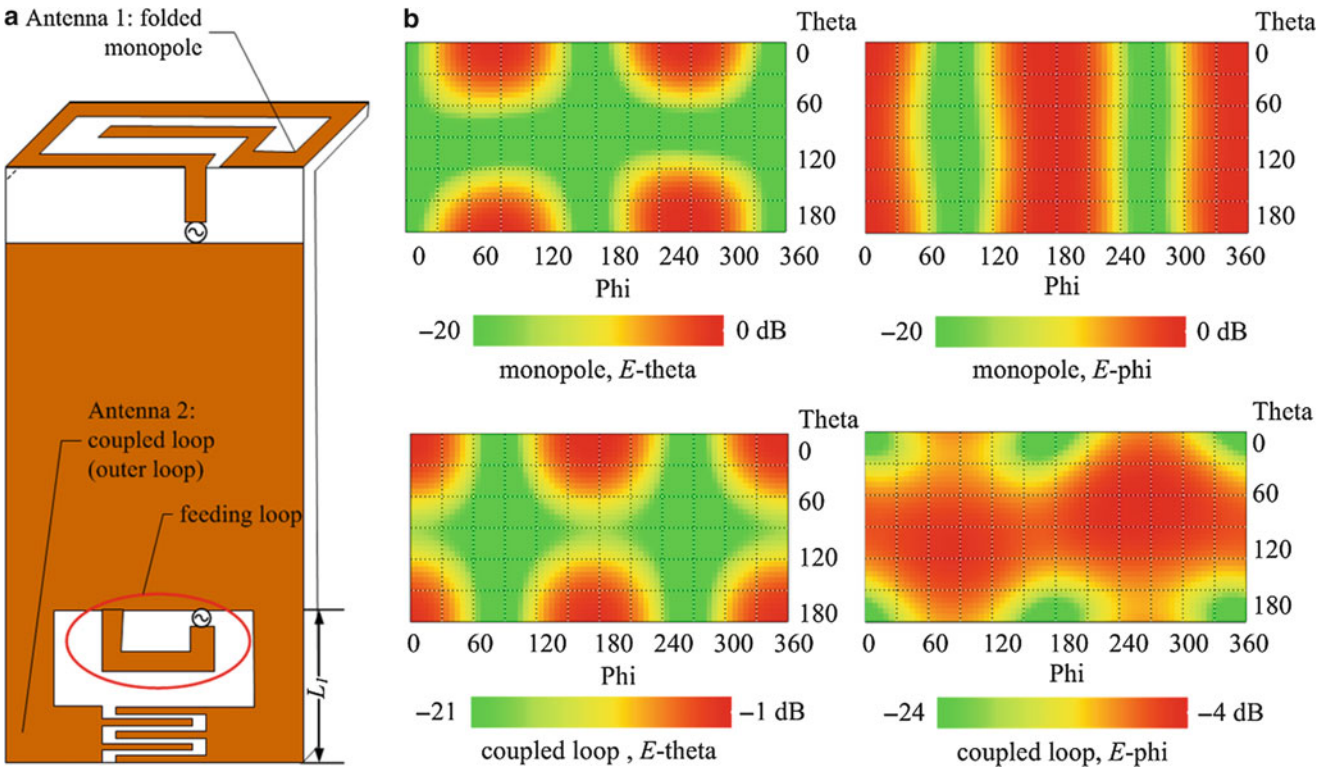


Fig. 19 Orthogonal dual-antenna design using monopole and coupled loop. (**a**) Antenna structure; (**b**) far field patterns (Li et al. 2012b)

DOI 10.1007/978-981-4560-75-7_99-1

a magnetic dipole was designed at one shorter edge, as shown in Fig. 19a. The small half loop inside is used to match the outer loop to 50 Ω. On the other shorter egde, a folded monopole is used to excite the chassis mode and create an electric dipole like pattern. Thus, the polarizations of the antennas are different and the patterns are orthogonal to each other, as shown in Fig. 19b. The isolation between the antennas is above 30 dB in simulation. With some modifications, the two antennas can be co-located at the same edge of the chassis to save more space on the PCB.

The main drawback of the two above decoupling methods is that the bandwidth of the non-chassis-exciting antenna is usually limited due to it being electrically small and not making use of the electrically larger chassis to radiate. Therefore, it is difficult to obtain low mutual coupling and correlation as well as adequate antenna bandwidth at the same time.

To solve the problem of limited bandwidth, an entirely different decoupling approach has been proposed. In particular, minor modifications of the chassis can be used to generate multiple resonant modes at low frequencies (i.e., lower than 1 GHz). Once multiple modes are achieved, they can be exploited in a similar manner as was described in the previous section for the case of higher frequencies.

The first example of chassis modification is to load the chassis with a metal bezel along its periphery, with the bezel connected to the center of one shorter chassis edge through a shorting pin (Li et al. 2013b). The geometries of the bezel-loaded chassis are shown in Fig. 20a. It is easy to practically integrate the bezel and the shorting pin onto the mobile casing. In fact, it should be noted that metal bezel is a design feature of some existing smartphone models, including Apple iPhone 4, which means that an existing structure of these phones can be used opportunistically to implement this decoupling technique. The eigenvalues of the bezel-loaded chassis are presented in Fig. 20b. At frequencies below 1 GHz, it can be seen that apart from the fundamental mode that was found earlier (denoted by λ_2), a new bezel mode (λ_1) is generated. The eigenvalue of this new mode is close to zero at frequencies around 0.81 GHz. The two

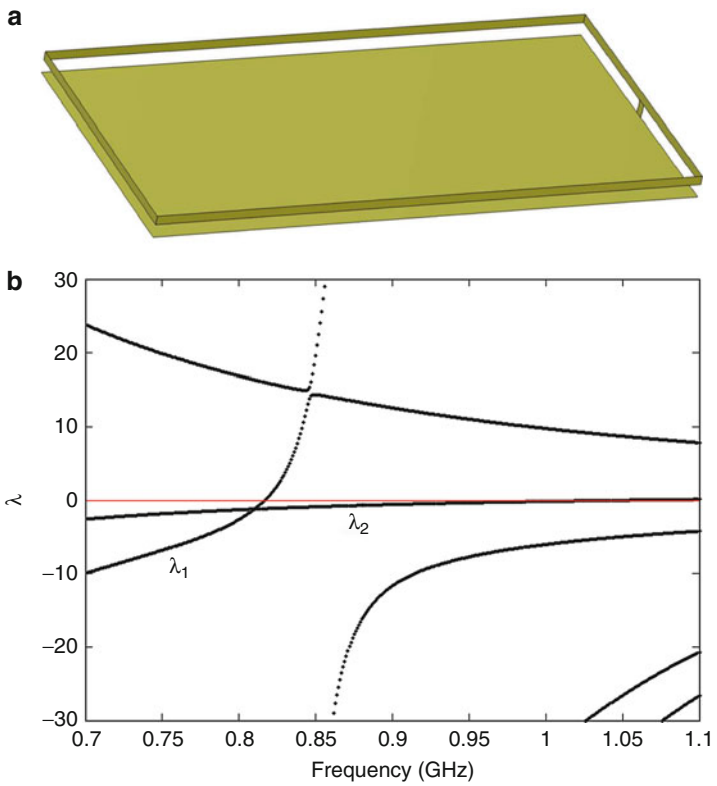


Fig. 20 (**a**) Geometries of the bezel-loaded chassis. (**b**) Eigenvalues of the bezel-loaded chassis (Li et al. 2013b)

DOI 10.1007/978-981-4560-75-7_99-1

orthogonal resonant modes at the same frequency can be used to design uncorrelated dual-element MIMO antennas through proper excitations. Detailed current distributions and characteristic patterns of the two modes are given in (Li et al. 2013b). High isolation of over 20 dB is achieved for the designed antennas. The eigenvalue curve also indicates that the fundamental mode has a large bandwidth, whereas the bandwidth of the bezel mode is relatively narrow. This is because a given mode becomes harder to excite as the eigenvalue increases, and a steeper slope in the eigenvalue around its resonant frequency indicates that modal excitation is fundamentally limited to a smaller bandwidth (termed as smaller "bandwidth potential" in the literature).

Another example of practical chassis modification is to load the chassis with T-strips along the longer edges (Li et al. 2014). The geometries of the T-strip loaded chassis are shown in Fig. 21a. Two metal strips along the length of the chassis are connected to the chassis through shorting pins at the center of each strip. The eigenvalues in Fig. 21b show that there are three resonant modes at frequencies below 1 GHz. Mode 1, which corresponds to the T-strip mode, and mode 3, which is the chassis' fundamental mode, are exploited to radiate due to their larger bandwidth potential (i.e., more gentle slopes around the resonant frequencies).

Through analyzing the characteristic electric and magnetic fields of the T-strip loaded chassis for modes 1 and 3, the feeding methods of both modes are determined. As shown in Fig. 22, a small narrow plate parallel to the chassis is used to create capacitive coupling to excite the T-mode, whereas a conventional coupled fed monopole is used to excite the fundamental mode. The detailed description on how to implement the feeds can be founded in (Li et al. 2014). The measured results show that both antennas cover LTE Band 8 with an isolation of above 10 dB. The envelope correlation between the two antennas in

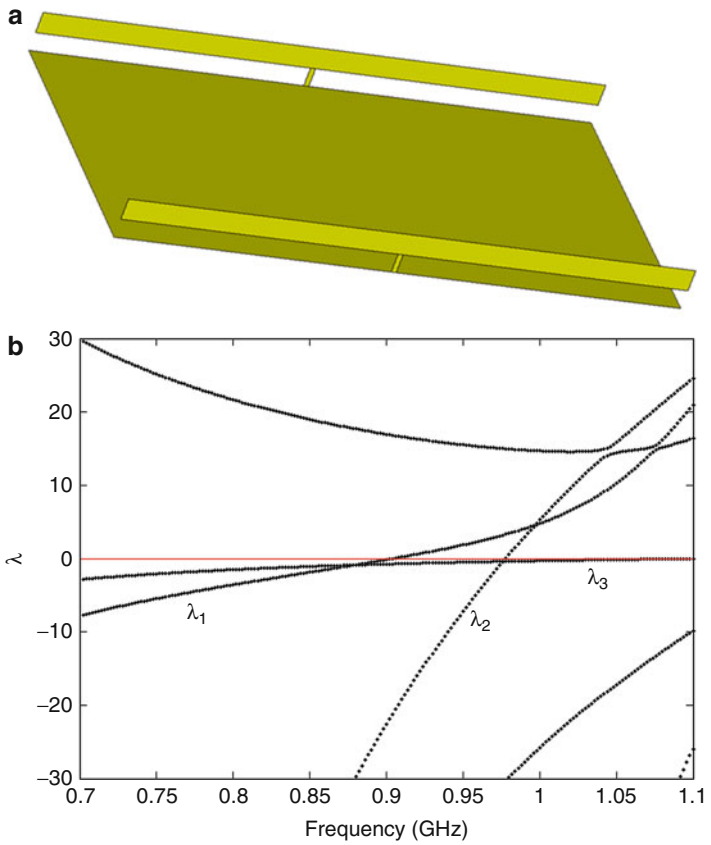


Fig. 21 (**a**) Geometries of the T-strip loaded chassis; (**b**) Eigenvalues of the T-strip loaded chassis (Li et al. 2014)

DOI 10.1007/978-981-4560-75-7_99-1

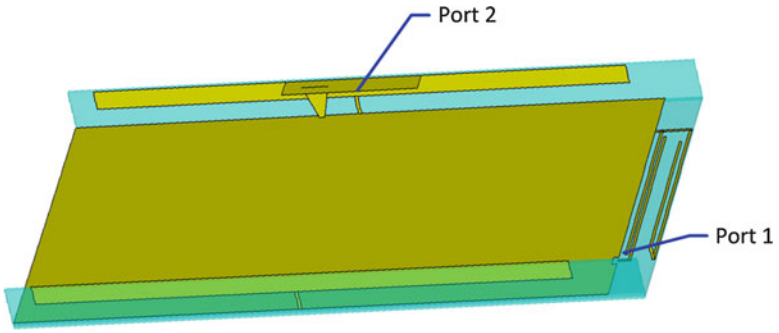


Fig. 22 Geometries of the T-strip loaded antenna system (Li et al. 2014)

free space is below 0.1 over the operating band. Due to the above reasons, a higher capacity was achieved, relative to existing designs. Another advantage of the antenna is that the S parameters and the envelope correlation coefficients were found in full-wave simulations to be only marginally influenced by the hands and the mobile casing.

The design of characteristic mode based MIMO antennas has been developed further to enhance the bandwidth and to allow multiband operation, two aspects which are critical for mobile terminal applications (Miers et al. 2013). The design framework is based on the correlation of characteristic currents (or characteristic near fields) between different modes. The correlation is used to identify if there are other modes having similar current characteristics in the vicinity of existing feed locations. By making opportunistic use of existing feeds to excite additional mode(s) at a different but close-by frequency, the bandwidth of the antenna can be enhanced due to multimode resonances. Multi-band operation is achieved if the chosen frequency is far away from the existing frequency band. In most cases, further (but minor) chassis modifications are needed.

For example, the eigenvalues of the T-strip loaded chassis were modified (as shown by the solid lines in Fig. 23a) when the two original shorting pins were displaced rotationally by 22 mm from the center. The chassis with the displaced shorting pins is shown Fig. 23b. The characteristic currents for mode 1 and mode 2 were highly correlated at the shorting pins, though their currents and near fields were significantly decorrelated across the remaining parts of the chassis. Consequently, simultaneous excitation of both λ_1 and λ_2 is possible through locating a single gap feed at one of the two shorting pins. This step extended the structure's bandwidth potential for the resonance below 1 GHz. For multi-band operation, cross-correlation was carried out between the currents around the shorting pins at the resonances of λ_1 and λ_2 and the characteristic currents of all the significant modes from 1.3 to 2.0 GHz. As a result, λ_2 and λ_4 showed significantly high correlation with the feed currents in the low band. However, neither of the two modes was resonant near 1.8 GHz, as can be seen in Fig. 23a. Thus, some structural modification was needed. Through shortening the shorter ends of the T-strips, the eigenvalues are modified as shown by the dashed lines ("Adapted modes") in Fig. 23a, where λ_2 and λ_4 are now close to zero (i.e., modal resonance) at the frequencies around 1.8 GHz. Thus, multiple bands have been achieved, along with enhanced bandwidth at the lower band. Following similar principles, multi-band operation of the bezel loaded antenna was also obtained, and the details can be found in (Miers et al. 2014).

DOI 10.1007/978-981-4560-75-7_99-1

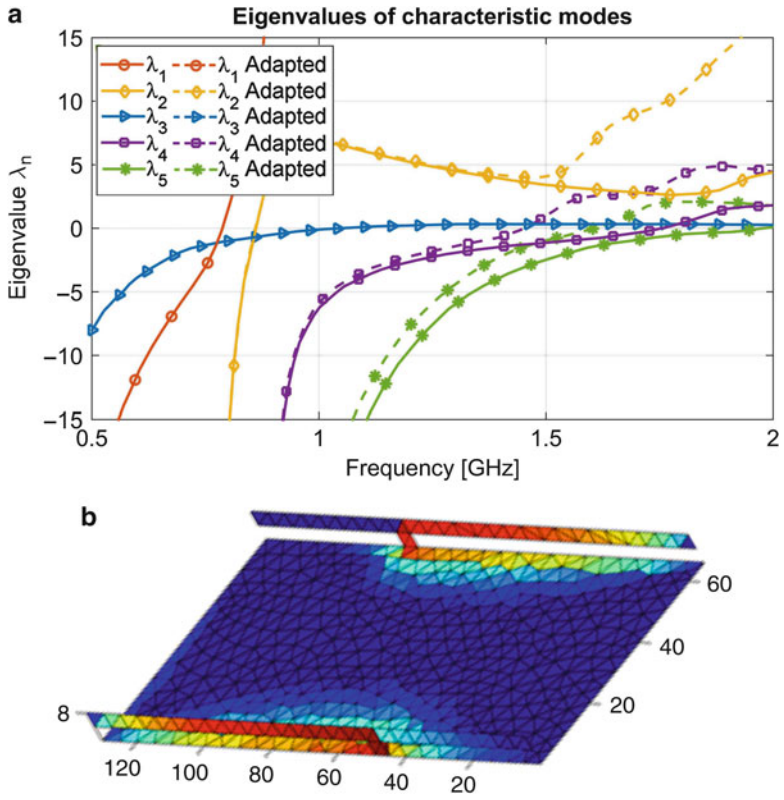


Fig. 23 (**a**) First five eigenvalues of the modified chassis with shorting pins at 22 mm rotationally symmetric offset. Adapted modes were formed through shortening the shorter ends of the capacitive strips by 10 mm; (**b**) T-strip loaded structure with shorting pins at 22 mm rotationally symmetric offset (Miers et al. 2013)

Future Directions and Open Problems

As detailed in this chapter, significant progress has been made in addressing the challenging problem of designing efficient multi-antennas for MIMO terminals. Apart from useful metrics that have been devised to better quantify MIMO performance, a number of decoupling and decorrelation techniques have been proposed to counteract the problem of limited implementation space in small terminals. In particular, the recently proposed approach of applying TCM analysis for designing MIMO terminal antennas has been shown to be very effective, since the characteristic modes have the desirable property that their farfield patterns are inherent orthogonal. Furthermore, since TCM analysis takes the entire terminal chassis into account, it can also provide decorrelated multi-antenna solutions with larger bandwidths than other existing methods.

Table 1 provides a quick overview of the different decoupling techniques described in this chapter, including their mechanisms, advantages and disadvantages. The concise information in the table can serve as guidelines for antenna designers on available options in designing decoupled and decorrelated MIMO terminal antennas.

Notwithstanding, the terminal antenna research community is facing even tougher challenges on the horizon. For example, nearly 40 cellular bands that range from 700 to 3.8 MHz have now been specified (3GPP TS36.101 2014). Even if simultaneous coverage of all bands is beyond reach, operators are demanding multi-antenna solutions with much larger bandwidths than those achieved today. Recent results have demonstrated that a tunable, chassis-exciting antenna solution can provide continuous

DOI 10.1007/978-981-4560-75-7_99-1

Table 1 Overview of different decoupling techniques

Decoupling techniques	Decoupling mechanisms	Advantages	Disadvantages
Decoupling network	Mitigation of mutual impedance or coupling coefficients using a coupled network	Decoupling of antenna ports without changing the antenna structure	Loss in the network; can be physically large if distributed elements used; can be hard to implement the network
Modified ground plane	Choking of coupling current by means of bandstop filter	Applicable to various kinds of ground plane based antennas; can be simple to implement	Change in the ground plane structure required
Neutralization line	Creation of an additional (direct) decoupling path through the line to cancel out existing coupling	Intuitive and simple to implement	Change in the antenna structure needed
Parasitic scatterer	Creation of an additional coupling path between two antenna elements through a scatterer	Simple to implement; high total efficiency demonstrated	A scatterer must be added between the coupled antennas
Current localization	Localization of the current to the near proximity of the excited antennas	Simple to implement; no additional structure needed	The decoupling method is antenna-specific
Characteristic mode	Excitation of different orthogonal modes of an antenna structure to achieve very low correlation	Systematical design method providing information on feeding method and location; larger bandwidth than other methods by exploiting the ground plane	Minor modification of the antenna structure (including ground plane) may be needed

coverage over the entire frequency range from 750 to 2500 MHz (Valkonen et al. 2013). However, this solution relies on the excitation of the relatively wideband fundamental mode of the chassis, especially at the lower frequency end. It will be very challenging to design a second antenna based on mode(s) that can be excited as well as being orthogonal to the first antenna.

To further complicate the design problem, the carrier aggregation (CA) feature in LTE requires simultaneous coverage across two cellular bands that can be very widely spaced. For example, inter-band CA is specified for LTE Band 1 (1920–2170 MHz) and Band 5 (824–894 MHz). This means that any tunable or frequency reconfigurable solutions should allow for simultaneous coverage even across large frequency spacings.

Furthermore, as LTE-Advanced (or "true 4G") with even higher data rates than LTE (up to 1 Gbits per second) is starting to appear on the market, terminal antenna designers will also need to take into account the coverage of both uplink and downlink frequencies for each band, as it allows for uplink MIMO transmissions. On top of even higher bandwidth requirements than LTE, LTE-Advanced also provides for up to 8×8 MIMO and 4×4 MIMO on the downlink and uplink, respectively.

Other major future tasks for the terminal antenna research community include the need to provide intelligent antenna solutions that will optimize the antennas to the immediate environment, which includes the user as well as the propagation channel. For example, measurement results indicate that reconfigurable radiations pattern can enable up to 30 % improvement in capacity in an interference-limited outdoor macrocellular environment (Plicanic et al. 2012).

DOI 10.1007/978-981-4560-75-7_99-1

Cross-References

- ▶ Antennas in Hand-Held Devices
- ▶ Broadband and Multiband Planar Antennas
- ▶ Commercial Antenna Design Tools
- ▶ Loop Antennas
- ▶ Low-Profile Antennas
- ▶ Metamaterials and Antennas
- ▶ Small Antennas (PIFA/PILA/Loading Antenna/etc)

References

3GPP TS 36.101 (2014) LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception. Version 11.9.0, Release 11

Ahn D, Park JS, Kim CS, Qian Y, Itoh T (2001) A design of the low pass filter using the novel microstrip defected ground plane structure. IEEE Trans Microwave Theory Tech 49:86–93

Alsath MGN, Kanagasabai M, Balasubramanian B (2013) Implementation of slotted meander-line resonators for isolation enhancement in microstrip patch antenna array. IEEE Antennas Wirel Propag Lett 12:15–18

Antonino-Daviu E, Cabedo-Fabres M, Ferrando-Bataller M, Herranz-Herruzo JI (2004) Analysis of the coupled chassis-antenna modes in mobile handsets. IEEE Antennas Propag Soc Int Symp 3:2751–2754

Antonino-Daviu E, Cabedo-Fabres M, Ferrando-Bataller M, Valero-Nogueira A, Martinez-Vazquez M (2005) Novel antenna for mobile terminals based on the chassis-antenna coupling. IEEE Antennas Propag Soc Int Symp 1A:503–506

Ban YL, Yang S, Chen Z, Kang K, Li JL (2014) Decoupled planar WWAN antennas with T-shaped protruded ground for smartphone applications. IEEE Antennas Wirel Propag Lett 13:482–486

Best SR (2009) The significance of ground-plane size and antenna location in establishing the performance of ground-plane-dependent antennas. IEEE Antennas Propag Mag 51:29–42

Blanch S, Romeu J, Corbella I (2003) Exact representation of antenna system diversity performance from input parameter description. Electron Lett 39:705–707

Cabedo-Fabres M, Valero-Nogueira A, Antonino-Daviu E, Ferrando-Bataller M (2006) Modal analysis of a radiating slotted PCB for mobile handsets. In: 1st European conference on antennas and propagation (EuCAP 2006), Nice, France

Chaudhury SK, Schroeder WL, Chaloupka HJ (2007) Multiple antenna concept based on characteristic modes of mobile phone chassis. In: 2nd European conference on antennas and propagation (EUCAP2007), Edinburgh, UK

Chebihi A, Luxey C, Diallo A, Thuc P, Staraj R (2008) A novel isolation technique for closely spaced PIFAs for UMTS mobile phones. IEEE Antennas Wirel Propag Lett 7:665–668

Chen SC, Wang YS, Chung SJ (2008) A decoupling technique for increasing the port isolation between two strongly coupled antennas. IEEE Trans Antennas Propag 56:3650–3658

Chiu CY, Cheng CH, Murch RD, Rowell CR (2007) Reduction of mutual coupling between closely-packed antenna elements. IEEE Trans Antennas Propag 55:1732–1738

Chung K, Yoon JH (2007) Integrated MIMO antenna with high isolation characteristics. IET Electron Lett 43:199–201

Coetzee JC, Yu Y (2008a) Closed-form design equation for decoupling networks of small arrays. Electron Lett 44:1441–1442

Coetzee JC, Yu Y (2008b) Port decoupling for small arrays by means of an eigenmode feed network. IEEE Trans Antennas Propag 56:1587–1593

Costa KQ, Dmitriev V, Nascimento DC, Lacava JCS (2007) Broadband L-porbe fed patch antenna combined with passive loop elements. IEEE Antennas Wirel Propag Lett 6:100–102

Diallo A, Luxey C, Thuc PL, Staraj R, Kossiavas G (2006) Study and reduction of the mutual coupling between two mobile phone PIFAs operating in the DCS1800 and UMTS bands. IEEE Trans Antennas Propag 54:3063–3074

Diallo A, Luxey C, Thuc PL, Staraj R, Kossiavas G (2008) Enhanced two-antenna structures for universal mobile telecommunications system diversity terminals. IET Microw Antennas Propag 2:93–101

Dossche S, Blanch S, Romeu J (2004) Optimum antenna matching to minimise signal correlation on a two-port antenna diversity system. Electron Lett 40:1164–1165

Dossche S, Blanch S, Romeu J (2005) Decorrelation of a closely spaced four element antenna array. IEEE Antennas Propag Soc Int Symp 1:803–806

Elsheakh DA, Elsadek HA, Abdalllah EA, Iskander MF, Elhenawy H (2010) Low mutual coupling 2×2 microstrip patch array antenna by using novel shapes of defect ground structure. Microw Opt Technol Lett 52:1208–1215

Ethier J (2008) MIMO antenna design using characteristic mode concepts. Master's thesis, University of Ottawa

Falconer DD, Adachi F, Gudmundson B (1995) Time division multiple access methods for wireless personal communications. IEEE Commun Mag 33:50–57

Gao Y, Chen X, Ying Z, Parini CG (2007) Design and performance investigation of a dual-element PIFA array at 2.5 GHz for MIMO terminals. IEEE Trans Antennas Propag 53:3433–3441

Garbacz RJ (1968) A generalized expansion for radiated and scattered fields. PhD thesis, Department of Electrical Engineering, Ohio State University

Garbacz RJ, Turpin R (1971) A generalized expansion for radiated and scattered field. IEEE Trans Antennas Propag 19:662–668

Gesbert D, Shafi M, Shiu DS, Smith PJ, Naguib A (2003) From theory to practice: an overview of MIMO space-time coded wireless systems. IEEE J Sel Areas Commun 21:281–302

Guha D, Biswas M, Antar YMM (2005) Microstrip patch antennas with defected ground structure for cross polarization suppression. IEEE Antennas Wirel Propag Lett 4:455–458

Guha D, Biswas S, Joseph T, Sebastian MT (2008) Defected ground structure to reduce mutual coupling between cylindrical dielectric resonator antennas. Electron Lett 44:836–837

Hallbjörner P (2005) The significance of radiation efficiencies when using S-parameters to calculate the received signal correlation from two antennas. IEEE Antennas Wirel Propag Lett 4:97–99

Harrington RF, Mautz JR (1971) Theory of characteristic modes for conducting bodies. IEEE Trans Antennas Propag 19:622–628

Harrington RF, Mautz JR, Chang Y (1972) Characteristic modes for dielectric and magnetic bodies. IEEE Trans Antennas Propag 20:194–198

Huang SY, Lee YH (2009) A compact E-shaped patterned ground structure. IEEE Trans Microwave Theory Tech 57:657–665

Hsu CC, Lin KH, Su HL (2011) Implementation of broadband isolator using metamaterial-inspired resonators and a T-shaped branch for MIMO antennas. IEEE Trans Antennas Propag 59:3936–3939

Iglesias ER, Teruel OQ, Sanchez LI (2008) Mutual coupling reduction in patch antenna arrays by using a planner EBG structure and a multilayer bioelectric substrate. IEEE Trans Antennas Propag 56:1648–1655

Iglesias ER, Teruel OQ, Sanchez LI (2009) Planar soft surface and their application to mutual coupling reduction. IEEE Trans Antennas Propag 57:3852–3859

Inclan-Sanchez L, Roy JLV, Iglesias ER (2007) High isolation proximity coupled multilayer patch antenna for dual-frequency operation. IEEE Trans Antennas Propag 56:1180–1183

Kermoal JP, Schmacher L, Pedersen KI, Mogensen PE, Frederiksen F (2002) A stochastic MIMO radio channel model with experimental validation. IEEE J Sel Areas Commun 20:1211–1226

Kin CS, Lim JS, Nam S, Kang KY, Ahn D (2002) Equivalent circuit modeling of spiral defected ground structure for microstrip line. Electron Lett 38:1109–1120

Kishor KK, Hum SV (2013) A two-port chassis-mode MIMO antenna. IEEE Antennas Wirel Propag Lett 12:690–693

Kishor KK, Hum SV (2014) A pattern reconfigurable chassis-mode MIMO antenna. IEEE Trans Antennas Propag 62:3290–3298

Kobayashi K, Chiba K, Takami T (1994) 1.5 GHz band portable telephone for personal digital cellular system. In: Proceedings of the IEEE 44th vehicular technology conference, Stockholm, Sweden, vol 1, pp 508–511

Kokkinos T, Liakou E, Feresidis AP (2008) Decoupling antenna elements of PIFA arrays on handheld devices. Electron Lett 44:1442–1444

Lau BK, Andersen JB (2007) An antenna system and a method for operating an antenna system. PCT Filed (Pub. No. WO/2008/030165)

Lau BK, Andersen JB (2009) Unleashing multiple antenna systems in compact terminal devices. In: International workshop on antennas technology, Santa Monica

Lau BK, Andersen JB (2012) Simple and efficient decoupling of compact arrays with parasitic scatterers. IEEE Trans Antennas Propag 60:464–472

Lee CH, Chen SY, Hsu P (2009) Integrated dual planar inverted-F antenna with enhanced isolation. IEEE Antennas Wirel Propag Lett 8:963–965

Lee B, Harackiewicz FJ, Wi H (2014) Closely mounted mobile handset MIMO antenna for LTE 13 band application. IEEE Antennas Wirel Propag Lett 13:411–414

Li H, Xiong J, He S (2009a) A compact MIMO antenns system of four elements with similar radiation characteristics and isolation structure. IEEE Antennas Wirel Propag Lett 8:1107–1110

Li H, Xiong J, He S (2009b) Extremely compact dual-band PIFAs for MIMO application. Electron Lett 45:869–870

Li H, Xiong J, Ying Z, He S (2010) High isolation compact four-port MIMO antenna systems with built-in filters as isolation structure. In: European conference on antennas and propagation (EUCAP), Barcelona, Spain

Li H, Tan Y, Lau BK, Ying Z, He S (2012a) Characteristic mode based tradeoff analysis of antenna-chassis interactions for multiple antenna terminals. IEEE Trans Antennas Propag 60:490–502

Li H, Lau BK, Ying Z, He S (2012b) Decoupling of multiple antennas in terminals with chassis excitation using polarization diversity, angle diversity and current control. IEEE Trans Antennas Propag 60:5947–5957

Li H, Lin X, Lau BK, He S (2013a) Equivalent circuit based calculation of signal correlation in lossy MIMO antennas. IEEE Trans Antennas Propag 61:5214–5222

Li H, Miers Z, Lau BK (2013b) Generating multiple characteristic modes below 1 GHz in small terminals for MIMO antenna design. In: Proceedings of the IEEE international symposium on antennas and propagation, Lake Buena Vista

Li H, Miers Z, Lau BK (2014) Design of orthogonal MIMO handset antennas based on characteristic mode manipulation at frequency bands below 1 GHz. IEEE Trans Antennas Propag 62:2756–2766

Lim S, Ling H (2010) Design of electrically small, pattern reconfigurable Yagi antenna. Electron Lett 43:1326–1327

Lin KC, Wu CH, Lai CH, Ma TG (2012a) Novel dual-band decoupling network for two-element closely spaced array using synthesized microstrip lines. IEEE Trans Antennas Propag 60:5118–5128

Lin X, Li H, He S (2012b) A decoupling technique for increasing the port isolation between two closely packed antennas. In: Proceedings of the IEEE antennas and propagation society international symposium, Chicago, US

Ling X, Li R (2011) A novel dual band MIMO antenna array with low mutual coupling for portable wireless devices. IEEE Antennas Wirel Propag Lett 10:1039–1042

Mak ACK, Rowel CR, Murch RD (2008) Isolation enhancement between two closely packed antennas. IEEE Trans Antennas Propag 56:3411–3419

Martens R, Safin E, Manteuffel D (2010) Inductive and capacitive excitation of the characteristic modes of small terminals. In: EUCAP2010, Barcelona

Miers Z, Li H, Lau BK (2013) Design of bandwidth enhanced and multiband MIMO antennas using characteristic modes. IEEE Antennas Wirel Propag Lett 12:1696–1699

Miers Z, Li H, Lau BK (2014) Design of bezel antennas for multiband MIMO terminals using characteristic modes. In: Proceedings of the 5th European conference on antennas and propagation (EUCAP), The Hague, Netherlands

Min KS, Kim DJ, Moon YM (2005) Improved MIMO antenna by mutual coupling suppression between elements. In: The European conference on wireless technology, Paris, France pp 125–128

Minz L, Garg R (2010) Reduction of mutual coupling between closely spaced PIFAs. Electron Lett 46:392–394

Paulraj A, Nabar R, Gore D (2003) Introduction to space-time wireless communications. Cambridge University Press, Cambridge

Peng HL, Tao R, Yin WY, Mao JF (2013) A novel compact dual-band antenna array with high isolations realized using the neutralization technique. IEEE Trans Antennas Propag 61:1956–1962

Perahia E (2008) IEEE 802.11n development: history, process, and technology. IEEE Commun Mag 46:48–55

Peterson HO, Beverage HH, Moore JB (1931) Diversity telephone receiving system of RCA communications, Inc. Proc IRE 19(4):562–584

Plicanic V, Lau BK, Derneryd A, Ying Z (2009) Actual diversity performance of a multiband diversity antenna with hand and head effects. IEEE Trans Antennas Propag 57:1547–1556

Plicanic V, Asplund H, Lau BK (2012) Performance of handheld MIMO terminals in noise- and interference-limited urban macrocellular scenarios. IEEE Trans Antennas Propag 60:3901–3912

Ranvier S, Luxey C, Le Thuc P, Staraj R, Kossiavas G, Vainikainen P, Icheln C (2006) Mutual coupling reduction for patch antenna array. In: Proceedings of the 1st European conference on antennas and propagation (EUCAP 2006), Nice, France

Rao AM, Weber A, Gollamudi S, Soni R (2009) LTE and HSPA: revolutionary and evolutionary solutions for global mobile broadband. Bell Labs Tech J 13:7–34

Salonen I, Icheln C, Vainikainen P (2005) The dependency of pattern correlation on mutual coupling and losses in antenna arrays. Microw Opt Technol Lett 47:145–147

Schroeder WL, Famdie CT, Solbach K (2005) Utilisation and tuning of the chassis modes of a handheld terminal for the design of multiband radiation characteristics. In: Proceedings of the IEE wideband multi-band antennas arrays, Birmingham, UK pp 117–121

Schroeder WL, Vila AA, Thome C (2006) Extremely small, wide-band mobile phone antennas by inductive chassis mode coupling. In: Proceedings of the 9th European conference on wireless technology, Manchester, UK pp 407–410

Shaker G, Rafi G, Safavi-Naeini S, Sangary N (2008) A synthesis technique for reducing mutual coupling between closely separated patch antennas. In: Proceedings of the IEEE antennas and propagation society international symposium, San Diego

Stein S (1962) On cross coupling in multiple-beam antennas. IRE Trans Antennas Propag 10:548–557

Sonkki M, Daviu EA, Fabres MC, Bataller MF, Salonen ET (2012) Improved planar wideband antenna element and its usage in a mobile MIMO system. IEEE Antennas Wirel Propag Lett 11:826–829

Stjernman A (2005) Relationship between radiation pattern correlation and scattering matrix of lossless and lossy antennas. Electron Lett 41:678–680

Su SW (2010) High-gain dual-loop antennas for MIMO access points in the 2.4/5.2/5.8 GHz bands. IEEE Trans Antennas Propag 58:2412–2419

Sung YJ, Kim M, Kin YS (2003) Harmonics reduction with defected ground plane structure for a microstrip patch antenna. IEEE Antennas Wirel Propag Lett 2:111–113

Taga T, Tsunekawa K (1986) A built in diversity antenna for 800 MHz band portable radio units. IEEE Antennas Propag Soc Int Symp 24:705–708

Taga T, Tsunekawa K (1987) Performance analysis of a built-In planar inverted-F antenna for 800 MHz band portable radio units. IEEE J Sel Areas Commun 5:921–929

Taga T (1990) Analysis for mean effective gain of mobile antenna in land mobile radio environment. IEEE Trans Veh Technol 39:117–131

Telatar IE (1999) Capacity of multi-antenna Gaussian channels. Eur Trans Telecommun 10:585–595

Thaysen J, Jakobsen K (2006) Envelope correlation in (n, n) mimo antenna array from scattering parameters. Microw Opt Technol Lett 48:832–834

Tian R, Lau BK, Ying Z (2011) Multiplexing efficiency of MIMO antennas. IEEE Antennas Wirel Propag Lett 10:183–186

Valkonen R, Kaltiokallio M, Icheln C (2013) Capacitive coupling element antennas for multi-standard mobile handsets. IEEE Trans Antennas Propag 61:2783–2791

Volmer C, Weber J, Stephan R, Blau K, Hein MA (2008) An eigen-analysis of compact antenna arrays and its application to port decoupling. IEEE Trans Antennas Propag 56:360–370

Wallace JW, Jensen MA (2004) Mutual coupling in MIMO wireless systems: a rigorous network theory analysis. IEEE Trans Antennas Propag 3:1317–1325

Wang Y, Du Z (2014) A wideband printed dual-antenna with three neutralization lines for mobile terminals. IEEE Trans Antennas Propag 62:1495–1500

Weber J, Volmer C, Blau K, Stephan R, Hein MA (2006) Miniaturized antenna arrays using decoupling networks with realistic elements. IEEE Trans Microwave Theory Tech 54:2733–2740

Winters J (1987) On the capacity of radio communication systems with diversity in a Rayleigh fading environment. IEEE J Sel Areas Commun 5:871–878

Xiong J, Zhao M, Li H, Ying Z, Wang B (2012) Collocated electric and magnetic dipoles with extremely low correlation as a reference antenna for polarization diversity MIMO applications. IEEE Antennas Wirel Propag Lett 11:423–426

Yang F, Rahmat-Samii Y (2003) Microstrip antennas integrated with electromagnetic band-gap (EBG) structures: a low mutual coupling design for array applications. IEEE Trans Antennas Propag 51:2936–2946

Zhang S, Khan SN, He S (2010) Reducing mutual coupling for an extremely closely-packed tunable dual-element PIFA array through a resonant slot antenna formed in-between. IEEE Trans Antennas Propag 58:2771–2776

DOI 10.1007/978-981-4560-75-7_99-1

Zhang S, Lau BK, Tan Y, Ying Z, He S (2012a) Mutual coupling reduction of two PIFAs with a T-shape slot impedance transformer for MIMO mobile terminals. IEEE Trans Antennas Propag 60:1521–1531

Zhang S, Lau BK, Sunesson A, He S (2012b) Closely-packed UWB MIMO/diversity antenna with different patterns and polarizations for USB dongle applications. IEEE Trans Antennas Propag 60:4372–4380

Zhu FG, Xu JD, Xu Q (2009) Reduction of mutual coupling between closely-packed antenna elements using defected ground structure. Electron Lett 45:1–4

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_100-1

Antennas in Handheld Devices

Cyril Luxey* and Aykut Cihangir
Laboratoire Electronique pour Objets Connectés, Université Nice Sophia Antipolis, Sophia Antipolis, France

Abstract

This chapter gives a general overview of the antennas used in mobile phones for wireless cellular communications. The evolution of the antenna structures is discussed in combination with the evolution of cellular communication standards. Some common antenna types widely used in mobile terminals are presented, some of them being examples from commercial products. The antenna design techniques that are used to achieve broader operating bandwidths with smaller antenna dimensions are addressed. Tunable and reconfigurable antennas are mentioned and their advantages and disadvantages versus passive antennas explained. The most important metrics that quantize antenna performance in mobile phones are presented. Some simulation and measurement results as well as data from previously published papers are used to support the claims and information in the chapter. The chapter is finalized with some forecast for the future cellular standards, and some challenges that antenna designers might face for the fifth generation are mentioned.

Keywords

Mobile phone; Handheld terminal; Antenna; PIFA; Multi-band; Matching networks; Reconfigurable; SAR; LTE; 4G; Cellular communication; Wireless communication

Introduction

This chapter is dedicated to the antennas used in handheld devices. As known, there is a very wide variety of electronic handheld devices, used for communication purposes. This chapter will confine itself to mobile phones (also referred to as mobile terminals in the text) and will aim to inform the user about the antennas used in mobile phones.

For this purpose, first a summary of the cellular communication standards, starting from the first generation of analog systems until the fourth generation, will be done. Concerning this evolution in the communication standards, the changes and improvements on the terminal antennas will be discussed, starting with external antennas. The migration to internal antennas will then be presented and the challenges faced for internal antenna design explained. Some common antenna types will be explained and some examples of antenna from commercial mobile phones given.

Some techniques used to obtain broadband antennas and to miniaturize them will then be presented with some emphasis on matching networks. Some tunable/reconfigurable antenna topologies will be mentioned, with their advantages/disadvantages over passive prototypes.

Interactions between the antenna and the user will then be given. Both, the effects of the user on the antenna performance and the radiation hazard risk on the user's head, will be discussed. Possibility of

*Email: cyril.luxey@unice.fr

DOI 10.1007/978-981-4560-75-7_100-1

utilizing antennas with multiple feeds and the advantages/disadvantages compared to single-feed antennas will be presented in the next section.

Finally, some problems that might occur during mobile phone antenna measurements will be discussed, some important metrics that should be measured in a commercial product antenna will be explained, and some common antenna manufacturing techniques will be presented. The chapter will be finalized with a conclusion section.

It should be noted, at this point, that MIMO and diversity antenna configurations are out of the scope of this chapter since they are discussed in different chapters dedicated for them. Multiantenna systems for mobile phones will only be mentioned as multifeed antennas that operate at different frequency bands and not in a MIMO configuration.

Evolution of Cellular Communications

Wireless communication is an indispensible benefit of technology used in every part of the modern life, with a wide range of applications like navigation, voice communication, urgency communication, air traffic, etc. Probably the most common wireless communication method in the consumer sense is the cellular mobile communication, which was once launched to give the users the freedom to make voice calls in a wireless sense. However, it now enables a diversity of applications like high-speed Internet access, video calls, video/audio streaming, etc.

In a typical cellular communication network, the land that will be serviced is divided into a number of cells and at least one transmitter (a base station) located to serve the users in each cell. The user can move between different cells, and the handover from cell to cell is automatically handled, transparent to the user. In general sense, a cellular communication system is composed of three layers: the core, the edge, and the access subsystems as stated in Toh (2011). The core is the layer handling the voice calls, doing the traffic control and user authentication, as well as keeping track of the usage/charges. The edge layer is the air interface between the core and access layers. The access layer (also called Node B or base station) is then responsible to establish the wireless link between the base station and the user's mobile phone and also to monitor the link parameters.

The wireless cellular communication requires a two-way link, meaning the user needs to both receive and send information. The transmission of information from the user to the base station is named "uplink," whereas the reverse direction is called "downlink." To enable this two-way communication, duplexing schemes are necessary, which can be either in time or frequency domain. In time division duplexing (TDD), the user transmits and receives in the same frequency band but in different time slots allocated. On the contrary, in frequency division duplexing (FDD) transmit and receive functions occur using different carrier frequencies for uplink and downlink. In order to serve multiple users from a base station, multiple access schemes (like frequency division multiple access or time division multiple access) are used as well.

As also mentioned in Akyildiz et al. (2010), Jaloun and Guennoun (2010), and Kumar et al. (2010), the cellular mobile communication has undergone several improvements since the first-generation (1G) technology was introduced in 1979 in Tokyo, followed by Europe in 1981 and USA in 1982 (Fig. 1). The 1G cellular networks were totally analog, able to deliver only voice through frequency modulation. Both 450 MHz and 900 MHz frequencies were used in Europe whereas the AMPS (Advanced Mobile Phone System) was used in USA at 850 MHz with a 40 MHz bandwidth. Each user was assigned a specific carrier frequency resulting in FDMA. The major disadvantage of the first-generation systems was the lack of interoperability between countries, resulting in a mobile terminal that could be used in that specific country only.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_100-1

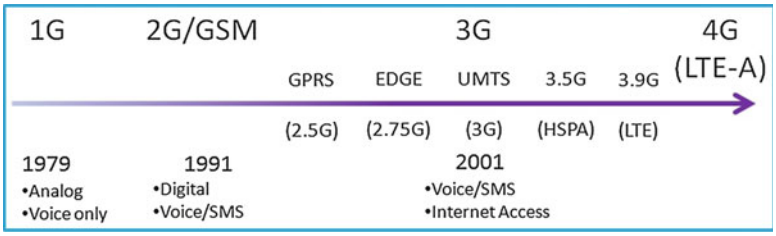


Fig. 1 Evolution of cellular communication technologies

Approximately 10 years later, the second generation of cellular mobile communications was launched in Finland. The most common 2G technologies are GSM (Global System for Mobile Communications, used globally), cdmaOne (mostly used in USA), iDen (mostly used in USA), and PDC (mostly used in Japan). These technologies used either TDMA or CDMA modulation schemes to handle multiple users. Among these technologies, GSM is the most commonly used, with a high market share. The biggest difference of this second generation compared to the first is the modulation schemes, which are completely digital like GMSK (Gaussian minimum shift keying) used in GSM technology, for instance. This led the way to higher spectral efficiency and better quality of services. In addition to the voice, the second generation allowed the user to receive/send text messages with the short message service (SMS), which can be counted as low-speed data transfer. Global roaming was also made possible, where the same phone number can be used in different countries, by means of the newly introduced subscriber identity module (SIM) cards. The GSM technology uses FDD as the duplexing scheme and both TDMA and FDMA to serve multiple users, where eight users share the same time frame. The channel bandwidth allocated for each user is 25 kHz. There were two main evolutions based on the GSM technology to supply users' higher data rates. Among these, GPRS (general packet radio service) is a technology that is referred to as 2.5G, increasing the maximum data rate up to 150 kbps. With GPRS technology, the circuit-switched data transmission was changed to packet-based type. To further increase the maximum data rate, EDGE (Enhanced Data rate for GSM Evolution) technology referred to as 2.75G was introduced, allowing up to 384 kbps in optimum conditions. One of the main reasons to increase the data rate was the utilization of 8-PSK (instead of GMSK) as the digital modulation technique.

The requirements for the third-generation cellular communication were determined by the 3GPP (3rd Generation Partnership Project) in Release 99 in the year 2000 in scope of the IMT-2000 project. In 2001, the first deployment of 3G was started in Japan, followed by South Korea in 2002 and Europe and USA in 2003. There are mainly two dominating technologies: UMTS (called also W-CDMA) used in Europe and CDMA2000 used in USA. The main target of 3G technologies was to propose to the user high data rate enabling video calls, high-quality video/audio streaming, and high-speed Internet benefiting also the introduction of smartphones with touchscreens. New frequency bands were allocated for the 3G communications in addition to the previously existing bands between 1990 and 2170 MHz. The maximum bandwidth for a user was increased to 5 with W-CDMA technology whereas CDMA2000 uses one or more available 1.25 MHz channels. A further improvement to the 3G was done with the introduction of HSPA (High-Speed Packet Access), called 3.5G and increasing the data rate to 14.4 Mbps and even higher with HSPA+ (21 Mbps, 42 Mbps in dual-carrier mode and 84 Mbps in dual-carrier mode + 2 $\times$ 2 MIMO).

In accordance with the improvements through 4G, 3GPP announced Release 8 through the end of 2008, describing the first LTE (Long-Term Evolution) specifications. The target peak data rate for LTE technology is 100 Mbps for downlink in a full-IP network. Since LTE standard could not reach the data rate levels necessary for 4G as strictly defined by the International Telecommunication Union, LTE is generally referred to as 3.9G. In this technology, various modulation techniques (like QPSK, 16QAM,

DOI 10.1007/978-981-4560-75-7_100-1

Table 1 LTE FDD frequency bands

LTE band	Area	DL frequency	UL frequency	Bandwidth (MHz)
1	All	2110–2170 MHz	1920–1980 MHz	5,10,15,20
2	NAR	1930–1990 MHz	1850–1910 MHz	1.4,3,5,10,15,20
3	All	1805–1880 MHz	1710–1785 MHz	1.4,3,5,10,15,20
4	NAR	2110–2155 MHz	1710–1755 MHz	1.4,3,5,10,15,20
5	NAR	869–894 MHz	824–849 MHz	1.4,3,5,10,15,20
6	APAC	875–885 MHz	830–840 MHz	5,10
7	EMEA	2620–2690 MHz	2500–2570 MHz	5,10,15,20
8	All	925–960 MHz	880–915 MHz	1.4,3,5,10
9	APAC	1845–1880 MHz	1750–1785 MHz	5,10,15,20
10	NAR	2110–2170 MHz	1710–1770 MHz	5,10,15,20
11	Japan	1476–1496 MHz	1428–1448 MHz	5,10
12	NAR	729–746 MHz	699–716 MHz	1.4,3,5,10
13	NAR	746–756 MHz	777–787 MHz	5,10
14	NAR	758–768 MHz	788–798 MHz	5,10
17	NAR	734–746 MHz	704–716 MHz	5,10
18	Japan	860–875 MHz	815–830 MHz	5,10,15
19	Japan	875–890 MHz	830–845 MHz	5,10,15
20	EMEA	791–821 MHz	832–862 MHz	5,10,15,20
21	Japan	1496–1511 MHz	1448–1463 MHz	5,10,15
22		3510–3590 MHz	3410–3490 MHz	5,10,15,20
23	NAR	2180–2200 MHz	2000–2020 MHz	1.4,3,5,10,15,20
24	NAR	1525–1559 MHz	1626.5–1660.5 MHz	5,10
25	NAR	1930–1995 MHz	1850–1915 MHz	1.4,3,5,10,15,20
26	NAR	859–894 MHz	814–849 MHz	1.4,3,5,10,15
27	NAR	852–869 MHz	807–824 MHz	1.4,3,5,10
28	APAC	758–803 MHz	703–748 MHz	3,5,10,15,20
29	NAR	717–728 MHz	–	3,5,10
30	NAR	2350–2360 MHz	2305–2315 MHz	5,10
31	CALA	462.5–467.5 MHz	452.5–457.5 MHz	1.4,3,5

64QAM) can be used for downlink, depending on the channel conditions. The OFDM (orthogonal frequency-division multiplexing)-type waveforms are utilized to improve link performance in urban multipath environments and the maximum channel bandwidth increased to 20 MHz. In MIMO (multiple-input–multiple-output) techniques, up to four transmit and four receive antennas are also proposed to reach the 100 Mbps maximum data rate. Both TDD and FDD duplexing schemes can be used as duplexing methods. Table 1 and 2 present the frequency bands, their region of utilization, and the available channel bandwidths for FDD and TDD LTE respectively. It can be seen that two new frequency bands have been introduced with LTE (in addition to existing UMTS bands), between 698 and 824 MHz in the low-band and between 2.5 and 2.69 GHz in the high-band.

To fulfill the requirements of 4G cellular communications, 3GPP published Release 10 in the beginning of 2011, named as LTE-A (LTE Advanced), which is the "real" 4G technology. The maximum data rate is targeted as 1 Gbps for low-mobility users (slow-moving like pedestrians) and 100 Mbps for high-mobility users (as in a car). The MIMO utilization is increased up to eight transmit and eight receive antennas. As mentioned before, the maximum single-channel bandwidth is up to 20 MHz for higher carrier frequencies and up to 10 MHz for lower carrier frequencies. The usable channel bandwidth can be increased up to

DOI 10.1007/978-981-4560-75-7_100-1

Table 2 LTE TDD frequency bands

LTE band	Area	Frequency	Bandwidth (MHz)
33		1900–1920 MHz	5,10,15,20
34	EMEA	2010–2025 MHz	5,10,15
35	NAR	1850–1910 MHz	1.4,3,5,10,15,20
36	NAR	1930–1990 MHz	1.4,3,5,10,15,20
37	NAR	1910–1930 MHz	5,10,15,20
38	China	2570–2620 MHz	5,10,15,20
39	China	1880–1920 MHz	5,10,15,20
40	China	2300–2400 MHz	5,10,15,20
41	All	2496–2690 MHz	5,10,15,20
42		3400–3600 MHz	5,10,15,20
43		3600–3800 MHz	5,10,15,20
44	APAC	703–803 MHz	3,5,10,15,20

NAR North American Region
EMEA Europe Middle East and Africa
APAC Asia and Pacific
CALA Central Latin America

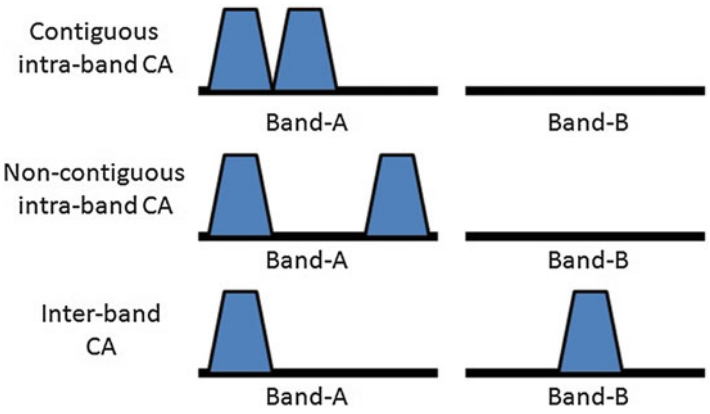


Fig. 2 Different carrier aggregation scenarios

100 MHz, using multiples of 20 MHz channels, defined as carrier aggregation (CA). Different CA scenarios are possible as presented in Fig. 2. When two adjacent frequency channels are assigned to a single user, it is named as contiguous intraband CA, whereas it is called noncontiguous if the two channels are separated in the same band. Likewise, depending on the frequency licenses the operator possesses, a user can be assigned a channel from one band and simultaneously another channel from a different band. This scenario is then called interband CA.

During the evolution of mobile cellular communications since the 2G, new frequency bands were introduced to the existing ones (as shown in Fig. 3). For example, the frequency interval of 1.71–1.99 GHz was introduced to the 2G frequency bands with the UMTS technology. Likewise with LTE, two new bands 698–824 MHz and 2.5–2.69 GHz were added. These newly added frequency bands form a key aspect from the antenna design point of view because they increase the design challenges of the implementation of broader-frequency antennas within a limited space (mainly reduced height). The possibility of using CA techniques is also of high importance from the antenna side, since they might need to cover simultaneously two or more frequency bands which might not be possible with

DOI 10.1007/978-981-4560-75-7_100-1

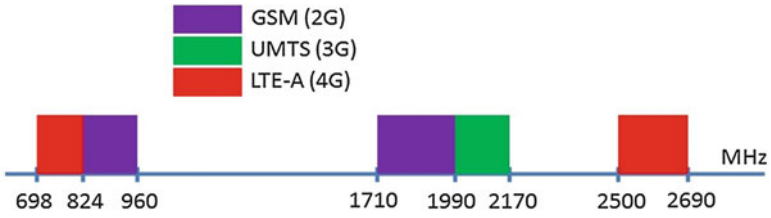


Fig. 3 Frequency bands used since 2G

band-switching reconfigurable antenna topologies. Finally, MIMO scenarios should also be considered in the design phase in order to situate the multiple numbers of antennas in a limited space with reasonable isolation.

Through the rest of this chapter, the frequency intervals of 698–960 MHz and 1.71–2.69 GHz will be called as 4G communication bands, the former referred to as low-band (LB) and the latter referred to as high-band (HB). In summary, the 698–824 MHz band will be called low-LTE band, 824–960 MHz as GSM band, 1.71–1.99 GHz as DCS/PCS band, 1.99–2.17 GHz as UMTS band, and 2.5–2.69 GHz as high-LTE band.

Evolution of Mobile Phone Antennas

The advancements on the cellular communication standards since 35 years had an absolute effect on the terminal antennas, as well as the other mobile phone components. The evolution of the mobile phone antennas was generally driven by two major forces. The first one is the user demand, which is mainly composed of aesthetical and ergonomic issues. This effect can be seen from the domination of the market by the mobile phones with internal antennas soon after the first internal antenna phone models were released. Mobile phones incorporating internal antennas had better appearances, were smaller in size, and thus were easier to carry so that the customer demand was immediately adhered to regardless of considering any disadvantages in antenna performance. The second major driving force on the antenna evolution is of course the newly added frequency bands with the introduction of new standard generations. This necessity put harder requirements on the antenna bandwidth in a limited volume, leading to different antenna design methodologies and different antenna layouts.

External Antennas

The mobile phones that were manufactured for the first generation and also in the beginning of the second generation had an external antenna placed on the top corner of the handset. Some examples of such antennas can be seen in Fig. 4. The most commonly used external antenna types are whip (monopole) and helical antennas. Some designs were made of both types in a proper combination arrangement.

Whip Antennas

Whip antennas were the most common type of antenna in the first mobile terminals. This type of antennas has a length close to quarter wavelength in the frequency of operation. They are fed similar to dipoles where the PCB ground plane is connected to one pole of the feed and the antenna is fed from the other pole. These kinds of antennas are very efficient radiators, especially when the phone has a metal chassis with a length close to quarter wavelength forming a full-structure dipole both with the whip antenna. For the ease of utilization, some whip antennas had the opportunity to be retracted into the phone when not in use, of course affecting the antenna performance due to the changing environment.

DOI 10.1007/978-981-4560-75-7_100-1

Fig. 4 External antenna examples for mobile phones

Helical Antennas

Helical antennas can achieve the coverage of the same frequency band with a shorter antenna element compared to whip antennas. This enables decreasing the total size of the device where they are used. For this reason, helical antennas have been largely used in many commercial mobile phones. In the normal mode (physical length smaller than one wavelength), helical antennas have a radiation pattern similar to a monopole element, mainly in the transversal direction with a null in normal (main axis) direction. It was possible to make helical antennas dual band as described in Ying (2000), Nevermann and Pan (2002). In Ying (2000), multiband behavior was achieved by using the total length of the helical antenna for low-band resonance and using nonuniform diameter and/or pitch angle for a resonance at higher frequencies. Another alternative to obtain multiband response was proposed in Haapala and Vainikainen (1996) by using two helical antennas on the two top corners of the mobile terminal. One of the antennas was tuned to operate in DCS/PCS frequency bands (1.7–2 GHz) and the other to UMTS frequency bands (2–2.2 GHz).

Combination of Whip and Helical Antennas

Common types of external antennas for mobile terminals were combinations of whip and helixes, enabling multiband coverage. A whip, which is placed in the vertical axis of a helix, is presented in Saldell (1997): the helical antenna is placed on the top corner of the device, and the whip is passing through the central axis and can be retracted when not in talk mode. Similarly, two helical antennas with different radii have been placed on the same axis, achieving dual-band response (Haapala and Vainikainen 1996).

The external antennas generally had a high efficiency and broadband matching despite making the total device bigger. One major disadvantage was the high SAR (specific absorption rate) due to the high efficiency of the antenna and the proximity of the radiating element to the user. With the introduction of mobile phones with internal antenna elements, external antennas started to disappear from the market and cannot be found in any major commercial phone anymore.

Internal Antennas

With the market trend toward internal antennas for mobile terminals, the space restriction about antenna placement started to be one of the main concerns from the antenna design point of view. Dealing with newer standards, another requirement was to obtain more broadband antennas in this limited space.

DOI 10.1007/978-981-4560-75-7_100-1

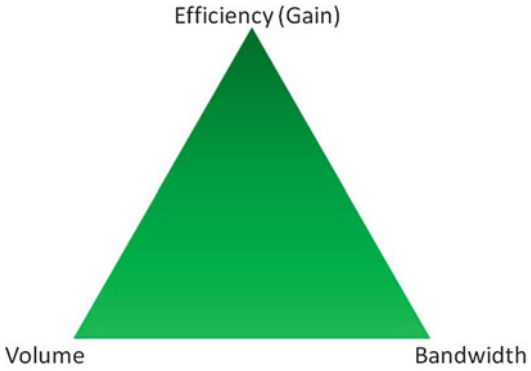


Fig. 5 Antenna efficiency, volume, and bandwidth trade-off

With the market trend to shrink the mobile phones as small as possible around 2000, this space limitation became even more challenging. With the introduction and the rise of smartphones, the device sizes started to increase, so did the space reserved for the antenna. But with the current trend of larger touchscreens and thinner smartphones, the antennas again need to be shrunk not really in terms of size but rather of height. This current constraint results in electrically small antennas mainly because of low LTE, though causing challenges in terms of antenna design (Fig. 5).

Electrically Small Antenna Challenges

The typical space allocated to the antenna in a 2G mobile phone or a smartphone is small compared to its free-space wavelength of operation. Equation 1 is the general formula to define an electrically small antenna, where r is the radius of the smallest sphere that can enclose this antenna.

$$\frac{2\pi r}{\lambda} \ll 1 \tag{1}$$

Considering a spherical volume having a maximum dimension of 60 mm ($r = 30$ mm) allocated to an antenna to operate at 900 MHz ($\lambda_0 = 333$ mm), the above inequality will be true. So, the antennas operating in GSM frequencies in a typical mobile phone can be considered as electrically small. It is known that for any antenna, the product of efficiency (or gain), volume, and bandwidth is a constant number which basically means for profiting one of the aforementioned parameters, the other two needing to be sacrificed. For example, to increase the bandwidth of an antenna, either a lower efficiency needs to be accepted or the volume of the antenna shall be increased. In the electrically small antenna case, since the antenna volume is limited, an optimum condition for the bandwidth and efficiency should be found. The main disadvantage of an electrically small antenna is the low radiation efficiency and the low real part of the input impedance which make a broadband input matching (to the front-end module or more generally to a fictitious 50 Ω source) hard and cause the antenna to be narrowband. Electrically small antennas and their fundamental properties or limitations have been studied in the literature (Chu 1948; Wheeler 1947; Collin and Rothschild 1964; Hansen 1981; McLean 1996; Yaghjian and Best 2005).

One of the most important figures of merit for an electrically small antenna is its quality factor (Q). This value is defined as the ratio of 2π times the energy stored in the near fields excited by the antenna to the energy radiated plus the energy dissipated per cycle. It is inversely proportional to the bandwidth potential. So Q should be minimized to maximize the matching bandwidth of an antenna. The lowest obtainable quality factor (also known as the Chu limit) for a linearly polarized antenna radiating in only one mode (the transverse electric or the transverse magnetic mode) can be calculated using Eq. 2.

DOI 10.1007/978-981-4560-75-7_100-1

$$Q = \frac{1}{(ka)^3} + \frac{1}{(ka)} \tag{2}$$

where k is the wavenumber and a is the radius of the smallest sphere that can enclose the antenna. It should be emphasized that this Q value is a theoretical low-bound scalar which is very optimistic to even get close in practice, even with complicated antenna structures. A formula to obtain the Q-factor of an antenna at any frequency using its input impedance was proposed in Yaghijan and Best (2005), given in Eq. 3.

$$Q(\omega) = \frac{\omega}{2R(\omega)} \sqrt{[R'(\omega)]^2 + \left[X'(\omega) + \frac{|X(\omega)|}{\omega}\right]^2} \tag{3}$$

where ω represents the angular frequency point, $R(\omega)$ and $X(\omega)$ stand for the real and imaginary parts of the input impedance at angular frequency ω, and $R'(\omega)$ and $X'(\omega)$ are the first derivates of $R(\omega)$ and $X(\omega)$ according to the frequency. This formula proposes, under some assumptions, a very fast and reliable calculation method, without the need for computing the stored energy and radiated energy of the antenna: it is valid for a single resonant antenna, close to its resonance frequency.

Using these Q-factor values, the obtainable frequency bandwidth could be calculated from Eq. 4 as also presented in Holopainen (2011), where S is the maximum VSWR target and T is the coupling coefficient as given in Pues and Capelle (1989).

$$FBW = \frac{1}{Q} \sqrt{\frac{(TS-1)(S-T)}{S}} \tag{4}$$

For a value of $T = 1$, Eq. 5 reduces to

$$BW_{cc} = \frac{1}{Q} \frac{(S-1)}{\sqrt{S}} \tag{5}$$

The bandwidth calculated by equating $T = 1$ is the critical coupling case, where the input impedance of the antenna at the center frequency is matched exactly to 50 Ω, ending in an infinitely small reflection coefficient. The bandwidth in this case is thus the frequency interval having a reflection coefficient below a predefined limit (generally −6 dB for mobile terminal antennas) around the center frequency. To illustrate the situation better, the input impedance of a critically coupled antenna is shown with red curves in Fig. 6 in a Smith chart (plotted between 0.6 and 1.1 GHz) as well as the reflection coefficient in dB form on right. It can be seen that for critical coupling case, the antenna has a very low reflection coefficient at a single frequency point and the impedance curve passes through the center of the Smith chart at this frequency. If −6 dB is chosen as limit, the antenna has a bandwidth of 100 MHz (between 740 and 840 MHz). An alternative coupling scheme to obtain more broadband antennas is the optimal overcoupling case where the antenna is matched with a worse reflection coefficient (still under the threshold) but in a broader bandwidth. The blue curve in Fig. 6 represents the optimally overcoupled case, where the bandwidth for −6 dB reflection coefficient is increased to 320 MHz, although the reflection coefficient is not as low as critical coupling case. Some examples of using a matching network to obtain optimally overcoupled antennas are presented in Antoniades and Eleftheriades (2010)and Selvanayagam and Eleftheriades (2010).

Considering the 4G-LTE frequency bands coverage in a typical recent smartphone, the antenna shall be able to cover between 700–960 MHz and 1.7–2.7 GHz with a profile generally smaller than 60 × 15 × 5 mm^3. It is obvious that the antenna will behave as an electrically small antenna in the low-band

DOI 10.1007/978-981-4560-75-7_100-1

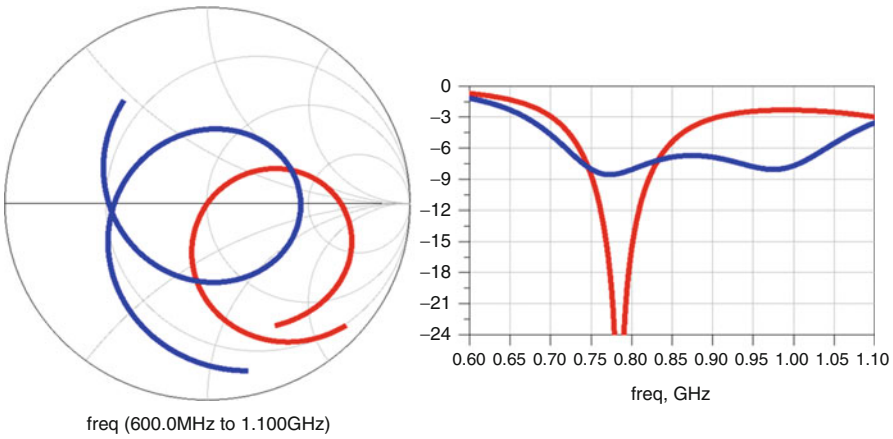


Fig. 6 Critical coupling and optimally overcoupled matching

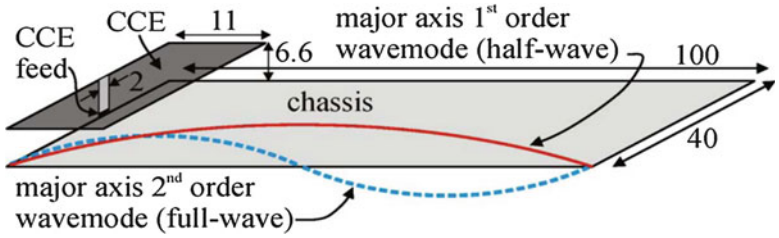


Fig. 7 First- and second-order wavemodes of a typical mobile phone's ground plane (© [2010] IEEE. Reprinted, with permission, from Holopainen et al. 2010)

(LB) especially around 700 MHz. Therefore, it is almost impossible to obtain the 30 % bandwidth coverage needed in the LB using only an electrically small radiating structure integrated in this volume. It was shown in Vainikainen et al. (2000, 2002, 2004)and Villanen et al. (2006) that the ground plane of the device plays a major role in the antenna performance in the LB and a minor role in HB. In fact, 90 % of the total radiation comes from the ground plane at 900 MHz (Villanen et al. 2006), regardless of the antenna type which is used. This is also evident from the dipole-type radiation shape of mobile terminal antennas in the LB, for almost all antenna types. So, the main role of the antenna in the LB is actually to excite electric currents on the system ground plane that will further compose the dominant part of total radiation in LB. Due to the large size of the ground plane which is not an electrically small structure anymore (thick rectangular conducting body), low-Q properties and thick dipole-type radiation can be obtained in the LB. This antenna-ground plane radiating structure can have a lower Q-factor if the ground plane is strongly excited. Therefore, the antenna element alone is not really an antenna in the LB but rather a coupling element for the ground plane. The obtained low Q-factor leads to higher bandwidth potential around the frequency region where the excitation is strong. Around the frequency corresponding to the $\lambda_0/2$ resonance of the ground plane, higher excitation and higher bandwidth potential can be observed if the coupling element efficiently excites the resonant mode of the ground plane. The length of typical ground planes in today's mobile terminals vary between 100 and 130 mm, which correspond approximately to $\lambda_0/2$ in LB. So, in a particular LB frequency, the first-order wavemode shown in Fig. 7, taken from Holopainen et al. (2010), can be strongly excited, depending on the orientation and type of the antenna coupling element. This excitation can increase the bandwidth potential of the antenna considerably, enabling the coverage of the 30 % bandwidth in LB. It should be noted that a matching network is needed to obtain the desired input impedance.

DOI 10.1007/978-981-4560-75-7_100-1

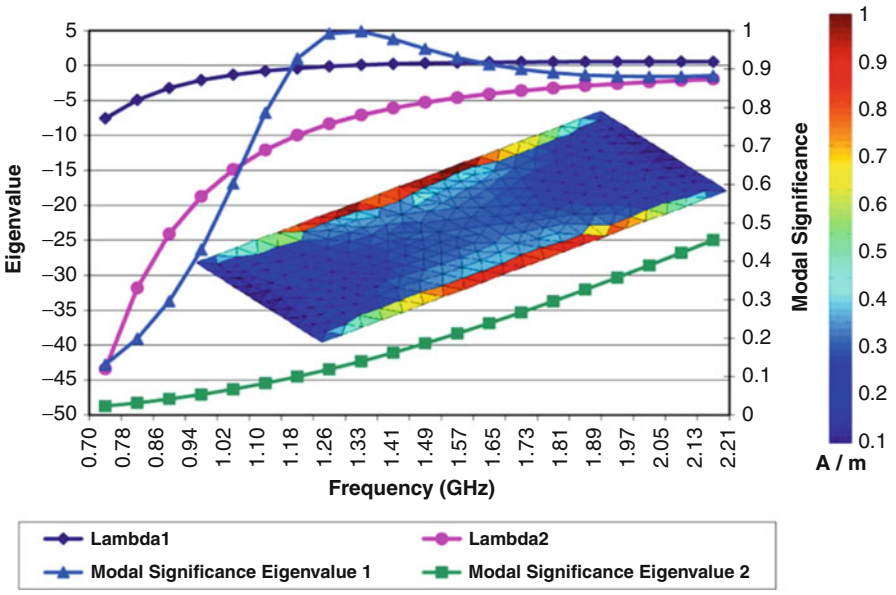


Fig. 8 Characteristic mode analysis of a 100 × 40 mm^2 (© [2011] IEEE. Reprinted, with permission, from Andujar et al. 2011)

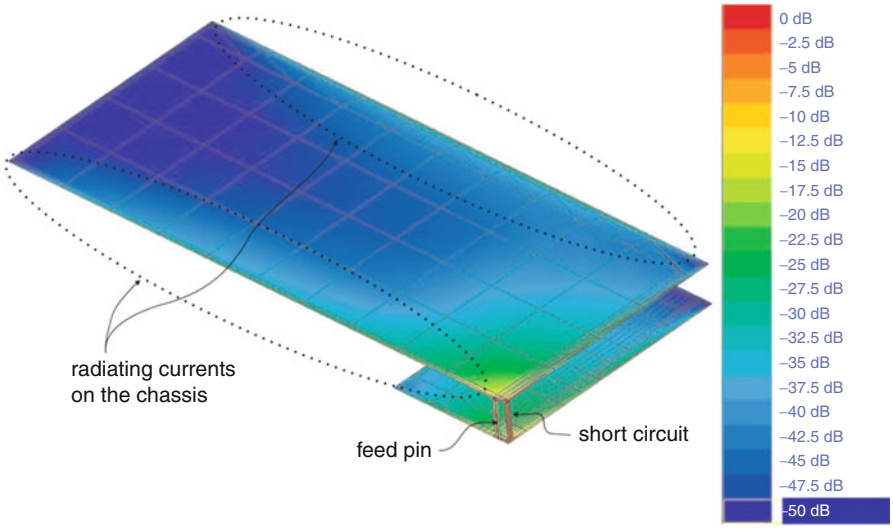


Fig. 9 Currents on the ground plane for a PIFA from Holopainen (2008)

For example, the characteristic mode analysis of a 100 × 40 mm^2 ground plane has been carried out in Andujar et al. (2011). As seen in Fig. 8, the first wavemode of this ground plane occurs around 1.25 GHz. At 892 MHz, the plotted currents are stronger through the long edges and weaker on the short edges of the ground plane which gives some guidelines for the design of an efficient coupling element.

A similar behavior was also presented in Holopainen (2008) for a PIFA (Planar Inverted-F antenna) placed over a mobile phone ground plane at 900 MHz (Fig. 9). The currents induced on the ground plane are concentrated on the long edges, forming the first wavemode.

DOI 10.1007/978-981-4560-75-7_100-1

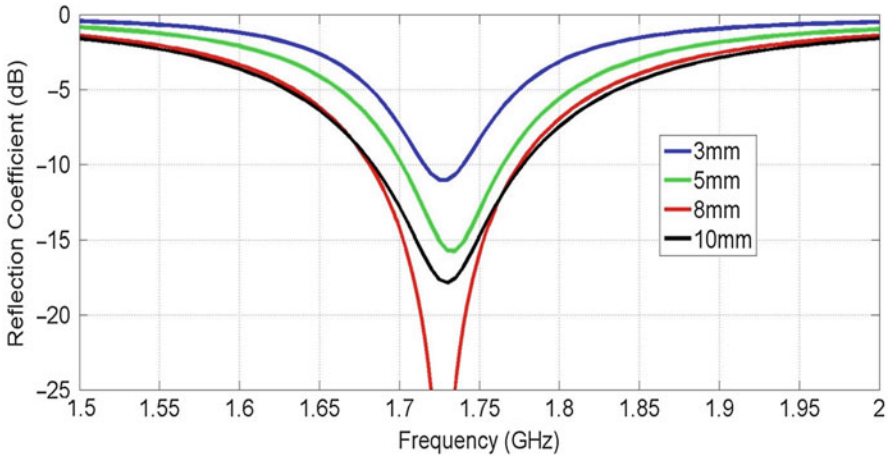


Fig. 10 Effect of the height of the PIFA on the antenna bandwidth

Common Antenna Types Used in Mobile Terminals

Planar Inverted-F Antennas (PIFAs)

PIFAs have been one of the most commonly used internal antennas in the mobile phone market. A PIFA is basically a quarter-wavelength patch antenna over a ground plane, having also one or more inductive ground connections, generally close to its feed. This ground connection acts as a shunt inductor and cancels out the high capacitance between the PIFA element and ground plane, at particular frequencies. This makes them resonant and narrowband antennas as a disadvantage. On the other hand, PIFAs are easy to manufacture, low profile, and might be low cost. Another main advantage is the ease of matching (if a narrow bandwidth is sufficient) by tuning the shorting strip and the feed distance, as well as tuning the thickness of the shorting pin.

In a mobile phone, the matching bandwidth of the PIFA is naturally increased in the LB (and not narrowband anymore) as the PIFA is also acting as a coupling element to excite the currents of the ground plane. Therefore, the bandwidth of the PIFA–ground plane combination depends heavily on the volume (and thus the height) of the antenna and the length of the ground plane. Figure 10 shows the effect of the antenna height (changing between 3 and 10 mm) on the bandwidth of a PIFA (tuned to the same resonance frequency at each height). The bandwidth for a reflection coefficient lower than −6 dB is 75 MHz for a 3 mm height and rises to 175 MHz for a 10 mm height.

PIFA antennas can be made multiband with various techniques like

- Etching slots on the antenna element to create alternative current paths of different lengths
- Adding resonant strips
- Usage of parasitic elements, either connected to ground plane or floating
- Adding multiple grounding strips

For example, a PIFA design with multiple radiating strips realized by etching slots is given in Korva (2007). A parasitic strip (connected to ground plane) is also used in this design to obtain multiband behavior. Another PIFA with three slots and two grounding strips to cover 880–980 MHz and 1.88–3 GHz is presented in Bhatti et al. (2008). The antenna in Ciais et al. (2004) utilizes a combination of these techniques to achieve multiband and wideband coverage (Fig. 11).

It is also of high importance to minimize the antenna dimension with negligible performance trade-off, as the area reserved for the antenna has generally a shrinking trend. To achieve this in a PIFA-type

DOI 10.1007/978-981-4560-75-7_100-1

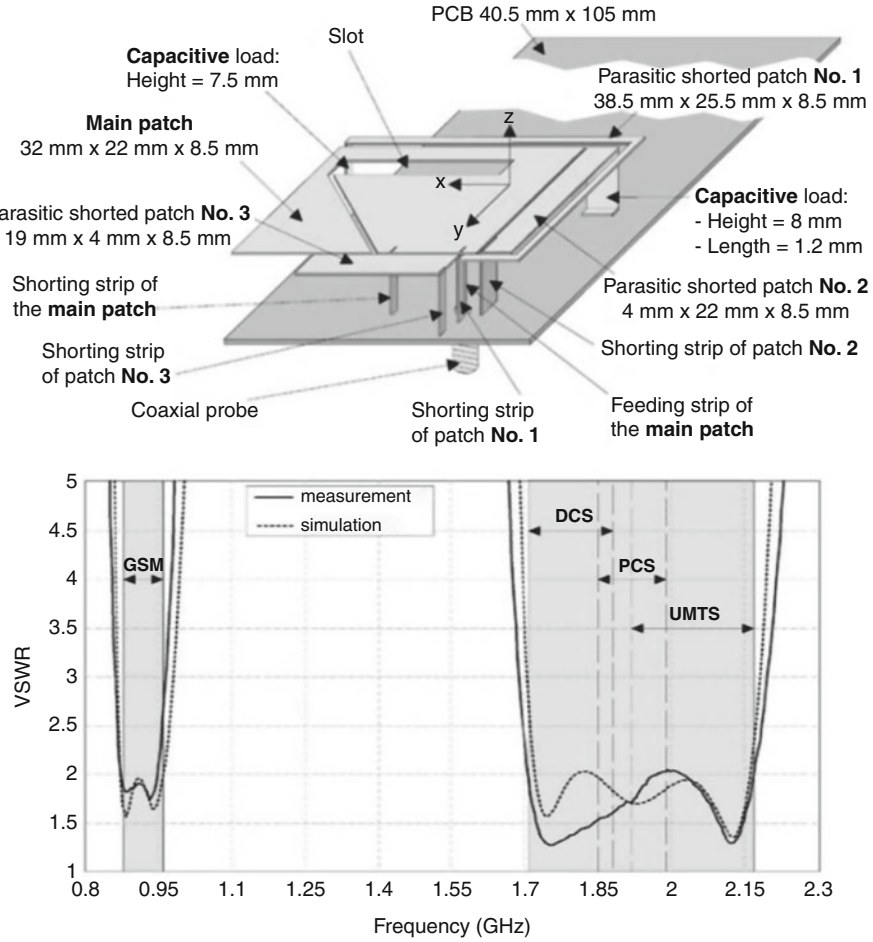


Fig. 11 An example of multiband PIFA design and associated VSWR performance (© [2004] IEEE. Reprinted, with permission, from Ciais et al. 2004)

antenna, techniques like etching slots (or meandering the main strip) and capacitive loading are used. For example, the antenna in Zaid et al. (1998) is shrunk in size by etching a slot and creating a longer path for the currents, although it is not exactly a PIFA. The capacitive loading technique was used in Guo et al. (2003)and Rowell and Murch (1997), where the main patch of the PIFA was extended down close to the ground plane to create a strong capacitance. In this way, the resonance frequency of the antenna could be decreased without increasing its size, at the expense of a reduced bandwidth.

To enlarge the operating bandwidth of a PIFA in a mobile terminal, the techniques used to obtain multiband antennas (described above) can be used. Another popular method for broader bandwidth is to etch slots on the ground plane. In this way, the wavemodes of the ground plane are pulled to lower frequency bands, and the length, width, and position of the slot can be arranged to set the wavemode of the ground plane to the frequency of interest: in fact, it lowers the Q-factor and increases the bandwidth potential at a specific frequency. Such a study on the slot parameters was carried out in Zhang and Zhao (2009) for a multiband PIFA. Slots on the ground plane of a PIFA were proposed also in Villanueva et al. (2013), Cabedo et al. (2009), andZhang (2013a, b) to obtain wideband and multiband performance. Slotted ground plane antennas could not get popular in the industrial sense due to practical manufacturing reasons. For example, components of high importance like the touchscreen or battery of a mobile terminal might need to fit over the slot and change completely the antenna behavior.

DOI 10.1007/978-981-4560-75-7_100-1

Monopole-Like Antennas

Another most commonly used antenna type in commercial mobile phones is the monopole-like antenna. The term *monopole* is used here to indicate that there is a ground clearance region under the monopole element and the driven strip of the antenna does not have a direct ground connection. Designs having also a ground connection forming a loop are also possible, so one pole of the feed port is connected to the strip, whereas the other pole is connected to the ground plane, to excite the antenna. Those types of antennas started to gain popularity especially after new frequency bands were added with 3G and 4G since they offer broader bandwidth when compared to PIFAs. One of the disadvantages for monopole-like antennas is the higher SAR (compared to PIFAs) due to the lack of shielding ground plane beneath the antenna. So, the possibility of locating the antenna at the bottom corner of the device also helped these antennas to be more popular as the SAR in this configuration is lower than the case of a top-corner monopole where the SAR might exceed the authorized limits.

To miniaturize monopole-like structures, the most popular and easiest method is meandering, to shift the resonance to lower frequencies within the same antenna volume (at the expense of a bandwidth reduction). Therefore, the bandwidth enhancement for this type of antennas can be achieved using multiple resonant arms or inserting capacitively excited parasitic elements. For example, in Wong and Chen (2010), the antenna element was formed from two metal strips connected together through an inductor to achieve dual-band coverage for the GSM and DCS/PCS/UMTS bands. Multiple resonant arms were used in combination with slots in the ground plane to achieve GSM/DCS/PCS/UMTS coverage in Liu et al. (2010). Parasitic elements (connected to ground plane or floating) excited by monopoles are commonly proposed in the literature to achieve wideband operation in mobile terminals. For this type of antennas, the main methodology is to capacitively excite a longer metal strip connected to the ground plane through a shorter driven strip. The lower frequency coverage is generally achieved by the excitation of the parasitic element and the higher frequency bands covered by the monopole together with the higher-order resonances of the parasitic. In Lin and Wong (2007), slots were etched on the ground plane to form two parasitic strips, excited by a monopole. In Chu and Wong (2012)and Yang et al. (2011), a monopole-driven strip is used to excite a longer parasitic element connected to the ground plane on one of its edge. The parasitic strip can also be connected to the ground plane on both edges (forming a loop) as in Chen et al. (2012)and Chiu et al. (2010). The loop structure was meandered to increase its electrical length and shift the resonance frequency down. In the antennas proposed in Chen and Wong (2010, 2011), Chu and Wong (2011), andLee and Wong (2010), monopole antennas with parasitic elements were proposed, taking into account a dielectric casing for the antenna (Fig. 12). In Cihangir et al. (2013), a parasitic element was exploited to cover low-LTE/GSM frequency bands (700–960 MHz) with a short ground plane (Fig. 13).

Capacitive Coupling Elements

The role of the system ground plane on the radiation and matching characteristics of a mobile terminal antenna was previously mentioned. Capacitive coupling elements (CEs) are inherently nonresonant structures that are designed and located with the purpose of exciting the wavemodes of the ground plane in a systematic manner (Vainikainen et al. 2002; Villanen et al. 2006; Holopainen et al. 2010). This way, the antenna system has a low Q-factor and thus high bandwidth potential especially close to the resonance frequency of the ground plane. The feeding mechanism of the CE is similar to one of a monopole with no inductive contact (or ground return) to the PCB ground plane. Since these structures are inherently nonresonant, they initially have poor matching, but the antenna can easily be tuned to the desired frequency band using a matching network (MN).

A typical CE configuration, its input impedance without any MN and the corresponding Q-factor can be seen in Fig. 14 taken from Holopainen (2011). It is worthwhile to note that in this figure, the CE is situated

DOI 10.1007/978-981-4560-75-7_100-1

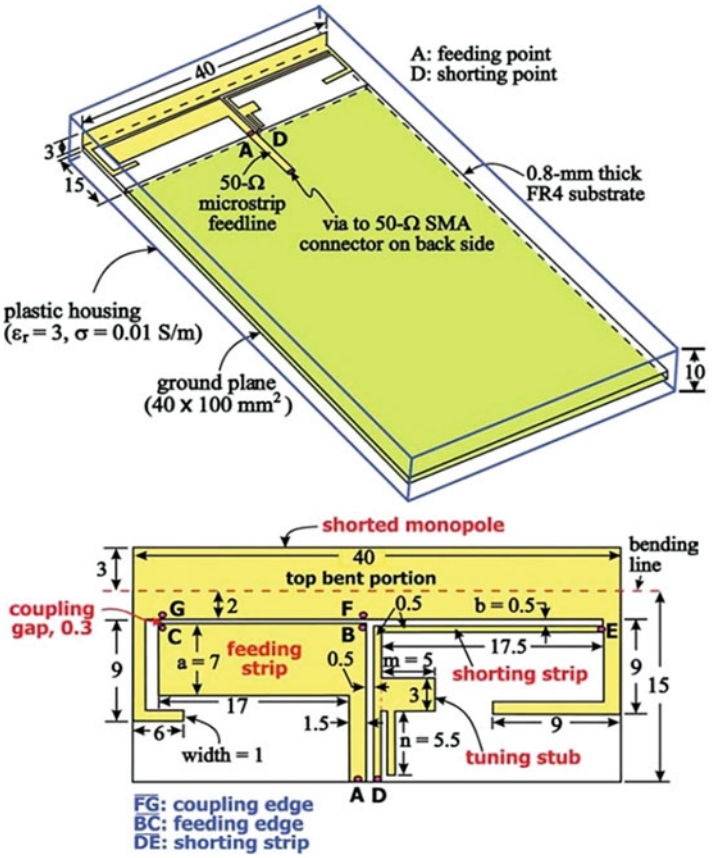


Fig. 12 Monopole antenna with parasitic element for 4G coverage from Chen and Wong (2010)

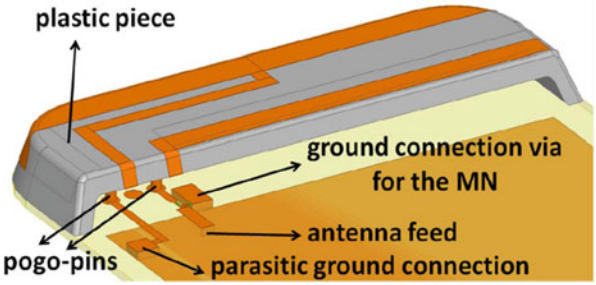


Fig. 13 Monopole antenna for low-LTE coverage from (Cihangir et al. 2013)

over the ground plane, which is not the common way in very recent publications. It was shown with later publications that leaving a ground clearance region under the CE leads to lower Q-factor and higher bandwidth potential (BP). The CE is fed similar to a monopole without any ground connection and excites the surface currents of the ground plane through capacitive coupling from the edges and corners, especially in lower frequencies. From Fig. 14, it is observed that the first wavemode of the ground plane occurs at 1.17 GHz, where the rate of change of the impedance (around this frequency) is slower compared to other frequencies. This effect can also be observed from the Q-factor, making a local minimum around 1.17 GHz. This phenomenon is beneficial especially for the low frequencies where the CE antenna element is itself electrically small. By benefiting from the currents on the ground plane, both the radiation efficiency and bandwidth potential are increased.

DOI 10.1007/978-981-4560-75-7_100-1

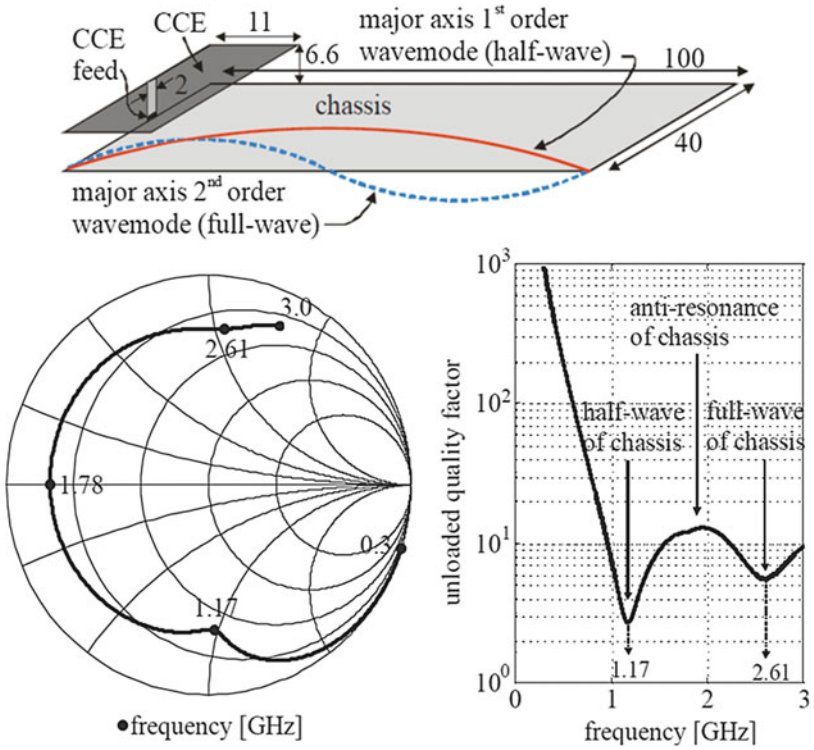


Fig. 14 Capacitive coupling element antenna and behavior from Holopainen (2011)

Antennas using capacitive CEs for mobile terminals have been widely proposed in the literature. In Villanen et al. (2007a), two capacitive CEs were placed on the two corners of the short edge of the PCB (as in Fig. 15), one covering the GSM frequency band (824–960 MHz) and the other the DCS/PCS bands (1.71–1.99 GHz). A matching network branch was designed for the feed of each CE and special attention given to make each branch behave as open circuit relative to the other branch. This way, the two branches could be connected into a single feed at the end. A similar strategy was used in Andujar et al. (2011) to use two CEs while the MN branches were connected together using notch filters.

In Valkonen et al. (2012), three different CEs were proposed, having different dimensions and orientations to investigate the effect of the user's hand on the matching and efficiency. The antennas were able to cover the low LTE (700–960 MHz) and DCS/PCS/UMTS bands (1.7–2.2 GHz) with a two-branch MN (nine SMD components with inductors and capacitors) connected to a single feed (Fig. 16).

An antenna system using a CE (and ground plane) to cover the LTE frequency bands (700–960 MHz and 1.7–2.7 GHz) was presented in Valkonen et al. (2013a). This antenna is like a dual-feed structure where the low-band feed uses the radiating element as a CE, whereas the high-band feed has an offset in the excitation region, using also the resonant properties of the radiating element (Fig. 17). These two feeds are combined together in a single feed after introducing a two-branch MN (eight lumped components) where each branch behaves as an open circuit at the operating frequency of the other.

Another design methodology using a CE to cover the LTE frequency bands was proposed in Cihangir et al. (2014). This design is a combination of a nonresonant CE with a resonant monopole antenna. The antenna is a dual-feed structure where one of the feeds is responsible for low-band (700–960 MHz) coverage and the other feed responsible for high-band (1.7–2.7 GHz) coverage. The LB feed is connected to a CE which is made hollow by removing the metal region inside, to be more space efficient. There is a MN of four lumped elements at the feed of the coupling element. The monopole, responsible for HB

DOI 10.1007/978-981-4560-75-7_100-1

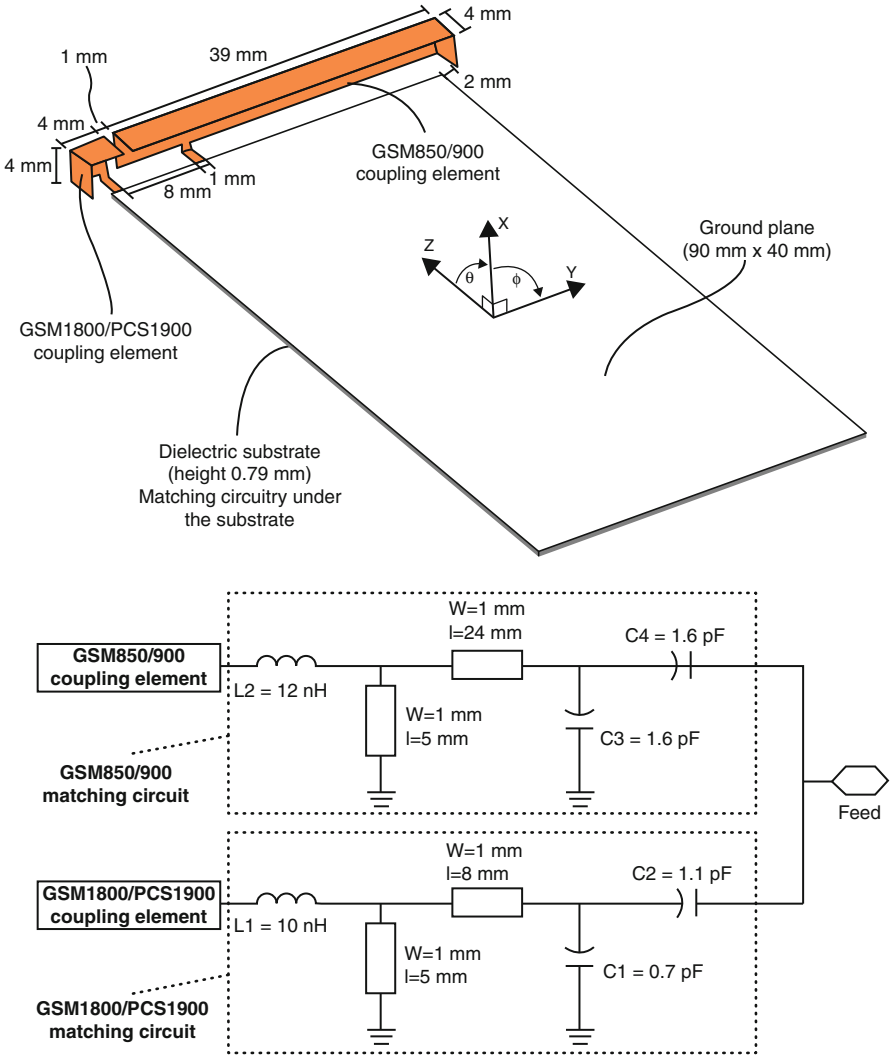


Fig. 15 Quad-band CE antenna structure (From Villanen et al. (2007a), with permission from John Wiley and Sons)

coverage, can be situated inside this available space (hollow CE) without perturbing the structure. However, it was printed on the substrate in this study for mechanical stability. The coverage of both bands was achieved with reasonable isolation between the feeds and with considerably lower number of MN components (Fig. 18).

The main advantage of using a CE-type antenna is the excellent bandwidth potential characteristics and the resulting broadband antennas that one can easily achieve. However, the absolute necessity of a MN at the feed is a disadvantage compared to other antenna types, in terms of cost. The disadvantage is more severe if a multiband antenna is targeted with a single antenna feed. In this situation, special attention shall be given to the MN design to be able to combine them into a single feed. Also the total number of SMD components in the MN will be high for this case, increasing the power losses in the components, the cost of the overall system, and the complexity of the MN that might cause unexpected measurement results considering the tolerance values of the lumped components (inductance or capacitance).

DOI 10.1007/978-981-4560-75-7_100-1

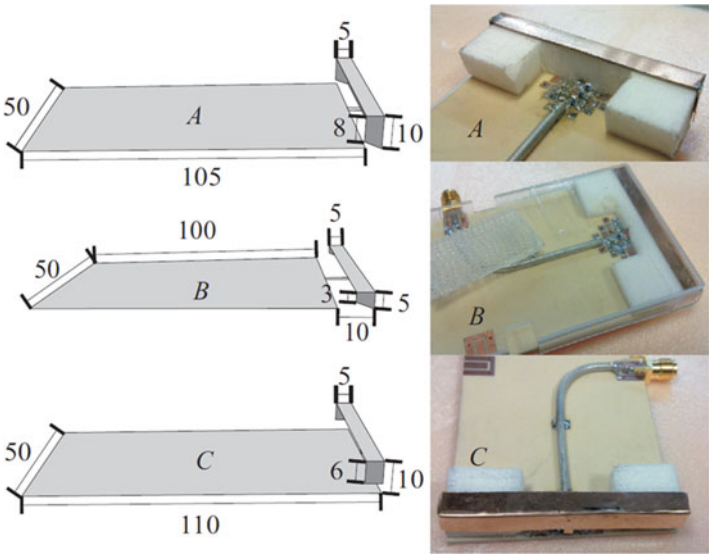


Fig. 16 CE structures proposed for user effect investigations from Valkonen et al. (2012)

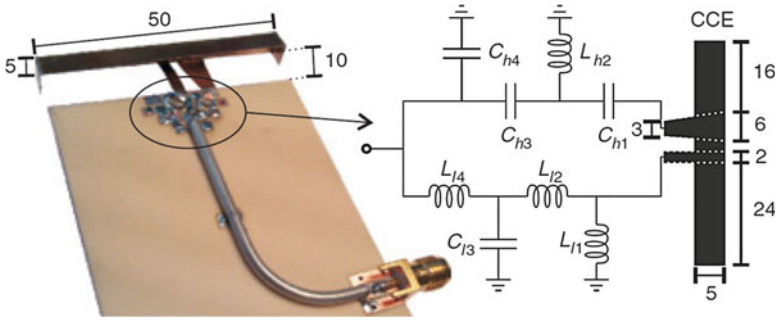


Fig. 17 CE antenna system for LTE coverage from Valkonen et al. (2013a)

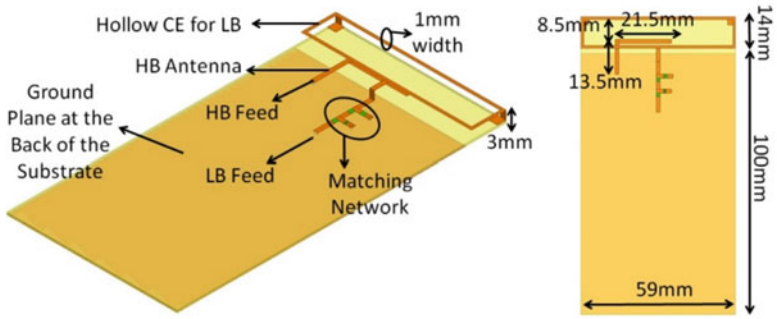


Fig. 18 Dual-feed antenna structure with a CE for LTE coverage from Cihangir et al. (2014)

Examples of Commercial Antennas

Since their introduction, mobile phones have gone under significant changes in terms of both size/shape and functionalities. The functionalities expected from the mobile terminals have always been in an increasing trend. For the size, until the release of smartphones, the general aim was to shrink them in

DOI 10.1007/978-981-4560-75-7_100-1

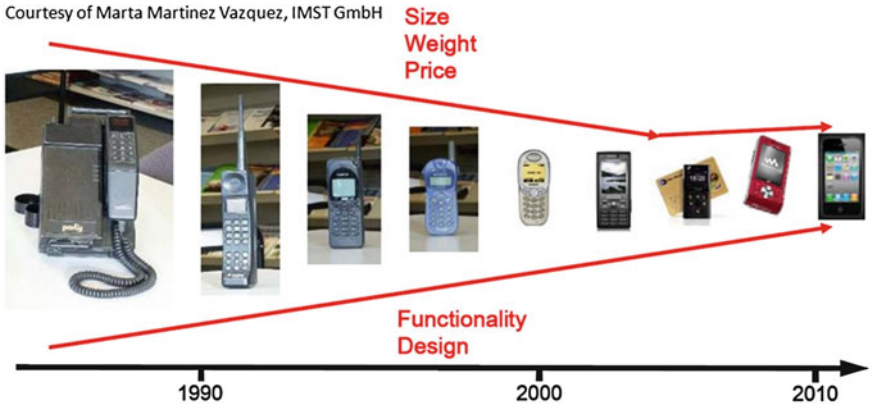


Fig. 19 The change in the size and functionality of mobile phones

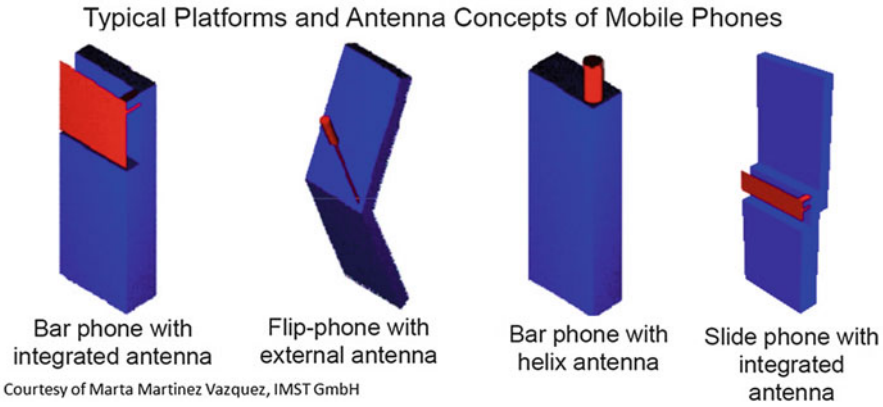


Fig. 20 CAD models for different mobile phone form factors

size for the ease of mobility. This returned to the antenna designer with the challenge of broadband antennas in a smaller volume. With the smartphone concept where a large touchscreen is the main component for device functionalities, the handheld size has more or less increased due to the touchscreen covering most of the device, but the space reserved for the antenna was not increased (Fig. 19). These kinds of constraints have led the antenna designers to look for new techniques and geometries to achieve antennas with broader bandwidth in limited volumes. Comprehensive studies about the evolution of mobile antennas were done in Ying (2005, 2012)and Rowell and Lam (2012). This section presents some examples of antennas integrated in commercial phones, together with some information about the most common device form factors.

Mobile Phone Form Factors

There are mainly three form factors that were widely used in commercial mobile phones as bar-type, flip-type, and slide-type (Fig. 20).

In the bar-type form factor, the device is formed by a rectangular prism without any movable part. Generally the keypad and the screen (or just the touchscreen) of the device are on the same face as seen in Fig. 21. Since the layout of the main components (like battery, screen, keypad, etc.) is stable and does not change, the antenna design for this form factor is more straightforward compared to other two form factors.

DOI 10.1007/978-981-4560-75-7_100-1

Fig. 21 Bar-type mobile phones

Fig. 22 Various flip-type mobile phones

For the flip-type phones, the device is formed by two sections that are connected to each other through hinges (Fig. 22). Those sections can be fairly modeled by full ground plane rectangles. When the device is in closed state, the total size shrinks into half, making it more ergonomic. Generally, the keypad and the screen are located in these two different sections and not visible when the phone is not opened. The antenna design for phone form factor is more tricky since the orientations of the main components relative to the antenna can be subject to change, depending on if the phone is open or closed. Moreover, the effective length of the chassis in the closed state is very short regarding the operating wavelength, and even the Q factor of this superposed state of PCBs is not low and could not lead to enough bandwidth potential at LTE LB. In Villanen et al. (2007b), the antenna performance for a flip-type phone was investigated, and it was shown that the radiation efficiency can decrease significantly when the phone is

DOI 10.1007/978-981-4560-75-7_100-1

Fig. 23 Various slide-type mobile phones

closed. In fact, in this situation, the currents excited on the two parts of the ground plane (strong especially in the lower frequencies) are in reverse direction and physically very close to each other that tend to cancel each other and not produce radiation in far field. This also affects the input impedance of the antenna decreasing the bandwidth potential as explained before. A reconfigurable matching network topology was suggested in the same study that will switch the matching network between two alternatives, one to match the antenna when the phone is closed and the other matching the antenna in open state.

The third common form factor is slide-type, whose examples are shown in Fig. 23. The device is again composed of two sections, but now, these sections can slide over each other on rails. Generally, one section that is in the front hosts the screen and the other the keypad. Also, special care should be taken in the antenna design for this form factor since the two conditions of the device may be different in terms of the surroundings of the antenna. The most important effect is probably the length of the ground plane and the wavemodes that can be excited on it, depending on the state. It can easily be told that almost the same design problems encountered with flip-type mobile phones exist for slide types, especially considering the closed state of the mobile phone.

Examples of Commercial Phones

The first generation of analog mobile phones used to have an external antenna placed on the top corner of the device. The preliminary models that were first released to the market used to have a whip antenna that is approximately a quarter wavelength long in the band of operation like the phone on left in Fig. 24. Some of the later models utilized retractable monopole antennas that were more ergonomic for the user (Fig. 24 on right).

External antennas continued to be used also in the former models of the 2G. The long-whip antennas switched to more user-friendly configurations like the retractable antenna. With dual-band operation being necessary, different design techniques were also used like inserting the retractable whip antenna in the center of a helical antenna or using a helical antenna with different pitch angles. The length of these antennas was considerably reduced by switching from whip antennas to helical antennas which also enabled shrinking the device form factors (Fig. 25).

The first commercial mobile phone with an internal antenna was the Hagenuk Global Handy, which was released in 1996. This antenna can be seen in Fig. 26, together with a CAD model taken from Rowell and Lam (2012). It is a slot antenna etched into the RF shielding operating at GSM900 band.

DOI 10.1007/978-981-4560-75-7_100-1

Fig. 24 Pictures of first generation of mobile phone antennas

Fig. 25 Pictures of external antenna examples

The internal antennas started quickly to become very popular due to both ergonomic and aestheticl reasons, with a lot of manufacturers switching to internal antennas in the newer models. Figure 27 shows three antenna models from the same manufacturer released in 3 consecutive years.

The detail of the internal antennas for the two mobile phones mentioned above can be observed in Fig. 28. The antenna on the left is a single-band antenna which uses a slot to increase the electrical length and achieve miniaturization. The antenna on the right is dual-band antenna placed on the top of the device, using a slot to obtain two patch sections and dual-band behavior.

One of the most common antenna types used in the 2G/3G mobile phones was the PIFA, with two examples shown in Fig. 29, taken from Rowell and Lam (2012). The antenna on the left is composed of two parts, the larger part responsible for GSM/DCS/PCS coverage and the smaller part responsible for UMTS coverage. Slots are used on the antenna elements to alter the current paths and enable both miniaturization and multiband coverage. The antenna on the right is a triband antenna (GSM/DCS/PCS), using a parasitic element which is capacitively excited by the main PIFA.

DOI 10.1007/978-981-4560-75-7_100-1

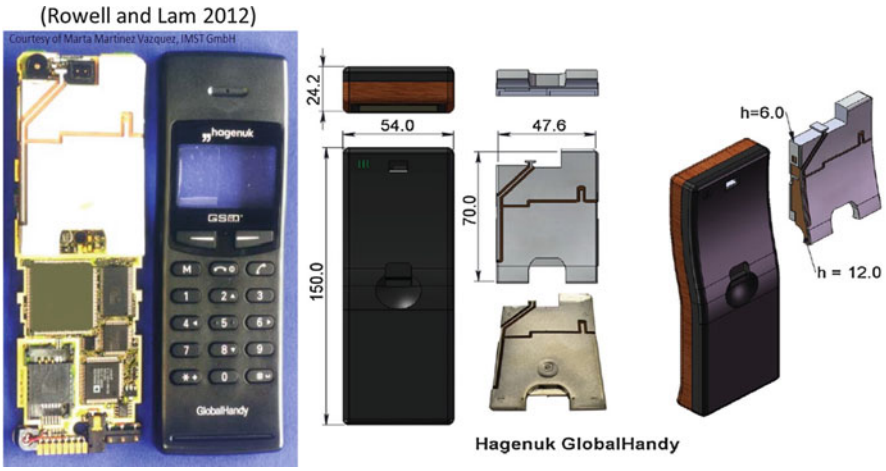


Fig. 26 First internal antenna in a mobile terminal (partially from Rowell and Lam (2012))

Fig. 27 Example of the transition from external to internal antennas during 3 years for the same manufacturer

Two other mobile phone models and their antennas are shown in Fig. 30. These antennas have no ground plane underneath them which make them behave as monopole-like antennas. In the model on the left, the antenna is placed at the bottom of the slide-type device, whereas it is placed in the middle (close to the hinge) for the flip-type phone on the right.

In the 4GS iPhone, the metallic portions of the phone casing (metallic frame) are used as CE to the PCB and also radiating structures. For this purpose, the metallic frame is split in two parts, the longer one forming the GSM/UMTS antenna and the shorter one forming the Bluetooth/Wi-Fi/GPS antenna. Very soon, it was revealed that this antenna was very sensitive to the user's hand (Antennagate), which degraded the received signal strength considerably when the finger is placed on the region separating the two metal frame portions (Fig. 31).

As the final of this section, the antennas used in Samsung Galaxy SII can be observed in Fig. 32. The main antenna responsible for GSM/UMTS coverage is placed at the bottom of the device. This layout typically decreases the SAR significantly since the antenna is usually set further away from the user

DOI 10.1007/978-981-4560-75-7_100-1

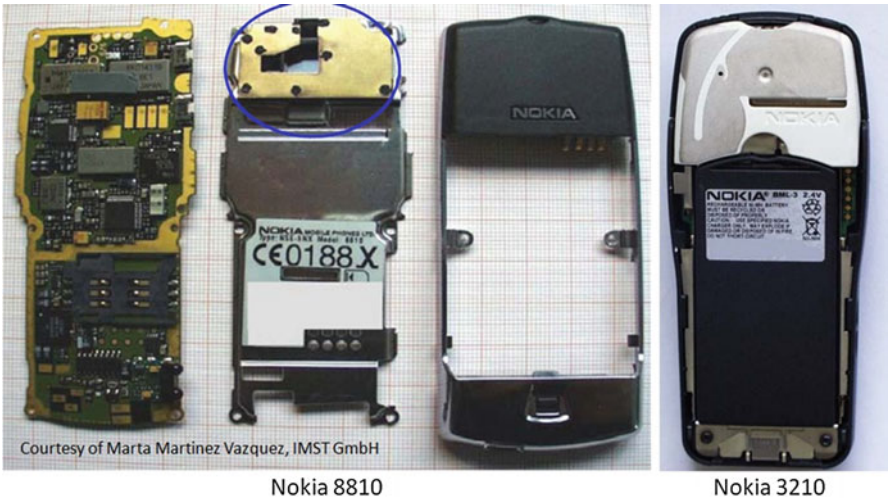


Fig. 28 Detailed view of internal antennas of two phone models

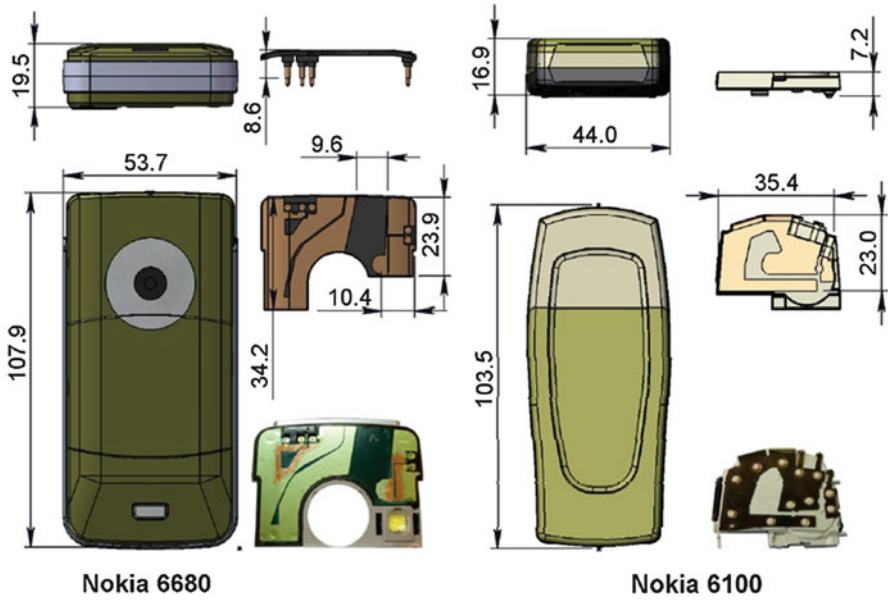


Fig. 29 Examples of PIFAs from commercial phones (© [2012] IEEE. Reprinted, with permission, from Rowell and Lam 2012)

(tissue) when compared to an antenna placed on top. A second antenna for Bluetooth/WiFi is also placed at the bottom of the phone. The GPS antenna, working in the receive mode only, is situated at the top of this device, just above the camera.

Antenna Feeding and Matching Methods

This section starts with the discussion of the two antenna feeding methods with their advantages or disadvantages. Next, the exploitation of parasitic elements is briefly introduced, with examples from literature. Information about the most common antenna matching methodology, namely, the utilization of

DOI 10.1007/978-981-4560-75-7_100-1

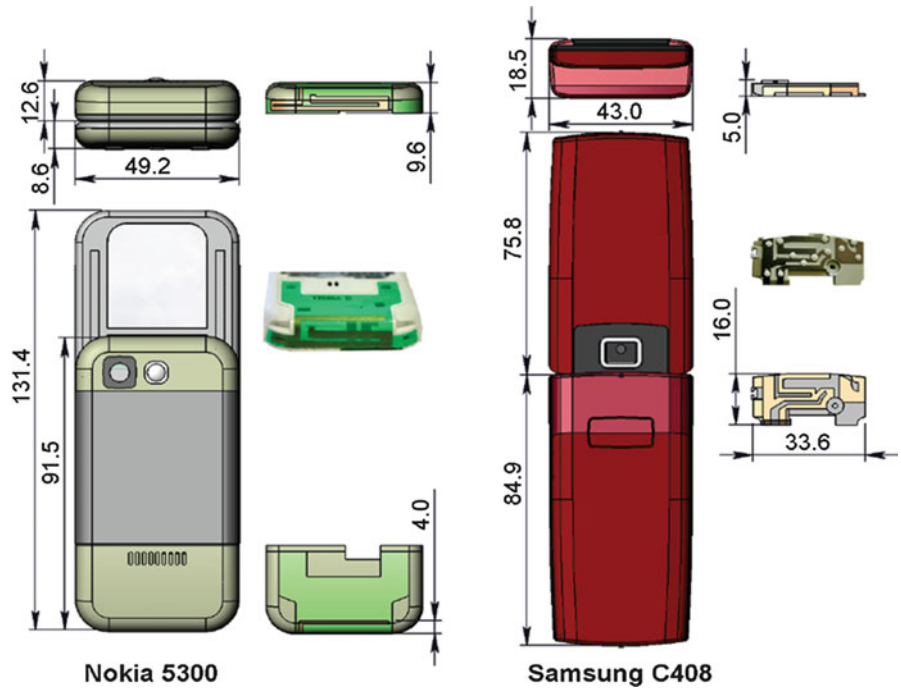


Fig. 30 Examples of monopole-like antennas from commercial phones from Rowell and Lam (2012)

Fig. 31 Coupling element-type antenna for a smartphone (iPhone 4S)

matching networks, is also given in this section with some basic topologies for different possible input impedances.

Antenna Feeding Methods

Inductive Feeding

In this feeding scheme, the power is coupled to the main radiating element through a direct metallic (galvanic) contact between the output of the front-end module (FEM) and the antenna. Examples can be any form of striplines, meander lines, or pogo pins.

DOI 10.1007/978-981-4560-75-7_100-1

Fig. 32 Picture of the antennas used in Samsung Galaxy SII

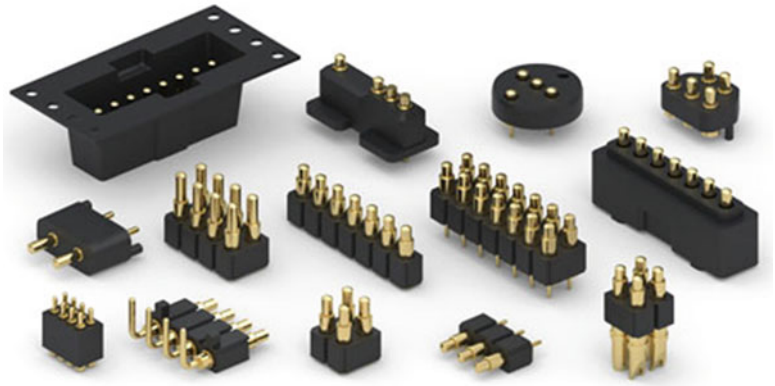


Fig. 33 Pictures of various pogo pins

Pogo pins are common inductive contacts. Since the mobile phone antennas generally have a height over the system PCB (rather than being directly printed on the PCB) for better bandwidth, the power transmission is necessary to/from the FEM. Pogo pins ensure good connection through spring-loaded pins (Fig. 33).

Meander lines are also an inductive feeding method that is commonly employed in mobile phone antennas to miniaturize the antenna element and still get the same operating (resonance) frequency. To illustrate this, simulations were performed with a monopole element, printed directly on an FR4 substrate as seen in Fig. 34. The resonance (defined here as the frequency having minimum reflection coefficient) of the monopole antenna (without any meanders) occurs at 1.8 GHz (Fig. 35) for a monopole length of 24 mm (PCB is 100 $\times$ 50 mm^2). The same resonance frequency can be obtained with a 14 mm monopole element when using meander lines, enabling approximately 50 % shrinkage in length. The main disadvantage is the lower real part of the input impedance of the antenna as seen in a Smith chart,

DOI 10.1007/978-981-4560-75-7_100-1

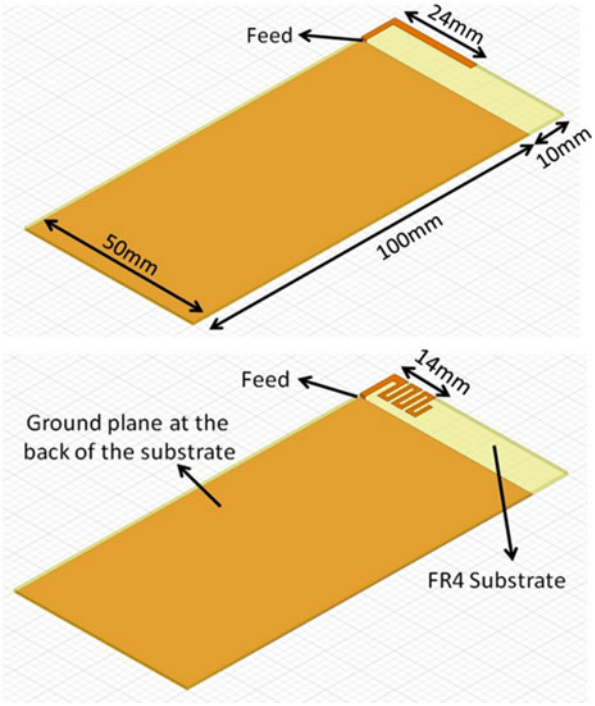


Fig. 34 Utilization of meander lines (geometry)

resulting in a worse reflection coefficient performance and less bandwidth potential compared to without meanders.

Capacitive Feeding

In the capacitive feeding scheme, the main radiating element is excited capacitively (with no direct metallic contact) through another driven element. This driven element is generally shorter and may cause some higher band resonances for the overall antenna. For a comparison between inductive excitation and capacitive excitation, the models shown in Fig. 36 are used. A monopole antenna printed on FR4 substrate has a 24 mm length. It is fed by inductive excitation (shown on left). This antenna is compared with a capacitively excited monopole, connected to ground plane on one end (shown on right). To excite this grounded monopole, a shorter feed is used close to the other end, which is left open. This last antenna has a total length of 20 mm.

The input impedances of these two antennas (plotted between 1.5 and 3 GHz) are presented in Fig. 37. It can be seen that a loop is formed in the Smith chart for the capacitive feed configuration. The bandwidth potentials calculated using a commercial software for a reflection coefficient below −6 dB are also given below for the two feeding methods. For this antenna layout, the capacitive excitation gives approximately double bandwidth potential around 2.3 GHz when compared to the inductive excitation, with a smaller profile.

Parasitic Elements to Increase the Bandwidth Potential

Utilization of parasitic elements to enhance the matching bandwidth is a popular technique in mobile terminal antenna design, since the antenna is electrically small especially in lower frequencies. Some antenna examples from the literature and from commercial phones that contain parasitic elements for bandwidth enhancement were presented in this chapter so they will not be repeated in this section. The effect of the parasitic element on bandwidth potential will be discussed here with an example.

DOI 10.1007/978-981-4560-75-7_100-1

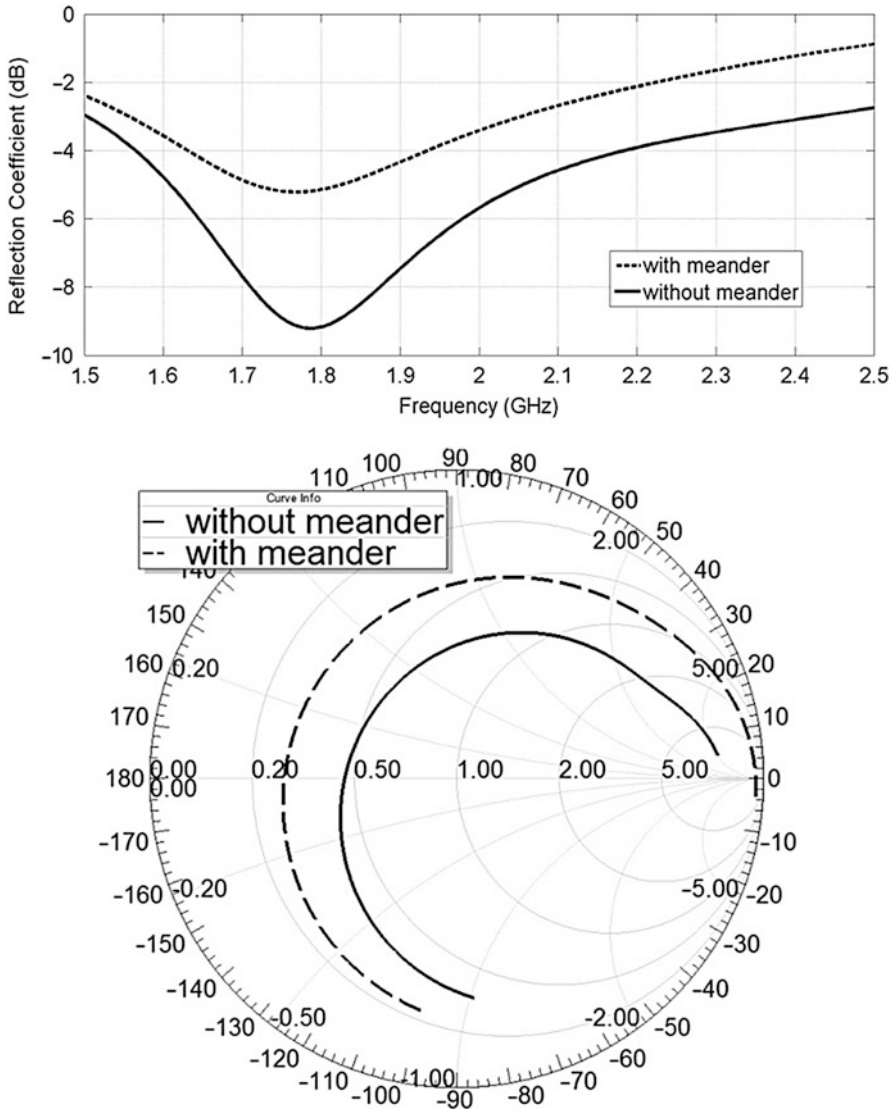


Fig. 35 Utilization of meander lines (reflection coefficient)

An antenna design that operates in the low-LTE/GSM frequency band (700–960 MHz) was presented in Cihangir et al. (2013). The design challenge in the study was the use of a short ground plane (75 mm) for the device. Longer ground planes (like 100 mm) have their first wavemode around 1.1 GHz; however, this resonance shifts up to around 1.4 GHz if this length is 75 mm. Keeping in mind the role of the ground plane wavemodes on the antenna bandwidth potential in the lower frequency bands, the bandwidth potential coming from this short ground plane will be very narrow between 700 and 960 MHz. For this reason, a parasitic element was used in combination with a monopole-like driven strip. The model of the antenna is presented in Fig. 38. The antenna rests on a plastic piece and is composed of two strips. The first strip is the driven strip where the antenna feed is connected. This strip capacitively excites a longer parasitic strip connected to the ground plane on one end. Pogo pins are used under the plastic piece for the connection of the antenna and the PCB.

The simulation results of this antenna with and without the parasitic strip are presented in Fig. 39. A new resonance that does not exist without the parasitic strip is seen to be created around 750 MHz with the introduction of this parasitic strip. When the input impedance in a Smith chart is observed, a loop is

DOI 10.1007/978-981-4560-75-7_100-1

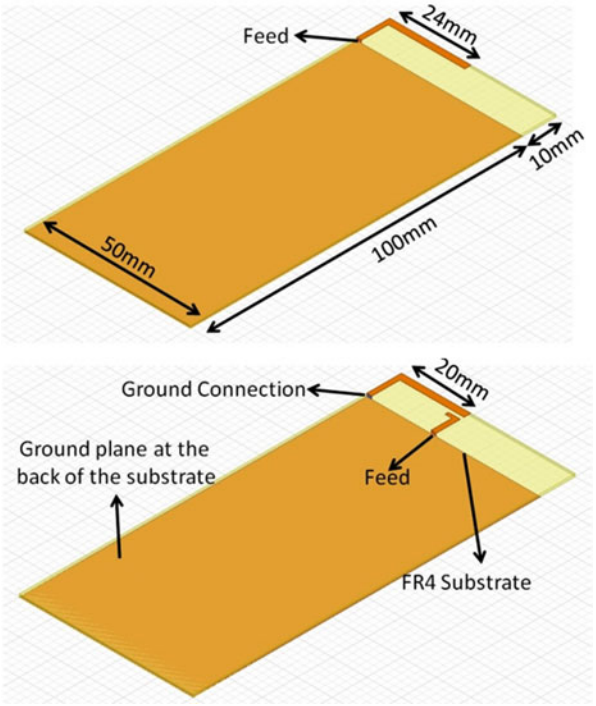


Fig. 36 Utilization of capacitive excitation (geometry)

seen in the presence of the parasitic strip which considerably increases the real part of the input impedance. To tune the antenna behavior, several design parameters can be changed. For example, increasing the length of the parasitic strip will lower the resonance frequency coming from it. Another parameter can be the coupling distance between the strips and/or the orientation of the two strips (affecting the coupling strength) that will change the shape of the loop in the Smith chart and in turn affect the bandwidth potential.

The simulated bandwidth potential with/without parasitic element can be seen in Fig. 40. To cover the target band (700–960 MHz), a bandwidth potential of nearly 260 MHz is necessary at 830 MHz (center frequency of this band). The bandwidth potential at this frequency is less than 50 MHz without the parasitic element. A significant increase in the bandwidth potential is observed with the introduction of the parasitic element, rising up to more than 200 MHz.

Utilization of Matching Networks for Impedance Matching

As mobile terminal antennas had to shrink due to available space and also due to the necessity of more broadband antennas, the usage of matching networks has come forward. These matching networks (MNs) can be placed at the antenna input, as well as distributed through the antenna like an inductor placed at the center of a metallic strip to increase its electrical length and decrease the resonance frequency. The MN can contain both lossless and lossy components, but the existence of lossy components will have a negative effect on the antenna efficiency so they are not widely preferred. For this reason, the MNs are generally composed of lumped inductors and capacitors. Figure 41 presents how the input impedance location in a Smith chart changes when a shunt/series inductor/capacitor is used at the antenna feed. Arrow-1 shows how the impedance moves in a Smith chart when the value of a series capacitor is decreased at the antenna input. On the contrary, increasing the inductance value of a series inductor will move the impedance in the direction of arrow-2. It should be noted that the components connected in series move the impedance on the “constant impedance circles” in the Smith chart. Arrow-3 shows

DOI 10.1007/978-981-4560-75-7_100-1

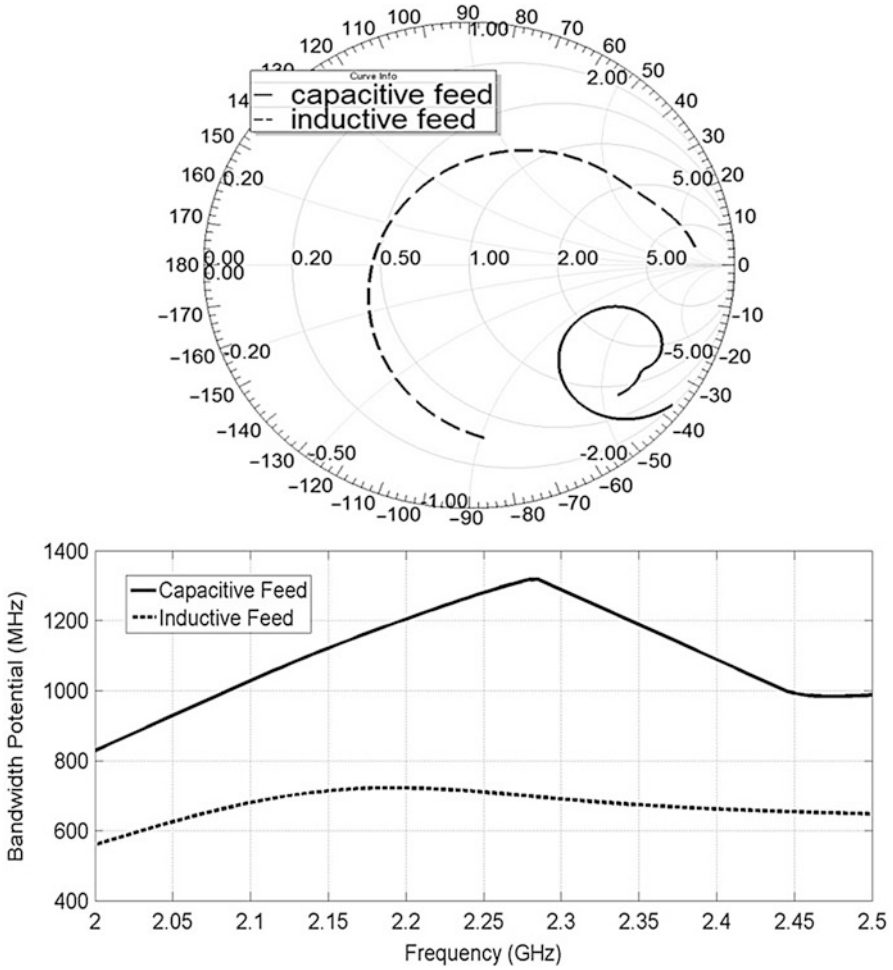


Fig. 37 Utilization of capacitive excitation (impedance and bandwidth potential)

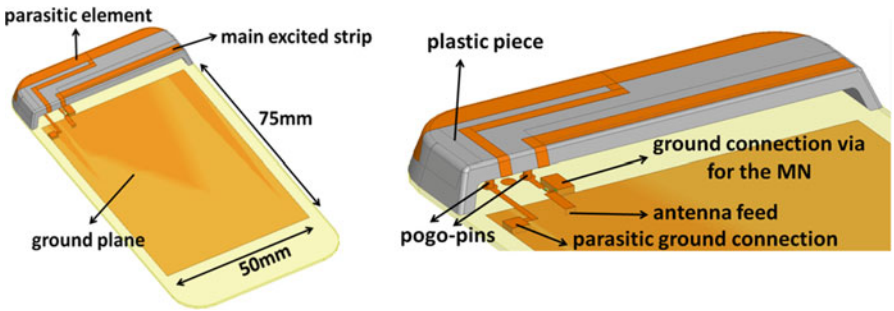


Fig. 38 Low-LTE/GSM antenna using a parasitic element from Cihangir et al. (2013)

the impedance behavior when the inductance of a shunt inductor is decreased, and arrow-4 presents the change when the capacitance of a shunt capacitor is increased. The impedance changes occur on the "constant admittance circles" when shunt elements are used.

One of the simplest MN topologies is L-type, placed at the input of the antenna. These types of MNs are composed of two lumped components, one of which is in series configuration and the other in shunt configuration. The eight possible L-type MN layouts are presented in Fig. 42. These different topologies

DOI 10.1007/978-981-4560-75-7_100-1

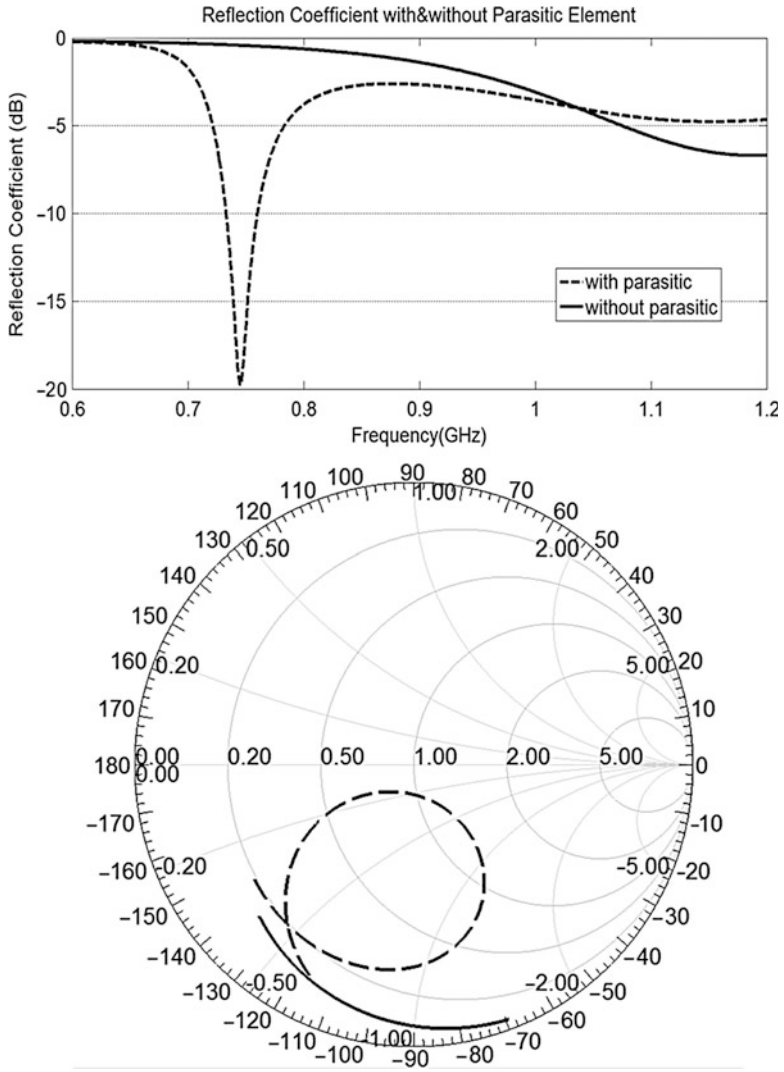


Fig. 39 Simulations of the effect of the parasitic element on the input impedance of the antenna from Cihangir et al. (2013)

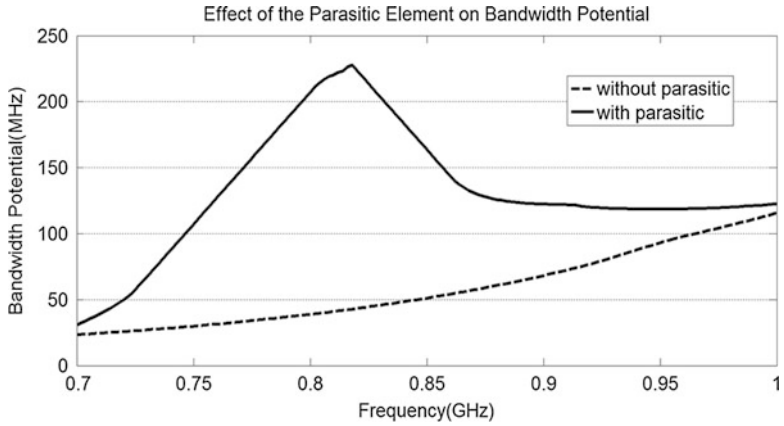


Fig. 40 Simulation of the effect of the parasitic element on bandwidth potential from Cihangir et al. (2013)

DOI 10.1007/978-981-4560-75-7_100-1

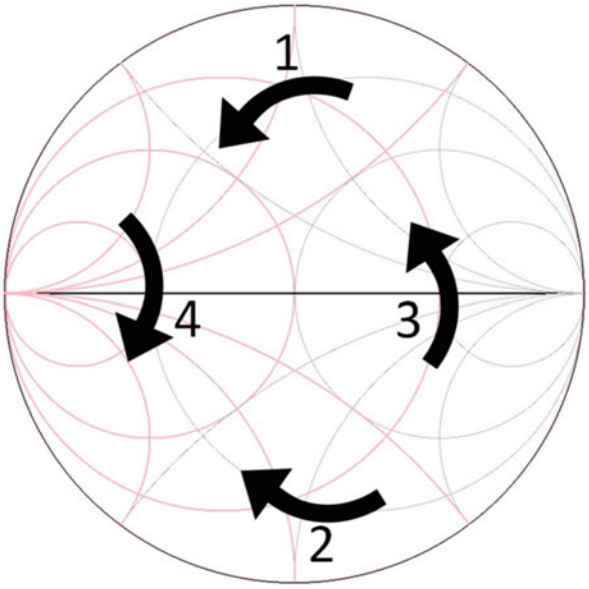


Fig. 41 Lumped inductor and capacitor effects on the antenna impedance

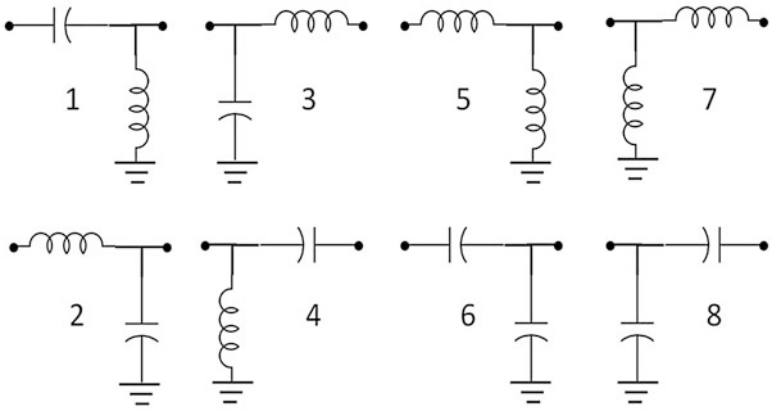


Fig. 42 L-type MN topologies

can be used to bring any input impedance to the center of the Smith chart, to obtain a good matching and low reflection coefficient. However, specific topologies can be used to match the impedances on specific regions of the Smith chart as shown in Fig. 43.

To further illustrate, two different MN topologies were used to match the same antenna in Figs. 44 and 45. The impedance of the antenna without any MN plotted between 1.7 and 2.2 GHz is shown with the red curve. Since the original impedance is in the lower hemisphere of the Smith chart, topology types 1, 3, 5, and 7 can be used (Fig. 43) to match this antenna. In Fig. 44, topology-3 was used. Using first the series inductor, the impedance is moved in clockwise direction on the constant impedance circle (black curve). Then using the shunt capacitance, the impedance is moved in clockwise direction on the constant admittance circle.

In Fig. 45, topology-5 was used to match the same antenna. Using the shunt inductor, the original antenna impedance is moved in counterclockwise direction along the constant admittance circle. Then

DOI 10.1007/978-981-4560-75-7_100-1

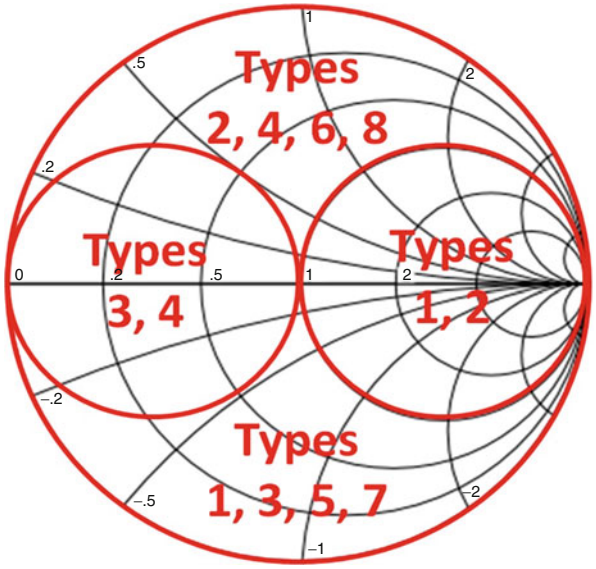


Fig. 43 Regions on Smith chart, according to the usable L-type MN topologies

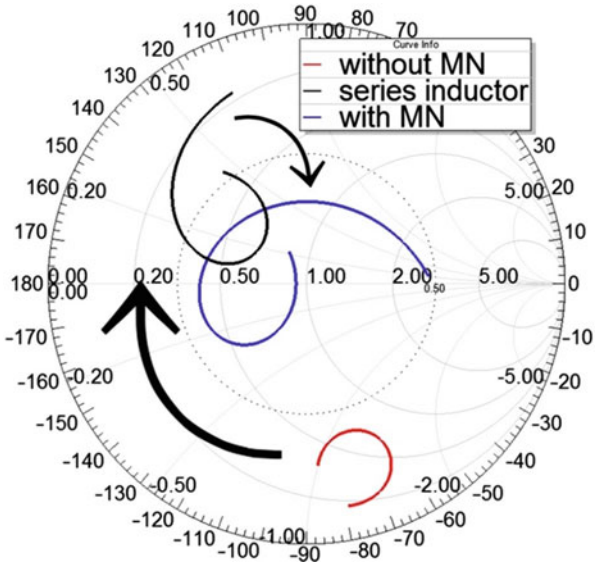


Fig. 44 Matching an antenna using L-type topology-3

using the series inductor, the antenna is matched by moving in clockwise direction on the constant impedance circle. Figure 46 presents the reflection coefficients for these three cases (without MN, with topology-3, with topology-5).

Tunable/Reconfigurable Antennas in Mobile Terminals

Tunability or reconfigurability is used for a wide variety of purposes in antennas like changing the radiation direction, the polarization, beamwidth, or even the operating frequency. For mobile phone antennas, it is by itself a big challenge to obtain acceptable level of matching like −6 dB in free space, but

DOI 10.1007/978-981-4560-75-7_100-1

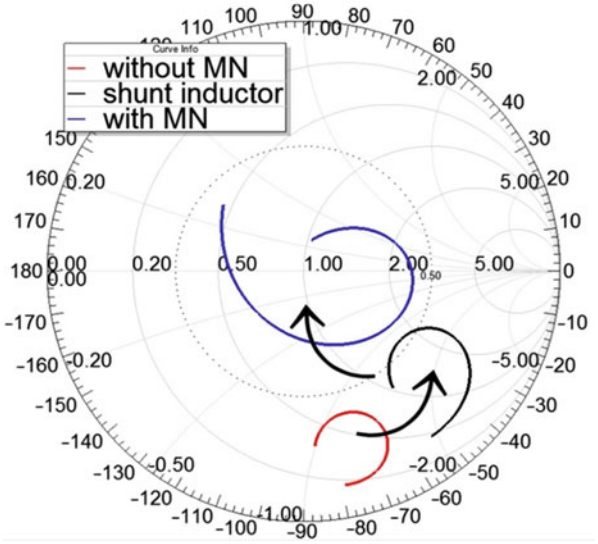


Fig. 45 Matching an antenna using L-type topology-5

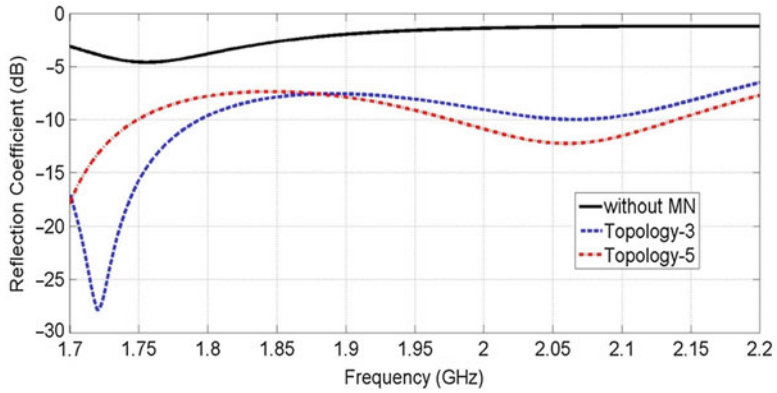


Fig. 46 Reflection coefficients for different MN topologies

if −10 dB is needed, the tunable/reconfigurable topologies have been already proposed to overcome the matching problem and decrease the reflection coefficient in a wider band.

A variety of tunable components are available for reconfigurable antennas. Switches that operate in the RF frequencies (like HEMT, pin diodes, or MEMS) are mainly used for matching an antenna in predefined frequency bands. On the other hand, tunable components (BST capacitors, MEMS variable capacitors, varicap diodes) are generally utilized to achieve a continuous but narrowband matching that can be swept through the target frequencies by changing the component values.

In Valkonen et al. (2007, 2010), matching networks (MNs) composed of multiple branches are used at the antenna feed. The active MN branch that feeds the antenna is changed using MEMS switches. For example, in Valkonen et al. (2010), a CE is placed on one edge of the PCB and excited through the MN seen in Fig. 47. According to the positions of the two switches, one of the two MN branches excites the antenna which tunes it to either the lower-frequency GSM band or the higher-frequency DCS/PCS band.

In Yamagajo and Koga (2011), a MN with switches is used again at the antenna feed. The antenna without MN covers the frequency ranges higher than 2 GHz but has poor matching below 2 GHz. The MN contains three switches which are connected to different shunt inductors. According to ON/OFF states of

DOI 10.1007/978-981-4560-75-7_100-1

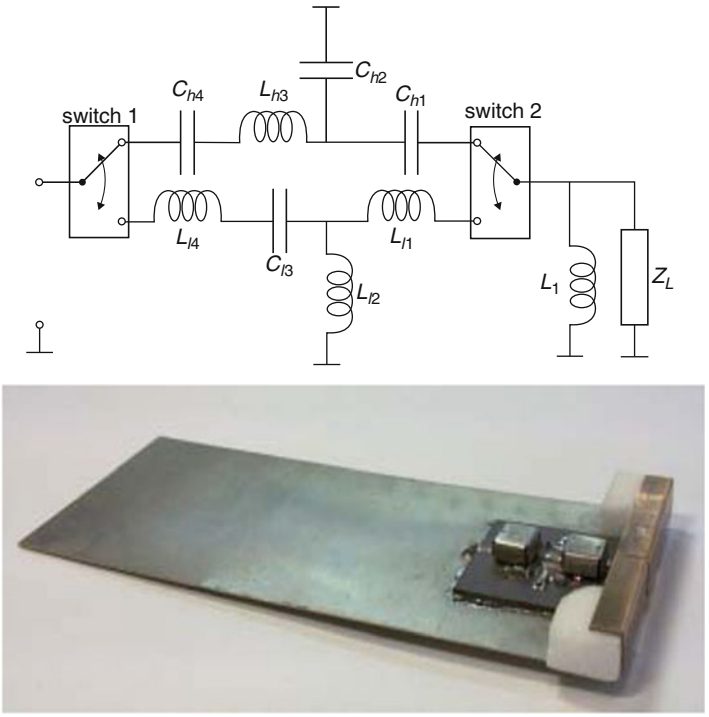


Fig. 47 MN and antenna topology for a tunable antenna using switches from Valkonen et al. (2010)

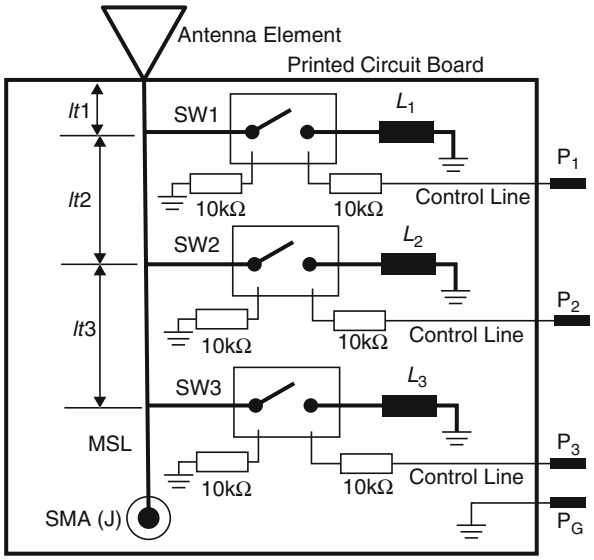


Fig. 48 MN composed of switchable shunt inductors (© [2011] IEEE. Reprinted, with permission, from Yamagajo and Koga 2011)

the switches, the overall shunt inductor value of the MN is changed and tunes the antenna to different operating bands between 640 MHz and 5.85 GHz (Fig. 48).

The tunable components are distributed through the antenna rather than placed at the feed in Mak et al. (2007) and Park and Sung (2012). For example, in Mak et al. (2007), two different topologies are evaluated, named as "switched feed design" and "switched ground design." In switched feed design, the feed of the antenna is either done from one end of a slotted strip or changed to the other end. In this way,

DOI 10.1007/978-981-4560-75-7_100-1

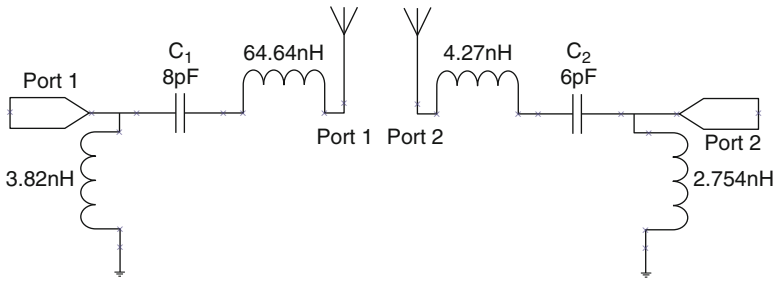


Fig. 49 Tunable MN with variable capacitors from Hu et al. (2010)

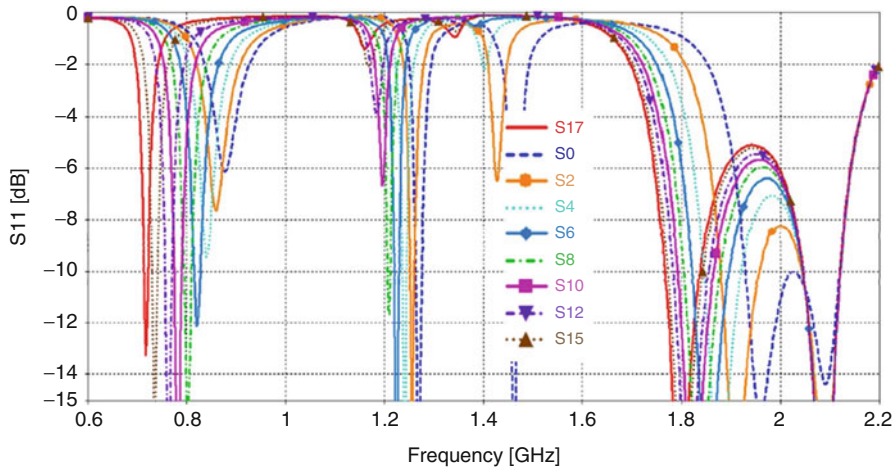


Fig. 50 Reflection coefficient for different capacitance values (© [2013] IEEE. Reprinted, with permission, from Ramachandran et al. 2013)

the reflection coefficient of the antenna operates in two states, the difference between the states being more in the higher band around 2 GHz. In switched ground state for the same antenna, one end of the metal strip is fed and the other end is either shorted to ground plane (forming a loop) or left open, according to the switch state.

In Park and Sung (2012), two pin diodes are used between the feed point and grounding point of a PIFA. According to the DC modes of the pin diodes, the antenna operates whether in PIFA mode or in loop mode to cover either the low-frequency GSM or DCS/PCS/UMTS bands.

Apart from these frequency switching antennas, tunable components can also be used to match continuously and instantaneously narrowband intervals for broader frequency bands. In Hu et al. (2010), two coupling elements are used with their own matching networks (Fig. 49). The MNs each consists of two lumped inductors and a variable capacitor. By changing the capacitance values between 0.2 and 8 pF, the first CE is tuned between 450 and 1360 MHz and the other one between 1660 and 2750 MHz.

In Boyle and Steeneken (2007) and Ramachandran et al. (2013), MEMS tunable capacitors are exploited at the ground connections of PIFAs to achieve tenability. Figure 50 displays the reflection coefficient for the antenna in Ramachandran et al. (2013) plotted for different capacitance levels of the tunable capacitor.

In Manteuffel and Arnold (2008), both switches and tunable components were used to obtain better matching with a simple CE, in a very wide frequency range going down to FM radio (100 MHz), DVB-H, GSM, and up to UMTS frequencies. Different MNs were used in different frequency bands, and these

DOI 10.1007/978-981-4560-75-7_100-1

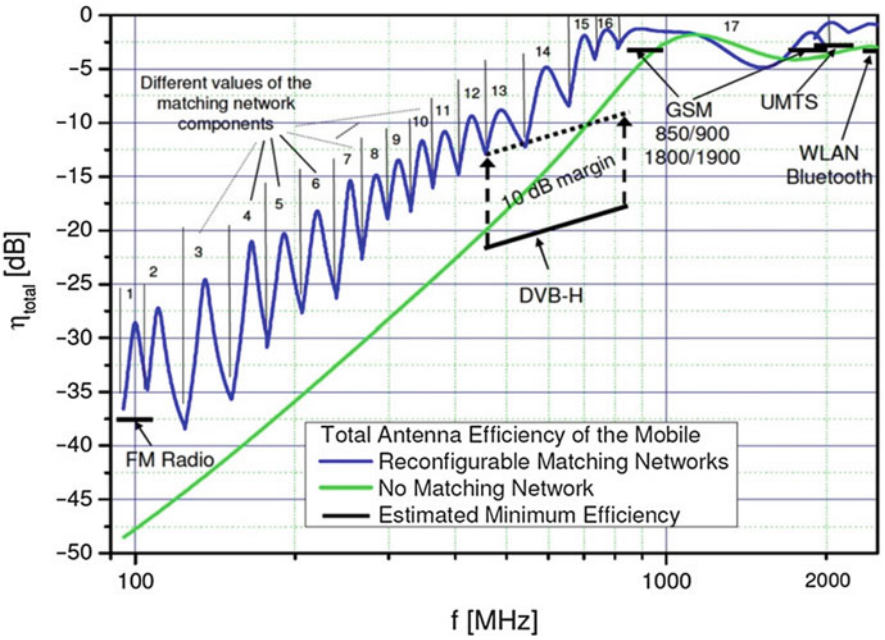


Fig. 51 Total efficiency for different MN states (© [2008] IEEE. Reprinted, with permission, from Manteuffel and Arnold 2008)

sub-MNs also contained tunable elements that enabled better matching in the sub-bands of interest. Figure 51 presents the total efficiency obtained from this antenna between 100 MHz and 2.2 GHz.

Tunable antennas have the advantage of enabling wider frequency coverage with a simpler antenna when compared to passive antennas. Tunable MNs can also be used to miniaturize antennas to obtain acceptable reflection coefficient and sufficient total efficiency. Another major advantage is the detuning effects (like the effect of the hand of the user covering the mobile phone) that can be compensated to some extent by adjusting dynamically the component values in a tunable MN. However, tunable/reconfigurable antennas also have some disadvantages compared to passive antenna layouts. The first and maybe the most significant one are the losses coming from the tunable components that will decrease the total efficiency of the antenna. Also, a real-time monitoring (closed-loop system) can be necessary to follow the healthy operation of the antenna and tunable components which increases both the complexity and the cost of the antenna parts of the mobile phone.

Antenna-User Interaction in Mobile Terminals

The interaction between a mobile phone antenna and the user can be categorized in two ways: the effects of the user on the antenna and the effects of the antenna on the user.

Effects of the User on the Antenna

In mobile phone antennas, the scenario which has the strongest interaction between the user and the antenna is the talk position. In this position, the antenna is held by the user's hand close to the cheek of the user. So, it can be told that the effect of the user on the antenna is limited by the hand and head of the user.

Practically, two effects of the user on the antenna can be mentioned. First is the detuning due to dielectric loading by the hand and head, since these tissues have high dielectric permittivity values. The second effect is the absorption of the radiation by the hand and head, due to their lossy characteristics. The

DOI 10.1007/978-981-4560-75-7_100-1

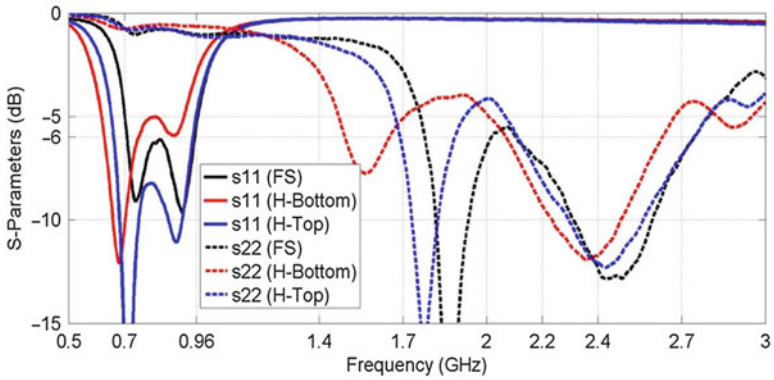


Fig. 52 Measured S-parameters with the user effects for the dual-feed antenna from Cihangir (2014)

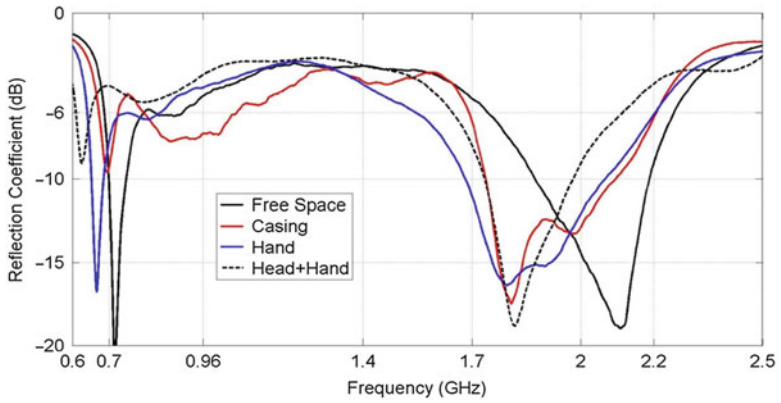


Fig. 53 Measured S-parameters with the user effects for the single-feed antenna from Cihangir (2014)

first effect can be observed from the reflection coefficient of the antenna whereas the second effect results in reduced total efficiency.

Detuning of the Antenna Due to User

The detuning of the antenna in the presence of the user's hand and head occurs due to the dielectric loading, by shifting the resonance frequency of the structure to lower frequencies. The effect of the user's hand was demonstrated with measurements with a dual-feed antenna in Cihangir (2014). The measured s-parameters of this study for three configurations are shown in Fig. 52. These three configurations are free-space, with hand when the antenna is placed on top of the PCB (facing the index finger), and with hand when the antenna is placed at the bottom of the PCB (facing the palm). In the hand-top configuration, a resonance shift toward lower frequencies is observed in both the lower and higher bands, but the shift is not as severe to cause a significant band detuning. When the antenna is placed at the bottom of the PCB with the hand, the resonance shift is stronger so the antenna gets detuned around 830 MHz and 1.9 GHz. It is expected to see a stronger detuning when the antenna is at the bottom of the PCB since the antenna in this position is totally covered by the lossy palm. However, when it is positioned on top, the antenna only faces the index finger which does not totally cover the antenna.

Another study about the user effect was done in Cihangir (2014), this time with a single-feed antenna and in four configurations as free space, with a dielectric casing, with hand, and with hand and head. In all

DOI 10.1007/978-981-4560-75-7_100-1

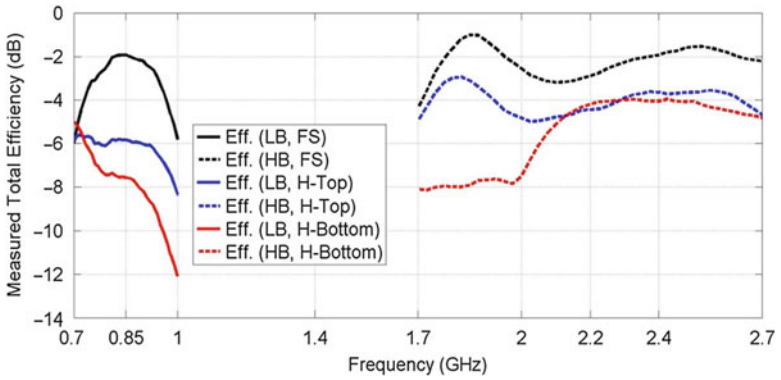


Fig. 54 Measured total efficiency with the user effects for the dual-feed antenna from Cihangir (2014)

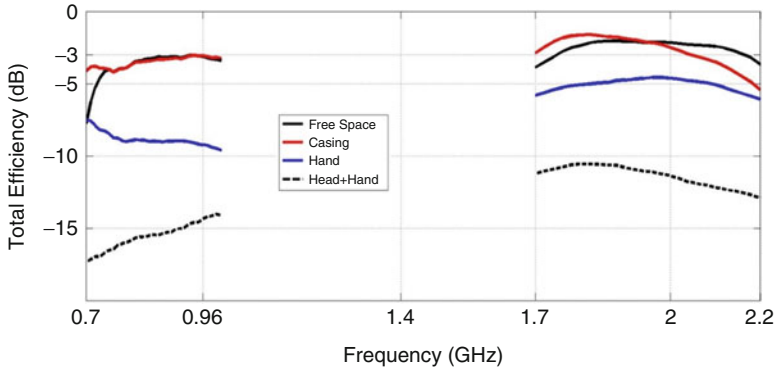


Fig. 55 Measured total efficiency with the user effects for the single-feed antenna from (Cihangir 2014)

these configurations, the antenna is placed at the bottom of the PCB, facing the palm. Figure 53 shows the measurement results from this study. The same behavior as before can be observed also here, with the resonance frequency shifting to lower frequencies with the presence of the casing, hand, and head. In this antenna, the detuning in the lower frequency band is more severe whereas the high-band detuning is less significant.

Total Efficiency Reduction Due to User

As also mentioned previously, the second major effect of the user on the antenna is the reduction of its total efficiency. The total efficiency can be reduced both due to the detuning (and thus higher return loss) and the absorption of the hand and head (that reduces the radiation efficiency). The total efficiency measurements for the two antennas presented above are given in Figs. 54 and 55. For the dual-feed antenna, the efficiency drops to around −6 dB and −8 dB in the low band for respectively "antenna on top" and "antenna on bottom" configurations. The efficiency in the high band is less affected compared to low band. Comparing, for example, the free-space (FS) and antenna on top with hand (H-top) configurations, there is a significant efficiency drop with the hand, although the reflection coefficient is not disadvantageous in Fig. 52. This suggests the decrease of the radiation efficiency owing to the lossy nature of the hand. As expected, the efficiency when the antenna is at the bottom is the least among all configurations.

DOI 10.1007/978-981-4560-75-7_100-1

Similar conclusions can be drawn from Fig. 55, representing the measured total efficiency for the single-feed antenna mentioned before. The efficiency with the hand is approximately around −9 dB in low band and around −5 dB in HB. The efficiency with also the head further decreases as expected.

The effect of the user on the antenna parameters was also investigated in a number of studies in literature. In Ilvonen et al. (2012), a shielding structure was proposed that minimizes the interaction between the hand and the antenna. Efficiency when the antenna is on top of the PCB in presence of the hand and head was measured to be −10 dB in lower band and −7.5 dB in higher band. In Valkonen et al. (2012, 2013b), antenna structures that were designed on purpose of preventing detuning when the antenna is on top of the PCB were proposed. The efficiency drop in the presence of the hand, compared to free space case, was around 7 dB for these antennas. The user effect on a narrowband PIFA was investigated in Pelosi et al. (2009), showing efficiency drop of 11 dB at 900 MHz and 8.5 dB at 1800 MHz in the presence of hand and head.

Effects of the Antenna on the User (Specific Absorption Rate)

It is a well-known fact that electromagnetic radiation can cause serious health risks for the human body. For the higher portions of the frequency spectrum (ultraviolet, X-rays, Gamma rays), the waves carry enough energy to ionize the atoms or molecules in the human tissues so they are called ionizing radiation. However for the microwave frequencies which mobile phones also use, the waves are categorized as nonionizing type. The main effect is thus confined to thermal that can cause damage through excessive heating of the tissues due to being exposed to radiation.

Specific absorption rate (SAR) is a quantity to calculate the RF power absorbed by the body. It is the accepted metric to quantize the potential risk on the user from the mobile terminal. The unit for SAR is (W/kg). It can be calculated through Eq. 6.

$$\mathrm{SAR} = \int \frac{\sigma(r)|E(r)|^2}{\rho(r)} dr \tag{6}$$

where σ is the conductivity of the tissue, $E(r)$ is the electric field amplitude, ρ is the mass density, and the integration is done over the volume where the SAR is measured.

It can be told that the SAR of a mobile phone antenna depends on the wave properties like frequency and polarization, the body characteristics like the shape and electrical properties of the tissue, the distance between the antenna and the tissue, and last but not least the antenna concept.

The SAR can be defined as full-body SAR or local SAR. For mobile phones, the only part of the body in close interaction with the antenna in talk mode is the head, so the local SAR in the head is used to characterize mobile phones. Two limit values were set by two standards (IEEE 2003), defining the limit in USA and (IEC 2005) in Europe. The US limit sets a maximum local SAR of 1.6 W/kg averaged over 1 g of tissue. The European limit is less strict compared to the US limit, being 2 W/kg averaged over 10 g of tissue. It can be told that a mobile phone compliant with the US SAR limit is almost always automatically compliant with the European SAR limit. Any mobile phones that exceed these limits are not authorized to be released on the market.

Different talk positions like cheek or tilt positions are defined for the simulation and measurement of the SAR. After the mobile phone is positioned in talk mode, the SAR is the maximum value seen in the head, after being averaged over 1 g (or 10 g) of tissue.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_100-1

Multifeed Antenna Systems

A lot of current mobile phones use a single-feed antenna that can cover all the required frequency bands for communication. This antenna connects to different front ends for different communication standards through a single-pole multithrow (SPMT) switch. According to the communication standard used at the instant, the antenna switch position is changed accordingly, to route the signal to/from the front end from/to the antenna. With the introduction of the LTE Advanced technology, it will be needed to receive and transmit data over two different frequency bands to achieve interband carrier aggregation. For example, a service provider can select offering the user an interband carrier aggregation scenario of aggregating LTE Band-4 (1.71–2.15 GHz) and LTE Band-12 (698–746 MHz). This means that the antenna should be able to operate in these two distinct frequency bands simultaneously and transmit/receive the RF signal to the two front ends for these two bands. This is not possible with the topology previously discussed since the SPMT switch will let the system transmit/receive only from a single band at a single instant. To overcome the problem, a switch that has logic to enable two or more inputs/outputs at an instant should be used. An alternative could be using a diplexer just before the antenna that will direct the higher and lower frequency portions of the carrier-aggregated signal to corresponding front ends. However, this will increase losses in the front end.

An alternative is to use multiple-feed antenna systems. Such topologies are composed of two or more antenna elements with their corresponding feeds covering narrower (compared to single-feed) different frequency bands. In this way, the constraints of a single-feed antenna are relaxed, and it enables the utilization of simpler switches. The major drawback is the challenge of minimizing the isolation between the different port-to-port antennas since it will degrade the total efficiency of the used antenna through coupling from one port to the other (if the isolation is not high enough).

An antenna of this kind for mobile terminals is presented in Ikonen et al. (2012), also shown in Fig. 56. Three monopole-like antennas are used, one to cover 700–960 MHz, one for 1.7–2.2 GHz, and one for 2.5–2.7 GHz. A matching network is used at the feed of each antenna. The S-parameter of the antenna system is presented in Fig. 57.

A similar strategy was applied to achieve LTE band coverage with multiple feeds in mobile terminals in Cihangir (2014). This time, only two antenna elements were used with a dual-feed configuration, one antenna to cover 700–960 MHz (larger antenna in Fig. 58) and the other (smaller one) to cover 1.7–2.7 GHz. The matching network used at the feed of the low-band antenna has a low-pass topology, and similarly the matching network at the feed of the high-band antenna has a high-pass topology. This configuration helps achieve higher isolation between the two antennas, although they are physically very close to each other. As seen in Fig. 59, the target bands can be covered with a reflection coefficient below −6 dB with a high isolation of more than 23 dB through the whole target band.

From the antenna design perspective, the main advantage of using multiple-feed antennas is the relaxed constraint on the bandwidth. Instead of targeting dual-band or triple-band behavior from a single antenna (meaning wider frequency band coverage), a single antenna element in a multifeed system shall only cover a smaller portion of the frequency band. This will ease the design of the antenna, help miniaturization, and also decrease the matching network complexity and hence the losses associated to them. Of course the main drawback is to get acceptable isolation between the multiple elements, especially at lower frequencies, using both antenna placement which complicates the antenna design and using band-reject matching networks at the feeds which complicates the MN design.

DOI 10.1007/978-981-4560-75-7_100-1

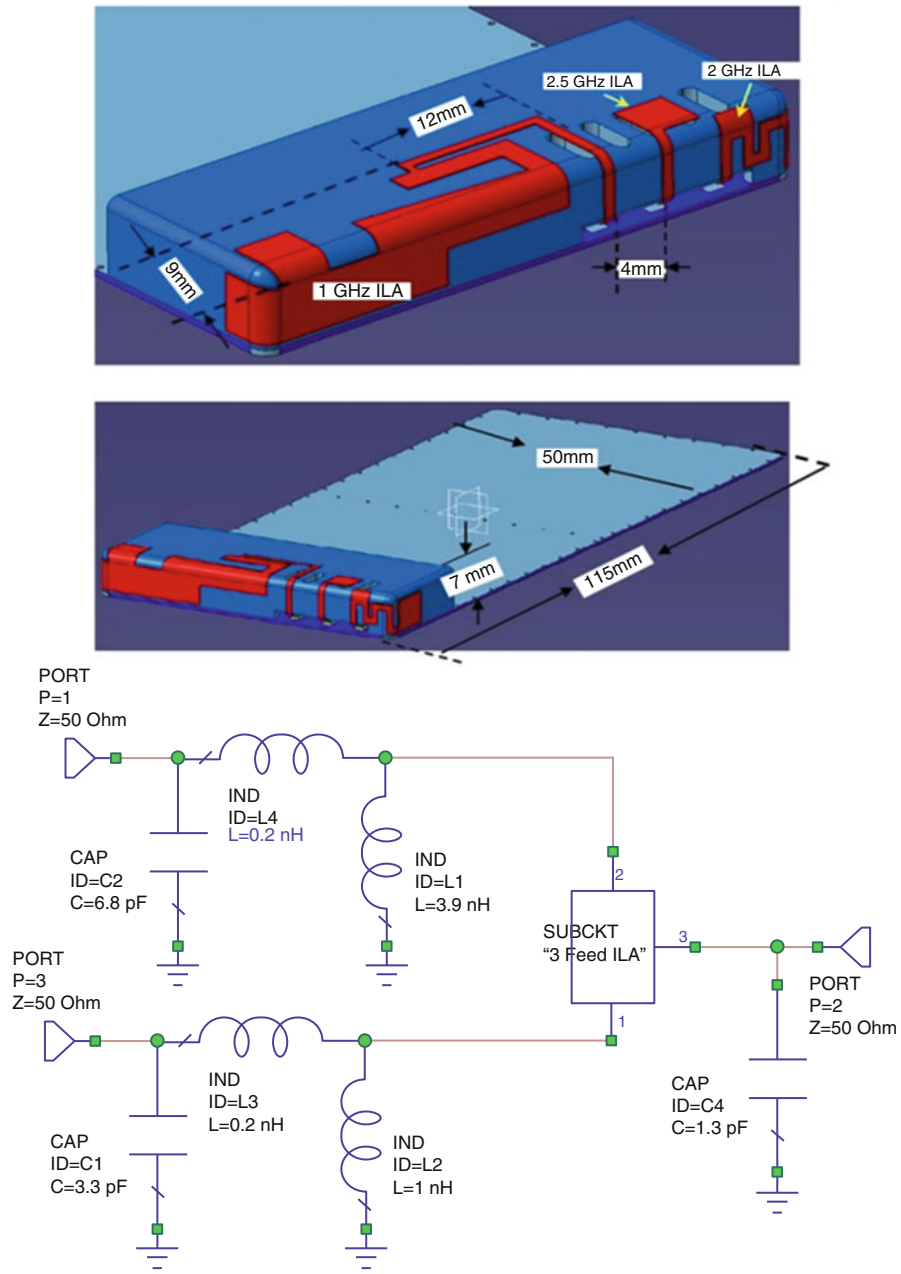


Fig. 56 Multifeed antenna system with three elements from Ikonen et al. (2012)

Handheld Device Antenna Measurements

This section includes some topics related to the measurement of mobile terminal antennas that have not been mentioned in the chapter before but deserve being mentioned. First, the effects of the feed cable on the S-parameter and efficiency measurements are discussed. The Wheeler Cap method which is a simple and easy method the measure the antenna efficiency is then mentioned. Finally, the definitions of TRP (total radiated power) and TIS (total *sotropic* power) are given with basic information about how they can be measured.

DOI 10.1007/978-981-4560-75-7_100-1

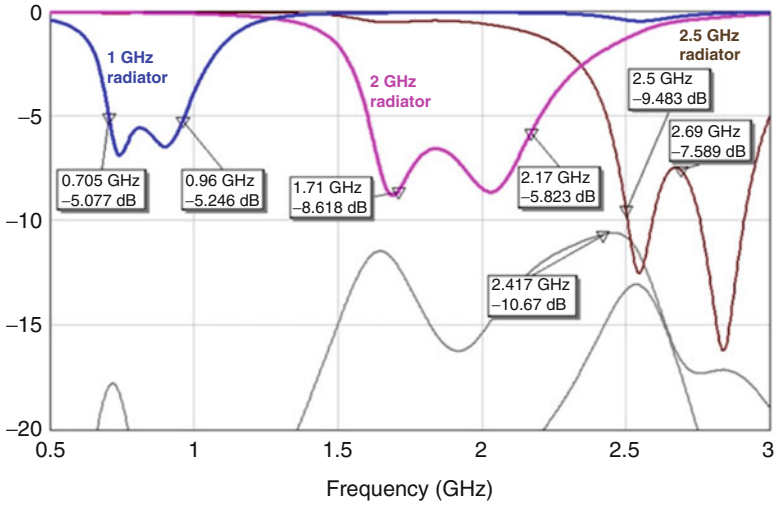


Fig. 57 S-parameters of the three-element antenna system presented in Fig. 56 from Ikonen et al. (2012)

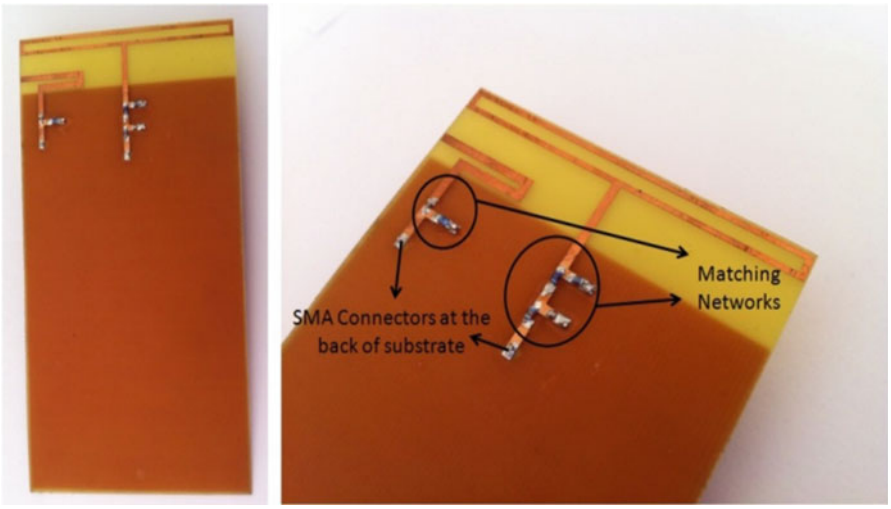


Fig. 58 Multifeed antenna system with two elements from Cihangir (2014)

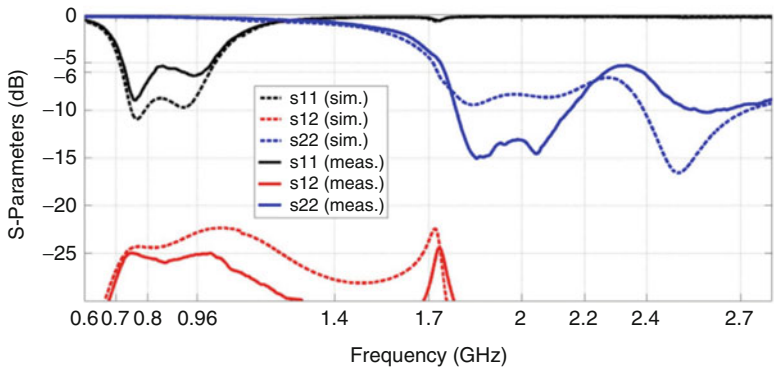


Fig. 59 S-parameters of the two-element antenna system from Cihangir (2014)

DOI 10.1007/978-981-4560-75-7_100-1

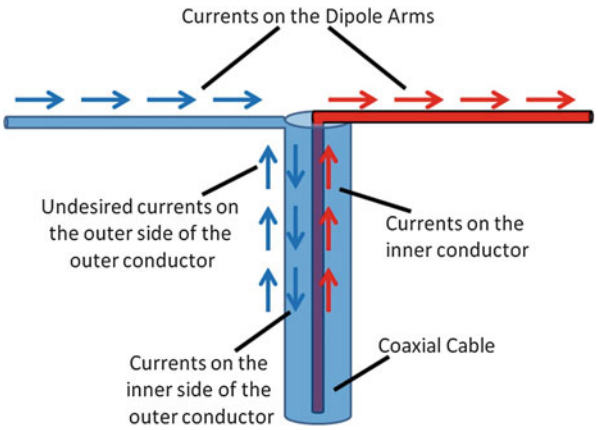


Fig. 60 Currents flowing on the feed cable of a dipole antenna

Effect of the Feeding Cable on Measurements

In most of the electromagnetic simulation tools where the mobile phone antennas are modeled and simulated, the antenna feed is modeled as a voltage difference between the feed point and the ground plane. However, in the manufactured prototype, generally an SMA connector is integrated to the prototype and the antenna fed through the inner conductor of this connector. The outer conductor of the SMA is then connected to the ground plane of the PCB. This difference between the feeding schemes in simulations and measurements causes some discrepancies between simulated and measured reflection coefficients, efficiency, and patterns. This situation has been investigated also in the literature as in Liu et al. (2012) and Chen et al. (2005).

The main problem causing this difference between simulations and measurements is the currents that flow on the outer face of the outer conducting part of the coaxial cable. This undesired effect can be explained for a dipole antenna as in Fig. 60. For the inner conductor of the feed cable, the current has nowhere else to go except the dipole arm. However for the outer conductor, the currents may flow both on the inner side or the outer side. The undesired currents that flow on the outer side of the outer conductor can be excited in two ways for mobile terminal antennas. First is due to the currents that are excited on the ground plane of the PCB and flow through the outer side of the cable. The other possible way is by the fields radiated from the antenna that excite surface currents on the cable without direct connection.

These currents can significantly change the input impedance of the antenna, causing errors in the measured reflection coefficient depending on the cable length. They also affect the antenna radiation pattern (and total efficiency), by creating ripples and nulls that normally should not occur. This phenomenon is more or less pronounced depending on the antenna concept which is measured, and it is very hard to predict if the undesired currents are going to be strongly excited or not before any measurement. A good trick to have an idea if a lot of currents are flowing on the outer part of the cable is to simply move your finger along the feeding cable during a VNA measurement and check if the reflection coefficient is strongly affected or not during these moves.

Some methods to minimize the disadvantages of this situation were also proposed. The first solution is to use quarter-wavelength long sleeve baluns on the outer face of the outer conductor. The balun is short circuited to the outer conductor on the balun's end so it presents an open circuit in the end of the cable owing to quarter-wavelength length. However, this technique blocks the currents in a limited bandwidth since the electrical length of the balun will change with frequency. Another alternative is to use EMI suppressant materials or place some ferrites on the outside of the cable. The ferrites operate well up to

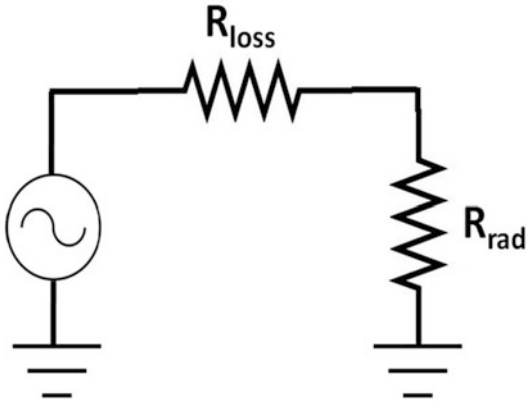


Fig. 61 Loss mechanism in the radiation of an antenna for Wheeler cap method

1 GHz, but their effectiveness is limited in the higher bands. These two solutions (EMI suppressant and ferrites) both suffer from reducing the efficiency of the antenna, since lossy materials are placed in its near field.

Wheeler Cap Method

The Wheeler cap method is the easiest and fastest way to measure the total efficiency of an antenna (Wheeler 1947). This method consists of placing the antenna under test inside a metallic cap (which can be spherical or cylindrical) and measuring the antenna reflection coefficient in free space and with this cap. If the metallic cap that encapsulates the antenna is well connected to the ground plane, it should reflect all the radiation from the antenna back to it, increasing the reflection coefficient. Figure 61 shows the radiation and loss mechanisms in a simple circuit representation. The first resistance R_{loss} stands for the radiation losses of the antenna at resonance which decreases its efficiency. R_{rad} is the radiation resistance of the antenna.

Hence the antenna efficiency (radiation efficiency) can be written as Eq. 7.

$$\eta = \frac{R_{rad}}{R_{rad} + R_{loss}} \tag{7}$$

When the reflection coefficient is measured with the antenna enclosed by the metallic cap, all the radiation will be reflected back and so $R_{rad} = 0$. Calculating R_{loss} from this measurement and R_{rad} from free space reflection coefficient, the antenna efficiency can be calculated without the need of any anechoic chamber or complex equipment. However, this method only gives information about the total efficiency and can't predict about the radiation pattern of the antenna. It can only be achieved in the prototype stage of the antenna. Several methods exist to extract the radiation efficiency from those two measurements. To the best of the authors' knowledge, the McKinzie method is a very accurate and reliable method (McKinzie 1997).

TRP/TIS

TRP (total radiated power) and TIS (otal isotropic sensitivity) are two metrics used to evaluate the radiation characteristics of a mobile phone when hardware is available. Indeed, the manufacturers need to evaluate compliance of their models with the standard values before releasing a phone on the market. Among these two definitions, TRP is related to the transmit characteristics of the phone, whereas TIS is related to receive characteristics.

TRP is the measure of the radiated power, when the transmitter, antenna, power amplifier, and their interconnections are in final state as the commercial phone. TRP measurements are active measurements,

DOI 10.1007/978-981-4560-75-7_100-1

where all the phone components shall run almost in final configuration, powered by the battery. For calculating the TRP, the EIRP (effective isotropic radiated power) shall be measured in all spherical directions. EIRP is basically the amount of equivalent input power that an isotropic antenna would need to have the same field strength as the device under test. To calculate the TRP, the time-averaged EIRP measured in all directions shall be integrated over the sphere.

TIS can be explained as the sensitivity of the receiver averaged over a 3D sphere in the basic sense. It is the integration of the EIS (effective isotropic sensitivity) over entire measurement sphere. The EIS is measured for a signal coming from a specific direction with a specific polarization and defines the minimum power level to resolve the data with a BER (bit error rate) lower than a predefined threshold. For this measurement, the mobile phone is placed in an anechoic chamber to obtain a measurement setup free from interferences: the phone uses a test SIM card and is linked to a base-station emulator. The power output level of the chamber antenna is decreased until the BER threshold is violated. In addition, broadband test antennas, positioning equipment, base station emulator, and head phantoms are also necessary.

Antenna Manufacturing Technologies

This section very briefly discusses some of the techniques used to manufacture mobile phone antennas. It is indeed difficult to describe every antenna manufacturing technology as the fabrication techniques evolve very quickly due to cost issues.

Metal Stamping

Metal stamping is the easiest and cheapest mobile phone antenna manufacturing technique. In this technique, the desired shape is given to a sheet of metal by stamping. The stamped metal is then integrated on a plastic piece, for example, by gluing. This technique does not offer the designer a lot of design freedom due to possible error sources like the placement of the metal piece on the plastic and the tolerance fabrication of the metal sheet.

Molded Interconnect Device (MID)

MID is a process that enables plastic parts that are selectively metallized. It allows integrating both electrical and mechanical elements enabling closely packed components.

One common technique is the two-shot molding, which uses two kinds of plastic materials molded together, differently from single-shot molding. For this technique, two resins are utilized, one of which is nonplatable and the other is platable. The platable material doped with catalyst is molded to the nonplatable material. This platable material shape also defines the traces of the metallization, intended to manufacture. The metallic traces are then realized with electroless plating.

Another technique used to obtain MID structures is the Laser Direct Structuring (LDS). This technology offers a lot of design freedom enabling the efficiency utilization of the available space. A thermoplastic material, which is doped with metal additives, is activated by laser according to the metal traces intended on it. In this way, a microscopically rough surface is created, forming the nuclei for the metallization step. This laser-activated material is then put into electroless copper baths to give it the final form. Using this technology, a high degree of freedom is obtained in terms of the metallization shape. An antenna manufactured using LDS technology can be seen in Fig. 62. Some thermoplastic material can also be chosen to allow reflow soldering which enables the possibility of directly integrating the MN into the plastic piece.

DOI 10.1007/978-981-4560-75-7_100-1

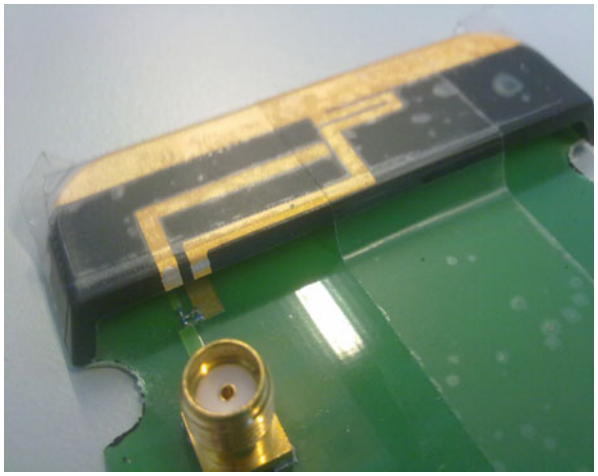


Fig. 62 An antenna realized using LDS technology from Cihangir (2014)

Future Directions and Open Problems

Antennas used in handheld mobile terminals were mentioned in this chapter.

General information about cellular communication standards and their evolution through time was summarized first. Through this evolution, some challenges arose for the antenna designer. With the announcement of each new generation, the antennas were needed to operate in broader frequency bands with generally smaller space dedicated to them.

In parallel to this evolution, the mobile terminal antennas have also undergone several changes, as explained in the second section. The initial antennas in former communication standards were external antennas, being generally of whip type, helical type, or a combination of these two. With the appearance of internal antennas in the market, there was a fast switch from external antennas to internal. Concerning this, different antenna types were investigated and used in the commercial products (some examples given in section "Examples of Commercial Antennas"), each having their own advantages and disadvantages over the other types.

As the internal antennas needed to be more broadband in a smaller size, some techniques have been used to make the antennas more broadband and miniaturized. Section "Antenna Feeding And Matching Methods" presented some of these methods. Different antenna feeding methods and the utilization of matching networks were discussed.

Active/reconfigurable antennas are also evaluated in industry and in academy to be used as mobile terminal antennas. These antennas can have some tunable components or switches at the antenna feed, as well as distributed over the antenna geometry. Some examples were given and their advantages/disadvantages over passive antennas discussed in section "Tunable/Reconfigurable Antennas in Mobile Terminals."

Section "Antenna-User Interaction in Mobile Terminals" was dedicated to the interaction between the user and the antenna of the handheld device. This interaction can occur in both ways. The user affects the antenna by detuning its reflection coefficient through dielectric loading and decreasing the total efficiency by absorption. The antenna has also effects on the user, launching electromagnetic fields on the tissues of the head, quantized by SAR.

The antenna topologies designed for use in mobile terminals can be single-feed or multifeed designs. Some multifeed design concepts with its advantages/disadvantages over single-feed antennas were mentioned in section "Multi-Feed Antenna Systems."

Section "Handheld Device Antenna Measurements" addressed some definitions and issues used in the measurements for mobile terminal antennas that were not discussed before in the chapter. One of the major problems in the reflection coefficient and efficiency measurements is the effect of the feeding cable that does not exist in a real-case scenario since the antenna is fed directly by a microstrip from the transceiver. Wheeler cap method was mentioned briefly as the simplest method to measure the efficiency. Definitions about TRP/TIS were also explained in this section.

The chapter was finalized with section "Antenna Manufacturing Technologies" mentioning the antenna manufacturing methods, commonly used in mobile terminal antennas.

Cross-References

- Antenna Design for Diversity and MIMO Application
- Broadband and Multiband Planar Antennas
- Impedance Matching and Baluns
- Radiation Efficiency Measurement of Small Antennas
- Reconfigurable Passive and Active Antennas/Tunable Antennas
- Small Antennas (PIFA/PILA/Loading Antenna/etc)

References

Akyildiz IF, Gutierrez-Estevez DM, Reyes EC (2010) The evolution to 4G cellular systems: LTE-advanced. Elsevier Phys Commun 3:217–244

Andujar A, Anguera J, Puente C (2011) Ground plane boosters as a compact antenna technology for wireless handheld devices. IEEE Trans Antennas Propag 59(5):1668–1677

Antoniades MA, Eleftheriades GV (2010) A multiband monopole antenna using a double-tuned wheeler matching network. In: Proceedings of the 4th European conference on antennas and propagation (EuCAP) 2010, Barcelona

Bhatti RA, Shin YS, Nguyen N, Park S (2008) Design of a novel multiband planar inverted-F antenna for mobile terminals. In: International workshop on antenna technology: small antennas and novel metamaterials (iWAT 2008), Chiba

Boyle K, Steeneken P (2007) A five-band reconfigurable PIFA for mobile phones. IEEE Trans Antennas Propag 55(11):3300–3309

Cabedo A, Anguera J, Picher C, Ribo M, Puente C (2009) Multiband handset antenna combining a PIFA, slots, and ground plane modes. IEEE Trans Antennas Propag 57(9):2526–2533

Chen S, Wong K (2010) Small-size 11-band LTE/WWAN/WLAN internal mobile phone antenna. Microw Opt Technol Lett 52(11):2603–2608

Chen S, Wong K (2011) Wideband monopole antenna coupled with a chip-inductor-loaded shorted strip for LTE/WWAN mobile handset. Microw Opt Technol Lett 53(6):1293–1298

Chen Z, Yang N, Guo Y, Chia M (2005) An investigation into measurement of handset antennas. IEEE Trans Instrum Meas 54(3):1100–1110

Chen W, Lee B, Liu Y (2012) A printed coupled-fed loop antenna with two chip inductors for the 4G mobile applications. Microw Opt Technol Lett 54(9):2157–2163

Chiu C, Chang C, Chi Y (2010) A meandered loop antenna for LTE/WWAN operations in a smart phone. Prog Electromagn Res C 16:147–160

Chu LJ (1948) Physical limitations of omnidirectional antennas. J Appl Phys 19(12):1163–1175

Chu F, Wong K (2011) On-board small-size printed LTE/WWAN mobile handset antenna closely integrated with system ground plane. Microw Opt Technol Lett 53(6):1336–1343

Chu F, Wong K (2012) Internal coupled-fed loop antenna integrated with notched ground plane for wireless wide area network operation in the mobile handset. Microw Opt Technol Lett 54(3):599–605

Ciais P, Staraj R, Kossiavas G, Luxey C (2004) Design of an internal quad-band antenna for mobile phones. IEEE Microwave Wireless Compon Lett 14(4):148–150

Cihangir A (2014) Antenna designs using matching circuits for 4G communicating devices. Dissertation, University of Nice-Sophia Antipolis

Cihangir A, Ferrero F, Luxey C, Jacquemod G, Brachat P (2013) A bandwidth-enhanced antenna in LDS technology for LTE700 and GSM850/900 standards. In: Proceedings of the 7th European conference on antennas and propagation (EuCAP), Chalmers, pp 2786–2789

Cihangir A, Ferrero F, Jacquemod G, Brachat P, Luxey C (2014) Integration of resonant and non-resonant antennas for coverage of 4G LTE bands in handheld terminals. In: Forum for Electromagnetic Research Methods and Application Technologies (FERMAT), 3(5). http://www.e-fermat.org/files/articles/1537fbf7ab0bf7.pdf. Last accessed 21 Apr 2015

Collin RE, Rothschild S (1964) Evaluation of antenna Q. IEEE Trans Antennas Propag 12(1):23–27

Guo Y, Ang I, Chia M (2003) Compact internal multiband antennas for mobile handsets. IEEE Antennas Wirel Propag Lett 2:143–146

Haapala P, Vainikainen P (1996) Helical antennas for multi-mode mobile phones. In: Proceedings of the 26th European microwave week, Prague, 9–12 Sept 1996

Hansen RC (1981) Fundamental limitations in antennas. Proc IEEE 69(2):170–182

Holopainen J (2008) Handheld DVB and multisystem radio antennas. Dissertation, Helsinki University of Technology. https://aaltodoc.aalto.fi/handle/123456789/44. Last accessed 21 Apr 2015

Holopainen J (2011) Compact UHF-band antennas for mobile terminals: focus on modelling, implementation, and user interaction. Dissertation, AALTO University. http://lib.tkk.fi/Diss/2011/isbn9789526040868/. Last accessed 21 Apr 2015

Holopainen J, Valkonen R, Kivekas O, Ilvonen J, Vainikainen P (2010) Broadband equivalent circuit model for capacitive coupling element–based mobile terminal antenna. IEEE Antennas Wirel Propag Lett 9:716–719

Hu ZH, Kelly J, Song C, Hall PS, Gardner P (2010) Novel wide tunable dual-band reconfigurable chassis-antenna for future mobile terminals. In: Proceedings of the 4th European conference on antennas and propagation (EuCAP), Barcelona

IEC (2005) Human exposure to radio frequency fields from hand-held and body- mounted wireless communication devices – human models, instrumentation, and procedures – part 1: procedure to determine the specific absorption rate (SAR) for hand-held devices used in close proximity to the ear (Frequency range of 300 MHz to 3 GHz).IEC 62209-1, 2005

IEEE (2003) IEEE recommended practice for determining the peak spatial-average specific absorption rate (SAR) in the human head from wireless devices: measurement techniques. IEEE Std 1528

Ikonen P, Ella J, Schmidhammer E, Tikka P, Ramachadran P, Annamaa P (2012) Multi-feed RF front-ends and cellular antennas for next generation smartphones. Available via. http://www.pulseelectronics.com/download/3755/indie_technical_article/pdf. Last accessed 02 July 2015

Ilvonen J, Valkonen R, Holopainen J, Kivekas O, Vainikainen P (2012) Reducing the interaction between user and mobile terminal antenna based on antenna shielding. In: Proceedings of the 6th European conference on antennas and propagation (EUCAP), Prague

Jaloun M, Guennoun Z (2010) Wireless mobile evolution to 4G network. Wirel Sens Netw 2:309–317

Korva H (2007) Internal multiband antenna. US Patent 7,256,743, 14 Aug 2007

Kumar A, Liu Y, Sengupta J, Divya (2010) Evolution of mobile wireless communication networks: 1G to 4G. Int J Electron Commun Technol 1(1):68–72

Lee C, Wong K (2010) Planar monopole with a coupling feed and an inductive shorting strip for LTE/GSM/UMTS operation in the mobile phone. IEEE Trans Antennas Propag 58(7):2479–2483

Lin C, Wong K (2007) Printed monopole slot antenna for internal multiband mobile phone antenna. IEEE Trans Antennas Propag 55(12):3690–3697

Liu C, Lin Y, Liang C, Pan S, Chen H (2010) Miniature internal penta-band monopole antenna for mobile phones. IEEE Trans Antennas Propag 58(3):1008–1011

Liu L, Cheung S, Weng Y, Yuk T (2012) Cable effects on measuring small planar UWB monopole antennas. In: Ultra wideband – current status and future Trends. InTech

Mak A, Rowell C, Murch R, Mak C (2007) Reconfigurable multiband antenna designs for wireless communication devices. IEEE Trans Antennas Propag 55(7):1919–1928

Manteuffel D, Arnold M (2008) Considerations for reconfigurable multi-standard antennas for mobile terminals. In: International workshop on antenna technology: small antennas and novel metamaterials (iWAT), Chiba, pp 231–234

McKinzie III, WE (1997) A modified wheeler cap method for measuring antenna efficiency. In: Proceedings of the IEEE antennas and propagation society international symposium (AP-S), Montreal

McLean JS (1996) A re-examination of the fundamental limits on the radiation Q of electrically small antennas. IEEE Trans Antennas Propag 44(5):672

Nevermann P, Pan S (2002) Multiband helical antenna. US Patent 6,501,438 B2, 31 Dec 2002

Park Y, Sung Y (2012) A reconfigurable antenna for quad-band mobile handset applications. IEEE Trans Antennas Propag 60(6):3003–3006

Pelosi M, Franek O, Pedersen G, Knudsen M (2009) User's impact on PIFA antennas in mobile phones. In: Proceedings of the IEEE 69th vehicular technology conference, Barcelona

Pues HF, Capelle AR (1989) An impedance-matching technique for increasing the bandwidth of microstrip antennas. IEEE Trans Antennas Propag 37(11):1345–1354

Ramachandran P, Annamaa P, Gaddi R, Tornatta P, Morrel L, Schepens C (2013) Reconfigurable small antenna for mobile phone using MEMS tunable capacitor. In: Loughborough antennas & propagation conference, Loughborough

Rowell C, Lam EY (2012) Mobile-phone antenna design. IEEE Antennas Propag Mag 54(4):14–34

Rowell C, Murch RD (1997) A capacitively loaded PIFA for compact mobile telephone handsets. IEEE Trans Antennas Propag 45(5):837–842

Saldell U (1997) Antenna device for portable equipment. US Patent 5,661,495, 26 Aug 1997

Selvanayagam M, Eleftheriades GV (2010) A compact printed antenna with an embedded double-tuned metamaterial matching network. IEEE Trans Antennas Propag 58(7):2354–2361

Toh CK (2011) 4G LTE technologies: system concepts, technology, white paper. http://www.alicosystems.com/4G%20LTE%20Technologies%20System%20Concepts.pdf. Last accessed 21 Apr 2015

Vainikainen P, Ollikainen J, Kivekas O, Kelander I (2000) Performance analysis of small antennas mounted on mobile handsets. In: Proceedings of the COST 259 final workshop-mobile and human body interaction, Bergen

Vainikainen P, Ollikainen J, Kivekas O, Kelander I (2002) Resonator-based analysis of the combination of mobile handset antenna and chassis. IEEE Trans Antennas Propag 50(10):1433–1444

Vainikainen P, Ollikainen J, Kivekas O, Kelander I (2004) Modular coupling structure for a radio device and a portable radio device. FI Patent 114260, 15 Sep 2004

Valkonen R, Holopainen J, Icheln C, Vainikainen P (2007) Broadband tuning of mobile terminal antennas. In: Proceedings of the 2nd European conference on antennas and propagation (EUCAP), Edinburgh

Valkonen R, luxey C, Holopainen J, Icheln C, Vainikainen P (2010) Frequency-reconfigurable mobile terminal antenna with MEMS switches. In: Proceedings of the 4th European conference on antennas and propagation (EuCAP), Barcelona, pp 1–5

Valkonen R, Ilvonen J, Vainikainen P (2012) Naturally non-selective handset antennas with good robustness against impedance mistuning. In: Proceedings of the 6th European conference on antennas and propagation (EUCAP), Prague, pp 796–800

Valkonen R, Ilvonen J, Icheln C, Vainikainen P (2013a) Inherently non-resonant multi-band mobile terminal antenna. IET Electron Lett 49(11):11–13

Valkonen R, Lehtovuori A, Icheln C (2013b) Dual-feed, single-CCE antenna facilitating inter-band carrier aggregation in LTE-A handsets. In: Proceedings of the 7th European conference on antennas and propagation (EuCAP), Chalmers

Villanen J, Ollikainen J, Kivekas O, Vainikainen P (2006) Coupling element based mobile terminal antenna structures. IEEE Trans Antennas Propag 54(7):2142–2153

Villanen J, Icheln C, Vainikainen P (2007a) A coupling element-based quad-band antenna structure for mobile terminals. Microw Opt Technol Lett 49(6):1277–1282

Villanen J, Mikkola M, Icheln C, Vainikainen P (2007b) Radiation characteristics of antenna structures in clamshell-type phones in wide frequency range. In: IEEE 65th vehicular technology conference, Montreal

Villanueva R, Miranda R, Mendez J, Aguilar H (2013) Ultra-wideband planar inverted-F antenna (PIFA) for mobile phone frequencies and ultra-wideband applications. Prog Electromagn Res C 43:109–120

Wheeler HA (1947) Fundamental limits of small antennas. Proc IRE 35(12):1479–1484

Wong K, Chen S (2010) Printed single-strip monopole using a chip inductor for penta-band WWAN operation in the mobile phone. IEEE Trans Antennas Propag 58(3):1011–1014

Yaghjian AD, Best SR (2005) Impedance, bandwidth, and Q of antennas. IEEE Trans Antennas Propag 53(4):1298–1324

Yamagajo T, Koga Y (2011) Frequency reconfigurable antenna with MEMS switches for mobile terminals. In: IEEE-APS topical conference on antennas and propagation in wireless communications (APWC), Torino

Yang C, Jung Y, Jung C (2011) Octaband internal antenna for 4G mobile handset. IEEE Antennas Wirel Propag Lett 10:817–819

Ying Z (2000) Multi-band non-uniform helical antennas. US Patent 6,112,102, 29 Aug 2000

Ying Z (2005) Progress of multi-band antenna technology in mobile phone industry. In: IET conference on wideband and multi-band antennas and arrays, Birmingham

Ying Z (2012) Antennas in cellular phones for mobile communications. Proc IEEE 100(7):2286–2296

Zaid L, Kossiavas G, Dauvignac JY, Papiernik A (1998) Very compact double C-patch antenna. IET Electron Lett 34(10):933–934

Zhang L (2013a) Quad-band internal antenna and mobile communication terminal thereof. US Patent 2013/0135155 A1, 30 may 2013

Zhang L (2013b) Penta-band internal antenna and mobile communication terminal thereof. US Patent 2013/0141298 A1, 6 Jun 2013

Zhang X, Zhao A (2009) Enhanced-bandwidth PIFA antenna with a slot on ground plane. In: PIERS proceedings, Beijing, 23–27 Mar 2009. https://piers.org/piersproceedings/download.php?file=cGllcnMyMDA5QmVpamluZ3wzUDNiXzEyNjgucGRmfDA4MDkyODA5NTkzNg. Last accessed 02 July 2015

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_101-1

Satellite Antennas on Vehicles

Stefan Lindenmeier* and Simon Senega
Institute of High Frequency Technology and Mobile Communication, Universität der Bundeswehr München, Neubiberg, Germany

Abstract

The mobile reception of satellite services on vehicles places high demands on the antennas in many regards. Due to the high path loss because of the great distances, low signal levels are experienced on the ground.

The following chapter gives an overview on antennas which can be used for the mobile reception on vehicles. The main areas of application in this regard are systems for global positioning and for satellite radio services. At first an overview of the requirements on the antennas imposed by the different services is given. Thereafter some basic antenna types are discussed regarding their advantages and disadvantages as far as the reception of satellite services are concerned including dipole and ring structures. More advanced antenna designs are also presented which are specifically optimized for different satellite systems.

In reception scenarios with severe signal impairments like multipath propagation resulting in deep signal fades, a single antenna is not sufficient for satellite reception. The mechanisms which lead to these scenarios are shortly introduced followed by a discussion of antenna diversity techniques which are an effective means to reduce these impairments. Special consideration is given to scan-phase diversity which efficiently combines the advantages of a simple system design with high signal quality improvements. Measurements obtained in real fading scenarios are presented for single antenna as well as scan-phase diversity systems. They show that antenna diversity can significantly improve the audio availability in adverse reception scenarios compared to single antenna systems. Furthermore, diversity can even allow for using antenna mounting positions which are unsuitable for single antennas like the dashboard or single side mirrors while still outperforming a rooftop mounted standard antenna.

Keywords

GNSS; Satellite radio; Ring antenna; Dipole; Scarabeus antenna; Multipath fading; Antenna diversity

Introduction

In this chapter an overview on car antennas and antenna diversity technologies for mobile reception of satellite radio signals is given. The special requirements for the design of satellite reception antennas on a vehicle are discussed in general and in detail for the very common satellite services for navigation and radio broadcast. Basic antenna types are discussed in comparison with each other with respect to meeting the requirements for mobile satellite radio reception on vehicles. Together with the well-known common structures, for example, patch antennas, new ring types of receiving antennas are explained leading to an improved reception quality and enabling a lean efficient design. Applications are shown for the examples

*Email: stefan.lindenmeier@unibw.de

DOI 10.1007/978-981-4560-75-7_101-1

of navigation and satellite radio services. Furthermore, ways of optimizing the reception quality of satellite radio significantly via antenna diversity are shown. This includes the discussion of appropriate diversity circuits and antenna structures which may be located together in one mounting volume which is as compact as former single antenna solutions. In real scenarios it is shown by experiment that in critical fading scenarios the number of mutes is reduced by more than one order of magnitude. The diversity technology enables the choice of new antenna mounting positions, which would not be chosen for a single antenna because of the strong decrease of reception quality. By means of diversity technology, successful reception of high quality is shown even for the use of inferior antenna positions like in side mirrors or on the dashboard.

General Requirements on Car Antennas for Satellite Reception

For satellite radio reception in cars, high demands are placed on the antennas with respect to gain, efficiency, volume requirements, and mounting conditions as well as reliable reception in critical reception scenarios like fading environment. Carrier frequencies to be considered are settled in the L-band and in the S-band. Since the wavelengths at these frequency bands are considerably shorter than those of the terrestrial radio reception services, only small antenna elements with a footprint of around 5 cm by 5 cm (2″ by 2″) and a height of less than 2.5 cm (1″) are needed. In the following, the main requirements for mobile satellite receiving antennas on vehicles are listed:

- Low mounting volume (typical footprint area: around 5 cm by 5 cm (2″ by 2″); typical height: $<$2.5 cm (1″))
- Low losses (typically 1.5 dB or less), high efficiency
- Power matching to 50 Ω
- Omnidirectional behavior with respect to azimuth
- Radiation pattern in elevation tailored to the satellite position to be expected from point of view of the receiver
- Easy fabrication and high reproducibility
- Circular polarization and high axial ratio
- Sufficient bandwidth (order of magnitude in MHz between 1 and 10 in case of navigation services and between 10 and 100 in case of satellite radio and mobile communication)
- Mounting position of low influence on the radiation pattern by the car's structure
- Low influence on the antenna performance when placing antennas for other services in close proximity

The angle ranges, in which certain gain values are required, differ very much with the satellite system. At LEO (low Earth orbit) and MEO (medium Earth orbit) systems, there is only a low height of the satellite above the sea level in comparison with the orbit radius, which means that the direction, in which the satellite might be expected from point of view of the receiving antenna, could be nearly arbitrary. With that, the radiation pattern of such systems should cover the complete upper half-space. In contrast in systems using a geostationary orbit (GEO) satellite, this satellite can be expected most probably in a range of diagonal elevation angles from point of view of the receiving antenna. HEO satellite systems use satellites where the transmission is active while its direction is of high elevation angle above the horizon. This means that the radiation pattern for such systems should have high gain values around the zenith. Additionally, there is often the requirement for reception of signals in very low elevation since terrestrial repeaters often support satellite services in urban regions where shadowing of high buildings might harm the satellite reception. As very common examples for satellite services are using orbits of MEO, GEO, and

DOI 10.1007/978-981-4560-75-7_101-1

HEO type, the following sections focus more closely on the requirements of navigation services and especially on the different aspects of requirements for satellite digital audio radio services (SDARS).

Antenna Requirements for Mobile Reception of Navigation Signals

For navigation, mobile satellite services such as GPS, GLONASS, and GALILEO are nowadays part of a typical equipment of vehicles for civil and military use. For these services the L-band is used with a center frequency of 1.575 GHz for GPS and Galileo, 1.65 GHz for GLONASS, and additional possible center frequencies at 1.23 GHz and 1.18 GHz for GPS as well as 1.19 GHz and 1.28 GHz for Galileo, each with a bandwidth of less than 10 MHz, typically 2–5 MHz (Hegarty and Chatre 2008).

Since the satellite orbits of the navigation systems are MEOs, the angle range of reception antennas for positioning services is very wide between zenith with theta = 0° and low elevation over ground with theta = 70°. In this angle range a constant gain is required. Often, a requirement for a certain axial ratio is reported. In some cases an axial ratio of less than 6 dB helps in fact to prevent first-order reflections of the radio waves in the environment since with the reflection the polarization is inverted. The problem is that for low elevation angles above the ground, only a high axial ratio is possible since in the horizontal plane above a very well conducting ground, an electromagnetic wave can only have a vertical polarization. In vertical direction a first-order reflection is only possible on the ground. In such a case it is more of interest how large the back lobe of the antenna is after being mounted onto a car. So the only case where a low axial ratio helps to prevent interference by reflections at high buildings is an elevation angle around 45°.

For these requirements, cheap patch antennas with a footprint area of less than 2.5 cm by 2.5 cm (1″ by 1″) have been commonly used. Nowadays other structures are considered too which also ensure easy fabrication combined with high efficiency and well covering of the required angle ranges.

Antenna Requirements for Mobile Reception of Satellite Digital Audio Radio Systems

In the last years, satellite radio systems gained more and more importance also for broadcast reception in the USA (Davarian 2002). Also for Europe, there have been plans to establish new satellite radio services in the S-band (Reding 2007) to follow an earlier system called WorldSpace which operated in L-band (Sallam et al. 2008).

In contrast to terrestrial systems, satellite radio services use satellites to broadcast the radio signals to a huge area of reception. By using satellites for transmission in satellite radio services, the reception area is extended to cover almost the complete continental USA as well as large parts of southern Canada. The frequencies usable in such a system are limited by the available transmission power as well as transmitting antenna size (and of course regulatory restrictions). American SDARS work in a frequency range from 2,320 to 2,345 MHz with a bandwidth of 25 MHz (Briskman and Prevaux 2004; Patsiokas 2001). While currently no European SDARS are commercially available, the frequency band between 2,170 and 2,200 MHz has been dedicated to this purpose (Reding 2007) and was auctioned to two buyers (Reding 2009). In contrast to the US systems, these will use a signal coding that is part of the DVB family and was standardized as DVB-SH by ETSI (ETSI 2010a, b, c).

An overview of the setup of a satellite radio system is depicted in Fig. 1. The broadcast signal is encoded in the broadcast center using all desired audio and data channels as well as supplementary information. After modulation and up-conversion in frequency, it is transmitted to the satellites. The satellites amplify and down-convert the signal to the desired frequency band and retransmit it after further

DOI 10.1007/978-981-4560-75-7_101-1

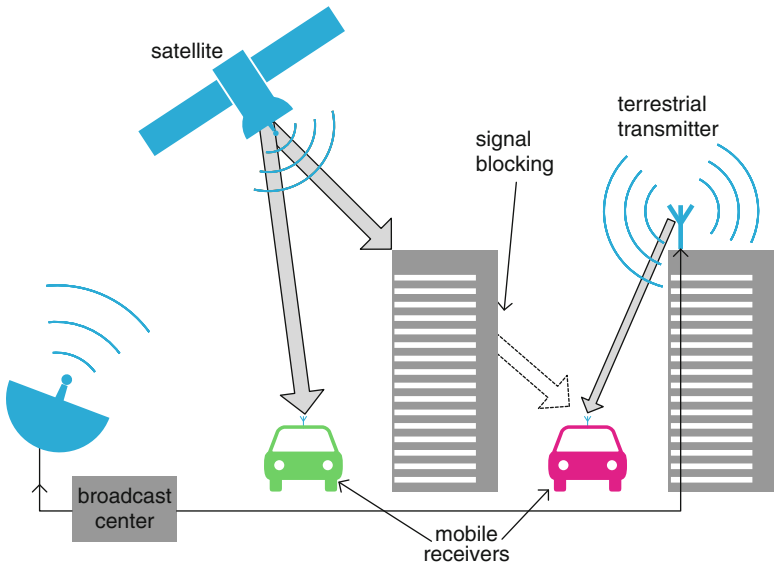


Fig. 1 SDARS system overview and signal paths (Senega 2013)

amplification (Briskman and Prevaux 2004). Mobile (and also stationary) receivers can then decode the transmitted information. In areas where a direct signal from the satellite is not available due to signal blocking – which can be the case due to natural or man-made attenuators in the signal path – supplementary terrestrial transmitters can be used. Of course these increase the overall system cost and are thus only sparsely used in areas with high numbers of service subscribers (e.g., in big cities).

Different satellite constellations are used by the (formerly two and now one) service provider(s) of US SDARS. One service started with (and still uses) two GEO satellites, while the other complemented its initial three HEO satellites with an additional GEO satellite. The HEOs' ground track – which describes the position of a satellite projected to Earth's surface over the course of one revolution – is a lemniscate. The ground tracks of three HEO and two GEO US SDARS satellites are depicted in Fig. 2. The GEO satellites always stand on the same position in the sky and therefore have an elevation angle that is almost constant unless the receiving antenna moves over great distances north or south. In most of the continental USA, the elevation angle of a GEO satellite is in the range of 30–45° above the horizon. The elevation angle of a HEO satellite strongly changes over time even for a stationary receiver. It can even have negative values when the satellite is below the horizon. Therefore at least three satellites are needed in order to ensure sufficiently high availabilities at any given time. The advantage of such a configuration is that one of the satellites is usually visible in very high elevation angles over ground of 60–90° (e.g., the red dot in Fig. 2) thus lowering the probability of a blocked signal path compared to a GEO transmitter. In Fig. 3 the elevation angles of the US SDARS satellites are plotted for a receiver located in a major US city over a time of 24 h.

The distance of the satellites to Earth is of course much higher than the distance between terrestrial transmitters and receivers. Due to propagation in the vastness of space, the free-space path loss of the transmission link gives a very convenient estimation of the attenuation that can be expected between the satellite and a receiver in line-of-sight (LOS) conditions. Free-space path loss a is calculated according to the following.

$$a = \left(\frac{\lambda_0}{4\pi d}\right)^2 \tag{1}$$

DOI 10.1007/978-981-4560-75-7_101-1

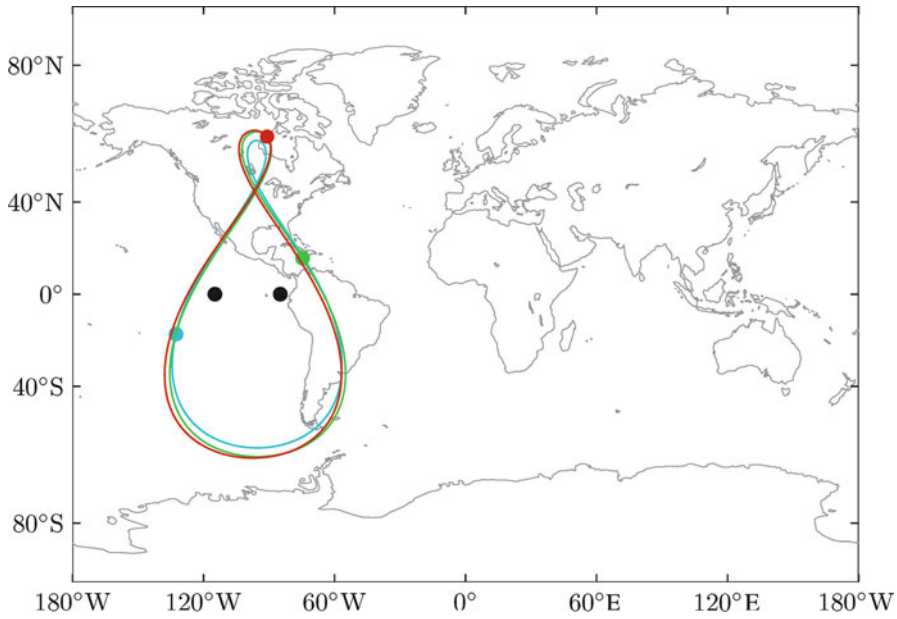


Fig. 2 Ground track of the satellites of XM Satellite Radio (GEO, *black dots*) and Sirius Satellite Radio (HEO, *blue*, *green,* and *red* lemniscate, GEO not depicted) (Müller 2010)

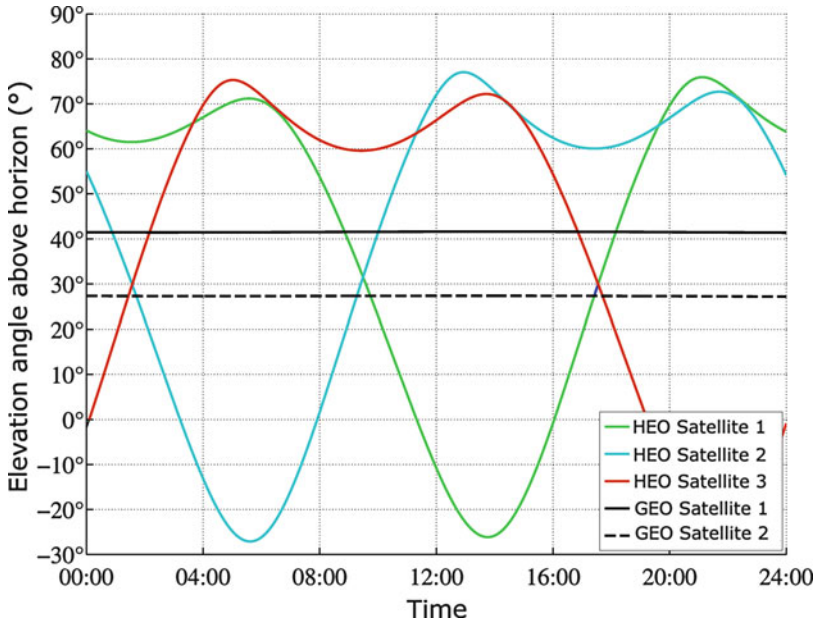


Fig. 3 Elevation angles over ground of GEO and HEO satellites over New York City, NY, USA, over a time period of 24 h (Müller 2010)

where λ_0 is the free-space wavelength and d is the distance between transmitter and receiver.

Due to the elliptical orbit, changes occur not only in the elevation angle but also in the distance of the satellite to the receiver. This means that the varying free-space path loss leads to signal levels that are dependent on the satellites position. A calculation of the free-space path loss of the signals from GEO and HEO using Eq. 1 and values given in Briskman and Prevaux (2004) leads to an attenuation in the range of −190 to −193 dB depending on orbit position as it is given in Table 1 (neglecting further attenuation due to atmosphere, weather, etc.).

DOI 10.1007/978-981-4560-75-7_101-1

Table 1 Free-space path loss from an SDARS satellite to Earth for different orbital positions

	Distance to Earth (km)	Free-space path loss (dB)
Perigee	24,469	(−187.55)
HEO 35° elevation	32,750	−190.09
Apogee	47,102	−193.24
GEO	38,192	−191.42

The modulation and digital coding of the satellite signals comprise a number of means to ensure high service availability in most reception scenarios (Briskman and Prevaux 2004). Of course any included redundant information for error correction or avoidance reduces the available number of information channels or the audio quality due to the limited frequency and power resources. Therefore the implemented measures are always a compromise between service variety, quality, and availability. While frequency, time, and spatial diversity as well as forward error correction schemes ensure high audio availabilities in many reception scenarios, they cannot avoid all signal outages, for example, in severe multipath fading scenarios.

The measure for quality of reception is the SNR at the receiver. If the SNR falls below a certain value – e.g., 5 dB – the decoding of the signal will suddenly be almost impossible due to bit errors. The exact value of the critical SNR is highly dependent on several parameters which among others include the modulation scheme, the number of bits per symbol, the method and parameters of the forward error correction, the influence of noise and disturbances on the transmission path and the receiver, etc. A critical decision during the design of communication systems is that of the link margin in typical reception scenarios. The link margin describes the typical difference between the signal level at the receiver and its sensitivity limit (the lowest signal power that it can decode quasi error free). In this regard the free-space path loss needs to be considered when the minimum transmit power of the satellites is specified. Any additional attenuation in the transmission path due to obstacles, weather, or multipath propagation will then reduce the remaining signal power until the SNR falls below the threshold, and decoding of the signal becomes impossible.

Special requirements on the radiation characteristics are made for SDARS signal reception (Haller 2001). Regarding the reception of US SDARS signals transmitted by HEO satellites, an angle range between zenith direction at theta = 0° and lower elevation at theta = 35° can be expected. In the last years transmission via GEO satellites gained more and more importance, so currently antennas are required which are appropriate for both the reception of signals from high elevation angles of HEO satellites and signals of medium elevation angles of GEO satellites. For GEO satellites the relevant angle range can be expected between low elevation at theta = 70° for northern regions and high elevation at theta = 45° for southern regions up to higher elevations at theta = 37° in a small area at the south border toward Mexico (if the receiver position's longitude is close to that of the GEO satellite). Each of the antennas still has to fulfill additional requirements for reception of terrestrial signals also at very low elevation angles. In parallel to this development, new ring type antennas have been created which can be made of panel without using lossy dielectric material. This results in a very high efficiency and hence to a strong improvement in signal to noise ratio.

Basic Antenna Structures for Mobile Satellite Signal Reception

In the following basic antenna structures are considered which are capable for the reception of satellite signals. All of these structures enable omnidirectional reception with respect to azimuth and a circular polarized field with a vertical main beam direction for HEO satellite signal reception or a diagonal main

DOI 10.1007/978-981-4560-75-7_101-1

beam direction for GEO satellite signal reception. All the structures considered fulfill the common requirement of a low antenna height in order to allow for easy integration underneath flat antenna radomes. In literature (e.g., Balanis 2005) there can be found a large number of other structures, for example, helical structures, which require a large antenna height or a large diameter with respect to the wavelength, so that normally such structures cannot be considered for integration in the small mounting volumes given in cars.

Crossed Dipole Antennas

A first very basic and well-known structure consists of crossed horizontal dipoles (Balanis 2005) which are positioned in a distance *d* above the ground plane. Figure 4 shows such a structure where the bold black lines represent the dipoles, while the gray structures represent the feeding network. If the distance *d* is one fourth of the wavelength, the reflection at the ground plane yields additional 3 dB in Gain in vertical direction, since a positive superposition of the waves propagating directly in vertical direction together with the waves which are reflected by the ground plane arises. For $\lambda/2$ dipoles the sinusoidal current distributions which are shown above the dipoles in the figure are obtained. In order to excite a circular polarized wave with vertical main beam direction, the two horizontal dipoles are connected with the feed with 90° of phase difference. This means that in one moment only the current distribution along one dipole is developed as shown in solid lines, while in one fourth of the time period, only the current distribution along other dipole is developed as shown in dashed lines.

If the branches of the dipoles were to be connected each with the same single-ended feed, it is obvious that a complicated network of phase shifters, power combiners, and baluns is necessary. In Fig. 4 an example of such a network is shown, where different phase shifters for each of the four branches yield a phase at these branches which is increasing in 90°steps in circular direction around the phase center.

For wideband applications such an antenna is advantageous since it can be expanded to a wideband crossed butterfly antenna structure as shown in Fig. 5. The wideband capabilities enable a use for several satellite services. Depending on the phase shifting and matching circuit, for different frequencies, LHCP and RHCP wave reception is possible at the same time. In Lindenmeier et al. (2001, 2002b), multifunctional antenna structures have been presented for different satellite services and terrestrial services. While for the satellite services the four vertical lines together with the horizontal elements act like a crossed butterfly antenna structure, they act in a common phase like a monopole with roof capacitor for terrestrial services.

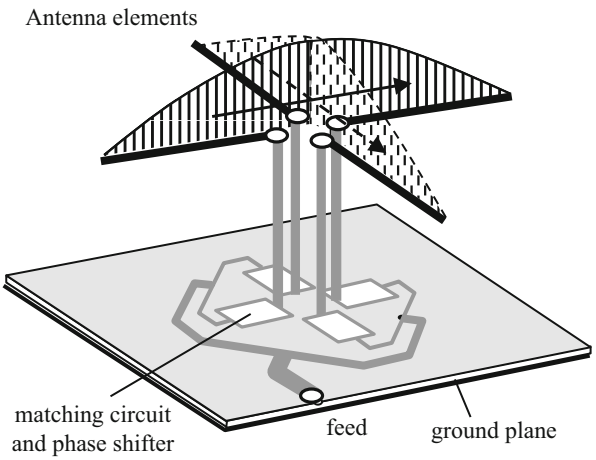


Fig. 4 Crossed dipole structure

DOI 10.1007/978-981-4560-75-7_101-1

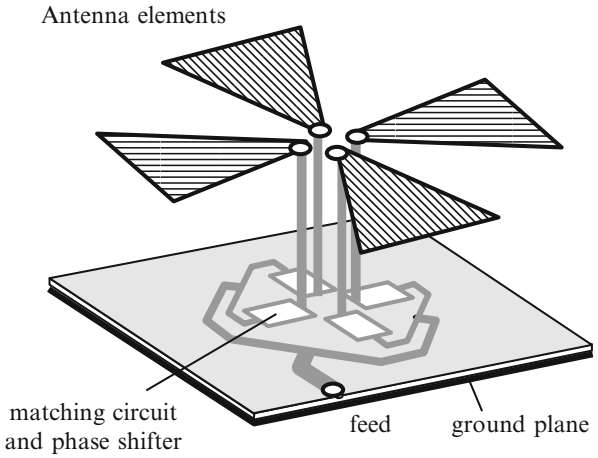


Fig. 5 Multifunctional crossed dipole structure

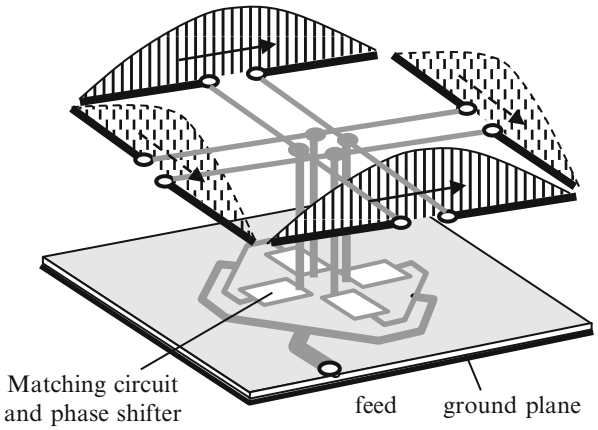


Fig. 6 Crossed pairs of dipoles

If further gain in the vertical direction is needed, the crossed horizontal dipoles can be expanded to two antenna arrays (Balanis 2005) as it is shown in Fig. 6. Off course, the up to 3 dB higher gain yields also a lower beam width which makes this structure more capable to reception of GEO satellite signals in southern regions. Since this structure is rather large, there is a strong requirement to shorten this structure. If the structure should be used for only one service, there is an unnecessary hardware effort to be paid for the feeding structure, which additionally decreases the antenna efficiency because of its circuit losses.

Patch Antennas

A very similar current distribution is occurring at the well-known patch antennas. Due to their small size, patch antennas are very common for mobile satellite signal reception and have been described in a large number of variations in literature, as, for example, in Sharma and Gupta (1983), Herscovici et al. (2003), Nasimuddin et al. (2007), and Pozar and Duffy (1997). As it is well known patch antennas represent a microstrip waveguide with an effective length of half of the wavelength for the first resonance mode. At such a short microstrip waveguide, there occurs a current distribution of a sinusoidal standing wave which has got maximum values along the opposite side edges of the microstrip line.

Patch antennas for a circularly polarized field are of roughly squared shape where the squared patch can be seen as a microstrip waveguide along two orthogonal horizontal directions. Hence, two orthogonal

DOI 10.1007/978-981-4560-75-7_101-1

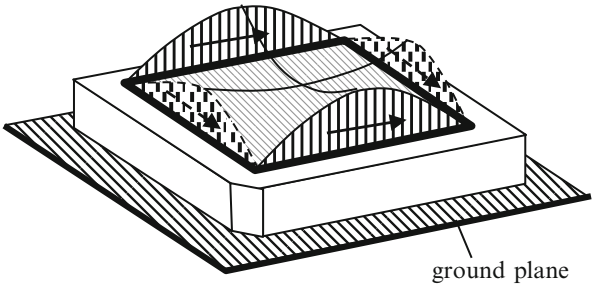


Fig. 7 Patch antenna

modes of resonance can be excited for the two orthogonal horizontal directions. In Fig. 7 a patch antenna is shown with the current distribution along the opposite edges of the patch. In between, the current distribution forms a saddle function with a minimum in the center of the patch.

If the two orthogonal resonant modes are excited with a phase distance of 90° between each other, current distributions are obtained at the edges of the patch which are similar to those of the pairs of crossed dipoles. The current distributions on opposite sides which are in phase with each other yield a strong gain in vertical direction, acting similar to a pair of λ/2 dipoles. On the other hand, the length of the edges is smaller than the length of the λ/2 dipole of Fig. 6 because of the shortening effect of the dielectric substrate. If ceramics substrate with high permittivity is used, the patch concept enables a small antenna size and hence an easy integration of the antenna in small mounting volumes.

The resonant modes can be coupled with each other via chamfered corners of the metal patch or the dielectric plate or a slightly rectangular shape together with an asymmetric feed, so that no phase shift circuit is needed and an easy fabrication is possible. This is why patch antennas are widely used for reception of GPS signals. The relative bandwidth of patch antennas, which is in a rather low range of a few percent, can be increased by the use of additional structures like stacking or insertion of slots in the patch (Herscovici et al. 2003; Nasimuddin et al. 2007; Pozar and Duffy 1997).

For the reception of satellite radio signals, patch antennas are often used either. Though the metal patch is oriented in horizontal plane, the substrate enables weak reception of a vertical field component at very low elevation which can be used for reception of signals of terrestrial repeaters.

Because of the relatively high gain in vertical direction, a patch antenna is not capable very well for reception of GEO satellite signals at the lower elevation angles. A general disadvantage of patch antennas is the loss in the dielectric material which decreases the signal to noise ratio of the received signal.

Crossed Frame Antenna

In order to obtain a radiation pattern which can be used for lower elevation angles, a crossed frame antenna has been introduced in Lindenmeier et al. (2002a). In the same way as a crossed dipole structure has been realized via phase shifters, two vertical frame antennas are crossed and excited with a phase distance of 90° to each other.

The basic principle is shown in Fig. 8 together with a sketch of the current distribution. Due to the vertical parts of the frame antennas, a considerable share of vertical polarized waves is received in horizontal plane and at low elevation angles. This enables also the reception of GEO satellite signals. Due to series capacitors being inserted in the frame antennas, the antennas are set into resonance and the radiation pattern can be adapted to special requirements. In Fig. 9 an example of such an antenna is shown together with its matching and phase-shift circuit. Like for the crossed dipole structure, the matching and phase-shift circuit required means an additional hardware effort, harming also the efficiency of the antenna. Nevertheless, due to its capability also for lower elevation angle ranges, this antenna has been used successfully for SDARS reception and in particular for GEO satellite reception.

DOI 10.1007/978-981-4560-75-7_101-1

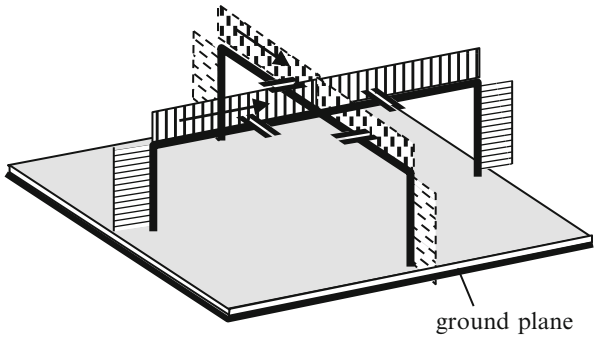


Fig. 8 Crossed frame antenna

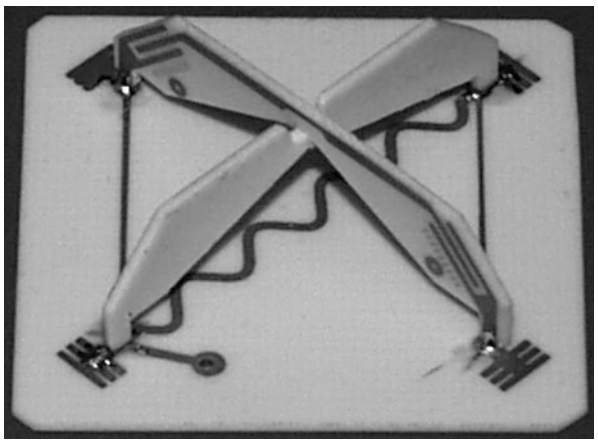


Fig. 9 Hardware example

Loop Antennas

In the following a very efficient antenna concept is shown for satellite signal reception. If a horizontal loop antenna has got a circumference equal to the wavelength, it is a resonant structure. It can be considered like an electromagnetic waveguide where the wave is guided in a circle. As long as such a wave is excited only for propagation in one direction, there will occur a current distribution which moves along the closed loop in positive direction as shown in Fig. 10. This current distribution yields a very pure circular polarization of the antenna structure. In Fig. 11 a radiation pattern for ideal conditions is shown for the left-hand and right-hand circular polarized field (LHCP and RHCP). While for the LHCP a gain of 8 dB in vertical direction is obtained, only a gain of −11 dB is occurring for RHCP.

An antenna design which is tailored to high efficiency, a wide angle range, and a small size is the "Scarabeus" antenna (Kammerer and Lindenmeier 2011), which is depicted in Fig. 12. It consists of a loop antenna which is connected to vertical elements. These vertical elements are connected to ground via capacitors. Due to the vertical elements, two advantageous effects occur: At first, the structure is even smaller than the loop structure of Fig. 10 because of a shortening effect of the capacitive vertical elements. Secondly, the current distribution along the vertical elements yields a nearly omnidirectional vertically polarized field in horizontal plane. In Fig. 13 the radiation pattern in the vertical plane is shown for the co-polarization LHCP and the cross-polarization RHCP. In Fig. 14 the radiation pattern is shown in the horizontal plane for ideal conditions.

DOI 10.1007/978-981-4560-75-7_101-1

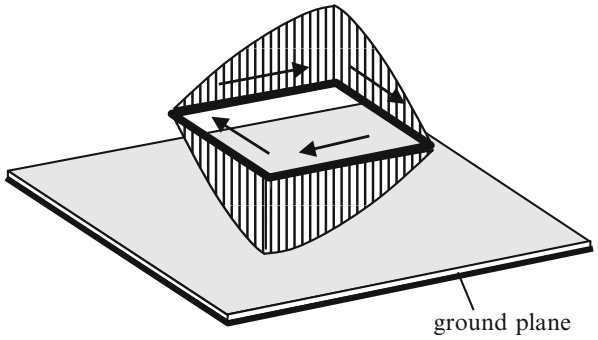


Fig. 10 Horizontal loop antenna

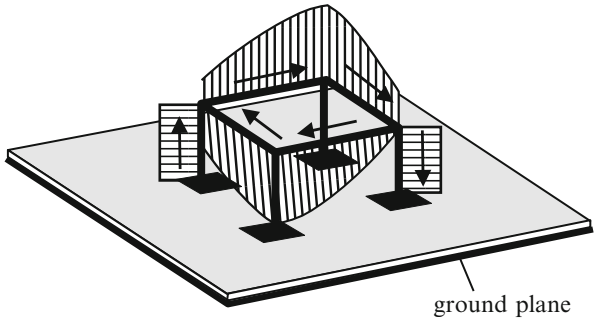


Fig. 11 Scarabeus antenna

Since manufacturing of these antenna structures is easy – the conducting structures are realized in panel and since dielectric material can be prevented – these structures have a very high efficiency with losses of only 0.5 dB. This is why even for a simple realization, as shown in hardware samples in Figs. 15 and 16, an antenna characteristic is achieved, which is very close to the ideal case, as is shown in the measured diagrams in Figs. 17 and 18.

As the radiation diagrams show, the antenna offers an omnidirectional behavior with respect to azimuth and a wide beam width in elevation of more than 90° which enables to use the concept for SDARS services with HEO and GEO satellites as well as for navigation services. Special characteristics of this antenna type are as follows:

- High efficiency and low losses of only 0.5 dB since the antenna structure can be realized by panel without the need to lead the electric field through material of high permittivity as it is necessary for patch antennas.
- Low footprint area of around 13 % of the wavelength.
- High axial ratio of more than 15 dB in zenith.
- Easy combination with terrestrial antennas which can be positioned in the phase center of the Scarabeus antenna.
- The wide bandwidth enables its use both for GPS and for GLONASS global positioning service.

Especially for situations where satellite antennas should be hidden invisibly, it might be necessary to embed the antenna into the skin of the car in a way that the antenna is not sticking out of the surrounding metal surface. This leads to an increased focus of the radiation pattern to higher elevation angles, which might harm its application for GEO satellite signal reception. In case of the Scarabeus antenna, the high

DOI 10.1007/978-981-4560-75-7_101-1

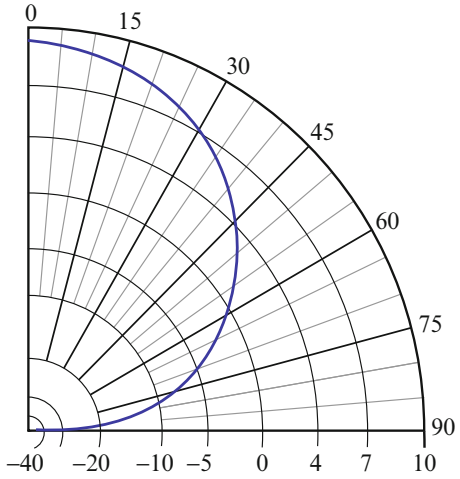


Fig. 12 Radiation pattern of loop antenna (*blue*: LHCP, RHCP not visible)

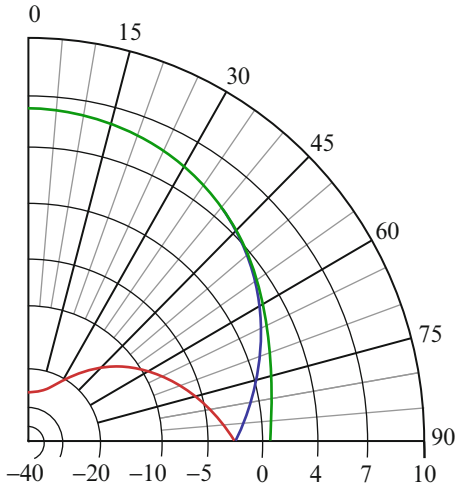


Fig. 13 Radiation pattern of Scarabeus antenna (*blue*: LHCP; *red*: RHCP, *green*: total)

efficiency nonetheless yields a sufficient reception quality even for SDARS signals as has been shown in Kammerer and Lindenmeier (2013b, a) even for an antenna combination of SDARS and GPS, hidden invisibly in a squared cavity of $40 \times 40 \times 10\ \text{mm}^3$. In numerous test drives in real reception scenarios, the Scarabeus antenna has proven a significantly larger reception quality due to its high efficiency together with an antenna characteristic which is tailored to the needs of GEO satellite signal reception while still enabling as well the reception of HEO satellite signals.

Due to its ring shape, Scarabeus antennas can be combined with each other for different services at different carrier frequencies (Kammerer and Lindenmeier 2012, 2013a). The design allows also easy combination with antennas for terrestrial signal reception as, for example, mobile phone antennas or even antennas for AM, FM, and DAB broadcast reception, which can be placed in the center of the Scarabeus antenna (Kammerer et al. 2012).

Circular Polarized Homogenous Field Loop Antenna

In the following an antenna structure will be considered which is useful for the reception of circularly polarized signals transmitted via GEO satellites (Saala et al. 2009, 2010; Saala and Lindenmeier 2010).

DOI 10.1007/978-981-4560-75-7_101-1

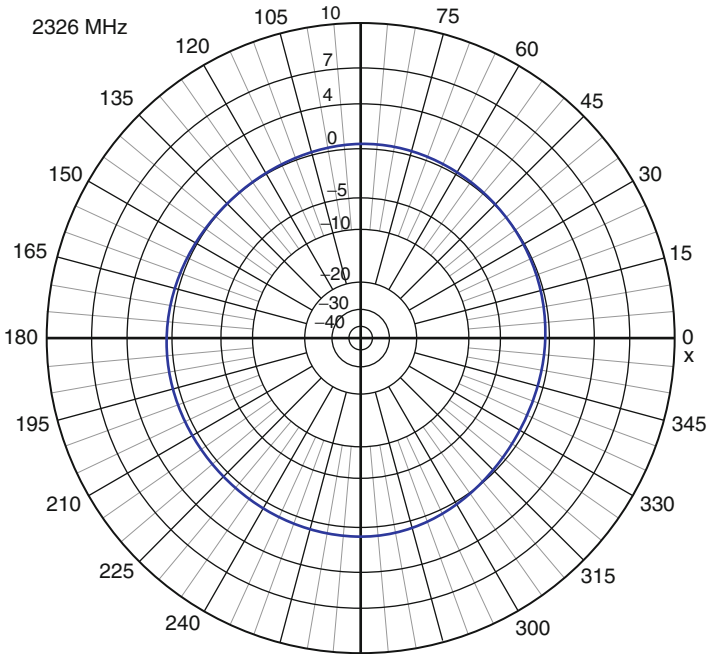


Fig. 14 Radiation pattern in the horizontal plane at 2.3 GHz (total field in vertical polarization)

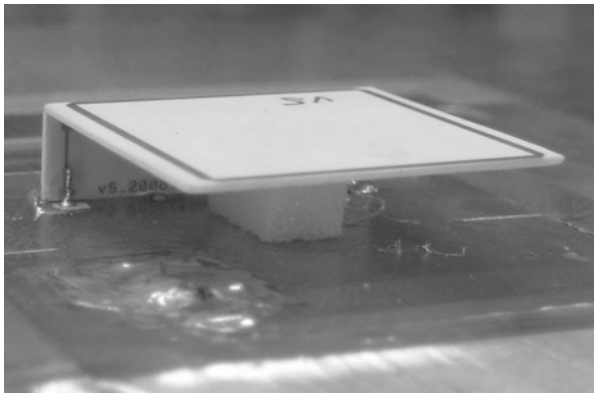


Fig. 15 Hardware sample of a loop antenna for 2.3 GHz; length and width: 34 mm; height: 10 mm

This structure consists of a ring structure which is combined with a vertically polarized monopole structure, as shown in Fig. 19.

In contrast to all the other antenna concepts shown here, the current distribution of this antenna is homogenous with respect to azimuth. While the current distribution of the monopole with its vertical direction is independent from azimuth anyway, an omnidirectional current distribution on the loop is achieved via insertion of series capacitors.

The monopole yields a vertically polarized field (theta direction) with omnidirectional characteristics and constant phase in azimuth and a horizontal main beam direction, as shown in Fig. 20a in blue. The loop antenna yields a horizontal polarized field (phi direction) also with omnidirectional characteristics and constant phase in azimuth and a main beam direction close to 45° in elevation, as shown in Fig. 20a in red.

DOI 10.1007/978-981-4560-75-7_101-1

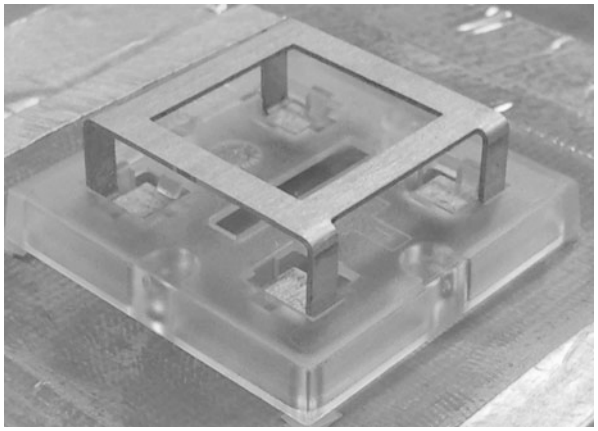


Fig. 16 Hardware sample of a Scarabeus antenna for 2.3 GHz; length and width: 16 mm; height: 10 mm

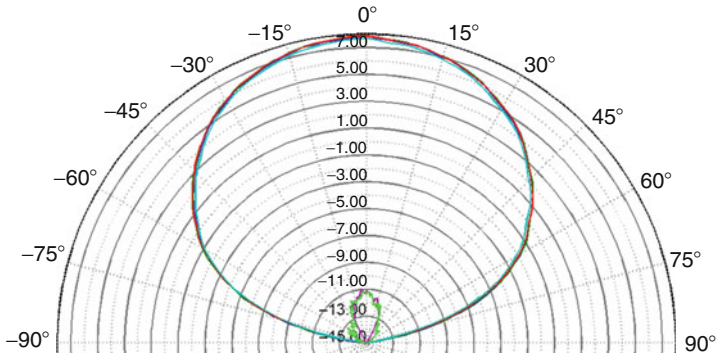


Fig. 17 Measured rad. pattern of loop antenna (LHCP and RHCP)

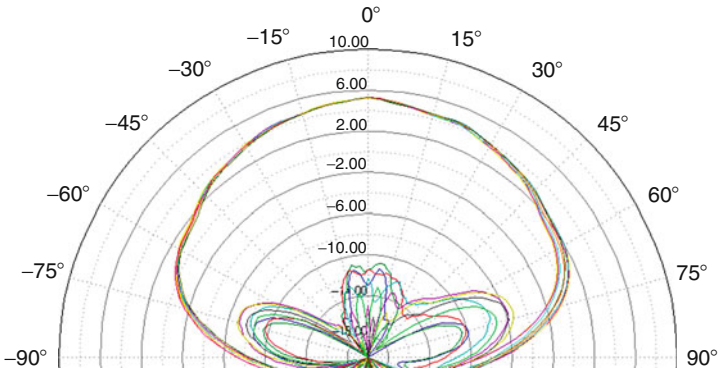


Fig. 18 Measured rad. pattern of Scarabeus (LHCP and RHCP)

Exciting the monopole and the loop antenna together with a chosen weighting and a phase distance of 90° to each other, a weighted and phased superposition of its characteristics is obtained. The resulting characteristics of a circularly polarized field are shown in the radiation patterns of Fig. 20b. The circular co-polarization (co-pol) of the complete antenna is shown in red, while the cross polarization (x-pol) is shown in blue. It can be seen that the cross-polarization ratio is dependent on the elevation angle. Depending on the power distribution between the horizontal loop and the vertical monopole, a minimum

DOI 10.1007/978-981-4560-75-7_101-1

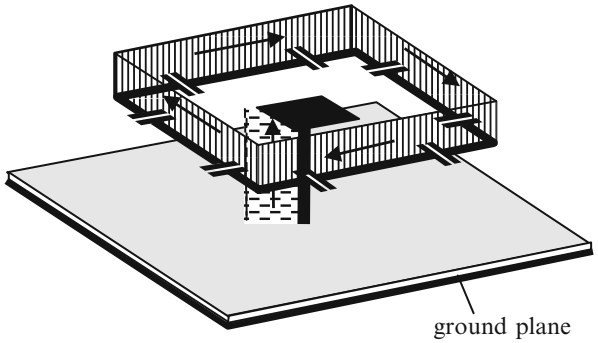


Fig. 19 Circular polarized homogenous field loop antenna

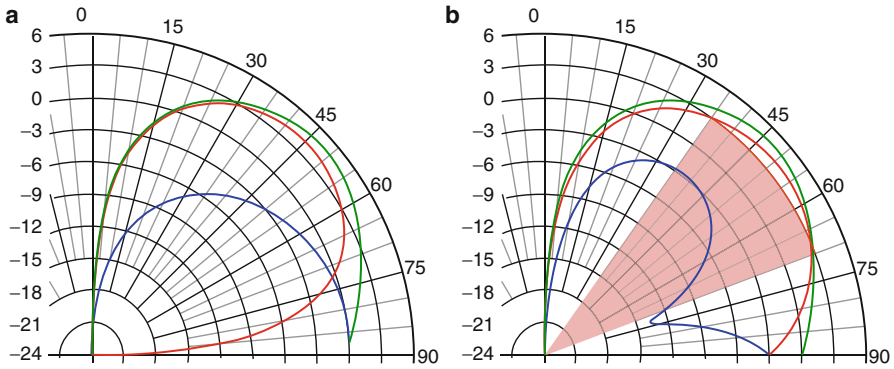


Fig. 20 Vertical radiation pattern of circularly polarized field loop antenna; (**a**) linear polarization vertical (*red*), horizontal (*blue*); (**b**) circular co-polarization (*red*), x-polarization (*blue*); total gain (*green*)

in cross polarization can be shifted to an elevation angle of choice. While in vertical direction there is no reception possible, for the circular co-polarization this antenna structure achieves an especially high gain for diagonal elevation angles around 45°.

Hence, this structure is ideally tailored to the reception of signals transmitted via GEO satellites. Within the required elevation angle range between theta = 65° and theta is 70°, a gain of more than 3 dB can be achieved in ideal conditions by the combination of those two antennas.

For real hardware structures, the efficiency is depending on the quality of the capacitors. For air-filled capacitors, which can be realized in an air slot technology, losses of less than 0.8 dB are achieved. In Fig. 21 a hardware sample of a circular polarized field loop antenna is shown. A typical measured radiation pattern is shown in Fig. 22. The 10 dB bandwidth achieved is 28 MHz and its efficiency is 1.5 dB.

According to the principle of duality, the abovementioned antennas can also be realized as slotted structures. Dual to the crossed dipole there, one can also have a crossed slot structure in the upper surface of a cavity (Balanis 2005). This realization is as well applicable to the ring structure in the form of a slot ring. Even the combination of dipoles and slots is possible as the next example will show.

Narrow Slot Dipole Antenna

In the following the combination of a slot antenna with a dipole antenna for mobile reception of circularly polarized waves is shown, as has been introduced in Müller et al. (2009). The structure is especially advantageous in mounting situations where the antenna has to be integrated into a very narrow finlike radome on the roof of a car.

DOI 10.1007/978-981-4560-75-7_101-1

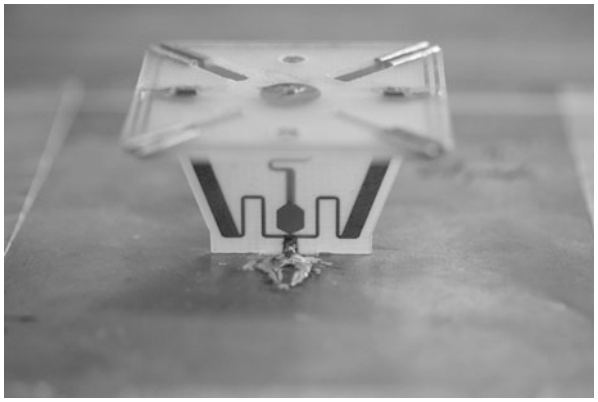


Fig. 21 Hardware sample of homogenous field loop antenna (length and width: 32 mm; height: 20 mm)

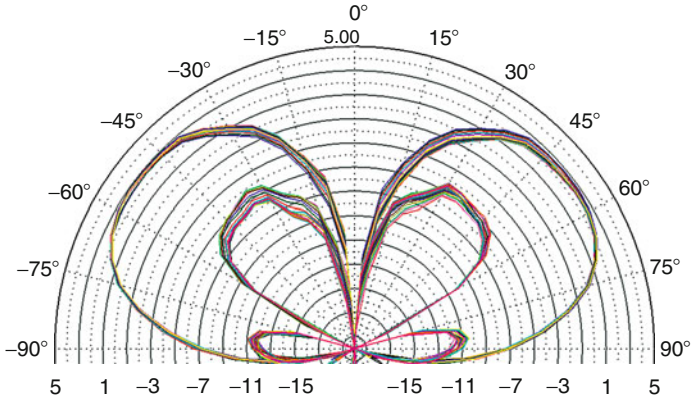


Fig. 22 Gain in dBic of the homogenous field loop antenna

The basic concept is shown in Fig. 23. Like at a crossed dipole structure, two antenna parts yield two shares of the field with linear polarization which are orthogonal to each other having a phase distance of 90° to each other. While in the common crossed dipole structure (see Fig. 4), two dipoles are contributing these linear polarizations, this is done here by a monopole and a slot antenna. While the dipole yields a current distribution corresponding to a field which is polarized in the same direction as the dipole, the slot antenna corresponds to a field with a polarization which is orthogonal to the slot direction. In order to obtain the required phase distance between the excitation of the two orthogonal field shares, the dipole is fed with that phase delay via a feed line with a length of λ/4.

In Fig. 24 a hardware sample of this antenna is shown for SDARS at 2.3 GHz with a height of around λ/4 (32 mm), a similar length of 30 mm, and a thickness of 1.5 mm. Figure 25 depicts a typical radiation pattern achieved by this structure.

For the first applications in SDARS signal reception, crossed frame antennas and patch antennas have been commonly used. Especially for the transmission via HEO satellites, patch antennas have proven to fulfill the requirements quite well. This led to the situation that the most common antennas for this application were patch antennas which have been used in different variations. For combined GEO and HEO satellite systems, loop antennas like the Scarabeus antenna are increasingly used because of its high efficiency in combination with an easy fabrication. This is also true for the positioning services where these antennas offer a wide bandwidth, which offers its use for more than one service, as, for example, GPS and GLONASS, being covered by the same antenna.

DOI 10.1007/978-981-4560-75-7_101-1

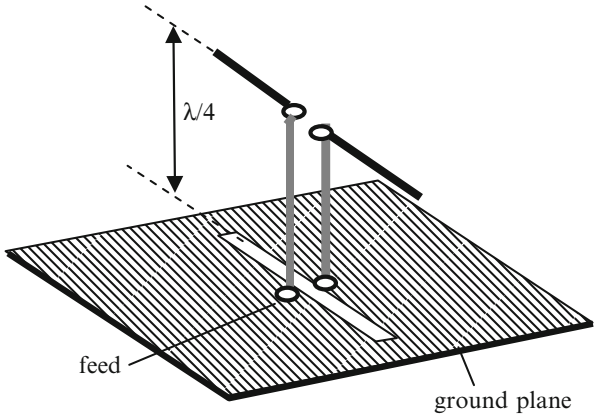


Fig. 23 Slot dipole antenna concept

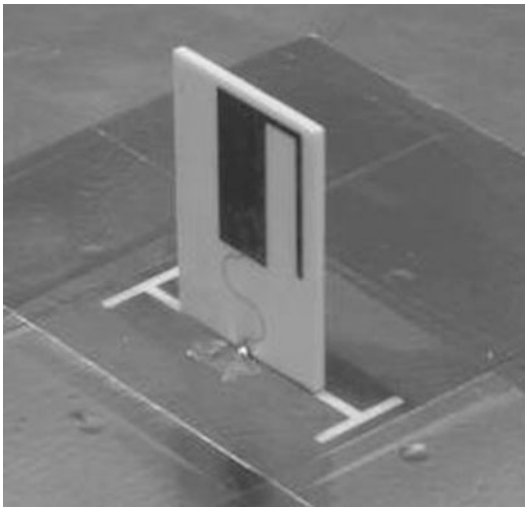


Fig. 24 Slot dipole antenna hardware sample

Environmental and Mounting Aspects, Antenna Combinations, and Multifunctional Structures

In the following the environment of satellite reception antennas is considered with respect to common integration technologies, common packaging and housing of multiple antennas and front-end parts, integration into narrow environments, and the investigation of coupling effects. In mobile applications the close integration of front-end modules and antennas in a narrow environment with a variety of possible electric and electromagnetic interferences is of high challenge. The small size of microwave antennas with low directivity enables a common integration and close packaging of antennas and circuitry for amplification of antenna signals or signal processing of multiple antenna signals in antenna diversity and smart antenna systems. In Fig. 26 a typical antenna combination is shown, in which terrestrial antennas for AM and FM broadcast and cell phone are combined with satellite antennas for SDARS and GPS signal reception underneath one common roof radome.

Coupling effects are prevented by inserting band-stop filters into a cell phone antenna. If terrestrial antennas can be positioned in the phase center of a satellite antenna, the modes of oscillation are

DOI 10.1007/978-981-4560-75-7_101-1

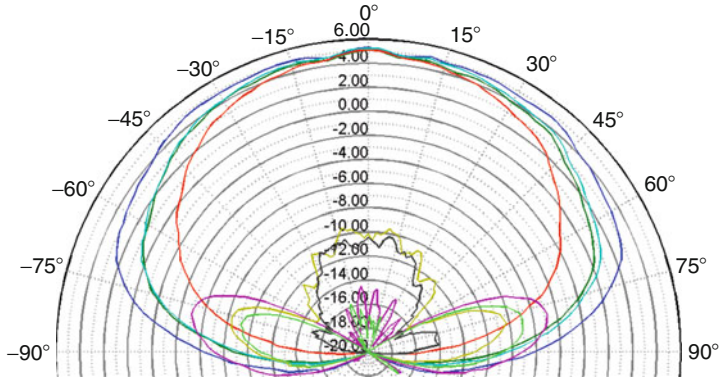


Fig. 25 Measured radiation pattern of the slot dipole antenna

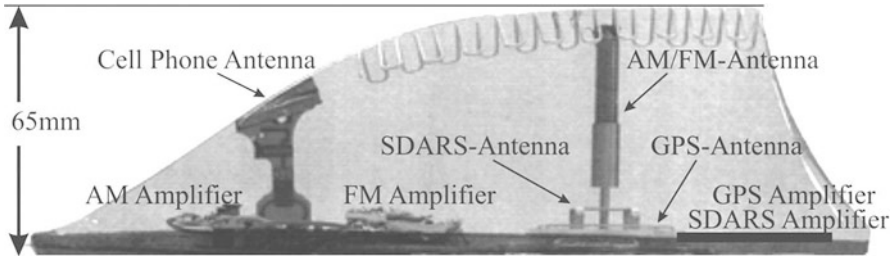


Fig. 26 Antenna combination of satellite antennas for SDARS and GPS with terrestrial antennas for AM, FM broadcast, and phone underneath one common roof radome (Kammerer et al. 2012)

decoupled from each other. This is possible with the Scarabeus loop antenna. As has been shown in Kammerer et al. (2012) even for the combination of Fig. 26, the radiation characteristics of two Scarabeus antennas – one for SDARS and one for GPS – whose phase centers are coincident with the phase center of a large AM-FM antenna are not harmed.

If a satellite antenna is integrated into plastic parts where no large ground plane is available and only a small plate can be added with a diameter of 7.5–10 cm (3″ to 4″), a ripple will occur in the vertical radiation diagram, as can be seen for different examples in Fig. 27. While the dashed line shows the diagram for a large ground plane in the solid blue lines, a superimposed ripple is observed whose envelope is mainly depending on the diameter of the ground plate. The number of ripples is depending on the height of the antenna position over ground.

In Fig. 28 a result is shown for a small ground plate with diameter of only 65 mm with a high ripple, leading to deviations from required gain by around 3–4 dB. Since the demands on the antenna gain are very close to that which is achievable for a large ground plane, it is obvious that in a situation with a small ground plate, the antenna will not be capable any more for SDARS satellite signal reception.

A similarly inferior situation where a single antenna would not fulfill the gain requirements in the regarded angle ranges also occurs if a satellite antenna is integrated in mounting positions, for example, inside the side mirror or on the dashboard of a car. In Fig. 29 the radiation pattern of an antenna is shown which is integrated in one side mirror. Not even on the right side of the diagram the required gain is fulfilled in all the angle ranges since values of gain below 0 dB at an elevation of theta = 60° are not acceptable. Because of shadowing and reflection effects, the gain is even further decreased on the left side of the diagram in the direction toward the car.

Generally spoken in situations of special mounting positions – where the antenna cannot be mounted on top of a large metallic plane – there are problems to keep the required reception quality of a single antenna.

DOI 10.1007/978-981-4560-75-7_101-1

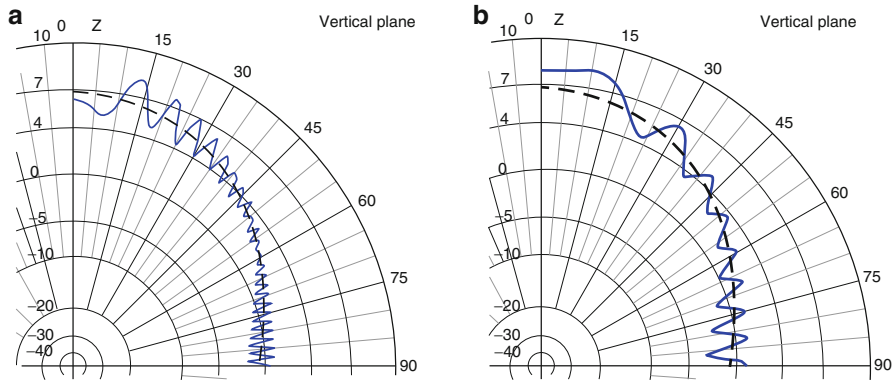


Fig. 27 Vertical radiation diagram of SDARS satellite antenna with small ground plane of 85 mm diameter for different height h over ground: (**a**) h = 1,5 m, (**b**) h = 0.5 m

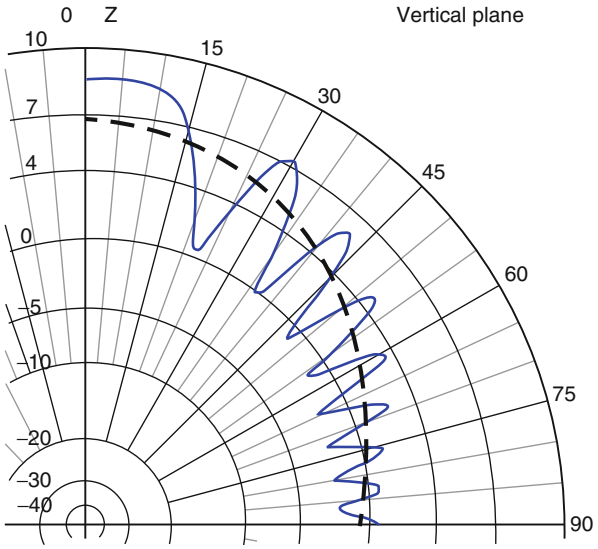


Fig. 28 Vertical radiation diagram of SDARS satellite antenna with small ground plane of 65 mm diameter at 0.5 m height over ground

In this case the use of antenna diversity is of great help, where critical mounting positions and changing environmental conditions can be treated by the use of alternative antennas in an adaptive way. In the next section antenna diversity for satellite radio reception is explained, which enables to keep a high reception quality also in critical mounting situations and also for critical environmental situations around the car.

Antenna Diversity for Mobile Satellite Reception

The reception of satellite radio signals using single antennas generally works flawless in conditions with an LOS propagation path from the satellite transmitter to a stationary or mobile receiver. Still it can be severely impaired in non-LOS scenarios (Parsons et al. 1975). The reasons for these impairments are a combination of attenuation due to signal shadowing and reflection, refraction, and scattering caused by objects in and close to the direct transmission path. This leads to a received signal which is a superposition of a multitude of waves with different directions, levels, as well as phases and causes statistically

DOI 10.1007/978-981-4560-75-7_101-1

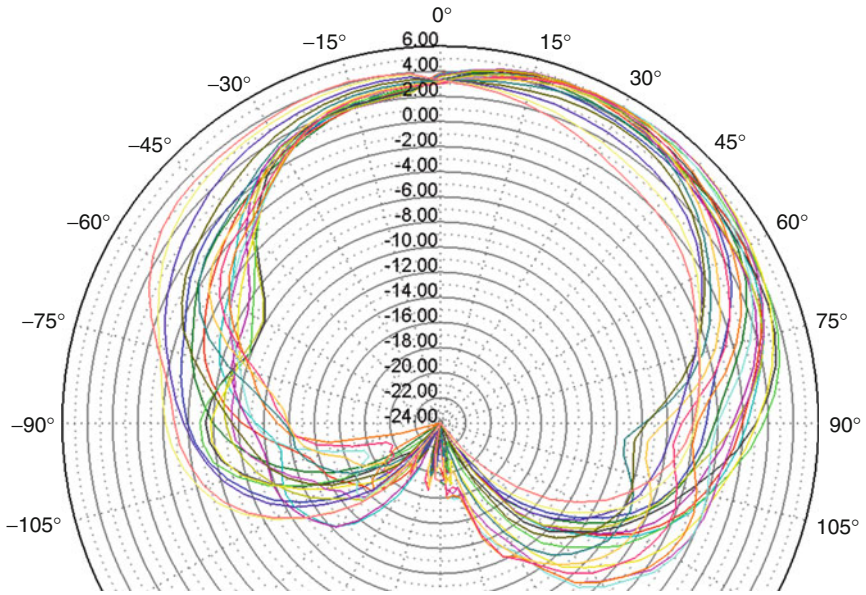


Fig. 29 Vertical radiation diagram of SDARS satellite antenna with small ground plane of 65 mm diameter at 0.5 m height over ground

distributed deep signal fades along the driving paths (Parsons et al. 1975; Brennan 1959). An example of such a scenario and the resulting signal levels is depicted in Fig. 30.

The probability of the deep fades is lower with a strong LOS component and increases with increasing attenuation of the direct signal. A very high probability of deep fades exists in the absence of an LOS signal component. Mathematically the probability distribution in a fading scenario can be described with a Rice distribution where an LOS component is present. In non-LOS conditions this reduces to a Rayleigh distribution with a much higher probability of low signal levels. The Rayleigh distribution of the probability f_R of signal amplitudes r in such a scenario can be described as

$$f_R(r,\Omega) = \frac{2r}{\Omega} \cdot e^{-\frac{r^2}{\Omega}}$$

where Ω denotes the total power of the signal. The resulting distribution in dependence of the signal power Ω is displayed in Fig. 31. The area below the probability distribution is 1 in all cases. When the total signal power is high ($\Omega = 4$), the probability distribution is flat meaning that a broad variety of signal amplitudes is experienced at the receiver. With smaller total signal power (e.g., $\Omega = 0.5$), the probability of high signal amplitudes is very low, whereas small amplitudes are very common. Assuming a signal outage when the amplitude is below r_{crit} (and amplitudes in the positive range including 0), the probability for outages is calculated using the cumulative distribution function

$$p_{r \le r_{crit}} = \int_0^{r_{crit}} f_R(r,\Omega)\, df = 1 - e^{-\frac{r_{crit}^2}{\Omega}}$$

Resulting values for $p_{r \le 0.5}$ for different values of Ω are given in Table 2.

Every single antenna will experience such a distribution of signal levels, and thus, deep signal fades along a driving path through this kind of fading scenario. This can be seen in Fig. 32 which shows two recorded antenna signal levels relative to the LOS level in a Rayleigh fading scenario. The signals have been recorded during test drives in a real fading scenario in the USA.

DOI 10.1007/978-981-4560-75-7_101-1

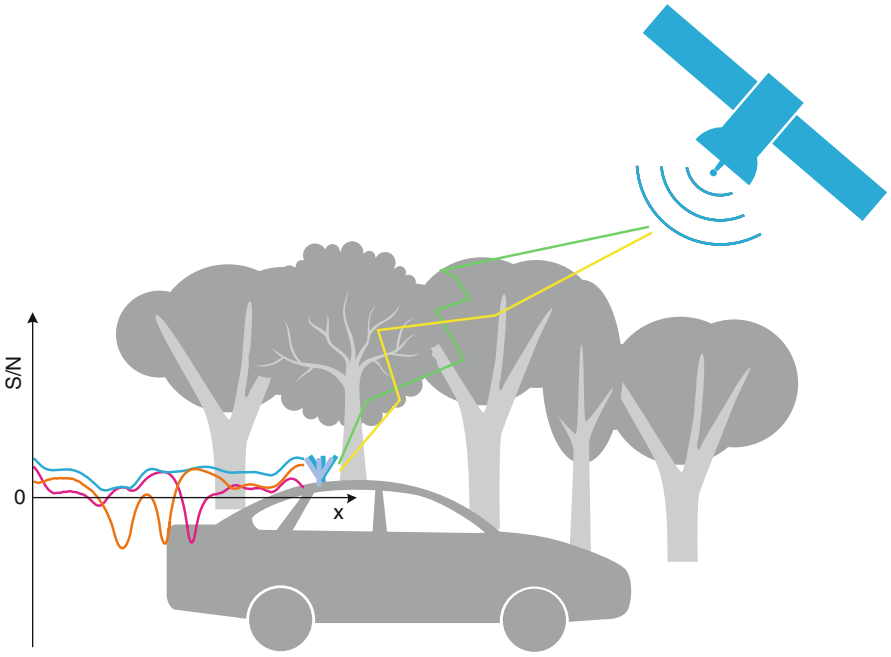


Fig. 30 Signal reception in multipath fading scenarios; *red* and *orange*: reception signals of two diversity antennas; *blue*: improved signal level with diversity

A possibility to avoid the deep fades and therefore increase the reception quality significantly is given by antenna diversity systems (Brennan 1959; Parsons et al. 1975). Assuming two antennas are completely uncorrelated and both show a mute probability of p_m, then the probability of a mute on both antennas simultaneously can be calculated as p_m^2. Therefore if each single antenna has a mute probability of 0.1, this means that two independent antennas have a mute probability of 0.01 and three antennas only have 0.001. Of course the reduction of the probability of mutes with an increasing number of antennas also increases the cost and mounting volume of the system so that a compromise has to be made.

The complexity of antenna diversity systems ranges from a lean switching diversity to the most complex form which is given by a maximum ratio combining diversity system. All systems have in common that their effectivity increases with antenna signals that are less correlated. This decorrelation can, for example, be ensured by a significant dislocation of the antennas' phase centers (Parsons et al. 1975). In practical applications it is much preferred to ensure decorrelated signals by using different polarizations, antenna patterns, or other means that do not necessitate in multiple mounting positions to be used.

Antenna Diversity Basic Concepts

Switching or scanning diversity simply avoids deep fades by switching between the available antenna signals (Barié et al. 2010; Lindenmeier et al. 2007; Lindenmeier 2007). A block diagram is depicted in Fig. 33. Switching is either initiated if the signal level of the currently selected antenna falls below a threshold or if another signal level is higher. In the case where always the best signal is selected, the diversity signal follows the highest of all input signal levels as it is depicted in Fig. 37.

Due to the phase modulation of the satellite radio signals and the decorrelation of the received antenna signals phase, differences occur between the antennas. Without further measures this would lead to bit errors and thus audio mutes after switching. In order to avoid these impairments, switching needs to be

DOI 10.1007/978-981-4560-75-7_101-1

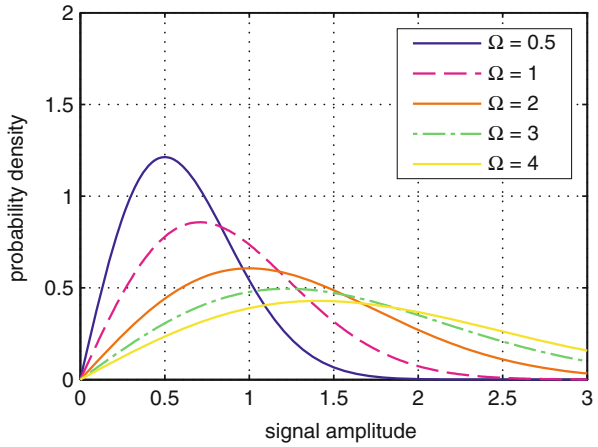


Fig. 31 Rayleigh distribution of signal amplitudes

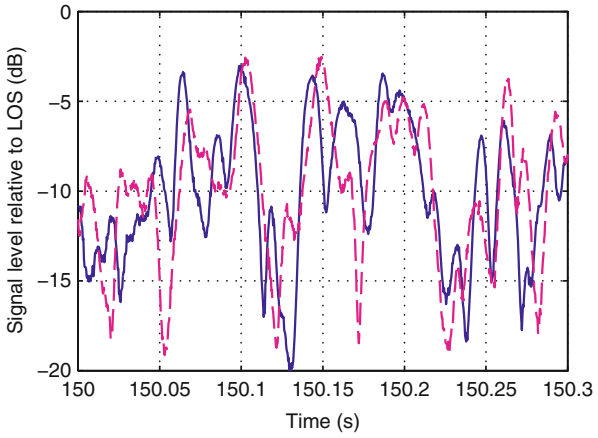


Fig. 32 Signal levels of two antenna signals in a Rayleigh fading scenario relative to the LOS signal level

Table 2 Probability of signal outages assuming a Rayleigh distribution. The critical amplitude $r_{\text{crit}} = 0.5$

Ω	$p_{r \le r\text{crit}}$ in %
0.5	39.4
1	22.1
2	11.7
3	8.0
4	6.1

synchronized to the phase reference symbols that are transmitted regularly (Barié et al. 2010). These are made to inform the receiver of the current absolute phase of the received signal thus allowing for the correct decoding of the transmitted information.

Overall, switching diversity can be realized in a quite lean diversity system, as it is shown in Fig. 33. Due to the frequency dependence of the transmission path, a channel-selective level detection circuitry is necessary. The signal does not need to be decoded in order to achieve this information; thus, a simple local

DOI 10.1007/978-981-4560-75-7_101-1

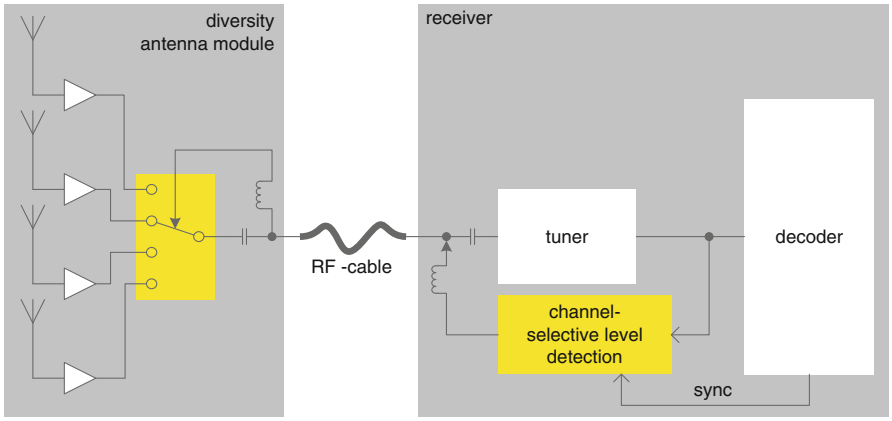


Fig. 33 Switching diversity system

oscillator suffices without the need of phase and frequency locking to the signal carrier. But due to the necessity of synchronizing the switching activity to the phase reference symbol, a cooperative receiver must deliver the respective timing information.

The ability of the diversity circuit to select another antenna if one of the antennas is in a deep fading leads to the prevention of the main adverse effect of Rayleigh fading. This is why without a phase alignment and superposition of antenna signals, the pure switching between the antenna signals and choosing the best antenna signal in each moment yield a considerable improvement in reception quality, as it has been shown in Lindenmeier et al. (2007), Lindenmeier (2007), and Barié et al. (2010).

Another diversity concept which is well known is phase diversity (or equal gain diversity), where there is no selection of antenna signals. Instead they are continuously superimposed after phase alignment. A block diagram is depicted in Fig. 34. Phase shifters in all antenna paths are needed to allow for the superposition of the signals. Also a channel-selective phase detection must be done for each path separately. Due to the complexity of this component, it is commonly situated in the receiver. Therefore separate RF cables and tuners are necessary for each antenna. Due to this restriction not more than two antennas have been usually considered for such a radio reception system. In cases where two input antenna signals have equal levels (in signal and noise, respectively), a phase diversity system allows for an improvement of the SNR of 3 dB. Lower improvements are achievable with signal levels differing stronger. A disadvantage of such a system is that the SNR of the combined signal can even be lower than that of the better input signal if the other signal has a very low signal level but a similar noise level (e.g., in a deep fade). This means that despite the very high effort to realize such a system, the improvement is statistically not better than the one of a switching diversity, which will be shown by example of signal levels depicted in Fig. 37.

A phase diversity system with maximum ratio combining phase-aligns the antenna signals before superimposing them. A (continuously valued) weighting factor is applied to each signal which is derived from its SNR. A block diagram is depicted in Fig. 35. By using phase alignment as well as signal weighting, negative contributions to the combined SNR by input signals with a low SNR are avoided. This allows for the highest possible SNR of the combined signal using the given input signals and explains the name of this system. Figure 37 shows signal levels of single antennas as well as of several diversity systems one of which is an MRC system. The signal levels of an (ideal) MRC diversity system are the highest to be achievable.

Nevertheless, the difference between the results of MRC and the other diversity systems is very low in critical moments of fading at an antenna. This means that the effect of improvement of reception quality is not much higher than the one of the other diversity concepts.

DOI 10.1007/978-981-4560-75-7_101-1

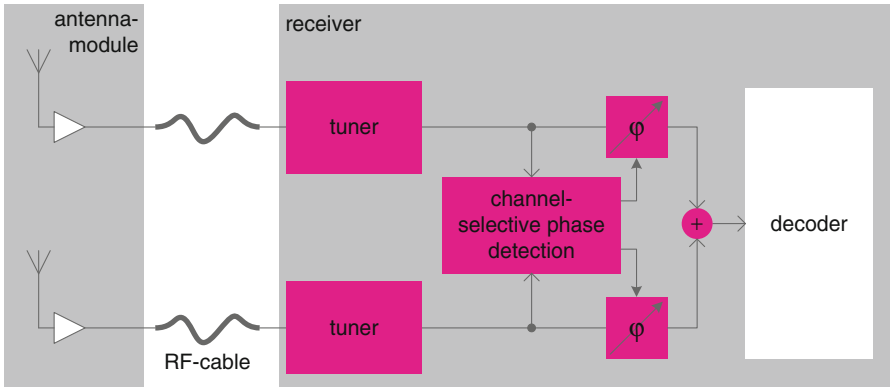


Fig. 34 Equal gain phase diversity system

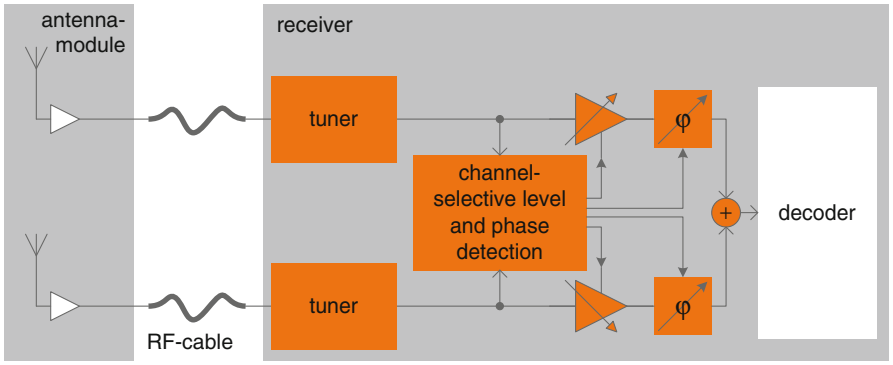


Fig. 35 Maximum ratio combining diversity system

On the other hand, a great disadvantage of this system is its very high complexity. Phase shifters as well as variable gain or attenuation stages are needed in every signal path. Also a complex SNR as well as phase measurement is needed for each signal independently. This leads to separate RF cables and tuners as well as an increased requirement in signal processing power in such systems.

A scan-phase diversity system combines the advantages of a simple switching diversity with the capabilities of a phase diversity (Lindenmeier et al. 2008; Barié 2008). This combination almost reaches the performance of an MRC diversity but with a much reduced hardware effort. The block diagram is given in Fig. 36. It shows the phase shifters and switches as well as the signal combining. In contrast to the MRC system, simple switches deactivate signals with a negative contribution to the SNR of the combined signal. Also phase shifters with discreet phase steps are sufficient due to the low order of the satellite signal's PSK modulation (usually QPSK is used) (Senega et al. 2010). Only a minor degradation of performance results from these steps that on the other hand significantly decrease the overall system complexity. The selection of antenna signals and the calculation of phase differences can be based on signal levels to further reduce the complexity of the detection circuitry by avoiding SNR and phase measurements.

The overall complexity is low enough to allow for the implementation of the complete system in the antenna module itself (Senega and Lindenmeier 2012). Only a single RF cable connects the antenna diversity module to the receiver. The diversity circuit is run independently of the receiver, requiring no further cooperation. Hence, regarding the interface to an SDARS receiver, it behaves like a single standard antenna. Therefore any receiver can be connected to such an antenna system without adaptation in order to improve the reception quality significantly compared to a single antenna.

DOI 10.1007/978-981-4560-75-7_101-1

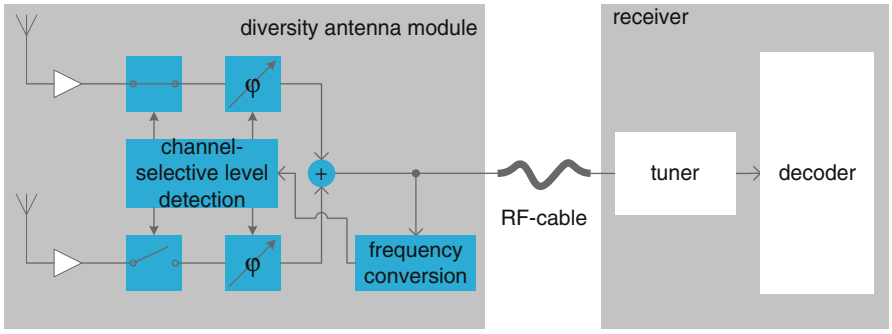


Fig. 36 Scan-phase diversity system

Signal levels of two single antennas as well as of a scan-phase diversity system are depicted in Fig. 37. The signal levels achieved with the scan-phase diversity system are approximately of the same values like those of an MRC diversity system with a very little average deviation of less than 0.4 dB.

Comparison of Diversity Systems

The antenna diversity systems described above show significant differences in their complexity. Table 3 gives a short summary of some of the most important and complex subcircuits which are needed in the different antenna diversity systems.

Regarding the improvements of signal reception which are possible using the presented diversity systems, Fig. 37 shows a short level history of recorded antenna signals as well as of simulated signal levels of the diversity systems. All of them avoid most of the deep signal fades because they use more than one antenna. Only in cases where all antennas experience a deep signal fade at the same time a deep fade in the diversity signal occurs.

Scanning diversity always shows the signal level of the best single antenna. Deep fades are avoided unless all input antenna signals fade at the same time (a total signal blockage). Due to phase differences of the uncorrelated antenna signals, a synchronization to the transmitted reference phase is mandatory in such a system which necessitates in a modified SDARS receiver. In contrast to a scanning diversity system, equal gain diversity can improve the signal level up to 3 dB above that of the two input signals if they show similar levels (e.g., around sample 10,150 in Fig. 37). But due to the lack of signal weighting, the SNR is lower than that of the best input signal in cases where one signal level is much lower than the other. In these cases an equal gain diversity shows worse output SNR than even a simple switching diversity system (e.g., around sample 10,125 in Fig. 37).

An ideal MRC system of course shows the highest improvement of all diversity systems – as it can be seen in Fig. 37 – due to its ability to weight the antenna signals before their phase-aligned combination. But although the hardware complexity of an MRC diversity is much higher, the improved signal quality of the scan-phase diversity system is comparable to that of an MRC diversity system. An analysis of recorded signal levels with calculated ideal diversity systems shows only a small deviation of less than 0.4 dB in average between the MRC system and a scan-phase diversity system (Lindenmeier et al. 2013).

While the results shown in Fig. 37 are true for the simulated diversity systems in reality, the described systems react to changing signal conditions on very different time scales. While switching and scan-phase diversity provide almost immediate responses to deep fades, the phase and SNR estimation invoked in equal gain and MRC diversity usually take much longer and therefore impede fast reaction times. This becomes increasingly problematic when higher driving speeds are considered. Investigations on switching speed and the resulting diversity efficiency show the importance of fast reaction times (Barié et al. 2010). Figure 38 shows the degradation of the effective number of antennas for switching time

DOI 10.1007/978-981-4560-75-7_101-1

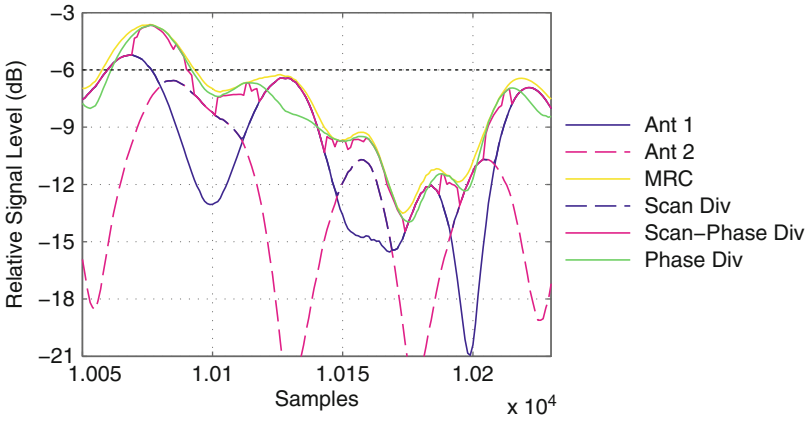


Fig. 37 Signal levels of two single antennas and different simulated ideal diversity systems

Table 3 Number of a selection of subcircuits needed in different diversity systems with *n* antennas (Senega 2013)

	Scanning div.	Equal gain div.	MRC div.	Scan-phase div.
Ant.-LNAs	*n*	*n*	*n*	*n*
RF cables	1	*n*	*n*	1
Tuners	1	*n*	*n*	1
Frequency conv.	–	–	–	1
Measured parameter	Level	Phase	Level + phase	Level

periods in the range of 137 μs to 2.2 ms. Up to a time period of 275 μs, almost constant numbers of effective antennas are achieved, while for 2.2 ms the diversity system shows no improvement of reception quality at all. This investigation was done for a speed of 104.6 km/h (65 mph) so that a distance of λ at SDARS frequencies is covered in 4.4 ms. Therefore reaction times equivalent to distances of less than λ/8 at the highest driving speed considered are necessary to ensure sufficient efficiency of diversity reception systems.

The necessity of fast reaction times to changing signal conditions is obvious from the signal level history depicted in Fig. 37 as well as by the effective number of antennas depicted in Fig. 38. A logical consequence of this fact is that slow beam steering of an antenna combination will not yield noteworthy improvements of signal levels in a fading scenario. The maximum gain which is possible by such a measure of course is 3 dB. While this might seem like a huge improvement at first glance, the fast-changing phases of the individual signals in such a scenario prohibit the combination of the antennas with slow reaction times so that the combined antenna will itself behave like a single antenna and experience deep signal fades which by far are much worse than the 3 dB gained by combination.

In contrast to a slow combination, the fast selection and superposition of antenna signals that is done in a scan-phase diversity system will significantly increase the SNR of the combined signal even in fading scenarios. While this gain in SNR could be seen as a bonus that improves the audio availability, another possibility is that such a gain could be considered in the system design because it is equivalent to an increased link margin of the transmission path. This increased link margin allows for the reduction of other error-correcting redundancies in the signal so that more information could be transmitted instead. Users would then benefit in the form of additional audio channels or higher audio quality.

DOI 10.1007/978-981-4560-75-7_101-1

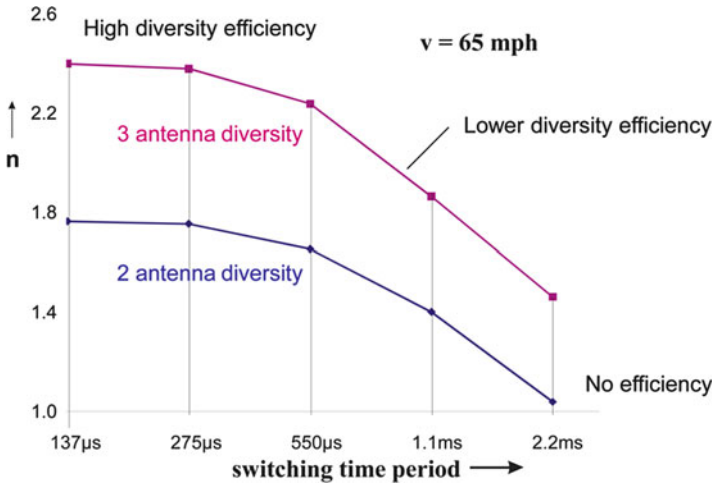


Fig. 38 Diversity efficiency is high for fast-switching systems

Satellite Antenna Structures for Diversity

In the following concepts for compact antenna, diversity structures are considered, in which the RF signal processing units and antennas structures are integrated together in a common mounting volume at one single position at the vehicle. Particular attention has to be paid to the decorrelation of the antenna signals. In the following examples for antenna, diversity structures are shown, which are advantageous for a phased superposition of their antenna signals with an optimized signal to noise ratio in order to increase the reception quality significantly in comparison to a common single antenna.

New fabrication technologies enable the integration of three antenna structures in one part in a mounting volume of $29 \times 29 \times 17$ mm^3 as shown in Fig. 39. The three antennas which are printed onto a common carrier consist of a horizontal loop antenna type together with a circular polarized homogenous field loop antenna and a vertically polarized monopole structure which yield radiation patterns being orthogonal to each other.

Such a set of three antennas being also presented in Müller et al. (2010a) provides three highly decorrelated reception signals for the diversity processor. After phase alignment of the antenna signals and superposition, a combined radiation pattern of the three antenna array (blue line) can be seen in Fig. 40 in comparison to the gain of a standard patch antenna (dashed black line). It shows that in LOS situations the three antennas together yield a gain which is overall around 2–3 dB higher than the one of a common patch antenna.

In a Rayleigh fading scenario, this additional gain would not help to increase the SNR if the main beam direction of the common antenna pattern would be only directed toward the satellite. Together with the diversity circuits shown above, the alternative antenna signals are combined in a way that even in a deep fading scenario – which is most critical for reception quality – the fading can be prevented. This is only possible if the antenna signals are strongly decorrelated, so that fading effects would not occur at the same time at all the alternative antennas of the diversity set. The radiation patterns of the different antennas show a sufficient orthogonality to each other, and hence, the antennas are decoupled from each other by values of more than 25 dB.

Similar values have been achieved with an antenna diversity structure as it is presented in Müller et al. (2010b) requiring a mounting volume of only $29 \times 29 \times 11$ mm^3 and consisting of a circular polarized homogenous field loop antenna together with a patch antenna. In Fig. 41 this antenna set is

DOI 10.1007/978-981-4560-75-7_101-1

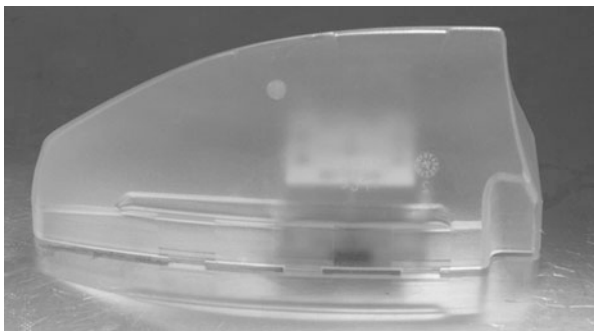


Fig. 39 Three-antenna diversity set (29 × 29 × 17 mm^3) fitting into an ESD cover

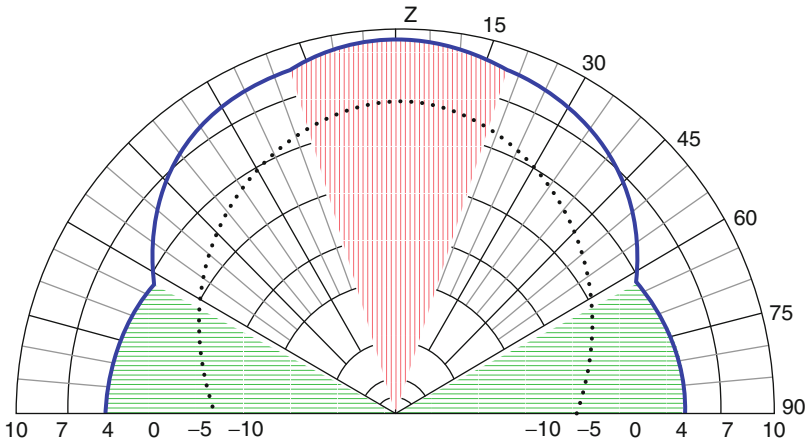


Fig. 40 Radiation pattern of the combined diversity antenna set (*blue line*) in comparison with the pattern of a typical single antenna for SDARS reception (*dashed black line*)

shown in a mounting position on the dashboard of a car. Figure 42 shows the measured radiation pattern of the combined diversity antennas in LOS conditions, which occurs depending on the phase constellation between the two antennas. The blue line shows the main lobe which can be steered onto any angle in azimuth via the phase distance between the two diversity antennas.

Measurements of Scan-Phase Diversity in Fading Scenarios

Test drives have been done in Rayleigh fading scenarios in the USA in order to assess the functionality of the scan-phase diversity system. For these test drives a hardware demonstrator was created and connected to a multitude of different antenna diversity sets located in various mounting positions. Also the reception of HEO and GEO satellites has been evaluated separately in order to show that both can benefit from such a diversity reception system. Pictures of two of the demonstrators used in these measurements are shown in Fig. 43.

Measurement results using a diversity antenna set to receive GEO satellite signals on the rooftop of a car show that the scan-phase diversity system with two antennas significantly improves the availability of the audio signal compared to a single standard patch antenna. Recorded audio levels of a typical result are depicted in Fig. 44. During a test time of 430 s, the mute time of 67.1 s with the single antenna is reduced to only 5.1 s using the scan-phase diversity system. Longer test drives over a duration of 15 min even show mute times as low as 10.9 s resulting in an availability of 99.1 % (Senega and Lindenmeier 2011).

DOI 10.1007/978-981-4560-75-7_101-1

Fig. 41 Diversity antenna function demonstrator in a mounting volume of 29 × 29 × 11 mm^3 on dashboard of a test car

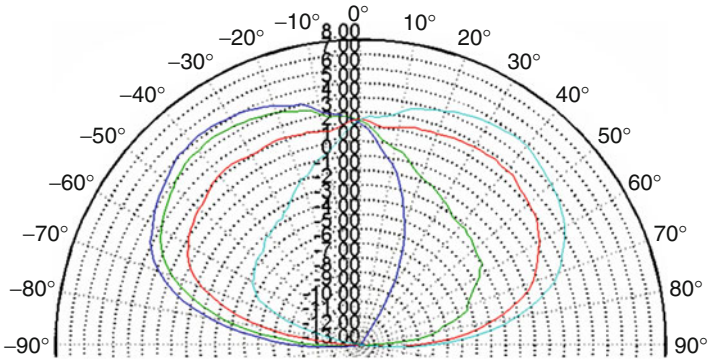


Fig. 42 Measured radiation pattern of combined diversity antennas; *blue line*: main lobe which can be steered onto any angle in azimuth via the phase distance between the two diversity antennas

Measurements of a single HEO satellite show an improvement of availability from 86.8 % with a single antenna to 97.8 % with diversity (Senega et al. 2009).

Apart from the significant improvement of audio availability when rooftop mounting is considered, diversity allows for the consideration of mounting positions that would be unthinkable of using a single antenna. A typical application of such mounting positions would be in convertibles (where rooftop mounting is impossible) or cars where the rooftop mounting space is otherwise restricted or unavailable. Test drives with a diversity set mounted on the dashboard of a car (see Fig. 41) show that diversity can even outperform a rooftop mounted single antenna. The availability is improved from 84 % with the single antenna to 86 % although the reference antenna was placed on the metallic roof of the same test car (Senega and Lindenmeier 2012).

Even mounting the antenna set in a single automotive side mirror has been investigated showing again even better results than a rooftop mounted reference antenna in the reception of a GEO satellite. The test drives have been conducted on a straight road in both directions east to west and west to east so that also in this test drive, the shadowing of the satellite by the car itself has been considered. The availability is improved from 96.7 % with the single antenna to 98.7 % with the diversity system (Senega et al. 2014).

DOI 10.1007/978-981-4560-75-7_101-1

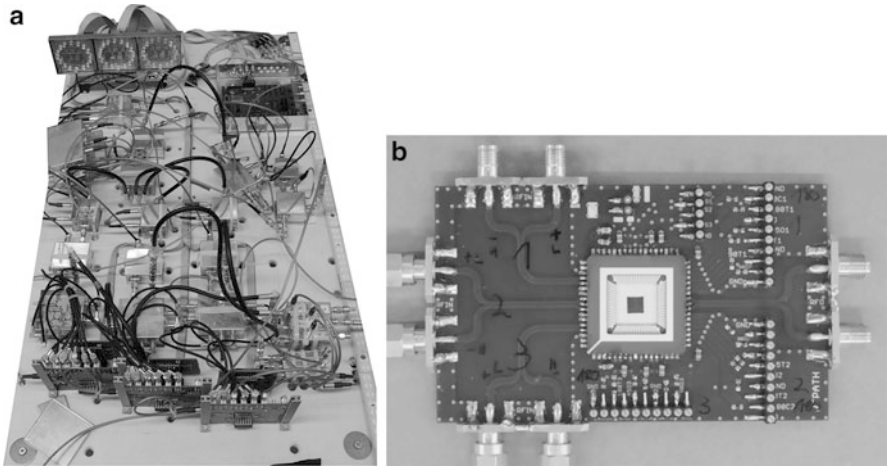


Fig. 43 Scan-phase diversity system demonstrators for up to three antenna signals: (**a**) first demonstrator board; (**b**) test board of actual 4 × 4 mm^2 diversity IC

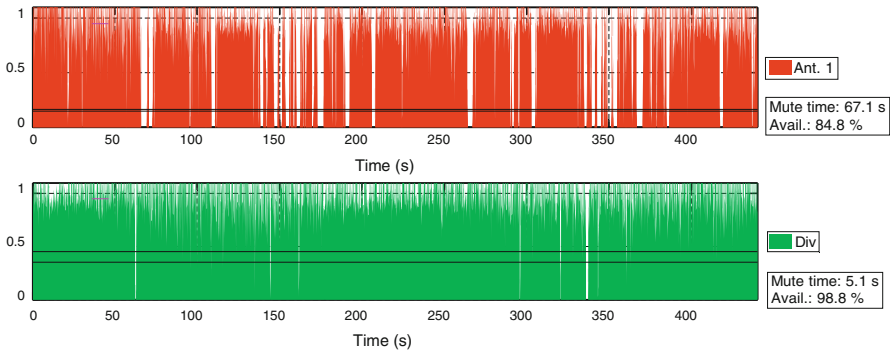


Fig. 44 Audio signal levels recorded in a Rayleigh fading scenario using a single antenna reception system (*red*) compared to audio levels from a scan-phase antenna diversity system (*green*)

Conclusion

Together with classical antenna concepts, new antenna types for reception of signals of HEO, MEO, and GEO satellite systems have been explained which yield high efficiency, low mounting volume, tailored radiation patterns, and easy manufacturing to meet the special requirements for satellite signal reception in a vehicle. Especially, loop antennas like the Scarabeus antenna yield an easy combination with terrestrial antennas in close proximity, which can be set into its phase center. For satellite digital radio services, where gain requirements are relatively high, diversity concepts have been shown which help to improve the reception quality in critical reception scenarios considerably. This is especially achieved with a concept of a scan-phase diversity which can be realized easily and which has got a simple interface to the radio which is equivalent to one of a common single antenna. Within a similar mounting volume as it is already used for single satellite reception antennas and its amplifiers, the complete diversity system can be integrated, consisting of up to three antennas and the diversity circuit. The use of diversity systems also enables the use of inferior mounting positions in the car, where no single antenna could achieve the required reception quality.

Cross-References

- ▶ Antenna Design for Diversity and MIMO Application
- ▶ Circularly Polarized Antennas
- ▶ Loop Antennas
- ▶ Low-Profile Antennas
- ▶ Microstrip Patch Antennas

References

Balanis C (2005) Antenna theory: analysis and design, 3rd edn. Wiley, Hoboken

Barié D, Senega S, Reiter L, Lindenmeier S (2008) Concept studies of scanning and combined scan/phase antenna diversity systems for SDARS. Frequenz 62(9–10):257–261. doi:10.1515/FREQ.2008.62.9-10.257.

Barié D, Reiter L, Lindenmeier S (2010) Fast switched diversity for optimization of S-band SDARS reception. In: Wireless Technology conference (EuWIT), 2010 European, pp 45–48

Brennan D (1959) Linear diversity combining techniques. Proc IRE 47(6):1075–1102

Briskman R, Prevaux R (2004) S-DARS broadcast from inclined, elliptical orbits. Acta Astronaut 54(7):503–518

Davarian F (2002) Sirius satellite radio: radio entertainment in the sky. In: 2002 I.E. aerospace conference, vol 3, pp 1031–1035

ETSI (2010a) Satellite Earth Stations and Systems (SES); Satellite Digital Radio (SDR) systems; part 1: physical layer of the radio interface; sub-part 1: outer physical layer. (ETSI European standard EN 302 550-1-1)

ETSI (2010b) Satellite Earth Stations and Systems (SES); Satellite Digital Radio (SDR) systems; part 1: physical layer of the radio interface; sub-part 2: inner physical layer single carrier modulation. (ETSI European standard EN 302 550-1-2)

ETSI (2010c) Satellite Earth Stations and Systems (SES); Satellite Digital Radio (SDR) systems; part 1: physical layer of the radio interface; sub-part 3: inner physical layer multi carrier modulation. (ETSI European standard EN 302 550-1-3)

Haller N (2001) Mobile antennas for reception of S-DARS. In: Antennas and Propagation Society international symposium, 2001 (APSURSI), vol 1, pp 426–429

Hegarty C, Chatre E (2008) Evolution of the global navigation satellite system (GNSS). Proc IEEE 96(12):1902–1917. doi:10.1109/JPROC.2008.2006090

Herscovici N, Sipus Z, Bonefacic D (2003) Circularly polarized single-fed wide-band microstrip patch. IEEE Trans Antennas Propag 51(6):1277–1280. doi:10.1109/TAP.2003.812241

Kammerer J, Lindenmeier S (2011) A compact car antenna with high efficiency for reception of HEO- and GEO-satellite signals. In: Antennas and Propagation Society international symposium, 2011 (APSURSI), pp 1205–1206

Kammerer J, Lindenmeier S (2012) A new compact antenna combination with high efficiency for reception of SDARS- and GPS signals. In: Antennas and Propagation Society international symposium, 2012 (APSURSI), pp 1–2

Kammerer J, Lindenmeier S (2013a) Invisible antenna combination embedded in the roof of a car with high efficiency for reception of SDARS – and GPS – signals. In: Antennas and Propagation Society international symposium, 2013 (APSURSI), pp 2075–2076

Kammerer J, Lindenmeier S (2013b) Invisible antenna embedded in the roof of a car with high efficiency for reception of satellite digital audio radio services (SDARS). In: Antennas and Propagation (EUCAP), 2013 7th European conference on, pp 1609–1611

Kammerer J, Reiter L, Lindenmeier S (2012) Automotive hexband antenna for AM/FM/GPS/SDARS and AMPS/PCS1900 cell phone in an only 65 mm high housing. In: Radio and Antenna Days of the Indian Ocean (RADIO 2012). IOP conference series: materials science and engineering, vol 44

Lindenmeier S (2007) Antenna diversity for the improvement of satellite radio reception in fading scenario. Inaugural session. In: International conference on wave propagation in communication, microwave systems and navigation, Chemnitz

Lindenmeier S, Luy J, Russer P (2001) A multifunctional antenna for terrestrial and satellite radio applications. In: 2001 I.E. MTT-S international microwave symposium digest (Cat. No.01CH37157). IEEE, pp 393–396

Lindenmeier H, Hopf J, Reiter L, Daginnus M, Kronberger R (2002a) A new design principle for a low profile SDARS-antenna including the option for antenna-diversity and multiband application. SAE International, Warrendale

Lindenmeier S, Olbrich G, Luy J, Russer P (2002b) A five-band antenna for terrestrial and satellite radio services. In: Proceedings of the XXVIIth General Assembly of the International Union of Radio Science, 1109–1112

Lindenmeier S, Reiter L, Barié D, Hopf J (2007) Antenna diversity for improving the BER in mobile digital radio reception especially in areas with dense foliage. In: Antennas, 2007. INICA '07. 2nd international ITG conference on, pp 45–48

Lindenmeier S, Barie D, Reiter L, Hopf J, Senega S (2008) Novel combined scan-phase antenna diversity system for SDARS. In: Antennas and Propagation Society international 2008, pp 1–4

Lindenmeier S, Reiter L, Kammerer J, Senega S (2013) Antenna technology for mobile satellite radio reception. In: Antenna Technology (iWAT), 2013 international workshop on, pp 163–166

Müller D (2010) Kompakte Diversity-Antennen für den mobilen Empfang von Satellitensignalen. Dissertation, Universität der Bundeswehr München, Fakultät für Elektrotechnik und Informationstechnik

Müller D, Hopf J, Lindenmeier S (2009) A cavity-backed thin combined slot-dipole antenna for mobile reception of satellite signals in automotive applications. In: Antennas and Propagation Society international symposium, 2009 (APSURSI), pp 1–4

Müller D, Senega S, Lindenmeier S (2010a) Compact 3-antenna diversity set for HEO and GEO satellite systems with terrestrial repeaters. In: Antennas and Propagation Society international symposium, 2010 (APSURSI), pp 1–4

Müller D, Senega S, Lindenmeier S (2010b) Novel 2-antenna diversity set for SDARS reception in GEO and HEO satellite systems. In: Wireless Technology Conference (EuWIT), 2010 European, pp 169–172

Nasimuddin, Esselle, Verma (2007) Wideband circularly polarized stacked microstrip antennas. Antennas Wirel Propag Lett 6(11):21–24. doi:10.1109/LAWP.2006.890749

Parsons J, Henze M, Ratliff P, Withers M (1975) Diversity techniques for mobile radio reception. Radio Electronic Eng 45(7):357–367

Patsiokas S (2001) XM satellite radio technology fundamentals. SAE International, Warrendale

Pozar D, Duffy S (1997) A dual-band circularly polarized aperture-coupled stacked microstrip antenna for global positioning satellite. IEEE Trans Antennas Propag 45(11):1618–1625. doi:10.1109/8.650073

Reding V (2007) Commission decision of 14 February 2007 on the harmonised use of radio spectrum in the 2 GHz frequency bands for the implementation of systems providing mobile satellite services. Off J Eur Union L 43(C(2007) 409):32–34

DOI 10.1007/978-981-4560-75-7_101-1

Reding V (2009) Commission Decision of 13 May 2009 on the selection of operators of pan-European systems providing mobile satellite services (MSS). Off J Eur Union L 149(C(2009) 3746):65–68

Saala G, Lindenmeier S (2010) Compact circular polarized antenna for mobile reception of radio signals transmitted by geostationary satellites. In: Microwave Conference (EuMC), 2010 European, pp 1465–1468

Saala G, Hopf J, Lindenmeier S (2009) Small satellite car antenna for simultaneous reception of LHCP and RHCP signals. In: European Conference on Antennas and Propagation EuCAP 2009. Proceedings: Estrel Convention Center, Berlin, 23–27 Mar 2009. VDE Verlag, Berlin, pp 2698–2700

Saala G, Müller D, Hopf J, Lindenmeier S (2010) Antenna with optimized pattern for simultaneous reception of terrestrial signals and signals of geostationary satellites. Adv Radio Sci 8:37–42

Sallam H, Abdel Nabi T, Soumagne J (2008) A GEO satellite system for broadcast audio and multimedia services targeting mobile users in Europe. In: Advanced satellite mobile systems, 2008, vol 4, pp 134–139

Senega S (2013) Mehrdienstfähiges Antennendiversity für den mobilen Satellitenrundfunkempfang. Dissertation, Universität der Bundeswehr München, Fakultät für Elektrotechnik und Informationstechnik

Senega S, Lindenmeier S (2011) A fast switching antenna diversity system for improved mobile reception of digital radio signals of a geostationary satellite. In: Antennas and Propagation (EUCAP), proceedings of the 5th European conference on, pp 262–264

Senega S, Lindenmeier S (2012) Antenna module with integrated scan-phase antenna diversity system for SDARS. In: Antennas and Propagation (EUCAP), 2012 6th European conference on, pp 2807–2810

Senega S, Müller D, Barie D, Reiter L, Hopf J, Lindenmeier S (2009) Investigation on the combination of a scan/phase antenna diversity system with a novel diversity antenna set. In: Antennas and Propagation Society international symposium, 2009 (APSURSI), pp 1–4

Senega S, Müller D, Reiter L, Lindenmeier S (2010) A fast-switching diversity- and beam-forming-circuit for S-band satellite reception in fading scenarios. In: Microwave conference (EuMC), 2010 European, pp 648–651

Senega S, Kammerer J, Lindenmeier S (2014) Scan-phase antenna diversity for digital satellite radio (SDARS) in a single automotive side mirror. In: 2014 8th European Conference on Antennas and Propagation (EuCAP), pp 3255–3259

Sharma P, Gupta K (1983) Analysis and optimized design of single feed circularly polarized microstrip antennas. IEEE Trans Antennas Propag 31(6):949–955. doi:10.1109/TAP.1983.1143162

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_103-1

Smart Antennas for Satellite Communications

Qi Luo* and Steven Gao
School of Engineering and Digital Arts, University of Kent, Canterbury, UK

Abstract

Smart antennas are important for satellite communications because they can increase the channel capacity, spectrum efficiency, and coverage range of the communication systems. This chapter reviews the technology for smart array antenna design with examples. Applications of smart antennas for satellite ground stations and direct broadcast satellite systems are also presented in this chapter. A detailed list of references is given in the end of this chapter.

Keywords

Smart antenna; Satellite communication; Array antenna; Beam forming

Introduction

Smart antennas are antenna arrays incorporated with active microwave circuits, a beam-forming network and signal process units. Beam forming can be performed by controlling the complex weight of each array element or each subarray. If the direction of arrival (DOA) algorithm is used, the beam of the smart antenna can be adaptively adjusted by calculating the beam-forming vectors to form a beam in the desired direction and form nulls at any undesired directions; thus, the interference levels are reduced, and the system capacity is improved. In some literatures, smart antennas are also mentioned as adaptive array antennas or intelligent antennas (Bachman 1996; Chryssomallis 2000).

High data rate and seamless coverage have become the trend of the nowadays wireless communications. With the increasing number of wireless users, to maximize the system efficiency and avoid the co-channel interferences, smart antennas are used to replace the conventional antenna systems. With a smart antenna, a communication system can have better signal-to-noise ratio and cover much larger area than the system using a conventional array, especially in high-interference environments (Chryssomallis 2000). In addition, smart antennas can increase the overall capacity in a mobile communication network. Meanwhile, for satellite communications (SatCom), where conventional reflector antennas are deployed for the ground stations, there is also a trend to employ smart antennas. One smart antenna can be used to communicate with several satellites almost simultaneously instead of having several reflector antennas which are usually bulky and can only track the satellites through mechanical steering.

Depending on how the beams of the smart antennas are steered, generally speaking, the smart antennas can be categorized as beam-switching smart antenna and adaptive beam-steering smart antenna. If the complex weights of the array element are predefined and the antenna system has limited number of fixed beam patterns, the smart antenna is called beam-switching smart antenna. On the another hand, if the complex weight of each array element can be adjusted adaptively and the shaped beam of the smart antenna can be steered continuously within certain angular range, then such a system is called adaptive

*Email: q.luo@kent.ac.uk

DOI 10.1007/978-981-4560-75-7_103-1

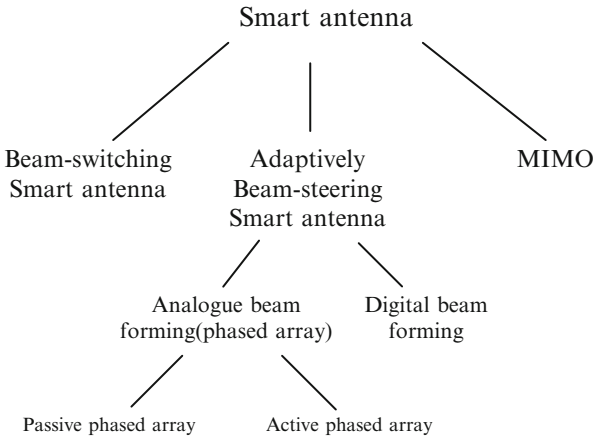


Fig. 1 Classification of smart antennas

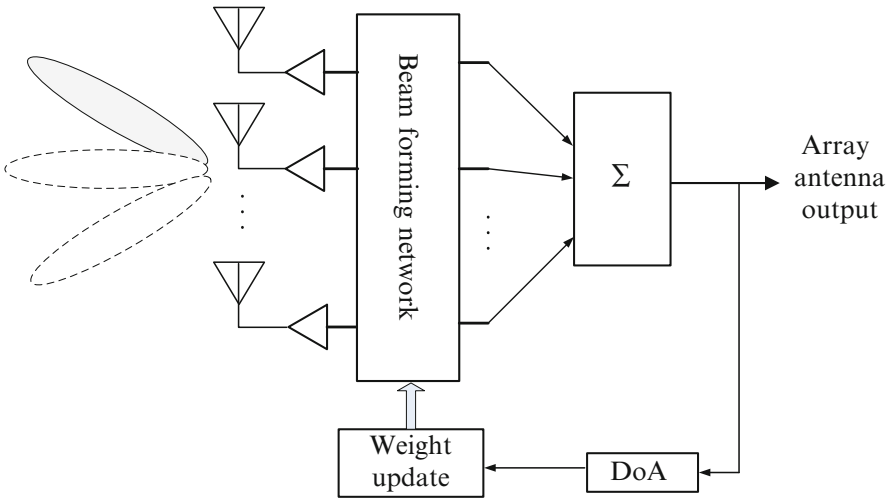


Fig. 2 The generic architecture of a smart antenna

beam-steering smart antenna. For both types of smart antenna, either analogue beam forming or digital beam forming can be employed. Multiple-input and multiple-output (MIMO) antenna is also a form of smart antenna. It utilizes the space diversity to achieve increased data throughput, and there is no dedicated beam forming required for such system. Figure 1 shows a classification for smart antennas. As can be seen, adaptive beam-steering smart antennas with analogue beam forming can be divided into passive phased arrays and active phased arrays. The active phased array refers to the antenna array that each of the radiating elements has its own transmit/receive (T/R) module and the active circuitry is integrated with the passive antenna element (Jenshan and Itoh 1994). In this chapter, only the beam-switching and beam-steering smart antenna will be discussed. For readers who are interested in MIMO antennas, they can refer to Kalis et al. (2014) for more details.

Key Parameters of a Smart Antenna

Figure 2 shows a general structure of a smart antenna. A smart antenna usually consists of an antenna array with a beam-forming network formed by microwave circuitry. Applying advanced signal processing

DOI 10.1007/978-981-4560-75-7_103-1

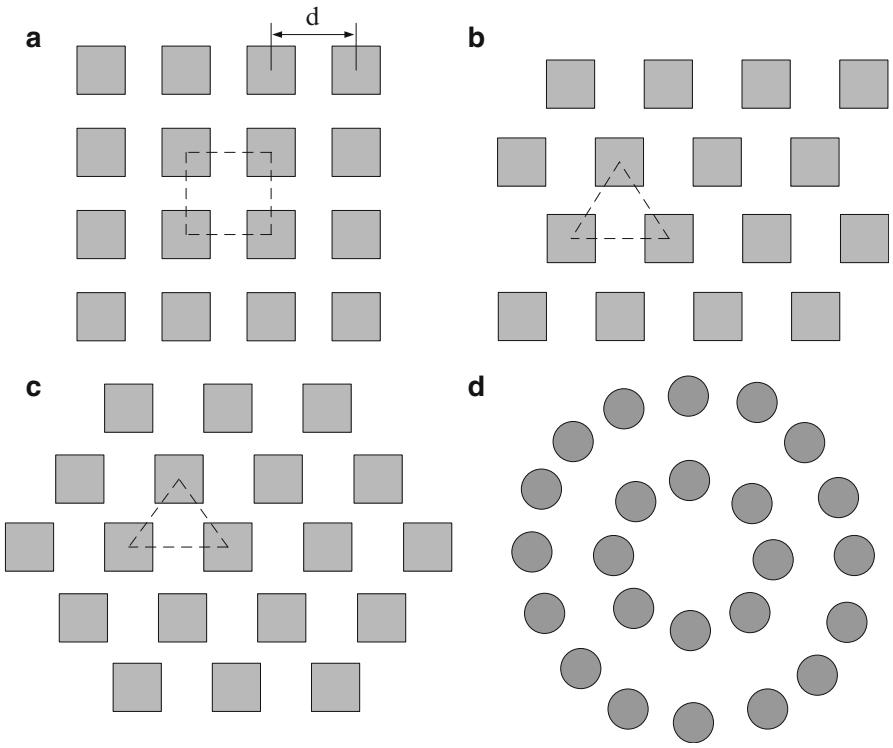


Fig. 3 Three typical topologies of the planar array antenna: (**a**) rectangular lattice, (**b**) triangular lattice, (**c**) hexagonal lattice (**c**) concentric circular array

algorithms to the digital baseband signals of all single antenna elements of the array, the radiation pattern of an array can be shaped even in real time.

The array antenna is a critical component for the smart antenna. It determines the beam-steering performance (e.g., largest beam-scanning angle), directivity, and radiation efficiency of the antenna system. The beam-forming network and the microwave circuit determine the beam-steering or beam-switching capability of the phased array antenna. Therefore, the key parameters of the smart antenna can be described from two aspects: the array antenna and the microwave circuit. Since this chapter is more focused on the designing of antennas and it is not feasible to cover the parameters of circuitry design, so the parameters presented below are only related to the performance of the array antenna.

The antenna array can be either a linear or planar array, depending on the requirements of the specific applications. In general, planar array is preferred for the design of smart antennas so it can benefit from the advanced beam-forming network and signal processing algorithm. Figure 3 shows four typical topologies of the planar array antenna.

The radiation pattern of an array antenna can be expressed as the product of the array factor (AF) times the radiation pattern of a single radiating element:

$$E_{\text{total}} = E_{\text{single\_antenna}} \times AF \tag{1}$$

where E_{total} represents the total electrical field of the array and $E_{\text{single\_antenna}}$ is the electrical field of a single array element. For a two-dimensional planar array with M elements placed in the X-direction and N elements placed on the Y-direction, the array factor can be expressed as (Balanis 2005):

$$AF = S_{xm} \times S_{yn} \tag{2}$$

DOI 10.1007/978-981-4560-75-7_103-1

where:

$$S_{xm} = \sum_{n=1}^{M} I_{m1} e^{j(m-1)(kd_x \sin\theta \cos\varphi + \beta_x)} \tag{3}$$

$$S_{yn} = \sum_{n=1}^{N} I_{1n} e^{j(n-1)\left(kd_y \sin\theta \cos\varphi + \beta_y\right)} \tag{4}$$

The β is the progressive phase in X- or Y-direction, and the (Θ, φ) present the angle of the beam in the elevation and azimuth planes, respectively. k is the propagation constant, and I is the excitation coefficient of each element. To avoid the grating lobes, with a rectangular lattice planar array, the space (d) between the array elements is required to be:

$$d \leq \frac{\lambda_0}{1 + \sin\theta_{\text{Max}}} \tag{5}$$

where θ_{Max} is the maximum scan angle of the array and λ_0 is the free space wavelength at the frequency of interest. When hexagonal or triangular lattice is employed, with the same requirement for grating lobe free scan angle range, the distance d can be larger, and the number of radiating element can be reduced by about 15 % (Hansen 2009).

The directivity (D) of the array antenna is decided by the number of radiating elements when the distance between each array element is fixed (Mailloux 2005):

$$D = \frac{4\pi A}{\lambda^2} \varepsilon_A \cos\theta \tag{6}$$

where A is the size of the radiating aperture, and ε_A is the aperture efficiency of the array. If the beam has a symmetrical pattern in both E- and H-planes, the 3-dB beamwidth of the array antenna in radian with uniform illumination taper can be estimated by (Mailloux 2005):

$$\theta_{3dB} = \sqrt{4\pi 0.866^2 / D} \tag{7}$$

The beam angle of the array antenna is determined by the complex weight on each array elements, which is controlled by the vector modulator when analogue beam forming is used. The vector modulator can perform the phase shifting and control the amplitude of the RF signal through variable attenuators. In case of digital beam forming, there is a digital to analogue converter for each radiating element, and there is no need of using any phase shifters.

It is desirable to have continuous phase control over 360° and continuous amplitude control. Figure 4 shows the amplitude and phase response of an ideal vector modulator.

However, practically the phase can only be provided discretely, and the phase quantization errors are related to the bits of phase shifters. For example, with a 6-bit phase shifter, 64 sets of phase can be provided with 5.625° level differences. Moreover, random phase and amplitude errors also exist. With the phase and amplitude errors, there will be pointing error, increased sidelobe level, and directivity decrease. When phase and amplitude errors are present, the directivity of a smart antenna and the variance of beam pointing deviation can be estimated by (Mailloux 2005):

DOI 10.1007/978-981-4560-75-7_103-1

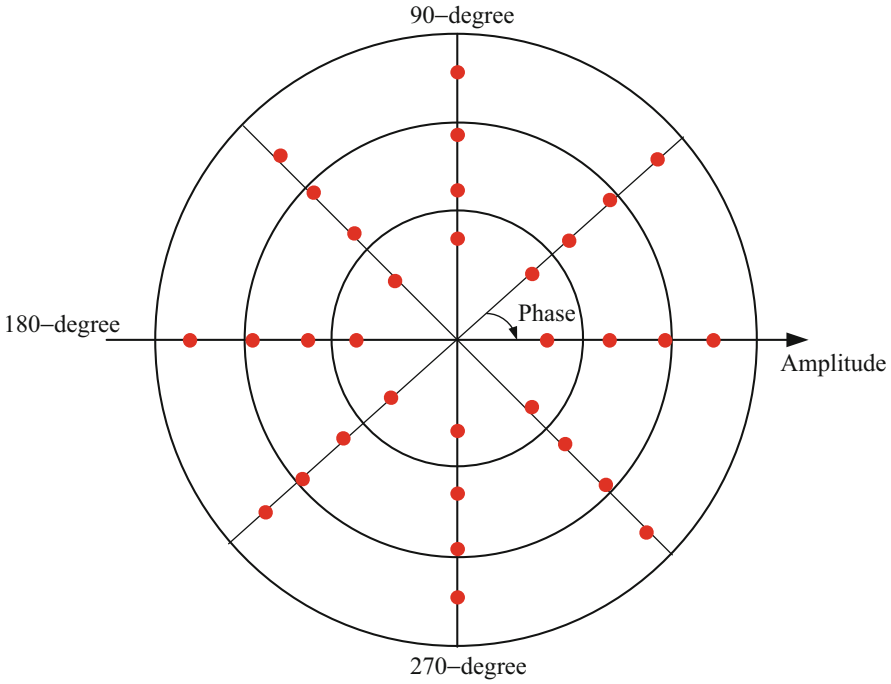


Fig. 4 The amplitude and phase response diagram of an idea vector modulator

$$\frac{D}{D_0} = \frac{P}{1 + \overline{\delta}^2 + \overline{\Phi}^2} \tag{8}$$

$$\overline{\Delta}^2 = \overline{\Phi}^2 \frac{\sum I_i{}^2 x_i{}^2}{\left(\sum I_i x_i^2\right)^2} \tag{9}$$

where D is the directivity of the array with errors, D_0 is the directivity of the array without any phase/amplitude errors, Φ is the phase error, δ is the amplitude error, $\overline{\Delta}^2$ is the variance of beam pointing deviation, I_i is the amplitude of ith element excitation, and x_i is the element position divided by element spacing d. Due to the existence of these errors, a smart antenna with analogue beam forming is required to perform calibrations before it is equipped on the communication system.

Review on the Smart Antenna Design

Beam-Switching Smart Antennas

Beam-switching smart antennas form a limited number of fixed beams in predefined directions. Different beams can be switched by employing a multi-port beam-switching network or introducing RF switches on the feed network. The multi-port beam-switching network can be realized by using the Butler matrix, which can be designed by cascading several branch couplers with RF crossover structure. Many beam-switching antenna arrays using the Butler matrix as the feed network have been reported, including the design of one-dimensional array with either linearly polarization (LP) (Han et al. 2012) or circularly polarization (CP) (Chao-Hsiung et al. 2008; Elhefnawy and Ismail 2009; Changrong et al. 2011). In Elhefnawy and Ismail (2009), a planar microstrip antenna array with a Butler matrix is designed to operate

DOI 10.1007/978-981-4560-75-7_103-1

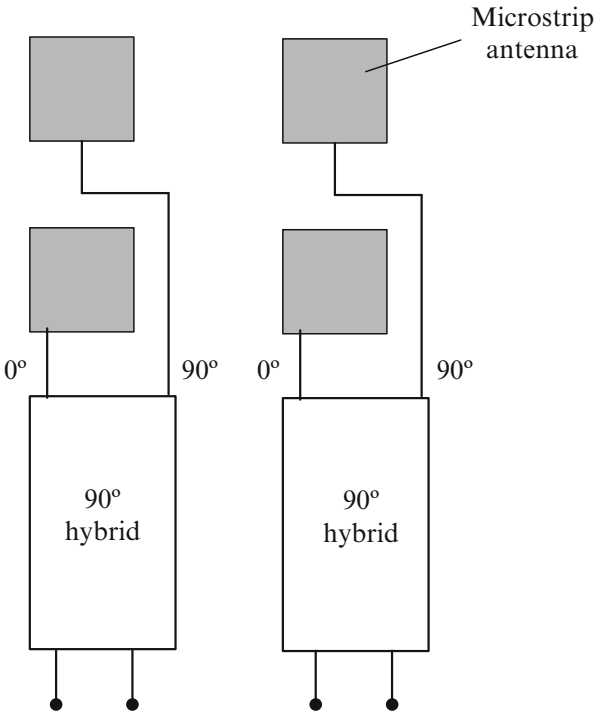


Fig. 5 Demonstration of using 90° hybrid to feed the array element in the same column as presented in Moulder et al. (2010)

at 2.4 GHz for indoor wireless dynamic environments. The CP radiation is achieved by employing one microstrip antenna array consists of four sequentially rotated inset-fed rectangular patch antennas.

A 60-GHz beam-switching smart antenna receiver subsystem based on substrate integrated wave (SIW) guide technology is presented in Fan Fan et al. (2012). Conventional 4 × 4 Butler matrix was used, and SIW bends were used to provide required phase delays. In this design, there is a high-density integration of front-end components into one single substrate, including the 4 × 4 slot array antenna, Butler matrix, band-pass filter, mixer, and local oscillator (LO). It is shown in Fan Fan et al. (2012) that four beams can be realized with the largest scan angle to 45°. Another integrated solution is presented in Patterson et al. (2012), where a 60-GHz active receiving switched-beam antenna array with integrated Butler matrix and gallium arsenide (GaAs) low-noise amplifiers is demonstrated. In this design, the beam of a 4 × 1 active array is controlled by GaAs single-pole-double-throw (SPDT) switches that incorporate with the Butler matrix. The antenna array consists of four quasi-Yagi dipole and can scan the beam up to 40°. The active receiving switched-beam array is printed on the liquid crystal polymer (LCP) substrate, and the measured results show that peak active gain of 27.5 dB is achieved.

While most of these designs can only switch the beam in one plane, a modified planar beam-switching network based on Butler matrix is presented in Moulder et al. (2010). This beam-switching array antenna is designed for 60-GHz short-range communication and is able to switch the beam two-dimensionally. The beam-switching array is comprised of a 2 × 4 planar array and a feeding network that includes two 4 × 4 Butler matrices and four hybrid couplers. Each of the hybrid couplers provides 90° phase shift between its two outputs which feeds the two antennas at the same column, as shown in Fig. 5. With this feed network, eight directional beams in the upper hemisphere can be obtained.

A hybrid solution, which introduces adaptive control in the beam-switching network, is developed in Chia-Chan et al. (2010), where an 8-way Butler matrix was used to feed a 1 × 8 printed patch array. The feeding network is realized by striplines, and the overall system has a multilayered structure. Different from the conventional design, phase shifters were introduced in the feed network, as demonstrated in

DOI 10.1007/978-981-4560-75-7_103-1

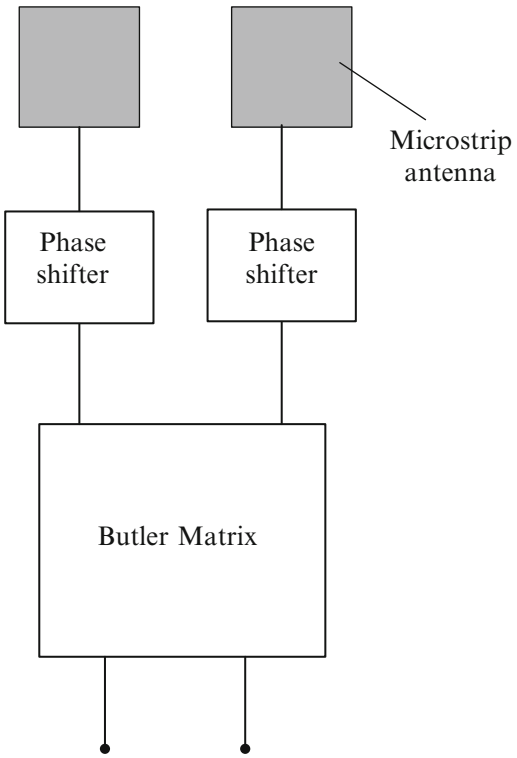


Fig. 6 Demonstration of introducing phase shifter in the Butler matrix as proposed in Chia-Chan et al. (2010)

Fig. 6. The beam of the array is initially switched to a certain direction through choosing the corresponding port of the Butler matrix, and then it can be slightly adjusted by controlling the phase shifters. Thus, the focused beam can be steered to cover more scan angles.

Recent progresses in the beam-switching smart antenna design with RF switches include employing RF PIN switches or RF MEMS switches on the feed network to provide a limited number of fixed beams up to 40° (Bayraktar et al. 2007; Jun 2011). However, this usually leads to a complicated feed network as the result of introducing additional DC biasing network, which limits its application in compact design especially when the number of radiating element is large.

Adaptive Beam-Steering Smart Antennas

The main disadvantage for the beam-switching smart antenna is that it can only provide limited number of fixed beams. Instead, the adaptive beam-steering antennas are able to continuously steer the beam to the desired directions and shape the beam to maximize the link budget with sophisticated signal processing algorithms. In order to adaptively shape the beam, an adaptive beam-forming network is required, which can be either an analogue beam forming (ABF) or digital beam forming (DBF). The ABF requires employing a large number of phase shifters and attenuators to provide complex weights to the array elements. The summed signals need to be down convert to baseband signal and then pass to the analogue/digital (A/D) converter before sending to the signal processing unit. Most of the beam-switching smart arrays steer the beam by controlling the input phase of the array elements only, and they belong to analogue beam forming. Instead of using large amount of microwave phase shifters, which are expensive and lossy, DBF uses A/D converter and digital down converter (DDC) on each array element, and then the beam forming can be realized digitally. However, DBF requires a large amount of A/D converters, which have high power consumptions and need real-time signal processing. To realize the adaptive beam forming, there are several well-defined algorithms available such as the least mean square (LMS)

DOI 10.1007/978-981-4560-75-7_103-1

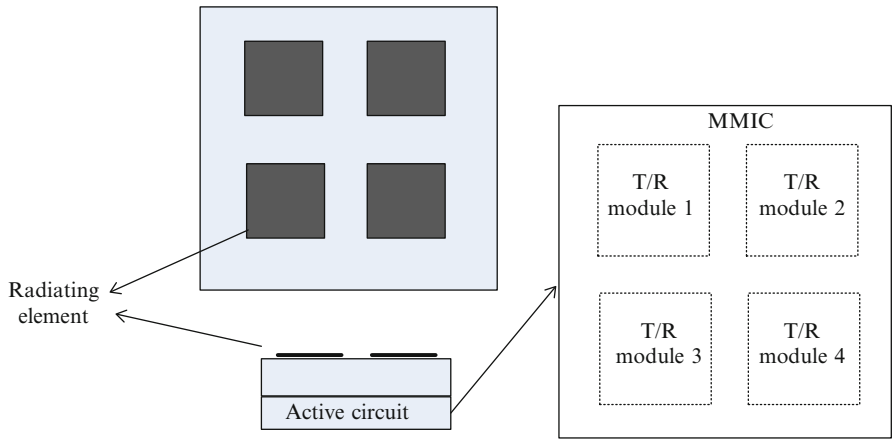


Fig. 7 Demonstration of using one MMIC integrated with four T/R modules to control four antenna elements as presented in Chaloun et al. (2014)

algorithm, the recursive least squares (RLS) algorithm, and the constant modulus (CM) algorithm. Reader can refer to Balanis and Ioannides (2007) for more details on these algorithms.

For mobile satellite applications, the smart antenna is required to scan the beam within large angle range with high efficiency. For such systems, one of the key components is an array antenna with large beam-scanning angles. Designing antenna arrays with large scan angles is challenging and has been attracted much research interest in recent year. Recent progresses in this field include using slot antenna (Zhang et al. 2012), patch (Zhou et al.), and stacked patch (Chaloun et al. 2014) as the array element, and the wide scan angle range is achieved by suppressing the surface waves and reducing the mutual coupling between adjacent elements. One state-of-the-art design in the field of smart antennas using ABF is reported in Chaloun et al. (2014), where a highly integrated Ka-band transmit/receive reflectarray (Pilz and Menzel 1998) with wide scan angle is presented. The reflectarray has 108 antenna elements, and the array can scan its beam to $\pm 60^{\circ}$ in the upper hemisphere. Plated through holes were introduced to suppress the surface waves and thus reduce the mutual coupling between adjacent elements. Because of the small size of the array element, which is 5 mm $\times$ 5 mm ($0.5\lambda_{30GHz} \times 0.5\lambda_{30GHz}$), four T/R modules are integrated on one Monolithic Microwave Integrated Circuit (MMIC), as shown in Fig. 7.

Recent development in the smart antennas with DBF includes electronically steerable receive and transmit antennas for satellite communications at Ka-band reported in Stark et al. (2009). In this work, the smart antenna is a highly integrated antenna with 15-layer low-temperature co-fired ceramic (LTCC) tiles and combines the antenna elements, the calibration network, the RF circuitry, and the cooling system. The architecture of this DBF smart antenna is shown in Fig. 8.

Since smart antenna is a complicated system especially for the adaptively beam-steering smart antenna, it is impossible to cover all the details for the design of each component including the RF circuit design in this chapter. Instead, this chapter will focus more on the design of the array antenna and its corresponding feed network. For the design of the RF circuits, readers can refer to Li (2012) for more details. This chapter is organized as follows. In section "Smart Antenna Design," the design techniques of array antennas for beam-switching and adaptive beam-steering smart antenna are demonstrated with examples. Then, in section "Application of Smart Antennas in SatCom," the applications of smart antennas in the field of satellite ground station and mobile direct broadcasting satellite systems are presented.

DOI 10.1007/978-981-4560-75-7_103-1

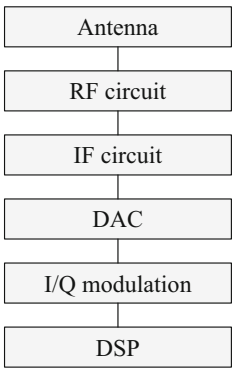


Fig. 8 The architecture of the smart antennas with DBF as presented in Stark et al. (2009)

Smart Antenna Design

As introduced in section "Introduction," smart antennas can be categorized into two main types: beam-switching smart antennas and adaptive beam-steering smart antennas. A beam-switching smart antenna has a beam-forming network that provides fixed phase shift to form shaped beams in predefined directions. After introducing either the RF switches or using couplers with multiple ports, the beam can be switched discretely. This leads to a low-cost solution because there is no need of using any phase shifters and amplitude attenuators. Due to the complexity of the feed network, most of the beam-switching smart antennas are only able to switch the beam in one direction.

In contrast to the beam-switching smart antenna, the adaptive beam-steering smart antenna is able to continuously steer the beam to the desired directions and shape the beam to maximize the link budget with sophisticated signal processing algorithms. The accurate control of the input phase and amplitude to each radiating elements increases the system complexity; thus, adaptive smart antennas exhibit higher cost.

Beam-Switching Smart Antenna

Beam-switching smart antennas provide a low-cost and low power consumption solution if the continuous beam steering is not required. It does not need to employ any phase shifters, and in most cases analogue beam forming is used. A beam-switching smart antenna has a beam-forming network that provides a set of fixed phase shifts to the array elements, and shaped beams in predefined directions can be formed. Thus, the beam-forming network is the most critical component of the beam-switching smart antenna.

Theory

Figure 9 shows the generic architecture of a beam-switching array. To realize a beam-switching feed network, multiple ports are usually employed, and the beams can be switched by selecting the corresponding RF port.

One technique for realizing such a beam-forming feed network is to introduce RF switches on the feed network. Figure 10 shows one example of introducing RF switches on the feed network of a one-dimensional microstrip patch transmitting array with four radiating elements.

As can be seen from Figure 10, two RF switches are mounted on the input of the microstrip feed lines, and by controlling the states of the RF switches, the patches can be fed by RF signals with two different phase delays. This is equivalent to a two-port beam-forming feed network. Introducing more RF switches increases the number of different phase delays for each radiating element, and then more radiation patterns can be obtained. One of the most common and cost-effective RF switches is the RF PIN diode. RF PIN

DOI 10.1007/978-981-4560-75-7_103-1

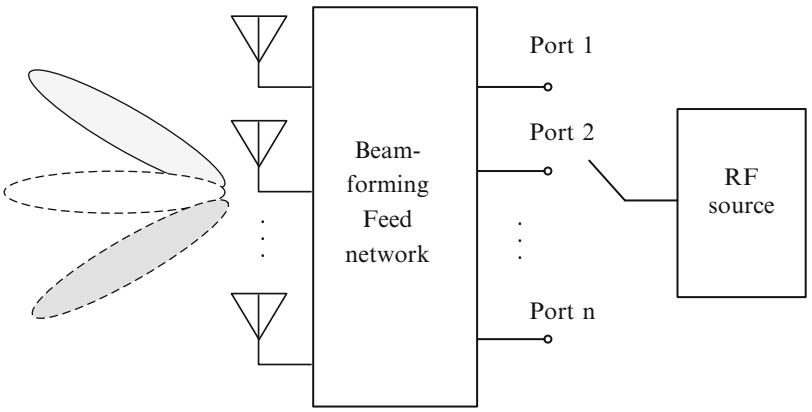


Fig. 9 The generic architecture of a beam-switching smart antenna

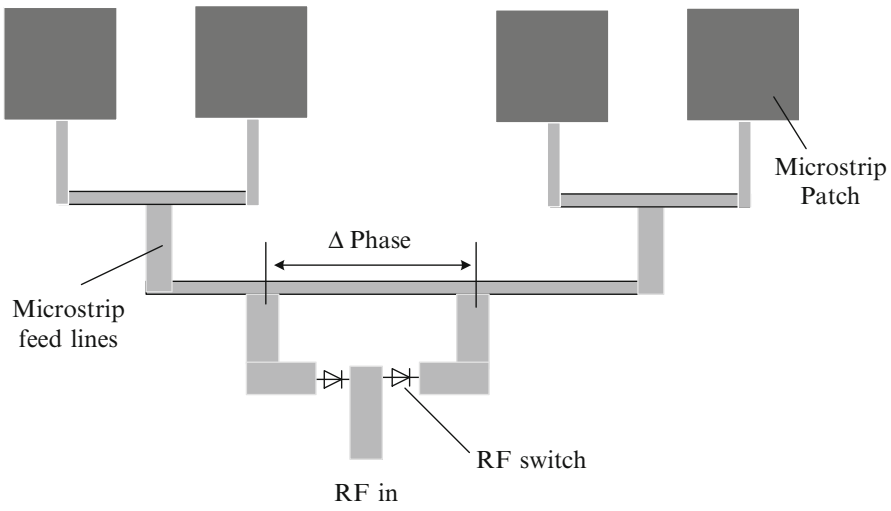


Fig. 10 Example of introducing RF switches on the feed network of an array antenna

diodes exhibit fast switching time with reasonable insertion loss and isolation. A RF MEMS (microelectromechanical system) exhibits a better performance with respect to low insertion loss and high operation frequency (e.g., millimeter wave frequencies) but with higher fabrication cost and assembly complexity.

A biasing circuit is required in the feed network in order to control the ON and OFF states of the RF switches. The biasing circuit must ensure that there is a high RF/DC isolation at both the RF port and the DC power source, while there is low insertion loss and impedance mismatch caused by the lumped elements introduced. Figure 11 presents a circuit model of the biasing circuit for an antenna application. The capacitor is employed to block the DC power, while the inductor is used to block the RF signals (RF chock). For printed microstrip antenna design, the DC bias line can be realized by microstrip lines of narrow width, which exhibits high impedance at RF frequencies. To ensure proper impedance matching, the capacitor needs to have a small reactance at the lowest microwave frequency of interest, and its self-resonant frequency must be much higher than the highest frequency of interest (AVX Corporation). A microstrip line of a quarter wavelength and radial stubs can be used to improve the impedance matching of the feed network.

Another well-known approach to design a beam-switching feed network is to use a Butler matrix (Butler and Lowe 1961). The Butler matrix has N input ports and N output ports, where:

DOI 10.1007/978-981-4560-75-7_103-1

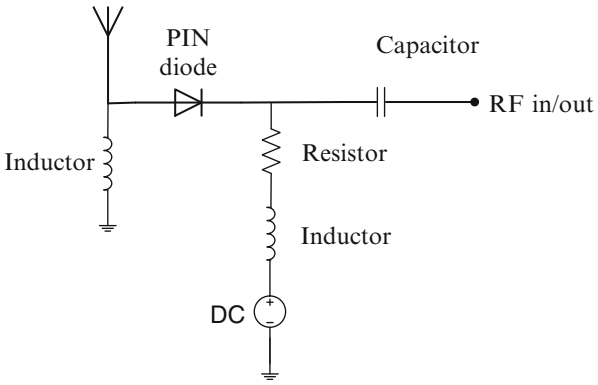


Fig. 11 A circuit model of the biasing circuit for RF switches in antenna application

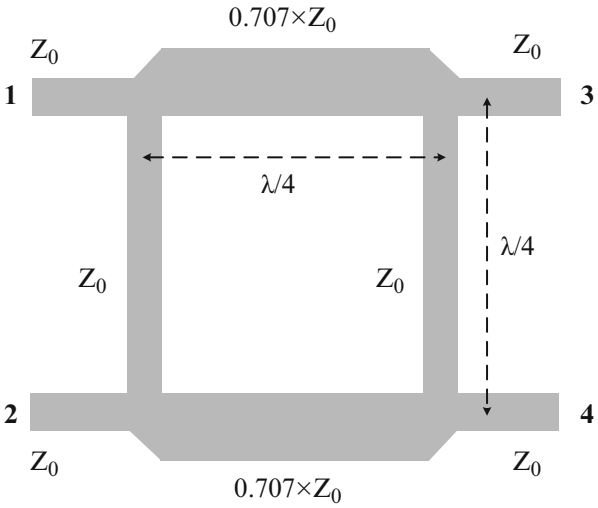


Fig. 12 A four-port branch-line couplers

$$N = 4n, \quad n = 1, 2, 3, \ldots$$

Figure 12 shows the layout of a four-port Butler matrix realized by the branch-line coupler. As shown in Figure 12, each transmission line is a quarter wavelength long at the frequency of interest, and the horizontal microstrip lines have impedance of $0.707 \times Z_0$.

With a four-port Butler matrix, the scattering matrix can be expressed as:

$$\begin{bmatrix} Output\ 1 \\ Output\ 2 \\ Output\ 3 \\ Output\ 4 \end{bmatrix} = B \times \begin{bmatrix} Input\ 1 \\ Input\ 2 \\ Input\ 3 \\ Input\ 4 \end{bmatrix} \tag{10}$$

where

DOI 10.1007/978-981-4560-75-7_103-1

$$B = \frac{1}{\sqrt{2}} \times \begin{bmatrix} 0 & 0 & 1 & -j \\ 0 & 0 & -j & 1 \\ 1 & -j & 0 & 0 \\ -j & 1 & 0 & 0 \end{bmatrix} \tag{11}$$

Assuming the port 1 (as indicated in Figure 12) is the input port with voltage of V_{in}, which means:

$$\begin{bmatrix} Input\ 1 \\ Input\ 2 \\ Input\ 3 \\ Input\ 4 \end{bmatrix} = \begin{bmatrix} Vin \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{12}$$

Solving the above equations, it can be derived that:

$$Output\ 1 = 0$$

$$Output\ 2 = 0$$

$$Output\ 3 = \frac{Vin}{\sqrt{2}} \times \exp(\varphi)$$

$$Output\ 4 = \frac{Vin}{\sqrt{2}} \times \exp\left(\varphi - \frac{\pi}{2}\right) \tag{13}$$

This means that 90° phase difference is obtained at the ports 3 and 4, and port 2 is isolated. By changing the input port from port 1 to port 2, the output phase of the ports 3 and 4 is swapped.

Design Examples

Figure 13 shows the schematic diagram of using branch-line couplers and power dividers to form a feeding network for a beam-switching array. As presented in Figure 12, the branch-line coupler can split the input RF signal with equal amplitude and 90° phase difference. Then, microstrip power dividers can be introduced to the output ports of the branch-line couplers. Microstrip lines of varied length can be used at the outputs of the power dividers to introduce the true time delays and thus control the input phases for each radiating elements.

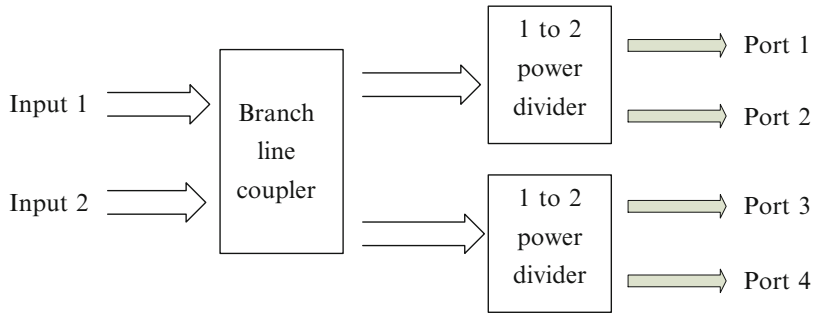


Fig. 13 The schematic diagram of using branch-line couplers and power dividers to design a feed network for a beam-switching array

DOI 10.1007/978-981-4560-75-7_103-1

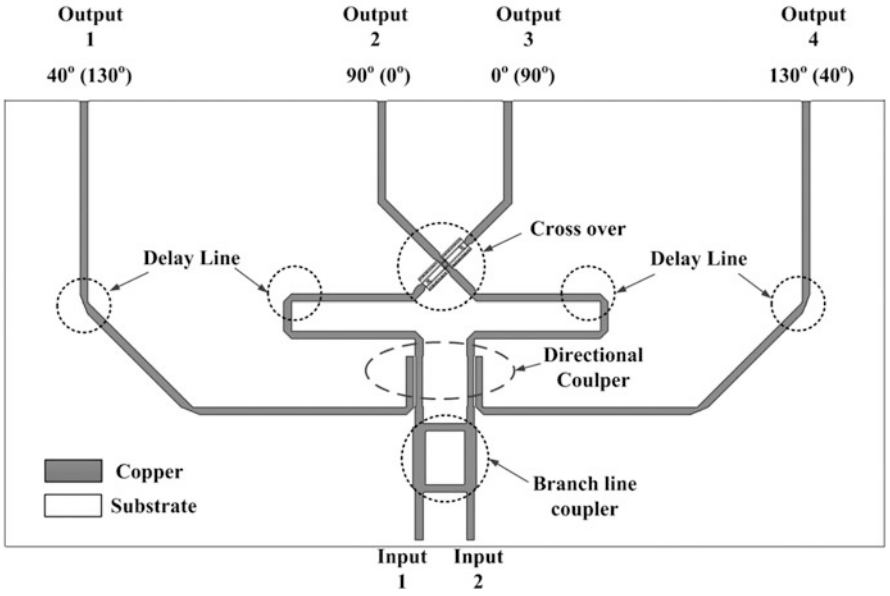


Fig. 14 The configuration of the feeding network presented in Han et al. (2012)

As an example, Figure 14 shows one feeding network based on the branch-line coupler presented in Han et al. (2012). This feeding network is designed for a beam-switching array that consists of four patch elements that operates at 5.1–5.3 GHz. A branch-line coupler is employed at the first stage of the feed network, and then the two coupled lines are introduced to split the power into four outputs, where the microstrip patch is fed by probe feeding. The length of the microstrip lines after the power divider can be varied to produce different phase delays, which is determined by the required beam-switching angle. The radiating elements are fed with two different sets of phase distributions when the RF signals are input from port 1 or port 2. The corresponding phases of each radiating element are indicated in Figure 14. Thus, the beam of this array can be switched by selecting different input ports.

Figure 15 shows the measured radiation patterns of this beam-switching array when port 1 and port 2 are selected. As seen, when port 1 is selected as the input port, this linear array has its beam pointing at 22.5° while the beam is switched to −22.5° when port 2 is excited. This design is a very classic example of applying branch-line coupler and power dividers to obtain a beam-switching feed network. In this example, the beam can only be switched between two different angles. More beam-switching angles can be obtained by cascading more couplers, and then the number of input port is also increased.

Besides using a Butler matrix, another approach for designing a beam-switching array is to introduce RF switches on the feeding lines. Figure 16 shows the layout of a feeding network with PIN diode switches embedded on the feed lines for the array antenna in Maqsood et al. (2014). The DC biasing network is required to provide proper DC bias for the PIN diode switches. To ensure that there is high isolation between DC and RF source, inductors of 100 nH are used as the RF chocking and capacitors of 10 μF are used as the shunt capacitors at the biasing points.

As shown in Figure 16, the feed network has four output states, which means that the phase of each radiating element can be adjusted with four different sets of phase; thus, the beam of the array can be switched to four different angles. Four dual-band circularly polarized stacked annular ring antennas are used in this array, and the resonating frequencies are designed to be 2.454 and 3.15 GHz. Figure 17 shows the measured radiation pattern of this beam-switching array with four different sets of phase distributions when different RF switches were forwarded biased. It is observed that at both frequencies, the beam of the array can be switched to ±25° with moderate gain and low sidelobes.

DOI 10.1007/978-981-4560-75-7_103-1

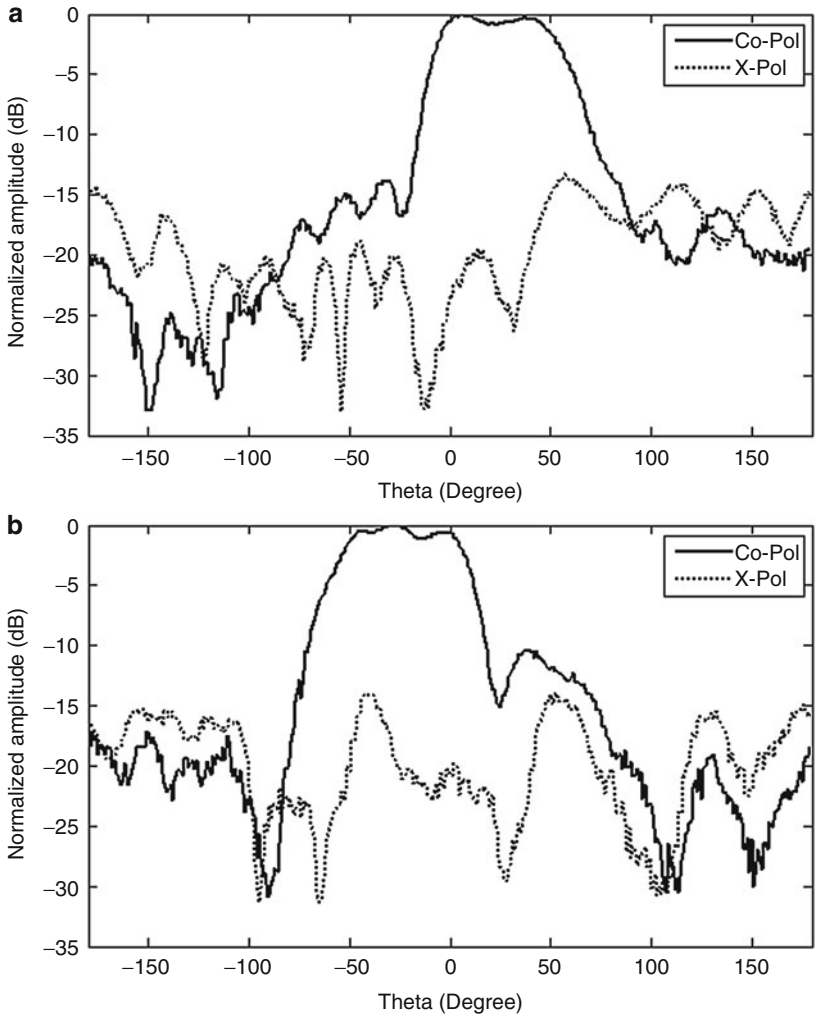


Fig. 15 The measured radiation patterns of the beam-switching array presented in Han et al. (2012) when (**a**) port 1 is excited and (**b**) port 2 is excited

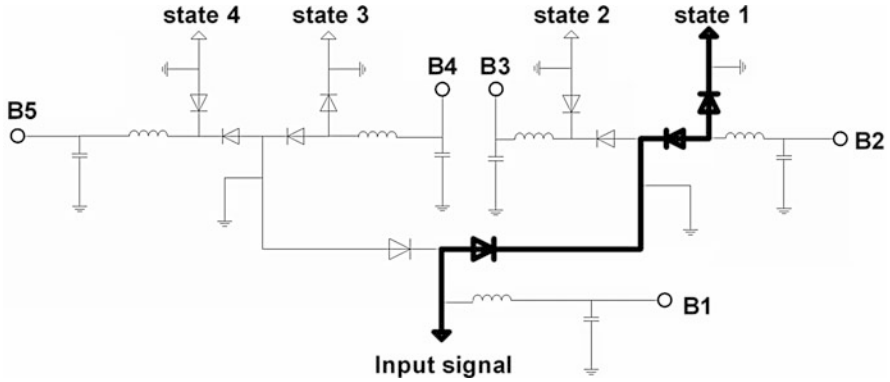


Fig. 16 The layout of the feeding network with PIN diode switches (Maqsood et al. 2014)

From these two design examples, it can be seen that with a Butler matrix, the beam switching of an array can be realized by selecting the corresponding input ports. The number of the ports increases with increasing number of antennas. The array is a multiple-input and multiple-output system and can also be

DOI 10.1007/978-981-4560-75-7_103-1

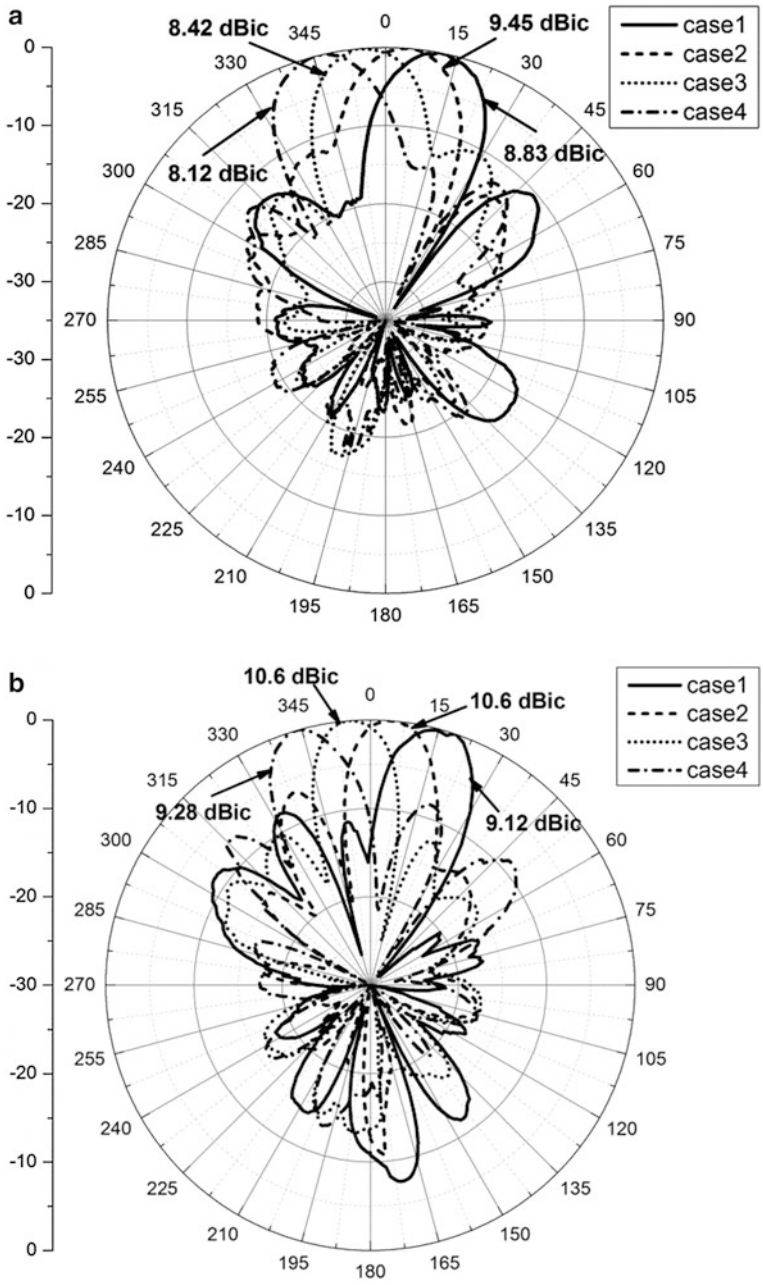


Fig. 17 The measured radiation patterns of the dual-band CP array at (**a**) 2.454GHz and (**b**) 3.15GHz (Maqsood et al. 2014)

applied to the multi-beam applications. With the second approach, which is introducing RF switches on the feeding lines, the phase distribution of the array can be adjusted by controlling the states of the RF switches, and only one input port is required. However, additional DC-biased circuitry is required to control the RF switches, which increases the overall system complexity and reduces the efficiency of the antenna system (due to the loss of the RF switches) if a large number of RF switches are employed.

DOI 10.1007/978-981-4560-75-7_103-1

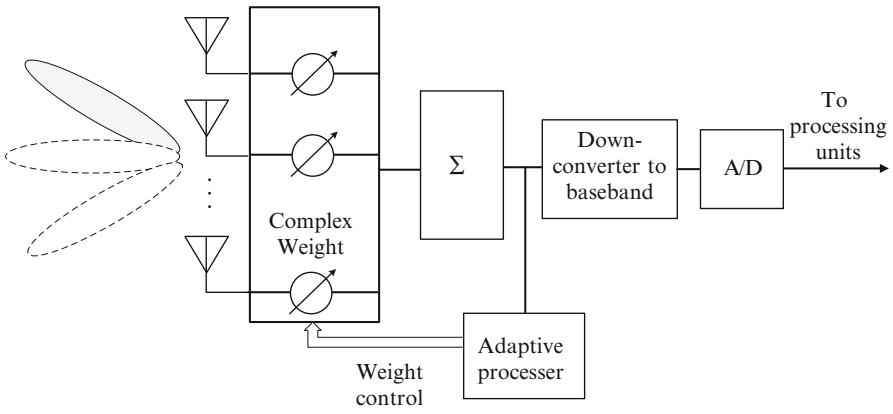


Fig. 18 The general architecture of a smart antenna with analogue beam forming

Beam-Steering Smart Antenna

Theory

Adaptive beam-steering smart antennas are able to continuously steer the beam to the desired directions and shape the beam to maximize the link budget with sophisticated signal processing algorithms. This makes it more attractive compared to the beam-switching smart antennas, which can only provide limited number of beams. For mobile satellite applications such as SatCom on the move, the adaptive beam-steering smart antenna is required to scan the beam within large angle ranges with high efficiency.

The beam forming of the adaptive smart array antennas can be achieved by employing a phased array with controllable amplitude and phase. Figure 18 shows the general architecture of a smart antenna with analogue beam forming (Luo et al. 2014).

A two-dimensional antenna array with large scan angle is usually required to be equipped on an adaptive smart antenna. When the size of the array is large, the radiating elements in the central region have similar characteristics as an element in an infinite array. Then, the array can be analyzed by using the Floquet modes with the assumption that the array has an infinite size. With Floquet modes, an antenna array with infinite sources can be represented as a superposition of Floquet modal functions, instead of solving Maxwell's equations with infinite source excitation functions. It has been proven that using the results from the infinite array, the performance of a finite array (e.g., mutual coupling, active reflection coefficients) can be determined with good accuracy (Bhattacharyya 2006).

When the array elements are excited with uniform amplitudes and linearly progressed phase, using the Floquet modal function, the impedance of an array element can be calculated by (Bhattacharyya 2006):

$$\frac{Z^{FL}\left(k_{x0},\ k_{y0}\right) = -<\vec{E},\ \vec{I}>}{I_0^2} \tag{14}$$

where k_{x0} is the phase constants in X-direction, k_{y0} is the phase constants in Y-direction, I_0 is the input current, $\vec{E}$ is the electric field at the current source location, $\vec{I}$ is the current distribution function, and $<\vec{E}\ ,\ \vec{I}>$ is the self-reaction of the current source in the unit cell and is given by:

$$<\vec{E}\ ,\ \vec{I}>=\iint \vec{E}\times\ \vec{I}\ dx\ dy \tag{15}$$

DOI 10.1007/978-981-4560-75-7_103-1

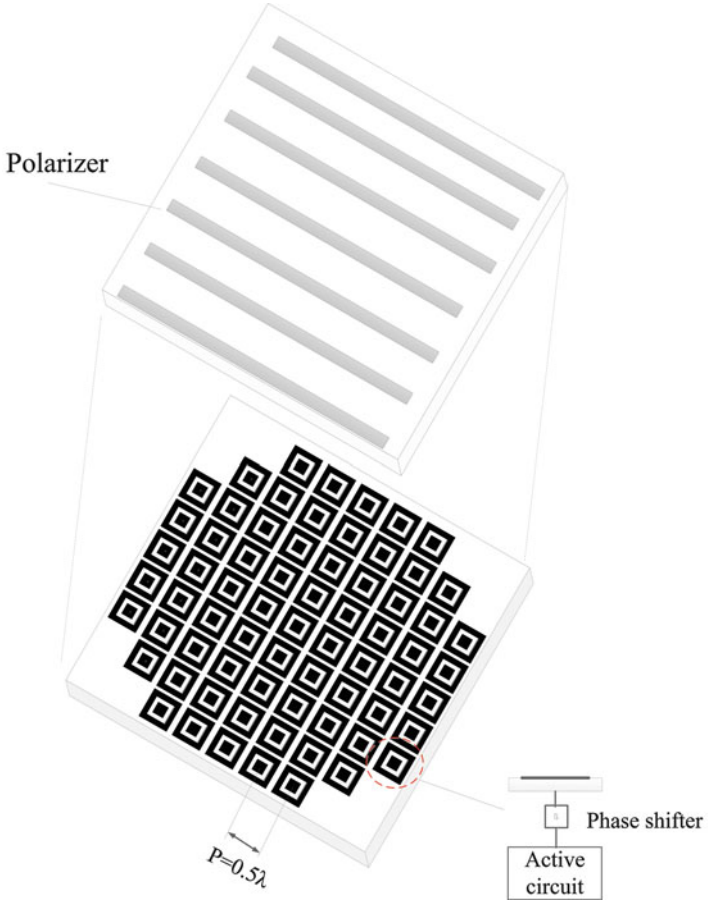


Fig. 19 The configuration of the folded reflectarray presented in Luo et al. (2013)

After computing the Floquet impedance of the array element with different scan angles, the scan blindness angle can be identified by locating the angle where the input impedance of the antenna exhibits a large value (poor impedance matching). At the scan blindness angle, the antenna array does not radiate, and almost all of the input RF energy is reflected back.

Design Example

One Ka-band folded reflectarray with large scan angle for SatCom on the move has recently been presented in Luo et al. (2013). This reflectarray can scan its beam to ±60° in the hemisphere with an appropriate active microwave circuitry. The configuration of this folded reflectarray is shown in Figure 19. As seen, this reflectarray consists of a planar microstrip slot antenna array and a dielectric polarizer. Behind each radiating element, its interface allows it to incorporate the transmit/receive unit with phase/amplitude control.

The unit cell of this reflectarray is based on a dual-polarized square-ring slot antenna fed by striplines, as shown in Figure 20a. The dual linear polarized array unit cell is used to rotate the E-field of the reradiated waves from the slot antenna by 90°, as required for the folded reflectarray design (Menzel and Kessler 2009). In order to integrate with the MMIC that lies below each of the radiating element, the RF signal is transmitted through a stripline to coaxial-like structure with shielding vias. For the purpose of characterizing the array elements, passive prototypes were developed, and stripline to microstrip line transitions are introduced, as shown in Figure 20b.

DOI 10.1007/978-981-4560-75-7_103-1

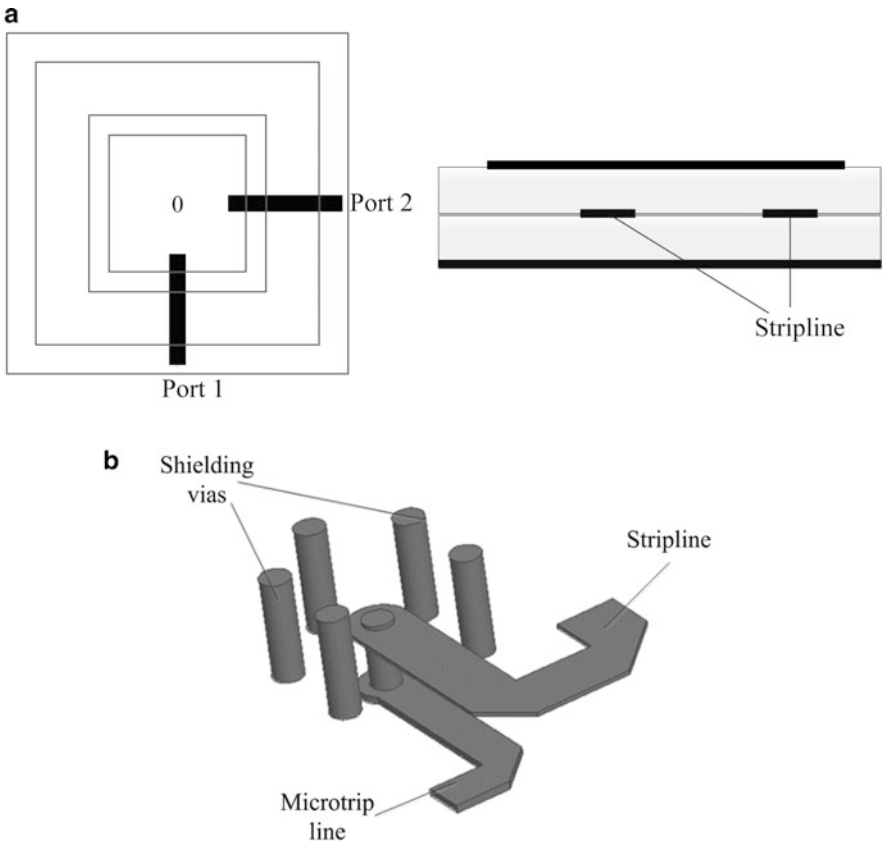


Fig. 20 (**a**) The top and side view of the reflectarray unit cell; (**b**) the stripline to microstrip line transition (Luo et al. 2013)

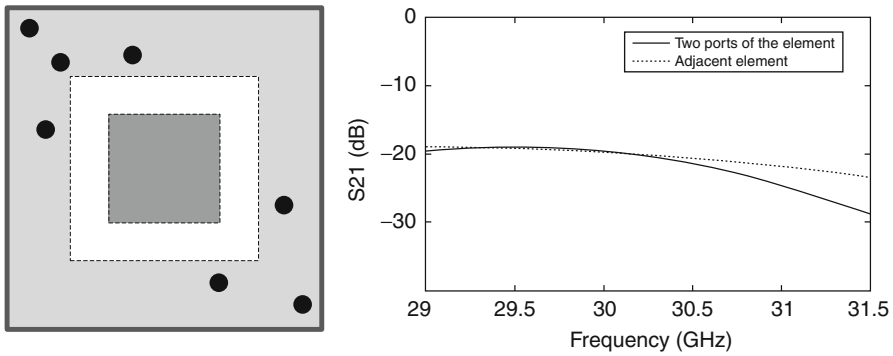


Fig. 21 The positions of the via on the array element and the simulated isolation between the two ports of the antenna as well as the lowest isolation between two adjacent elements in an infinite size array

The distance between each unit cell is kept to half of its free space wavelength at 30 GHz, in order to avoid the grating lobes during the beam scanning. Meanwhile, to make sure the array can steer the beam to large angles, it is important to suppress the propagation of the surface wave and keep a high isolation between the two input ports of the array element as well as the ports between the adjacent radiating element. To achieve this, vias were introduced near the radiating slots. Figure 21 shows the positions of the vias on the radiating element and the simulated isolation between the two ports of the array unit cell as well as the isolation between adjacent antenna elements analyzed by using the Floquet modes. As can be seen, the isolation between the two ports of the antenna is always higher than 20 dB, and the lowest

DOI 10.1007/978-981-4560-75-7_103-1

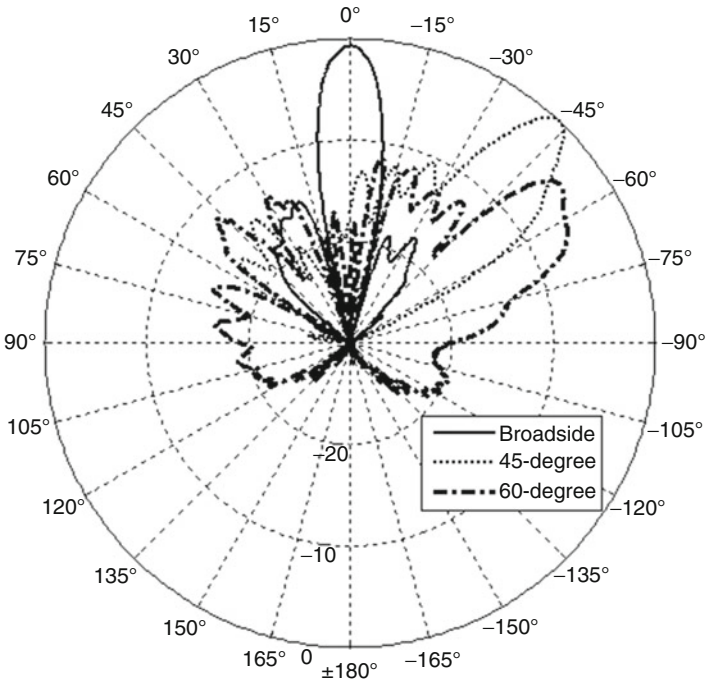


Fig. 22 The measured radiation patterns of the folded reflectarray with beam pointing at 0°, 45°, and 60° in E-plane

isolation between two adjacent elements is higher than 19 dB. These high isolations ensure that this array can scan its beam to large angles without any scan blindness.

The required phase for each radiating element in order to have its focused beam steered to certain angle is calculated by using the formula (Huang and Encinar 2007):

$$\Phi_R = k_0(d_i - (x_i \cos\varphi_b + y_i \sin\varphi_b)\sin\theta_b)$$

where Φ_R is the phase of the reflection coefficient of the antenna element i, k_0 is the phase constant in vacuum, (x_i, y_i) are the coordinates of the array element i, d_i is the distance from the phase center of the feed to the antenna unit cell, and (θ_b, φ_b) is the expected scan angle of the beam in the spherical coordinate system.

Several passive demonstrators were fabricated. The measured radiation patterns of the passive reflectarray at 30GHz with fixed beam steered to different angles are shown in Figure 22. As can be seen from these measurement results, the radiation performances of the passive reflectarray demonstrators exhibit low sidelobes with frozen beam pointing at different scan angles up to 60°. The measured maximum gain at broadside is 21.7 dBi, which represents 42 % aperture efficiency. Because of having a symmetrical structure, similar results were observed for the measured radiation patterns in H-plane.

To achieve an active phased array, it is important to incorporate the active microwave circuit with beam-forming network. MMIC devices are normally employed for such applications. Due to the small size of the Ka-band array element, which is only 5 mm × 5 mm, it requires combining several transmit/receive (T/R) modules on one chip. It is impossible to address the details of the microwave circuit design in this chapter, but readers can refer to Valenta et al. (2012) for more information.

From this example, it can be seen that the complexity of designing an adaptive beam-steerable smart antenna is high, for both the array antenna and the microwave circuit design. Since a large number of radiating elements are usually employed for such application, using Floquet modal functions provides an

DOI 10.1007/978-981-4560-75-7_103-1

Table 1 Common frequency bands for satellite communication (Amyotte and Martins Camelo 2012)

	Uplink	Downlink
L-band	1.62–1.66 GHz	1.52–1.56 GHz
S-band	~2 GHz	~2.2 GHz
C-band	5.5–6.5 GHz	3.5–4.2 GHz
Ku-band	13–14.5 GHz	10.7–11.7 GHz
Ka-band	29.5–31.2 GHz	19.5–21.2 GHz

Table 2 Specifications of some non-geostationary satellites systems (Radiocommunications-Agency-UK)

Constellation	Lugos	Skybridge
Approximate number of satellite	161	64
Uplink frequency band	29.1 GHz	14 GHz
Downlink frequency band	19.3 GHz	11 GHz
Uplink terminal antenna gain	32.26 dBi	32.2 dBi
Downlink terminal antenna gain	33.69 dBi	30.6 dBi

effective method to analyze the radiation performance of the array. In the next section, the applications of smart antenna in satellite communications will be presented with practical design examples.

Application of Smart Antennas in SatCom

Smart Antennas for Satellite Earth Station Systems

Satellites can be categorized as geostationary and non-geostationary satellites according to the orbits that they are operating. The geostationary satellites, which are in a circular orbit directly above the equator of the earth, have positions in the sky that remains the same for the ground station stationary observer on earth. In this case, the antenna for the ground station does not need to track the satellites, and the beam steering for the satellite ground station is not required as the satellites are in relative fixed locations. For such application, conventional reflector antennas can provide a low-cost solution with promising radiation performance. For the non-geostationary satellites, such as satellites in elliptical geosynchronous orbits or in low Earth orbit (LEO), it is required that the antenna systems for the ground stations are able to track the satellites.

Therefore, the antennas for the satellite earth station application are usually required to satisfy the following technique specification:

1. Have high gain with low sidelobes: to ensure good data connection and minimize the unwanted interference
2. Be able to track the non-geostationary satellites and provide wide angular coverage
3. Use circularly polarized antenna

Table 1 lists some of the frequency bands that have been defined for the satellite communication.

Table 2 summarizes the characteristics of some non-geostationary satellite systems.

Intelsat has defined many standards for the satellite earth stations. Each standard corresponds to different types of services provided by the earth station. Some of the Intelsat Earth Station Standards (IESS) are listed in Table 3. More standards with detailed definitions can be found in Intelsat (2005, 2006).

DOI 10.1007/978-981-4560-75-7_103-1

Table 3 Intelsat Earth Station Standards (Intelsat 2005, 2006)

Earth station standard	Frequency band	G/T
A	6/4 GHz	35 dB/K
B	6/4 GHz	31.7 dB/K
C	14/11 GHz and/or 14/12 GHz	37 dB/K
E-1	6/4 GHz	25 dB/K
E-2	6/4 GHz	29 dB/K
K-2	14/11 GHz and/or 14/12 GHz	19.8 dB/K
K-3	14/11 GHz and/or 14/12 GHz	23.3 dB/K

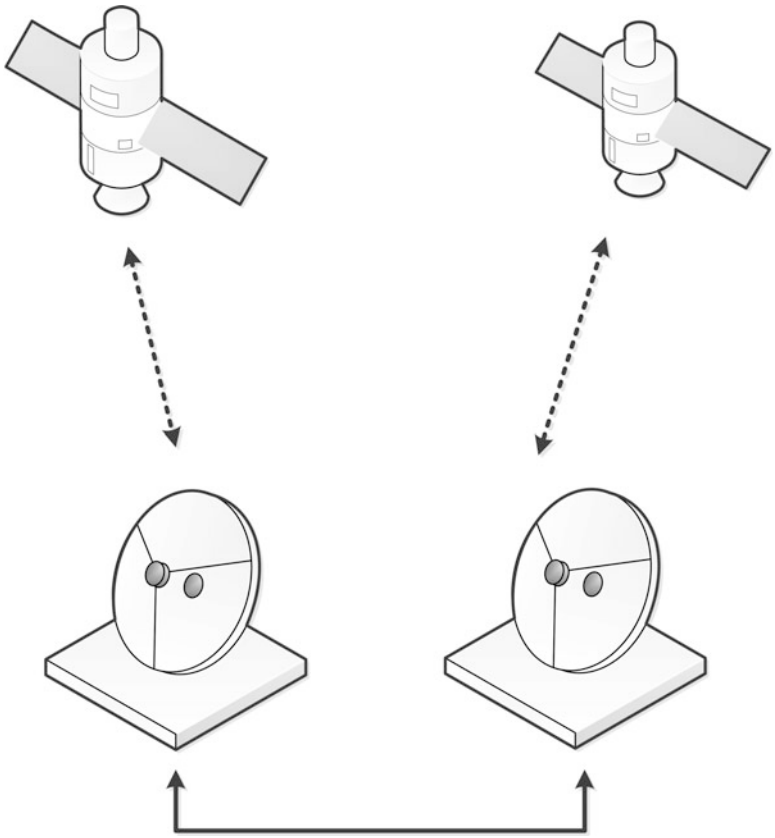


Fig. 23 Using multiple reflector antennas to communicate with multiple satellites

Conventionally, reflector antennas are employed for the satellite earth station, as its advantages of high gain, broad operation bandwidth, and low cost. However, they are always bulky and can only track the satellites by mechanical movement, which makes it impossible for them to communicate with multiple satellites simultaneously or perform seamless handover unless more than one reflector are employed, as shown in Figure 23.

Compared to the reflector antennas, smart antennas have the capability to scan the beam electronically with high pointing accuracy. The beam of the array can be steered in microseconds; thus, with only one smart antenna, different satellites can be tracked, and the handover can be done seamlessly. Figure 24 illustrates the scenario that uses one smart antenna array to communicate with multiple satellites.

It is known that with a planar phased array, the gain of the array decreases at larger scan angles in the elevation plane, as the results of the reduction in the effective radiation aperture, and it cannot cover the

DOI 10.1007/978-981-4560-75-7_103-1

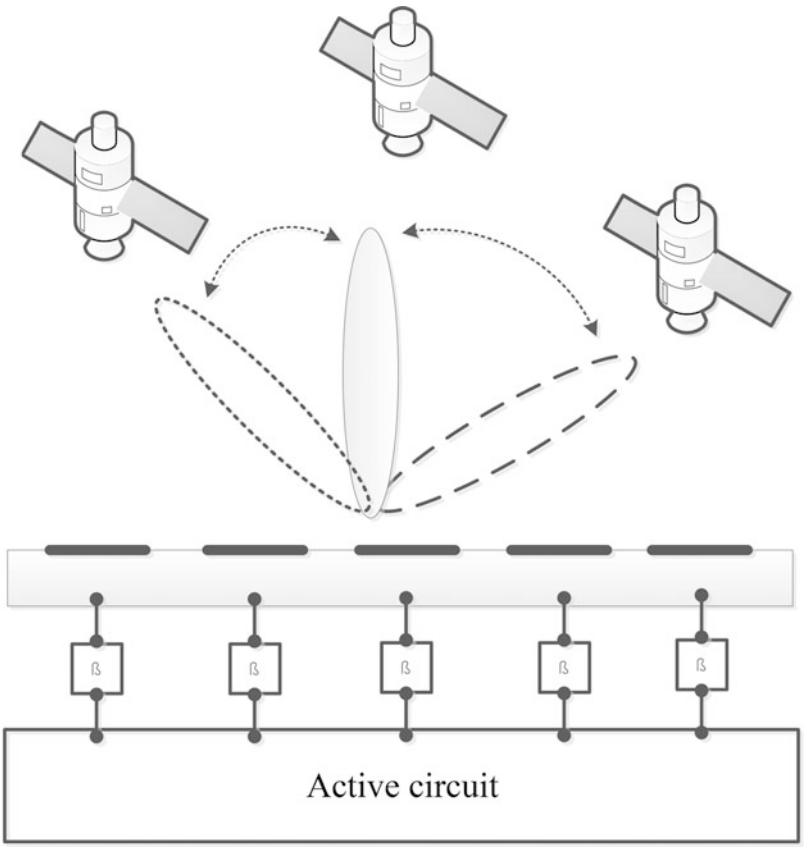


Fig. 24 Illustration of one smart antenna array communicates with multiple satellites

entire azimuth plane. Therefore, in order to have a smart antenna that can steer the beam in the entire hemisphere, faceted array and hemispherical array can be employed (Josefsson and Persson 2006). The advantage of using such arrays is that the array facets can be selected in order to enhance the antenna performance in certain parts of the overall coverage region. For either the hemisphere array or the faceted array, the number of active radiating elements can be controlled so the array with different beamwidth and directivity can be formed based on the needs of the mission. Narrower beamwidth can be achieved at small elevation angles because more radiating elements can be activated due to the geometrical configuration. Figure 25 shows the calculated radiation patterns of a hemisphere array and a faceted array with beam pointing at different elevation angles (Yao et al. 2012). It can be seen that a hemisphere array and a faceted array can steer the beam over large angle ranges with similar radiation patterns. The beam steering of the array is achieved by taking advantage of its physical geometry. During the beam steering, only a small number of the antenna elements are radiating, and the rest of array elements are deactivated, which results in low aperture efficiency of the array antenna.

One important issue associated with the fully electronic beam-steering smart array antenna is the complexity of the RF circuit, especially when the number of the radiating element is large. Besides the complexity in the circuit design, the use of large amount phase shifters, the size of the feed network, and effective system thermal dissipation can increase the cost and reduce the radiation efficiency of the overall antenna system. Thus, until now there are few available commercial products equipped with a fully electronic beam-steering smart array antenna.

To reduce the cost while maintaining reasonable system efficiency, a hybrid system provides a compromised solution. Figure 26 shows the configuration of an m × n microstrip antenna array. Instead of incorporating one phase shifter for each of the antenna element, the array is divided into *m* linear

DOI 10.1007/978-981-4560-75-7_103-1

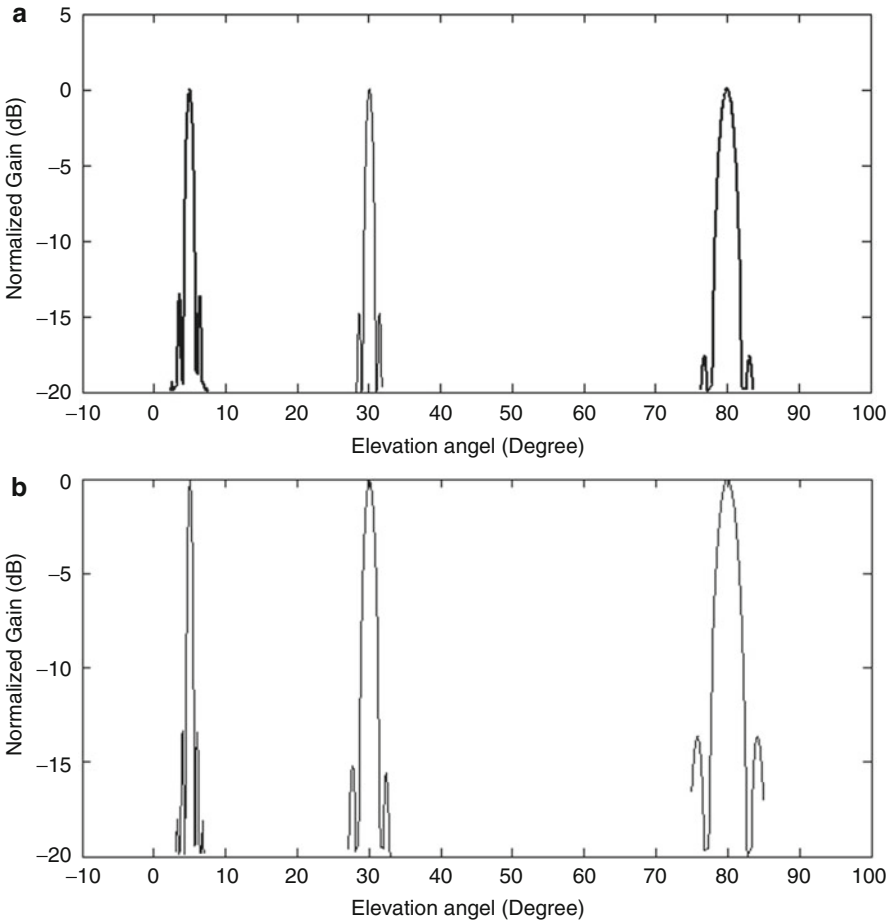


Fig. 25 The calculated radiation patterns of the (**a**) hemisphere array and (**b**) facet array with beam pointing at 10°, 30°, and 50°. There are totally 82,177 radiating elements in the hemisphere array and 40,100 radiating elements in the facet array (Yao et al. 2012)

subarray arrays, and the phase shifting is implemented at the subarray level. Thus, the amount of the phase shifters is reduced from $m \times n$ to m. However, with this configuration the array is only able to scan in one direction. By introducing a mechanical steerable platform, beam scanning within the entire hemisphere would be possible.

In the applications where fast tracking is not required, using mechanically beam-steerable smart antenna provides the most economical solution. Figure 27 demonstrates one planar array placed at a platform that can be rotated in both azimuth and elevation plane. In this configuration, there is no need of using any phase shifters, and the beam steering is controlled by the corresponding motors.

Figure 28 shows a photo of the commercial available terminal for the aero mobile applications. This array is designed for Ka-band application, and dual-polarized (RHCP and LHCP) waveguide horn is used as the radiating element. It operates at 28.1 to 30.0 GHz for transmission and 18.3 to 20.2 GHz for receiving. The antenna system consists of an antenna array, control unit, and programmable servomotors that can be used for azimuth and elevation position controlling. It is reported that this antenna has elevation coverage from 0° to 75° and 360° continuous coverage in azimuth plane. More than 2,500 radiating elements are used, and the antenna array has EIRP (equivalent isotropically radiated power) of 43.5 dBW after taking account of the radome loss.

DOI 10.1007/978-981-4560-75-7_103-1

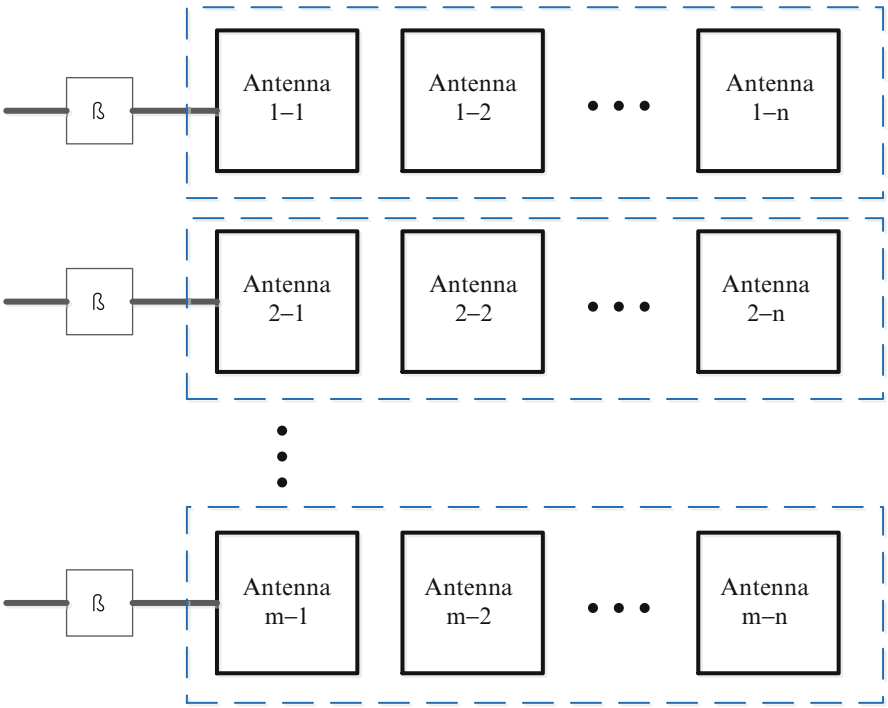


Fig. 26 Illustration of using reduced number of phase shifters for a planar array

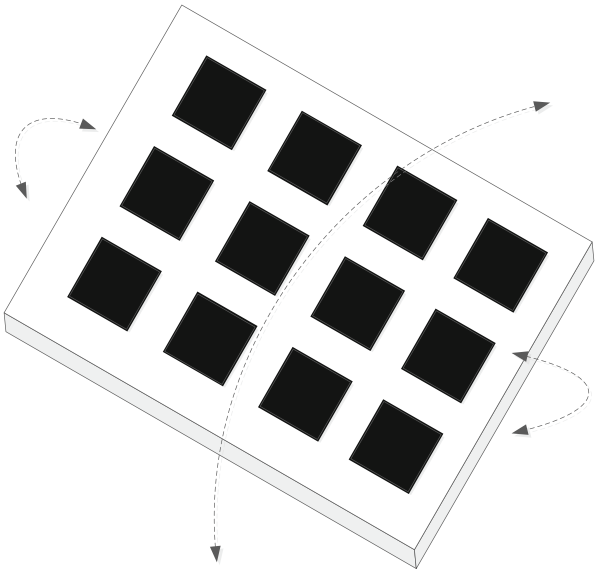


Fig. 27 A planar array placed at a platform that can be rotated in both azimuth and elevation plane

Smart Antennas for Mobile Direct Broadcast Satellite Systems

The International Telecommunication Union (ITU) authorized the use of high-power Ku-band for the satellite-to-home video program during 1979 and 1989. These services are the so-called direct broadcast satellite (DBS) services (Meadows 2008). DBS provides high-powered broadcast services, including television and radio services, to users in large geographic areas (Thorburn 2012). Figure 29 shows the scenario of the direct broadcasting satellite system, where one broadcast center uplinks signals to the satellite and then the satellite retransmits the signals to multiple users in the areas covered by the satellite

DOI 10.1007/978-981-4560-75-7_103-1

Fig. 28 The photo of the aero mobile terminal for satellite communication at Ka-band developed by Viasat ©2014 Viasat (Reprinted with permission from Viasat)

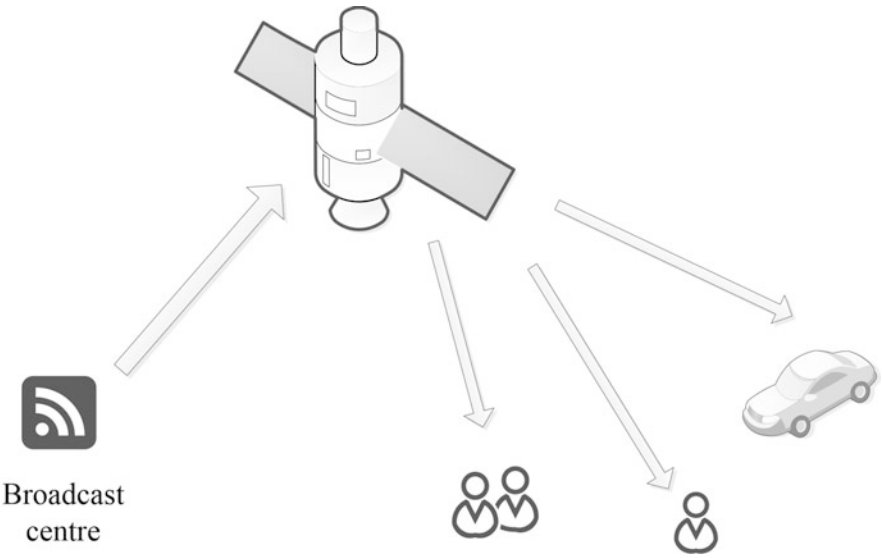


Fig. 29 Illustration of the direct broadcasting satellite system

antenna system. The frequency bands allocated for the DBS are 11.7–12.2 GHz in ITU Region 3 (Asia and Australia), 10.7–12.75 GHz in ITU Region 1 (Europe, Russia, and Africa), and 12.2–12.7 GHz in ITU Region 2 (North and South America).

To ensure good reception from the broadcasting satellites, the DBS antenna is required to have a gain of more than 21dBi and have a beam with a certain title angle (Hirokawa et al. 1995; Watanabe et al. 1996) in the elevation plane. For mobile users, such as moving vehicles, the antenna system is also required to be able to track the DBS satellites.

To reduce the system complexity, the antenna array can be designed to exhibit relatively broad beamwidth in the elevation plane and place the array with certain inclination angle (Soon-Ik et al. 2000). The beam steering in the azimuth plane can be realized by rotating the array mechanically (e.g., by a motor), as demonstrated in Figure 30. It is reported that with a beamwidth of about 12°, even without the beam steering, promising reception can still be reached in Japan (Hirokawa et al. 1995; Watanabe et al. 1996).

Mechanical steering is a low-cost solution, but it suffers from the low tracking speed. For the system presented above, to track the broadcasting satellite, about 2 s are needed (Hirokawa et al. 1995). It is shown that for the real-time road tests, the turning speed of a medium size car can be as fast as 60°/s with an angular acceleration of up to 85°/s^2 (Mousavi et al. 2009). Thus, more accurate tracking of the satellite

DOI 10.1007/978-981-4560-75-7_103-1

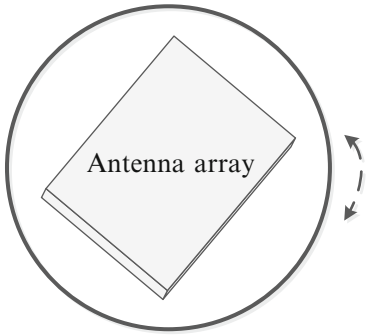


Fig. 30 Employing a planar array on a rotating base that steers the beam in the azimuth plane

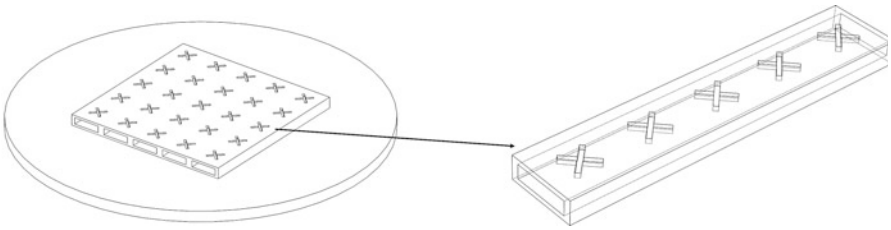


Fig. 31 Configuration of the smart antenna for DBS consists of wave guide slot array presented in Wang and Winters (2004)

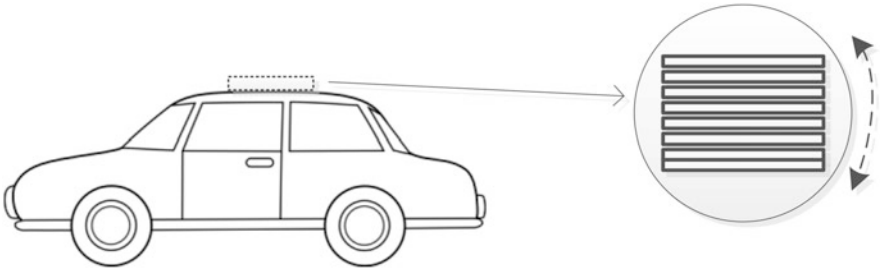


Fig. 32 Demonstration of mounting of the smart antenna presented in Wang and Winters (2004) on the roof of the car

is needed for better services. In this case, electronic beam steering is desired. However, similar to the smart antenna for ground station, a full electronic beam-steering array can lead to a complex and high-cost system. Therefore, the hybrid design that combines both the electronic and mechanical beam steering provides a relatively low-cost solution with improved system performance.

A vehicle-mounted satellite antenna for mobile DBS application is described in Wang and Winters (2004). The look angle of the antenna is set to about 40° in the elevation plane, and the beams can be electronically steered by ±15°. In the azimuth plane, mechanical steering is employed. There are two reasons for such configuration. Firstly, the tracking accuracy needs to be higher in elevation than azimuth plane. Secondly, it can keep a low profile for the overall antenna system, so it is suitable for mounting on the roof of the vehicle. This antenna consists of ridge waveguides, on which cross-shaped slots were uniformly spaced, as shown in Figure 31. By optimizing the positions and the length of the cross-shaped slot, circularly polarized radiation can be achieved. As a waveguide slot array, it can radiate different polarized waves in forward and backward direction (Gao et al. 2014). Thus, either LHCP or RHCP can be obtained.

Figure 32 demonstrates the mounting of this DBS smart antenna on the roof the car. This smart antenna has a low profile with height of less than 76 mm.

DOI 10.1007/978-981-4560-75-7_103-1

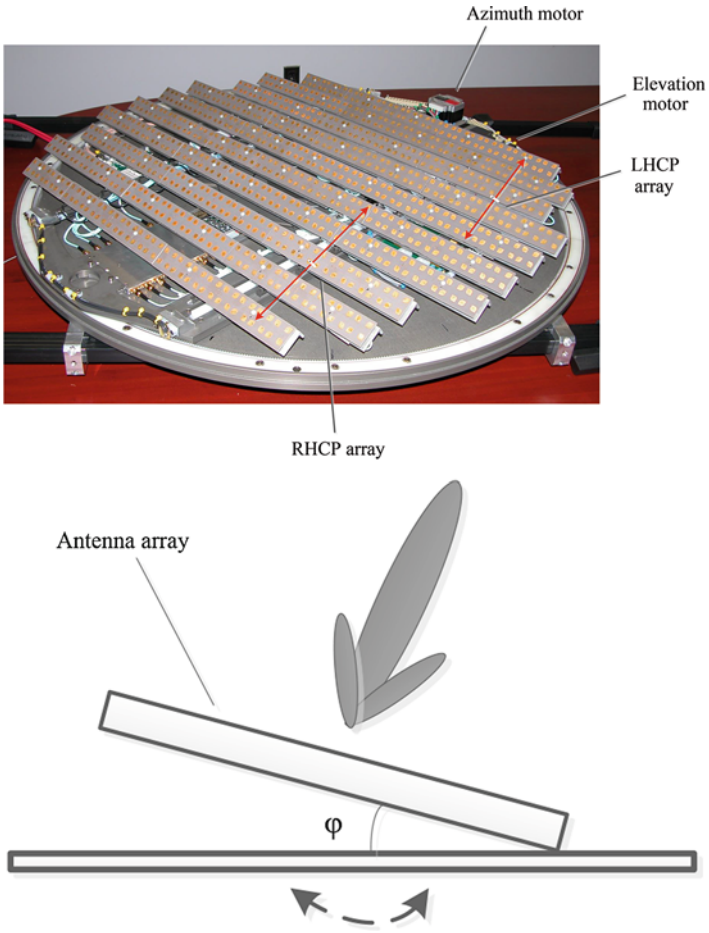


Fig. 33 The phased array antenna for DBS application reported in Mousavi et al. (2010) (Courtesy of Centre for Intelligent Antenna and Radio Systems (CIARS), Canada)

Recently, a smart antenna array based on microstrip printed antenna and subarray technique has been presented in Mousavi et al. (2008, 2010). Mechanical rotation at both azimuth and elevation planes is employed, while the array can additionally scan the beam electronically in the elevation plane. Thus, this smart antenna can provide a wide-angle reception for the DBS satellite reception. 2×2 linearly polarized patch was arranged in a sequentially rotated position to generate the CP radiation. There are 17 subarrays used for either RHCP or LHCP operation, and for each polarization, the antenna arrays are placed in five consecutive panels with certain tilt angle, as shown in Figure 33. Since the array antenna is fabricated by using PCB techniques, it exhibits lighter weight with less fabrication complexity compared to the waveguide-based arrays.

Figure 34 shows the system configuration of this smart antenna system. There are two motors that can be controlled to mechanically rotate the antenna array in both azimuth and elevation planes. Each subarray is connected to a voltage-controlled analogue phase shifter via an LNA and low-loss cable. The phase shifters can be controlled by the voltages, which can be adjusted by the beam-forming algorithm, applied through the digital to analogue converter (DAC) board. For either LHCP or RHCP polarization, a power combiner is used to combine the outputs of the phase shifters. Novel model-free beam-forming algorithms (Fakharzadeh et al. 2009) are employed to track the desired satellite and compensate the inaccuracy from the fabrication errors. It is reported in Mousavi et al. (2010) that with these algorithms, this smart antenna can maintain the received power level above the 90 % in any road test circumstances.

DOI 10.1007/978-981-4560-75-7_103-1

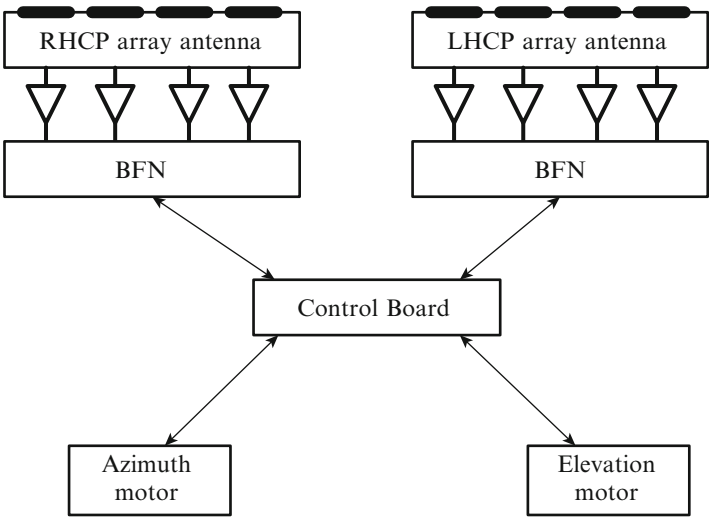


Fig. 34 The system configuration of the smart antenna reported in Mousavi et al. (2010)

Future Directions and Open Problems

Smart antennas are important for satellite communications. They can increase channel capacity, track satellites electronically, and avoid interference signals. These advantages over conventional antennas make them attract much research interest. In this chapter, a review on smart antenna technologies and design examples, including one beam-switching CP array and one reflectarray with large scan angle range, are presented. The beam-switching smart array provides a low-cost solution with a limited number of fixed beams, while the adaptive smart array can steer the beam continuously with higher system complexity and cost. The applications of smart antennas for satellite ground stations and mobile direct broadcast services are also introduced in this chapter. Different solutions with design examples as well as commercial applications are discussed and presented.

The future work for the smart antenna design involves the design of fully electronically scanned smart array antenna for satellite communications at Q-band (33–50 GHz), V-band (50–75 GHz), and W-band (75–110 GHz), as the need for more spectrum and satisfying the growing communication capacity requirements (Acosta 2012). Although 60-GHz antenna arrays with multilayered configuration and fabricated by using low-temperature co-fired ceramic (LTCC) technology have been reported recently (Hucheng et al. 2013; Lei et al. 2013), these presented arrays cannot perform beam scanning. The challenge at these higher frequency bands is the following: In order to keep the beam-scanning capability and maximize the largest scan angle of the smart antenna without any grating lobes and scan blindness, the maximum dimension of the array element is approximately half the free space wavelength at the frequency of interest. In order to make the array antenna be scalable, it is desirable to have an independent MMIC below each array element. Thus, each radiating element can be treated as an intelligent pixel. This also gives challenges to the MMIC and beam-forming network design as the size of the MMIC is constrained by the size of the radiating element. Moreover, the integration of the overall system and the heat dissipation need to be addressed as well.

Terahertz antennas also attract much research interests in recent years, and antenna array for operation at terahertz frequency is anticipated. Electronically controllable phase shifters operating at THz frequencies are critical components for a terahertz phased antenna array. Terahertz antenna phase shifter designs, including using electrically driven THz metamaterials (Chen et al. 2009) and integrally gated graphene transmission lines (Pai-Yen et al. 2013), have been investigated. Since terahertz frequencies are between

DOI 10.1007/978-981-4560-75-7_103-1

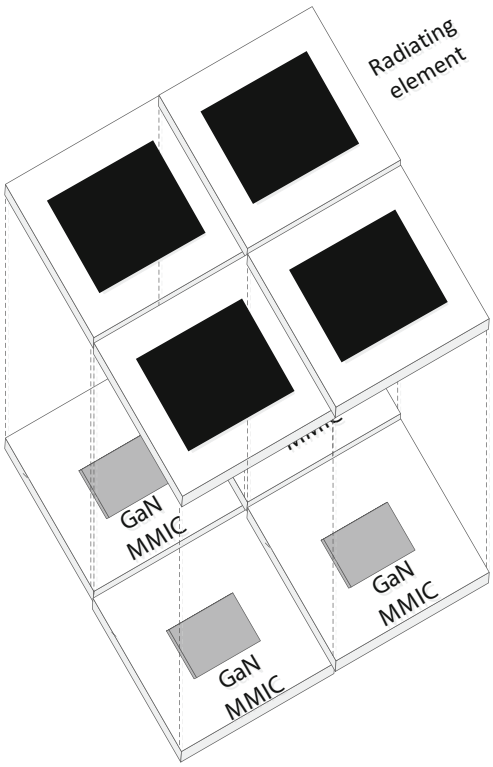


Fig. 35 The concept of the intelligent pixel with GaN-based MMIC

the microwave and optical frequencies, efficient fabrication techniques with low antenna loss need to be further explored, especially when many radiating elements are required to form a high-gain array antenna.

Another direction for the future work is to incorporate gallium nitride (GaN)-based MMIC with high power amplifiers (HPA) under each radiating element, and then the power of amplifiers can be combined in space by the smart array antenna, as shown in Figure 35. The material properties of GaN exhibit revolutionary performance for millimeter wave power amplifiers in many aspects compared to conventional gallium arsenide (GaAs) technology, including the output power, power-added efficiency (PAE), operation bandwidth, signal linearity, gain flatness, and noise level (Oki et al. 2013). In term of applying the GaN technology in the smart antenna design, one of the challenges for this concept is how to efficiently dissipate the heat generated by all the HPAs.

Using new materials, such as graphene/carbon nanotube (Wu et al. 2014) brings new perspectives to the future antenna and antenna array designs. One antenna-enhanced optoelectronic microscopy study on nanoscale devices has been presented recently (Mauser et al. 2014), but there are few works in using these new materials to design smart array antennas up to now. Thus, more investigations are expected in this new field.

References

Acosta R, Nessel J, Simons R, Zemba M, Morse J, Budinger J (2012) W/V-Band RF propagation experiment design. 8th Ka and broadband communication conference, Ottawa

Amyotte E, Martins Camelo L (2012) Antennas for satellite communications. In: Space antenna handbook. Wiley, Chichester, pp 466–510

AVX-corporation. Introduction to microwave capacitors. Retrieved April 2014

Bachman HL (1996) Intelligent antennas – an emerging technology. Southcon/96. Conference record
Balanis CA (2005) Antenna theory: analysis and design. Wiley, Hoboken
Balanis CA, Ioannides PI (2007) Introduction to Smart Antennas, Morgan & Claypool, USA
Bayraktar O, Topalli K, Unlu M, Civi OA, Demir S, Akin T (2007) Beam Switching Reflectarray using RF MEMS Technology. The Second European Conference on Antennas and Propagation, Edinburgh, UK, pp.1,6, 11–16
Bhattacharyya A (2006) Phased array antennas: floquet analysis, synthesis, BFNs, and active array systems. Wiley-Interscience, Hoboken
Butler J, Lowe R (1961) Beam-forming matrix simplifies design of electronically scanned antennas. Electron Des 9:170–173
Chaloun T, Menzel W et al (2014) A wide-angle scanning active transmit/receive reflectarray. IET Microwaves Antennas Propag 8(11):811–818
Changrong L, Shaoqiu X et al (2011) Circularly polarized beam-steering antenna array with butler matrix network. IEEE Antennas Wirel Propag Lett 10:1278–1281
Chao-Hsiung T, Chih-Jung C et al (2008) A low-cost 60-GHz switched-beam patch antenna array with butler matrix network. IEEE Antennas Wirel Propag Lett 7:432–435
Chen H-T, Padilla WJ et al (2009) A metamaterial solid-state terahertz phase modulator. Nat Photonics 3(3):148–151
Chia-Chan C, Ruey-Hsuan L et al (2010) Design of a beam switching/steering butler matrix for phased array system. IEEE Trans Antennas Propag 58(2):367–374
Chryssomallis M (2000) Smart antennas. IEEE Antennas Propag Mag 42(3):129–136
Elhefnawy M, Ismail W (2009) A microstrip antenna array for indoor wireless dynamic environments. IEEE Trans Antennas Propag 57(12):3998–4002
Fakharzadeh M, Jamali SH et al (2009) Fast beamforming for mobile satellite receiver phased arrays: theory and experiment. IEEE Trans Antennas Propag 57(6):1645–1654
Fan Fan H, Ke W et al (2012) Low-cost 60-GHz smart antenna receiver subsystem based on substrate integrated waveguide technology. IEEE Trans Microwave Theory Tech 60(4):1156–1165
Gao S, Luo Q et al (2014) Circularly polarized antennas. Chichester, UK
Grant AE, Meadows JH (2008) Communication technology update and fundamentals. Burlington, MA, USA
Han W, Zhijun Z et al (2012) A beam-switching antenna array with shaped radiation patterns. IEEE Antennas Wirel Propag Lett 11:818–821
Hansen RC (2009) Phased array antennas. Wiley, Hoboken
Hirokawa J, Ando M et al (1995) A single-layer slotted leaky waveguide array antenna for mobile reception of direct broadcast from satellite. IEEE Trans Veh Technol 44(4):749–755
Huang J, Encinar JA (2007) Reflectarray antennas. Hoboken
Hucheng S, Yong-Xin G et al (2013) 60-GHz circularly polarized U-Slot patch antenna array on LTC-C. IEEE Trans Antennas Propag 61(1):430–435
Intelsat (2005) Intelsat Earth Station Standards (IESS) Document IESS-101 (Rev.61)
Intelsat (2006) Intelsat Earth Station Standards (IESS) Document IESS-208 (Rev.6)
Jenshan L, Itoh T (1994) Active integrated antennas. IEEE Trans Microwave Theory Tech 42(12):2186–2194
Josefsson L, Persson P (2006) The shapes of conformal antennas. In: Conformal array antenna theory and design. Wiley, Hoboken, New Jersey, pp 49–71
Jun OY (2011) A circularly polarized switched-beam antenna array. IEEE Antennas Wirel Propag Lett 10:1325–1328

Kalis A, Kanatas AG, Papadias CB (2014) Parasitic antenna arrays for wireless Mimo Systems. Springer, New York

Lei W, Yong-Xin G et al (2013) Wideband high-gain 60-GHz LTCC L-probe patch antenna array with a soft surface. IEEE Trans Antennas Propag 61(4):1802–1809

Li RC-H (2012) RF circuit design. Wiley, Hoboken

Luo Q, Gao S et al (2013) Antenna array elements for Ka-band satellite communication on the move. Antennas and propagation conference (LAPC), 2013 Loughborough, pp 135–139

Luo Q, Gao S et al (2014) Intelligent antenna technology for mobile satellite communications. e & i Elektrotechnik und Informationstechnik 131(6):155–160

Mailloux RJ (2005) Phased array antenna handbook. Artech House, Norwood, MA

Maqsood M, Gao S et al (2014) Low-cost dual-band circularly polarized switched-beam array for global navigation satellite system. IEEE Trans Antennas Propag 62(4):1975–1982

Mauser N, Hartmann N et al (2014) Antenna-enhanced optoelectronic probing of carbon nanotubes. Nano Lett 14(7):3773–3778

Menzel W, Kessler D (2009) A folded reflectarray antenna for 2D scanning. German microwave conference, Munich, German, 2009

Moulder WF, Khalil W et al (2010) 60-GHz two-dimensionally scanning array employing wideband planar switched beam network. IEEE Antennas Wirel Propag Lett 9:818–821

Mousavi P, Fakharzadeh M et al (2008) A low-cost ultra low profile phased array system for mobile satellite reception using zero-knowledge beamforming algorithm. IEEE Trans Antennas Propag 56(12):3667–3679

Mousavi P, Fakharzadeh M et al (2009) A low cost 1K elements phased array antenna. IEEE MTT-S international Microwave symposium digest, Boston, MA

Mousavi P, Fakharzadeh M et al (2010) 1K element antenna system for mobile direct broadcasting satellite reception. IEEE Trans Broadcast 56(3):340–349

Oki AK, Wojtowicz M et al (2013) GaN technology for millimeter wave power amplifiers. ECS Trans 50(3):307–311

Pai-Yen C, Argyropoulos C et al (2013) Terahertz antenna phase shifters using integrally-gated graphene transmission-lines. IEEE Trans Antennas Propag 61(4):1528–1537

Patterson CE, Khan WT et al (2012) A 60-GHz active receiving switched-beam antenna array with integrated butler matrix and GaAs amplifiers. IEEE Trans Microwave Theory Tech 60(11):3599–3607

Pilz D, Menzel W (1998) Folded reflectarray antenna. Electron Lett 34(9):832–833

Radiocommunications-Agency-UK [online], study into phased arrays in NGSO earth station antennas, retrieved May 2014

Soon-Ik J, Young-Wan K et al (2000) A new active phased array antenna for mobile direct broadcasting satellite reception. IEEE Trans Broadcast 46(1):34–40

Stark A, Dreher A et al (2009) SANTANA: Advanced electronically steerable antennas at Ka-Band. Antennas and propagation, 2009. Berlin, German. 3rd European conference on

Thorburn MA (2012) System architectures of satellite communication, radar, navigation and remote sensing. In: Space antenna handbook. Wiley, Chichester, pp 76–105

Valenta V, Kaynak M, Liu G, Çagrı Ulusoy A, Purtova T, Trasser A, Schumacher H (2012) From enabling technology to applications: reconfigurable SiGe BiCMOS ICs with fully integrated RF MEMS switches for millimetre wave transceivers. ESA microwave technology and techniques workshop

Wang J, Winters JH (2004) An embedded antenna for mobile DBS. Vehicular technology conference, Los Angeles, CA, USA, 2004. VTC2004-Fall. 2004 I.E. 60th

Watanabe T, Ogawa M et al (1996) Mobile antenna system for direct broadcasting satellite. Antennas and Propagation Society international symposium, 1996. AP-S. Digest

DOI 10.1007/978-981-4560-75-7_103-1

Wu J, Shi W et al (2014) Optical properties of gold/multilayer-graphene/carbon nanotube hybrid materials. Carbon 68:708–717

Yao C, Nuan S et al (2012) Satellite ground stations with electronic beam steering. Satellite telecommunications (ESTEL), 2012 I.E. first Rome conference on

Zhang C, Zhou D et al (2012) Investigation of dual-polarised printed slot antenna for folded reflectarray unit cells at Ka band. Communication systems, networks & digital signal processing Poznan, Poland, 2012 8th international symposium on

Zhou D, Gao S et al (2011) Investigation of mutual coupling reduction for folded refelctarray unit cells using dual-polarized microstrip antennas

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_106-1

Antennas in Access Points of WLAN/WiFi

Hiroyuki Arai*
Yokohama National University, Yokohama, Kanagawa, Japan

Abstract

This chapter presents antennas used in access points of WLAN/WiFi systems. The frequency spectrum for these systems based on the standards is described first. Antenna parameters to understand the WLAN/WiFi antennas are also explained. The antennas are classified into outdoor, indoor, and built-in antennas in this chapter, which are presented in detail.

Keywords

WLAN; WiFi diversity; MIMO; Access point; Yagi-Uda array; Reflector antenna; Microstrip antenna; Collinear array; Sleeve antenna; Dipole antenna; Monopole antenna; L-shaped antenna; Inverted-F antenna

Introduction

This chapter presents antennas in access points of WLAN and WiFi. The antenna specifications, frequency and antenna gain, are summarized in the following section, and the definitions of antenna parameters are also described. Recent high-speed wireless terminal use MIMO (multiple input and multiple output) system and then its evaluation factors are also defined. Section "WLAN/WiFi Antennas" describes antennas used in WLAN and WiFi systems, classified into outdoor, indoor, and built-in antennas.

Specifications of WLAN/WiFi Antennas

Wireless local area network (LAN) is designated as WLAN, the local area network systems provided by electromagnetic waves. The WLAN specifications are created and maintained by IEEE LAN/MAN standards committee (IEEE802.org), so-called IEEE802 since 1997. WiFi is a name of nonprofit organization, Wi-Fi Alliance, to promote wireless LAN products and devices. This section presents major frequency spectrums of WLAN and antenna specification in use and the antenna parameters to find the definition of technical terms.

Frequency Spectrum Allocations

Major frequency bands for WLAN are 2.4 (2.417–2.497) GHz and 5 (5.15–5.35, 5.475–5.725) GHz. For a point to point wireless communication, 25 (24.77–25.23)-GHz band is also used for outdoor applications. Millimeter frequency band of 60 (59.0–66.0) GHz is released to provide high-speed data transmission protocols up to 7 GHz bps. This standard is designated as WiGig (Wireless Gigabit). The WLAN

*Email: arai@ynu.ac.jp

DOI 10.1007/978-981-4560-75-7_106-1

access points often require several antennas to mitigate multipath fading for the IEEE 802.11b/g at 2.45 GHz and 802.11a at 5 GHz. This scheme is called as diversity antenna utilizing two reception antennas and more. To increase the data transmission speed, MIMO (multiple input and multiple output) scheme is also used in WLAN systems. The number of MIMO streams, channels to transmit different data, is 4 (802.11n) and 8 (802.11ac), which is given by the product of the number of transmitting and receiving antennas $N_t \times N_r$.

Antenna Parameters: Bandwidth, Directivity, Half-Power Beam Width, and EIRP

Required frequency bandwidth of WLAN antennas is determined by voltage standing wave ratio (VSWR) at the antenna feeding point. The VSWR is defined by a transmission line impedance Z_o and an antenna impedance Z_i as follows:

$$\mathrm{VSWR} = \frac{1 + |\Gamma|}{1 - |\Gamma|}, \Gamma = \frac{Z_i - Z_o}{Z_i + Z_o} \tag{1}$$

where Γ is the reflection coefficient at the antenna input port and the Z_o is usually 50Ω for the matching with output or input impedance of transceivers. Examples of typical VSWR value for WLAN antennas are $\mathrm{VSWR} \leq 2$ or $\mathrm{VSWR} \leq 3$. Radiated power from the antenna given by $1 - |\Gamma|^2$ is 89 % of input power for $\mathrm{VSWR} = 2$ and 75 % for $\mathrm{VSWR} = 3$.

The antenna radiation performance is evaluated by antenna gains, which are directive gain and absolute gain. The directive gain does not include losses inside the antenna and input impedance mismatch. It defines the spatial distribution of radiated energy power into space from the antenna. The absolute antenna gain G_s including all the antenna loss factors specifies the transmission power in real operating environments. The absolute gain G_s is measured by comparing the received electric field strength between the standard gain antenna and the antenna under test (AUT). The absolute gain of a standard antenna is $G_s = 2.15$ dBi for a half wavelength dipole antenna.

The antenna pattern is also characterized by half power beam width and front to back (FB) ratio. The half power beam width H_b is the angle between the half power points of the main beam as shown in Fig. 1. The FB ratio is the directivity ratio between the front and back of antenna pattern. The front ($\theta = 0°$) is the maximum radiation direction and back one is taken as an angle of the maximum radiation level within the range of $\theta = 120(180 - 60)°$ to $240(180 + 60)°$ as shown in Fig. 1.

The antenna radiation level of WLAN antennas is generally defined by effective isotropic radiated power (EIRP). It is the amount of power for an isotropic antenna, which needs to radiate the same power level for a given angle of directive antenna. To keep this specification, the input power level of directive antenna should be decreased.

Diversity and MIMO

To reduce the fading effect in the real propagation environment, diversity reception is used for wireless systems to compensate for very weak signal reception caused by multipath fading. Antenna diversity techniques are classified into space, polarization, and pattern diversity. The space diversity uses several independent signals from different antennas spaced apart, and the polarization diversity is given by receiving the normal and cross-polarization components at the same position. In addition, the antenna pattern diversity utilizes the signals from antennas with nonoverlapping antenna radiation patterns.

The evaluation factor of diversity antenna is the correlation coefficient between different outputs of the antenna. When there are two outputs for a diversity antenna, the correlation coefficient of the detected signal envelope of e_1 and e_2 is defined as

DOI 10.1007/978-981-4560-75-7_106-1

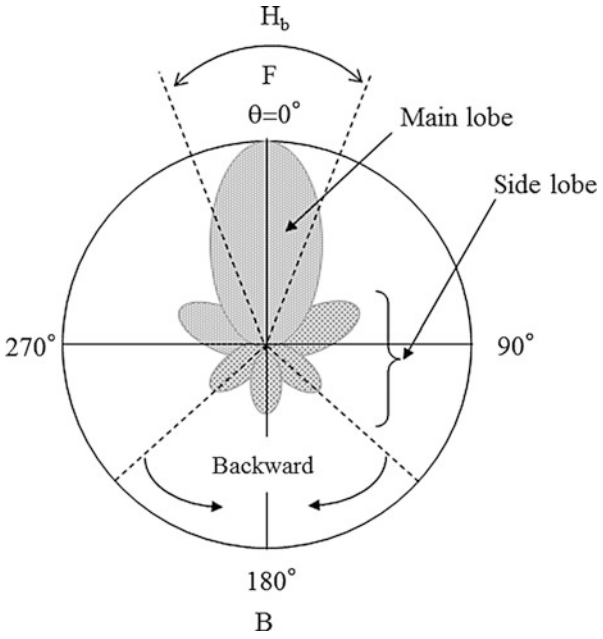


Fig. 1 Antenna pattern parameters

$$\rho = \frac{\frac{1}{2}\langle (e_1 - \langle e_1 \rangle)^* (e_2 - \langle e_2 \rangle) \rangle}{\frac{1}{2}\sqrt{\left\langle (e_1 - \langle e_1 \rangle)^2 (e_2 - \langle e_2 \rangle)^2 \right\rangle}} \tag{2}$$

where <x> denotes the mean value of parameter x and (y)* denotes the complex conjugate of the difference parameter (y). The correlation for antenna with low loss is calculated using S-parameters as follows (Thaysen and Jakobsen 2006):

$$\rho = \frac{\left| S_{11}^* S_{12} + S_{12}^* S_{22} \right|^2}{\left(1 - |S_{11}|^2 - |S_{21}|^2\right)\left(1 - |S_{12}|^2 - |S_{22}|^2\right)} \tag{3}$$

The correlation is also evaluated by complex radiation patterns of diversity antennas, which is time consuming by measuring patterns in the whole angles along azimuth and elevation coordinates. Then Eq. 3 is widely used to evaluate the diversity antennas.

The diversity performance is determined by the envelope correlation which is approximated by $\rho_e \cong \rho^2$. Small ρ_e increases the diversity gain, and its recommended critical value is $\rho_e = 0.6$ for effective diversity antenna design (Yamada et al. 1991).

MIMO performance characterized by channel capacity is approximated by McNamara et al. (2000)

$$C \cong C_{SNR} + C_{COR} = N_r log_2\left(\frac{\rho}{N_t}\right) + log_2\left\{det\left(\boldsymbol{H_n H_n^H}\right)\right\} \tag{4}$$

where N_t and N_r denote the number of the transmitting and receiving antennas and ρ is the averaged signal-to-noise ratio (SNR) at each receiving antenna. $\boldsymbol{H_n}$ is a normalized channel matrix and the superscript H represents a complex conjugate transpose. The above equation indicates that channel

DOI 10.1007/978-981-4560-75-7_106-1

capacity contains elements depending on SNR (C_{SNR}) and correlation (C_{COR}). The channel matrix characterized as det($\boldsymbol{H_n H_n}^{\boldsymbol{H}}$) is dominated by the spatial correlation coefficient γ_s and small γ_s enhances the channel capacity. To satisfy this requirement, high gains with low correlated radiation patterns are required in access point antenna designs. The channel capacity of 2×2 MIMO is saturated for $|\gamma_s| < 0.7$, and this is the critical value to evaluate the correlation (Uchida and Arai 2013).

WLAN/WiFi Antennas

This section presents WLAN antennas used in actual use. The antennas are classified into outdoor, external, and built-in antennas in this section. The antennas of WLAN/WiFi terminal antennas are discussed and terminal ones are not included.

Outdoor Antennas

The WLAN outdoor antennas are high gain and used for the point-to point wireless data communication systems. Typical high-gain antennas are Yagi-Uda array and parabolic reflector antenna. Yagi-Uda array consists of a radiating element, a reflector, and director elements as shown in Fig. 2. Its directivity gain is enhanced by increasing the number of director elements. The range of antenna gain of 2.4 GHz outdoor Yagi-Uda array is 10–19 dBi. The highest gain antennas for outdoor applications are provided by parabolic reflector antennas consisting of a grid reflector and a primary feeder as shown in Fig. 3. The grid reflector is used to reduce the load by wind pressure. The antenna gain is 24–30 dBi at 2.4 GHz and 5 GHz WLAN systems.

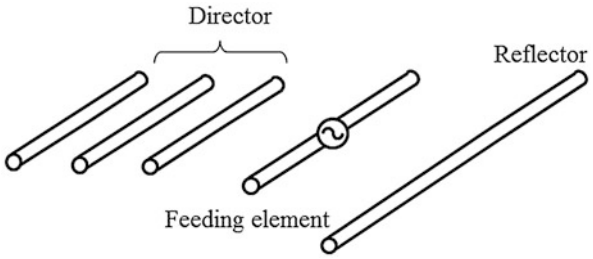


Fig. 2 Yagi-Uda array

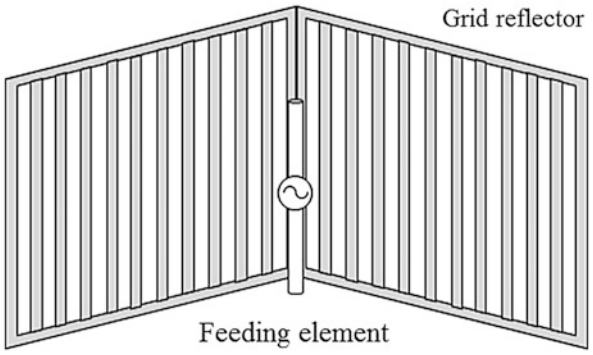


Fig. 3 Grid reflector antenna

DOI 10.1007/978-981-4560-75-7_106-1

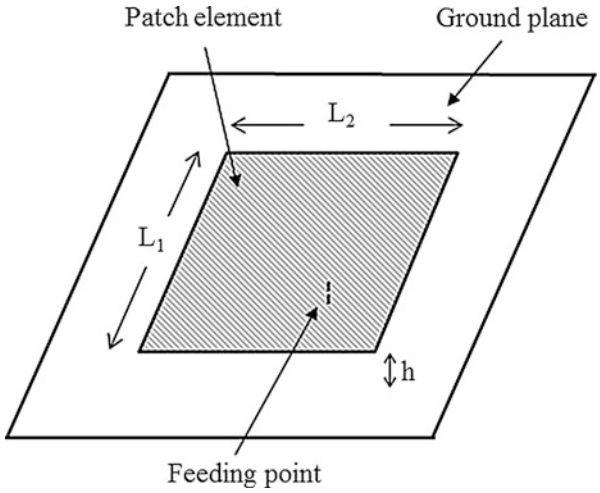


Fig. 4 Microstrip antenna

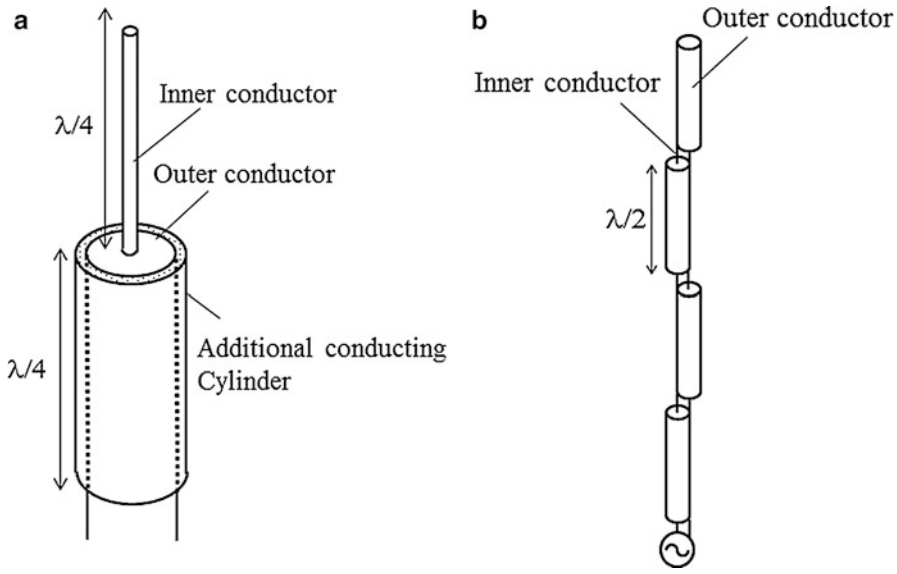


Fig. 5 Sleeve and collinear array, (**a**) sleeve antenna, (**b**) collinear array

For the multipoint WLAN applications, sector-shaped patterns and omnidirectional patterns in horizontal plane are used. The sector pattern is given by a microstrip (patch) antenna as shown in Fig. 4 and the gain range is 5–10 dBi. The omnidirectional pattern is provided by a sleeve antenna and collinear array as shown in Fig. 5. The sleeve antenna is a kind of a half wavelength dipole antenna involving the balun transformer which converts an unbalanced feeding cable like the coaxial line to a balanced feeding antenna. The collinear array is made up by several half-wavelength coaxial cables connecting the inner conductor of the lower element to the upper outer one to change the phase alternatively at the edge. The antenna gain of sleeve antenna is 2.15 dBi and that of collinear array is 5–10 dBi.

DOI 10.1007/978-981-4560-75-7_106-1

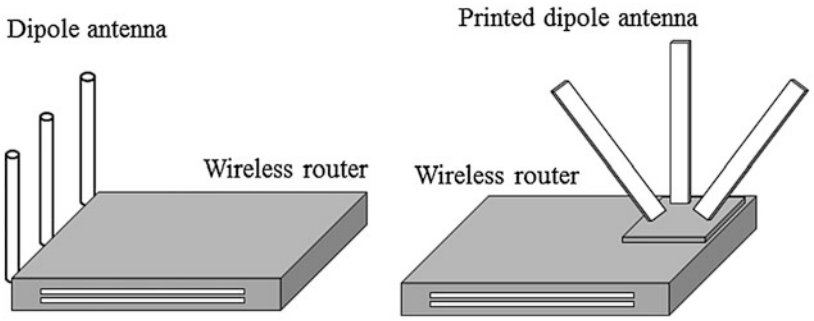


Fig. 6 External antennas

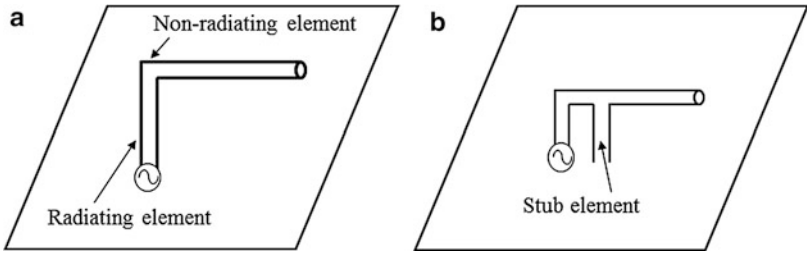


Fig. 7 Bent monopole antenna

External Antennas

The external WLAN antennas described in this section are the antennas protruding from the access point body. Most of the antennas are half-wavelength dipole antennas consisting of conducting pole or dielectric substrate as shown in Fig. 6. A dual-band printed antenna involves 2.4 GHz and 5 GHz dipoles etched on a dielectric substrate, and each antenna is fed by a thin coaxial cable. These antennas require balanced feeding, and then the balun like the sleeve antenna is built in at the feeding pint. A whip antenna is also used for the external antenna. It is a quarter-wavelength monopole antenna on a finite-sized ground plane given by a part of access point body.

Two external antennas are used for 2 × 2 MIMO access point; however, another one is often added for the diversity reception. Two antennas are used to reduce the fading effect at the reception, which is equivalent to select a pair of low-correlated antennas. MIMO channel capacity is also enhanced by low-correlated antennas, and the diversity reception by three antennas is useful for WLAN access point antennas.

Built-In Antennas

The external antennas are replaced with built-in antennas due to the miniaturization of access point and good outside appearance. A bent monopole antenna and an inverted-F antenna are often used for built-in antenna.

The bent monopole antenna is called as L-shaped antenna as shown in Fig. 7. A vertical part to the ground plane is radiating element and a horizontal one is called as a non-radiating element. Low profile L-shaped antenna with short radiating element is not easy to match the antenna input impedance with that of transmission line. Another vertical element to connect non-radiating one with ground plane is a stub element, which is used for impedance matching of the antenna. This antenna is called as inverted-F

DOI 10.1007/978-981-4560-75-7_106-1

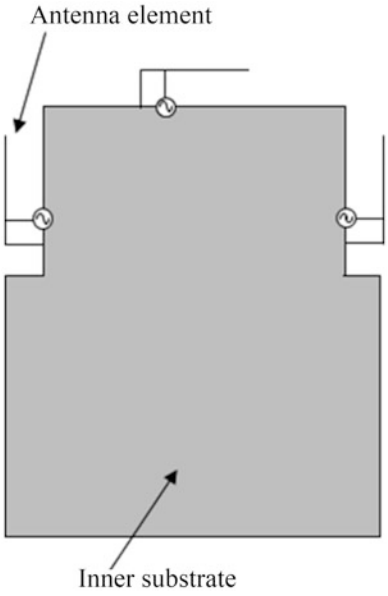


Fig. 8 Built-in antennas

antenna and is widely used for built-in antennas. Three inverted-F antennas are attached at the edge of internal substrate as shown in Fig. 8 for the MIMO system.

The integrated antenna is one of built-in antennas used in millimeter frequency band. A short wavelength in this band needs to remove transmission line and to integrate antennas into device package. These antennas are called as antennas in package. The aperture-coupled microstrip line-fed antennas are embedded on low-temperature cofired ceramic (LTCC) substrate operating in the 60-GHz frequency band. Air cavities processed inside the LTCC substrate are used to improve the bandwidth and gain of the antennas (Lamminen et al. 2008). Another antenna-in-package with high efficient radiation and high shielding effectiveness to electromagnetic noises is composed of U-shaped slots cut on the shielded ball grid array package. The measured antenna gain of the fabricated antenna is from 2.9 to 5.2 dBi in the 60 GHz band (Hashimoto et al. 2014). A multilayer parasitic microstrip antenna array on a multilayer Teflon substrate for millimeter wave is also proposed for system-on-package modules. This antenna achieves a radiation efficiency of greater than 91 % and antenna gain of 11.1 dBi at 60 GHz (Seki et al. 2005). The integrated WiGig antennas are still under development and high efficient and high gain antennas are required.

Summary

This chapter presented antennas used in access points of WLAN/WiFi systems. The antennas classified into three types were explained by showing example antennas. High-gain antennas like Yagi-Uda array and reflector antennas are often used to point-point wireless communication systems, and dipole antennas are major ones for the external access point antennas. As the built-in antennas, L-shaped antenna and inverted-F antenna are also explained. Further miniaturization of wireless devices and the use of millimeter wave need to develop another high efficient antennas for WLAN/WiFi systems.

DOI 10.1007/978-981-4560-75-7_106-1

Cross-References

- ▶ Antenna Design for Diversity and MIMO Application
- ▶ Antenna Systems for Cellular Base Stations
- ▶ Linear Wire Antennas (Including Yagi-Uda Antennas)
- ▶ Microstrip Patch Antennas
- ▶ MIMO Systems and Antennas for Terminals (Including Portable Devices Such as Handsets, iPad Laptops)
- ▶ Small Antennas (PIFA/PILA/Loading Antenna/etc.)

References

Hashimoto K et al (2014) Millimeter-wave band slot antenna on shielded BGA package. In: 2014 I.-E. antennas and propagation society international symposium, Memphis, pp 1280–1281, July

IEEE802.org. (2015) http://www.ieee802.org/. 15 Jan 2015

Lamminen AEI et al (2008) 60-GHz patch antennas and arrays on LTCC with embedded-cavity substrates. IEEE Trans Antennas Propag 56(9):2865–2874

McNamara DP et al (2000) Capacity variation of indoor multiple-input multiple-output channels. Electron Lett 36:2037–2038

Seki T et al (2005) Millimeter-wave high-efficiency multilayer parasitic microstrip antenna array on Teflon substrate. IEEE Trans Microw Theory Tech 53(6):2101–2106

Thaysen J, Jakobsen KB (2006) Envelope correlation in (N, N) MIMO antenna array from scattering parameters. Microw Opt Technol Lett 48(5):832–834

Uchida D, Arai H (2013) Requirement for spatial correlation reduction in 2 × 2 MIMO systems. IEICE Commun Expr 2(4):123–128

Yamada Y et al (1991) Diversity antennas for base and mobile stations in land mobile communication systems. IEICE Trans Comm E74-B(10):3202–3209

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_108-1

Antennas in Body-Centric Sensor Network Devices

Leena Ukkonen[a]* and Yahya Rahmat-Samii[b]
[a]Department of Electronics and Communications Engineering, Tampere University of Technology, Tampere, Finland
[b]University of California, Los Angeles (UCLA), Antenna Research, Analysis and Measurement Laboratory, Los Angeles, CA, USA

Abstract

This chapter concentrates on embroidered textile antennas for body-centric wireless identification and sensing systems. Methods for modeling and characterization of embroidered UHF RFID tag antennas, creating totally embroidered wearable patch-type antennas, and co-designing textile-based transmit antenna and mm-size implant antenna for challenging wireless brain-machine interface systems will be presented. Wireless body-centric identification and sensing systems hold an enormous potential to revolutionize wearable intelligence by extending the functionality of advanced garments. There are several specific requirements for on-body and implant antennas and their communication: on-body antennas have to be lightweight, conformal, easy to integrate into clothing, and as immune as possible to the performance-degrading effects of human body. When implanted systems are considered, communication link between the on-body and implant antennas has to be as efficient as possible, but at the same time SAR regulations must be obeyed. Moreover, implants and thereby implant antennas are intended to be as small as possible, and therefore the on-body textile antennas have to be able to efficiently communicate with mm-size implant antennas. Approaches to solve these challenges are presented in this chapter. Additionally a comprehensive and organized list of references is provided to assist the readers in identifying most pertinent publications about this very important and timely subject.

Keywords

Textile antennas; Body-centric wireless identification and sensing systems; Electro-textiles; Embroidered antennas; On-body antennas; Implant antennas; RFID systems; Wireless brain-machine interface systems

Introduction

Wireless body-centric identification and sensing systems hold an enormous potential to revolutionize wearable intelligence by extending the functionality of advanced garments. Wireless sensor networks integrated with garments allow inexpensive and continuous health state and environment monitoring. These sensor networks incorporate physiological sensors, low-power integrated circuits and computing devices, and wireless communication antennas for transmission and reception of recorded data. This relatively new interdisciplinary technology can be considered as a breakthrough technology in personal healthcare domain, ultimately making concepts such as telemedicine to become reality (Wang et al. 2013; Yang et al. 2008; Occhiuzzi et al. 2010; Vena et al. 2013). Antennas are one of the paramount components defining the reliability and efficiency of the wireless communication link between the body-worn electronic systems and the surrounding environment. Wearable passive ultrahigh (UHF) frequency

*Email: leena.ukkonen@tut.fi

DOI 10.1007/978-981-4560-75-7_108-1

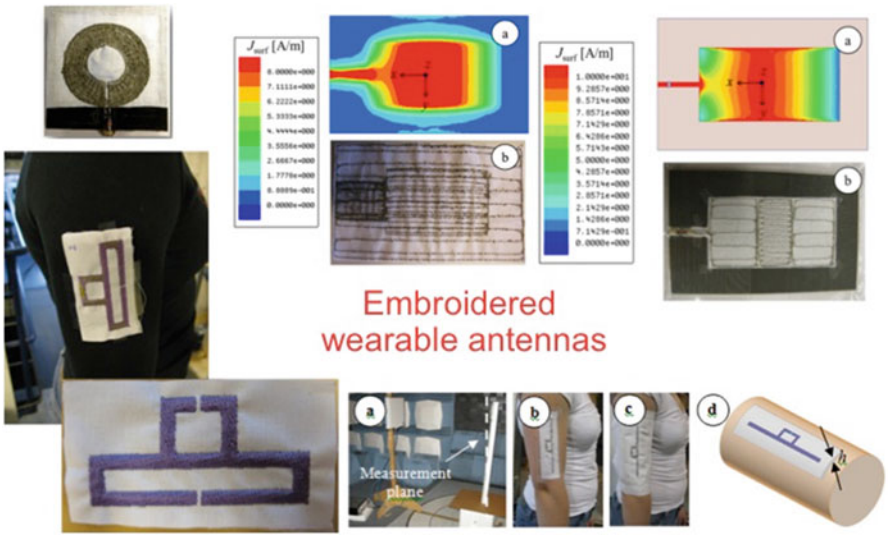


Fig. 1 Embroidered antenna structures for on-body use to be discussed in this chapter

radio-frequency identification (RFID) tag antennas, operating in the 860–960-MHz band, are of particular interest due to providing a predictable verification platform. They provide an invaluable contribution in the development toward an intelligent environment, where low-cost tag antennas are seamlessly integrated with daily garments to enable wireless body-centric communication everywhere and at any time (Polívka et al. 2009; Marrocco 2010; Koski et al. 2013a). Figure 1 presents various embroidered antenna structures, which are discussed in this chapter.

Biomedicine and personal healthcare systems demand information about sensed or measured biological parameters to be reliably and quickly sent over a wireless communication link for analysis purposes. Further, the communication system must be totally maintenance-free, comfortable to wear, low cost, and low power in order to meet widespread use and to provide an independent life for patients. Wireless body-centric communication sensing systems are promising candidates for the realization of such systems incorporating mobile control units and wearable and implantable antennas. Seamlessly integrated wearable and miniaturized implantable antennas have thus the potential to revolutionize the everyday life of people and contribute to independent living. Body-centric wireless communication is considered as an important part of the fourth-generation mobile communication systems and is expected to be part of the future convergence and personalization of the personal area networks (PANs) and body area networks (BANs) (Hall and Hao 2006).

The wireless brain-machine interface (BMI) system is a challenging future application in the field of body-centric communication systems. Real-time wireless interfaces between the human brain and mechanical prosthetics are highly desirable for patients suffering from severely disabling neurological conditions as these would allow them to use prosthetic limbs in the same manner as biological limbs. As a groundbreaking example, one can envision a technology that would convert a thought into action and let a person control a device just by thinking about it (Carmena 2012), or even further, one can imagine tiny brain-implantable devices which would restore the motor, sensing, or cognitive skills of an injured person. This system requires neural sensors transferring neural data to an external device to be decoded and used to control artificial actuators (Nicolelis 2001; Rabaey 2011). Thanks to the plasticity of the brain, the implanted BMI quickly becomes a seamless part of the brain allowing the direct brain control of smart prosthetics and computers (Nicolelis 2001; Koralek et al. 2012).

Currently, communication between the neural sensor and the computer is established via wires, which increases the infection risk of the patient and limits the usability of the system only to clinic environments

DOI 10.1007/978-981-4560-75-7_108-1

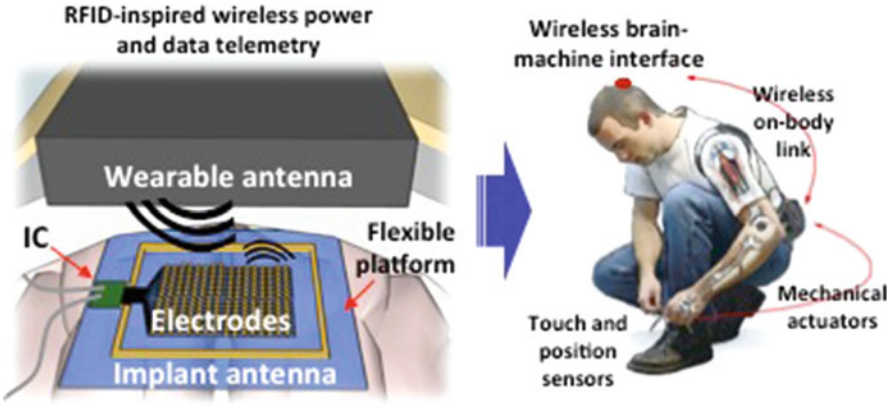


Fig. 2 Backscattering-based approach for wireless communication in brain-machine interface systems

(Collinger et al. 2013). Hence, to have a mobile BMI system and to be able to use the system in everyday life, data and power telemetry to the implanted neural sensor should be implemented wirelessly. There are two main challenges to reliably and accurately control prosthetic devices for long time periods: to realize viable neural interfaces that last a lifetime and to control the prosthetic devices in a naturalistic way (Carmena 2013). For this purpose, the neural implant must be biocompatible and minimally invasive to reduce tissue damages. Thus, it should be wirelessly powered, need minimum power, and support bidirectional data flow, i.e., "reading" and "writing" from/to the brain (Yakovlev et al. 2012).

The neural implant requires a transmitting antenna outside the human body to establish the wireless body-centric communication link. Figure 2 presents the backscattering-based approach for wireless BMI systems. This antenna should be comfortable to wear, low profile, and optimized for strongest possible coupling to the implant antenna. In this chapter, the feasibility of using flexible electro-textiles in the realization of transmitting loop antennas for wireless BMI systems is explained. For this purpose, both embroidered and commercially available textiles will be presented.

Design Considerations

Wearable antennas face many requirements. They should be comfortable to wear, lightweight, totally maintenance-free, and inexpensive for their widespread use to become reality (Koski et al. 2013a; Kellomäki and Ukkonen 2010; Bayram et al. 2010). Considering identification systems, RFID chip impedances typically have capacitive input impedance. To ensure proper impedance matching and thereby maximum power transfer at antenna-chip interface, the input impedance of the tag antenna has to provide the corresponding inductive component. Efficient impedance matching is particularly important in passive RFID systems, where all tag antenna available power comes from the remote reader unit (Marrocco 2008). Additionally, wearable antennas should show high immunity to the human body interaction, which otherwise may drastically detune the antenna input impedance and degrade the radiation efficiency. These issues are also common to the design of antennas operating in close proximity of metal objects (Ukkonen et al. 2005). In the UHF RFID band, the human body represents a high-loss dielectric object with a relative permittivity $\varepsilon_r \approx 50 - 60$ and loss tangent $\tan\delta \approx 0.5 - 1.2$ (Polívka et al. 2009). It is commonly known that in case of omnidirectional radiators, such as the dipole antennas, the presence of the human body causes considerable input impedance detuning and absorption of the received or radiated RF energy (Foster and Burberry 1999; Raumonen et al. 2003). As a result, the antenna radiation efficiency is degraded, and the identification distance becomes shorter.

DOI 10.1007/978-981-4560-75-7_108-1

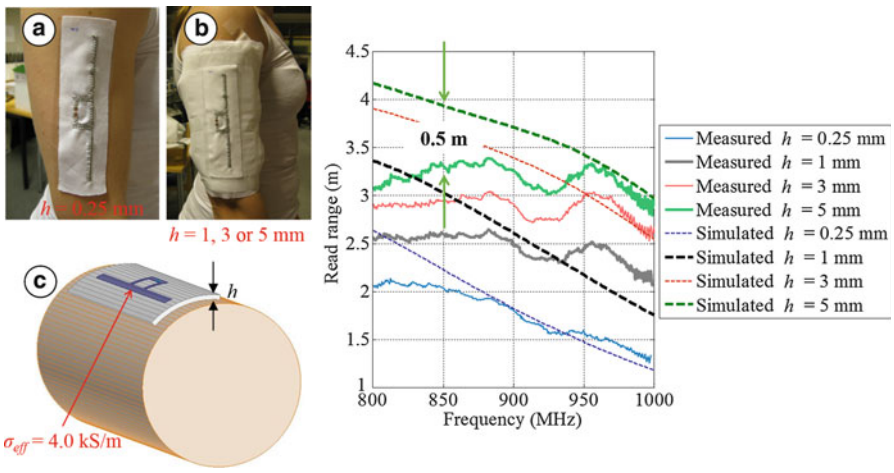


Fig. 3 Measured (*M*) and simulated (*S*) on-body read ranges of embroidered dipole antenna with different insulation layer thicknesses (*h*). (**a**) Embroidered antenna on the arm with 0.25-mm insulation. (**b**) Embroidered antenna on the arm with 1-, 3-, or 5-mm insulation. (**c**) Modeling approach with simplified arm model

However, reasonable read ranges can be achieved with embroidered dipole-type antennas, when they are attached on human body. Figure 3 presents a case study with a dipole RFID tag antenna embroidered utilizing conductive silver thread. The antenna has also been modeled utilizing an arm model (Koski et al. 2014a), and simulation and measurement results have been compared. Different thicknesses of insulating cotton fabric (0.25–5 mm) were placed between the antenna and human body. It can be observed that at around 900 MHz, the on-body read range varies between 1.5 m and 3.5 m, depending on the thickness of insulating layer, which is sufficient for short-range wireless identification and sensing applications.

Fabrication and Characterization of Embroidered RFID Tag Antennas

Fabrication and characterization of embroidered RFID tag antennas requires taking into account specific aspects, such as methods for modeling these antennas utilizing computational electromagnetic modeling methods and taking into account the physical behavior of current flows in the antenna structure. Different types of wearable RFID tags have been designed and explored in Choi et al. (2008), Yeonho et al. (2007), and Goojo et al. (2008). Authors of these studies have sewed the geometry of the tag antenna on textile with a computer-aided sewing machine using conductive threads. These studies explore the performance of sewed tag antennas and investigate the factors affecting the conductivity of the sewed antennas. According to the measurement results, the conductivity of sewed antenna improves by using highly conductive threads, increasing stitch and thread density, and also by choosing a sewing pattern that consists of sewed lines along the direction of the current flow. However, presented next is another approach to model and analyze the conductivity and performance of embroidered RFID tag antennas.

Fabrication and Read Range Performance

In designing an embroidery tag antenna with electromagnetic modeling tools, the complex embroidery pattern can be treated as a uniform conductive material layer when the pattern is sewed densely. Hence, it is the conductivity of an embroidery pattern that is the decisive factor in the simulation model of an embroidery antenna. This chapter focuses on wearable embroidered dipole-type UHF RFID tag antennas. The objectives of this chapter are to explain an estimation technique for the conductivity of embroidered RFID tag antennas and also to further investigate the factors affecting the conductivity and performance of sewed dipole-type tag antennas.

DOI 10.1007/978-981-4560-75-7_108-1

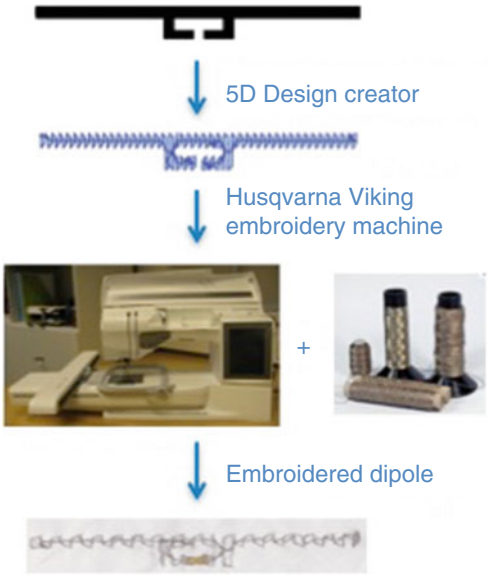


Fig. 4 Process of creating the embroidered antenna

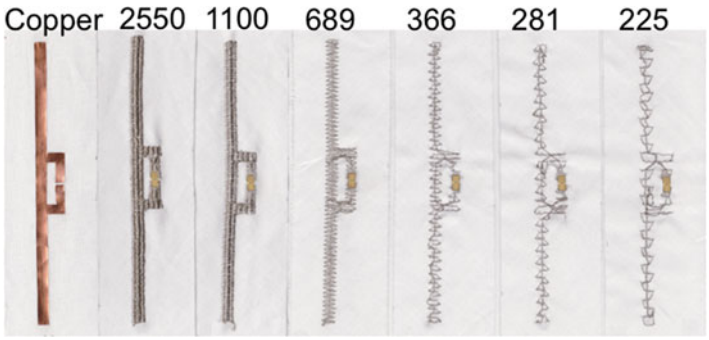


Fig. 5 Embroidered tags with different stitch densities and a copper reference tag. Numbers above the antennas refer to the number of stitches used to create the antenna

As presented in Fig. 4, the embroidery antennas are sewed with HUSQVARNA VIKING computer-aided sewing machine, using electric thread 110f34 dtex 2-ply HC. The thread has a weight of 110 dtex (dtex = g/10,000 m) raw yarn twisted with a second 110 dtex raw yarn to make a yarn of 220 dtex. Then the yarn is plated with silver to gain a weight of 275 dtex. The DC lineal resistivity of the thread is 500 ± 100 Ω/m, and the diameter is approximately 0.16 mm. The thread consists of 34 filaments in a yarn. The sewing pattern is created with 5D Embroidery System software. The software enables to control the stitching techniques and the stitching density when creating the sewing pattern. The file containing the created pattern is then transferred to the embroidery machine to be embroidered on cotton automatically. Figure 5 presents the embroidered dipole antennas. The created sewing pattern consists of a horizontal line along the length of the dipole and zigzags sewed vertically on it. The dipole with 2550 stitches has the highest zigzag density, and the dipole with 225 stitches has the lowest zigzag density. The rightmost tag is the copper-based reference.

The effect of the stitch and zigzag density on the performance of the antenna is investigated by embroidering six dipoles with different zigzag densities and measuring their performance in an anechoic chamber. The fully assembled tags are shown in Fig. 5 with the amount of stitches of each sewed tag written on top of it.

DOI 10.1007/978-981-4560-75-7_108-1

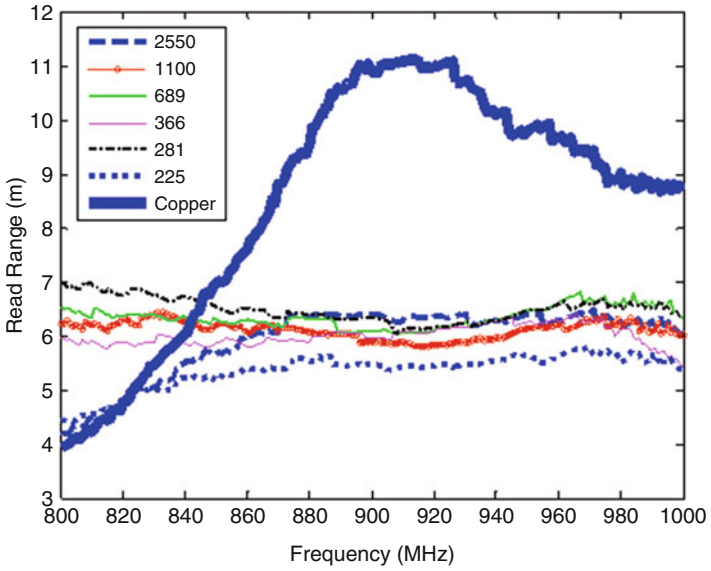


Fig. 6 Measured forward read range graphs of embroidered dipoles with different stitch densities

To characterize the performance of the embroidered dipoles, the maximum achievable read range (r_{max}) of the fully assembled tags using Voyantic Tagformance RFID measurement system was measured. In practice, the read range r_{max} is attained when the tag is located in the main beam of a reader antenna, which has its polarization perfectly aligned with the tag and is transmitting the maximum power allowed by the regulations. In this measurement configuration, the achievable read range is calculated based on Friis transmission equation as (Dobkin 2007)

$$r_{max} = \frac{\lambda}{4\pi}\sqrt{\frac{EIRP}{L_{fwd}P_{th}}}, \tag{1}$$

where λ is the wavelength, L_{fwd} is the measured linkloss factor from the output port of the generator to the input of a hypothetic isotropic radiator placed at the tag's location, P_{th} is the minimum transmitted power required to turn the tag IC on, and $EIRP$ is the regulated effective isotropic radiated power (EIRP). By definition, $EIRP$ is the amount of power that would be needed to put into an isotropic antenna to produce the same power density in the radiation field as was observed in the main beam of the directional antenna. In the European RFID band (865.7–867.6 MHz) $EIRP = 3.28$ W.

Figure 6 shows the measured read ranges of the tags in air shown in Fig. 5. Comparing the read ranges graph of the copper dipole with the corresponding graphs of sewed dipoles, it can be observed that in the case of sewed dipoles the frequency bandwidth is greatly increased. This is mainly due to the reduction of conductivity and quality factor of the sewed antennas. The achieved read ranges of the sewed tags range from 5.5 m to 7 m. This is competitive against the results reported in Choi et al. (2008), Yeonho et al. (2007), and Goojo et al. (2008) and without a doubt sufficient in many applications.

According to Fig. 6, the performance of sewed tags is fairly similar regardless of the zigzag density, and only when the amount of stitches is as low as 225, the performance starts to change noticeably. Thus, it can be deduced that antennas can be sewed sparsely to save materials and still achieve nearly as good performance as with antennas sewed with a high stitch and zigzag density.

DOI 10.1007/978-981-4560-75-7_108-1

Table 1 Thickness of the sewed antenna (upper face of each tag)

Number of stitches	2550	1100, 689	366, 281, 225
Thickness (mm)	0.425	0.2	0.1

Simulation Model and Effective Conductivity

For modeling the embroidery tag antennas, the effective conductivity of the embroidery pattern needs to be known. In this way the antenna can be modeled as a uniform conductor layer, and its performance can be optimized in a well-founded manner. This chapter presents a technique to evaluate the conductivity of sewed antennas based on the comparison between simulation and measurement results. This enables the design of antennas optimized for a given stitch pattern.

In order to find the conductivity of the embroidered antennas, the measured realized gain graphs of the embroidered antennas are compared with simulation results of the dipole with different conductivity values and thicknesses. The realized gain of the tag antennas can also be calculated based on link loss factor L_{iso}, IC sensitivity (i.e., the minimum input power required to activate the IC) $P_{ic,0}$, and the measured threshold power P_{th} results (Ukkonen et al. 2012):

$$G_{r,tag} = \frac{P_{ic,0}}{L_{iso}P_{th}}. \tag{2}$$

Realized gain takes into account the radiation efficiency of the antenna and power reflection coefficient, i.e., impedance matching with the RFID IC.

The thickness of embroidered antennas changes as a function of stitch density. When the amount of stitches and the thread density are high enough, in a sense that the sewed structure resembles a flat pattern, the thickness of the sewed structure can be talked about. But, when the sewed pattern is sparse, it is not convenient to talk about the thickness of it because the sewed structure is not flat and solid. Still, in simulator it is needed to define a thickness for the antenna material. Thus, the thickness of the sewed structure on the upper face of each tag is measured, and it is defined as the thickness of the conductive part of the antenna in ANSYS HFSS simulator. The measured thicknesses of the embroidered planar dipoles are listed in Table 1.

As illustrated in Fig. 6, the read range graphs of embroidered dipoles are located closely to each other. This might lead to a conclusion that the conductivities of embroidered antennas are almost the same too. However, when the varying thickness of the embroidered structure is considered, different conductivity values for each of them are achieved.

As illustrated in Fig. 7a, the simulated realized gain of the dipole with a thickness of 0.425 mm and conductivity of 2000 S/m matches with the measured realized gain of the sewed dipole with 2550 stitches. Thus, it can be deduced that the conductivity of this sewed dipole is 2000 S/m. Similarly, Fig. 7b shows simulated realized gain of the dipole with different conductivities and measured realized gain of sewed dipoles with 1100 and 689 stitches. In this case, in simulations, the thickness of the dipoles is set to 0.2 mm. According to Fig. 7b, the conductivity of sewed dipoles with 1100 and 689 stitches varies from 3000 S/m to 5000 S/m. Figure 7c shows simulation results and measurement results of sewed tag antennas with 366, 281, and 255 stitches. In this case, in HFSS model, the thickness of the dipole is 0.1 mm, which causes higher conductivity values than before. According to this figure, the conductivity of these tag antennas varies from 6000 S/m to 9000 S/m.

According to these results, the effective conductivity of the embroidered dipoles decreases as the thread density and the thickness of the sewed pattern increases. There is only a little performance deviation between the embroidered dipoles with 281–2550 stitches, while they have different thicknesses. Thus, the

DOI 10.1007/978-981-4560-75-7_108-1

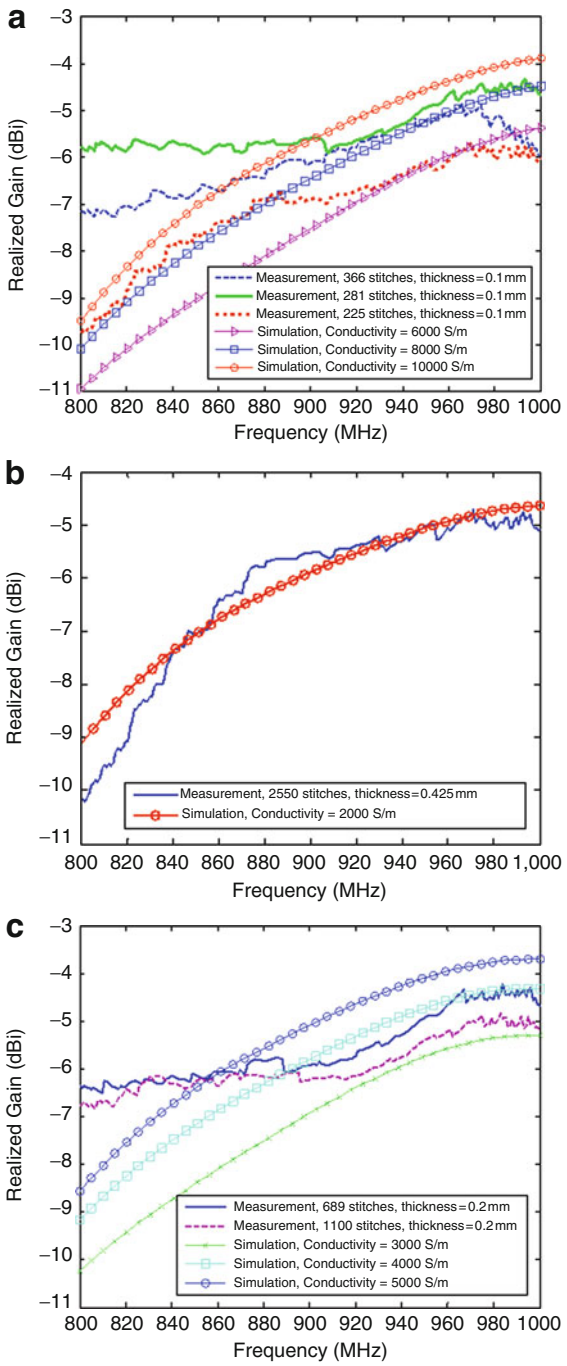


Fig. 7 Measured realized gain graphs of sewed dipoles and simulated realized gains. (**a**), (**b**), and (**c**) refer to different number of stitches and corresponding simulated conductivity values

thicker pattern (or conductive layer) must have less conductivity in order to have losses at the same level as the thinner one.

DOI 10.1007/978-981-4560-75-7_108-1

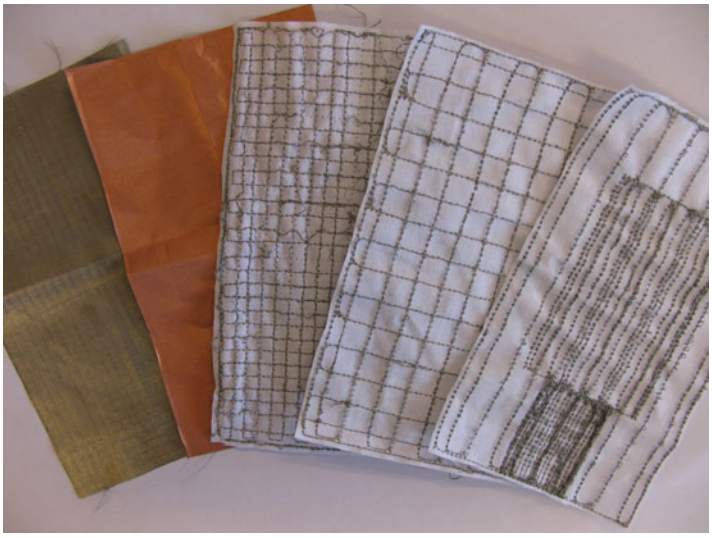


Fig. 8 Electro-textile ground planes for wearable patch antennas. From left: silver fabric, copper fabric, embroidery of small squares, embroidery of large squares, embroidery created for maximized antenna read range, and minimized back lobe level

Selected Case Studies

Embroidered Patch-Type Textile Antennas in Wireless Body-Centric Applications

Performance of dipole-type antennas is relatively sensitive to the presence of human body, due to absorption of electromagnetic energy by biological tissue (Ukkonen et al. 2012). Patch-type antennas with ground plane offer an alternative with less sensitivity to the materials the antennas are mounted on (Hall and Yang 2012; Ziai and Batchelor 2010). Moreover, utilization of different electro-textile approaches provides novel ways to construct flexible, garment-integrated antennas. Electro-textiles are becoming widely used in body-centric sensing systems as wearable antenna materials, thanks to their suitable radio-frequency (RF) performance, flexibility, and durability (Wang et al. 2013; Ouyang and Chappell 2008; Bayram et al. 2010). Electro-textiles typically refer to a group of electrically conductive textiles. These are created either by using conductive sewing threads in a computer-assisted embroidery machine (Ouyang and Chappell 2008; Koski et al. 2012; Wang et al. 2012) or by plating or interpolating a nonconductive fabric surface with an alloy of metals or with pure metal (Locher et al. 2006). The embroidered textile always results in an anisotropic pattern, in which the conductivity is strongly dependent on the direction of the current flow in the pattern, pattern geometry, and stitching density of the pattern (Koski et al. 2013a; Moradi et al. 2012). The different types of electro-textile ground planes for wearable patch antennas are presented in Fig. 8.

Electro-Textile Ground Plane and Top Radiating Patch Characterization

An efficient and proper ground plane provides important antenna-to-body isolation and reduces the body effects on the antenna performance (Hall and Yang 2012). A patch-type tag will concentrate its radiated energy away from the ground plane, and hence, less energy is lost in the human body and the overall radiation efficiency is enhanced. Therefore, a patch antenna is a potential choice for wearable antenna applications.

As new unconventional materials for wearable antennas are emerging, it is of paramount importance to characterize their applicability as antenna ground plane and top patch structures. From the modeling point of view, all electro-textiles may be considered as infinitely thin dissipative dielectric materials and characterized with the complex sheet impedance parameter $Z_S = R_S + jX_S$. Depending on the

DOI 10.1007/978-981-4560-75-7_108-1

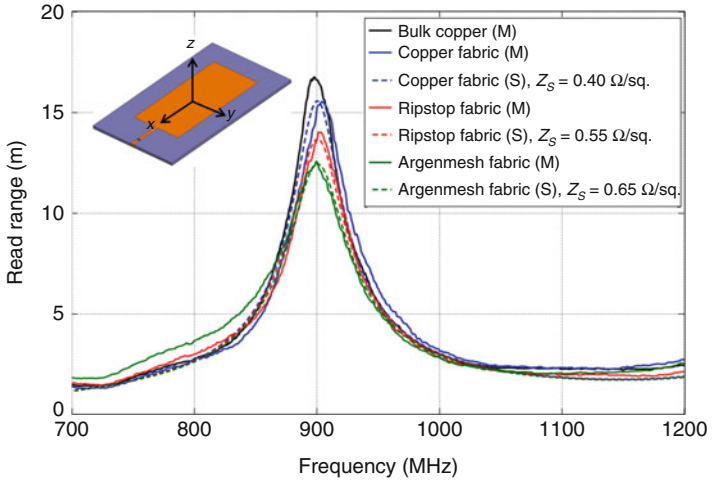


Fig. 9 Measured (*M*) and simulated (*S*) read range in air in + *z* direction for a patch tag antenna with conductive fabric ground plane. No impedance detuning is encountered; the conductive fabrics may be approximated as conductors whose sheet impedance is purely real. Also, the measured E-plane power patterns show low antenna back lobe levels for the conductive fabrics

electro-textile structure and on the implementation of the metallization, the sheet impedance is either purely real or complex. When the electro-textile does not affect the antenna input impedance, the sheet impedance imaginary part is negligible. This is the case for pure conductors and most conductive fabrics (e.g., the copper and silver fabrics in Fig. 1). As Fig. 9 shows, the copper and silver fabrics (ripstop and argenmesh) do not cause any impedance detuning compared to a bulk copper, and hence, their sheet impedance is purely real.

For embroidered structures, the sheet impedance is always complex. The imaginary part accounts for the increased antenna electrical length, which results from the increased current path length in the embroidery. Therefore, the embroidered structure increases the antenna input inductance, and as a consequence, the antenna resonance is shifted toward lower frequencies compared to a bulk copper antenna. The magnitude of the imaginary part is determined by the embroidered structure shape, size, and density. In Figs. 10 and 11, the sheet impedance for different embroidered ground planes and top patches, respectively, are determined.

When the embroidered structure shape and size are fixed, the imaginary part of the sheet impedance is directly proportional to the embroidery density. In Fig. 12, this relation is evidenced for the embroidered ground planes G1–G3 and top patches T1–T2.

When the current path length in the embroidered ground plane structure becomes long, the amount of back radiation is increased. This is demonstrated in Fig. 13.

Realizing Fully Wearable Embroidered Patch Antenna

The electro-textile sheet impedance is required in the design and optimization of wearable antenna to properly account for the electromagnetic effects induced by the material. In Fig. 14, a fully wearable embroidered antenna is implemented using the G6 and T2 embroideries. The bandwidth of the regularly shaped patch antenna is directly proportional to the substrate thickness and inversely proportional to the square root of its dielectric constant (Kumar and Ray 2003). Therefore, a thicker substrate with a low dielectric constant is generally preferred. For better radiation efficiency, a low-loss dielectric substrate is desired. The 3-mm-thick EPDM (ethylene propylene diene monomer) foam serves as a suitable wearable antenna platform.

DOI 10.1007/978-981-4560-75-7_108-1

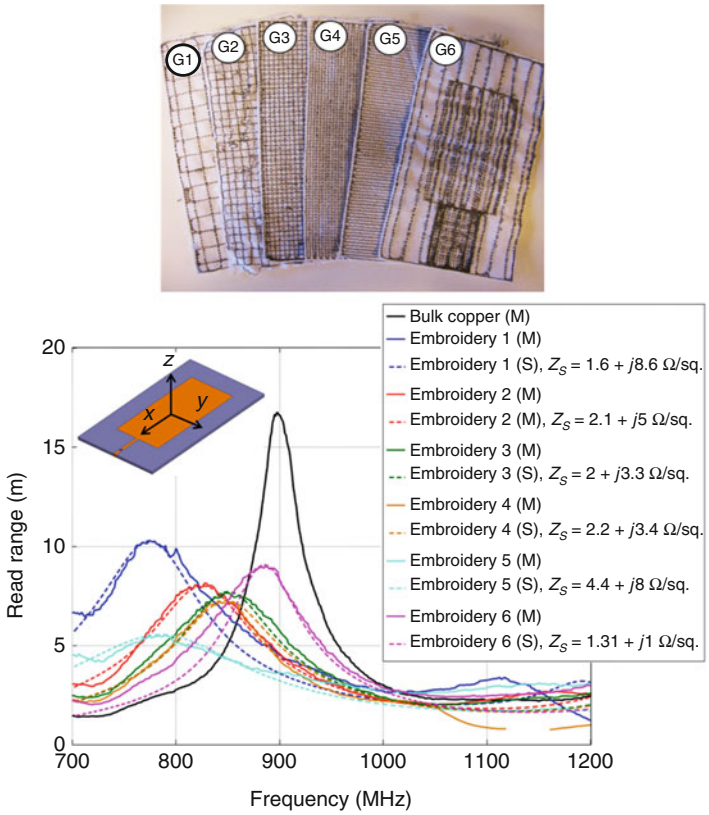


Fig. 10 Measured (*M*) and simulated (*S*) copper tag antenna read range in air in + *z* direction with embroidered ground plane. When the current path length in the ground plane is long, the impedance detuning becomes significant, and the electro-textile sheet reactance becomes large. The electro-textile sheet resistance and reactance values are minimized by optimizing the thread density according to the simulated surface current density in the copper tag antenna ground plane (embroidery 6)

The performance of the wearable embroidered antenna is presented in Fig. 15. The read range is relatively stable over the observed frequency range. At 900 MHz, the antenna achieves 3-m read range in air. When the antenna is operated near the human body, the high-permittivity human tissues are expected to shift the resonance back to the RFID band. The antenna achieves extremely low back lobe level, which is required for a good antenna-to-body isolation. Most importantly, Fig. 15 provides a strong indication that a complex sheet impedance is required to properly account for the electromagnetic effects induced by the conductive fibers.

In the future, embroidery techniques enable seamless and robust integration of antennas with clothing, which promote the industrial and public acceptance of the technology.

Wearable Embroidered Patch Antenna Performance in Close Vicinity of the Human Body

The desired wearable antenna performance must be guaranteed in close vicinity of the human body. Therefore, every design process of a wearable antenna should involve a verification step in which the antenna is measured on the human body. The embroidered antenna was measured on the human arm (Koski et al. 2014b). The antenna was separated from the arm with a 0.7-mm-thin tight shirt. The body effects are more prominent when the antenna-to-body separation is small. A thin shirt was chosen to evaluate a worst case scenario.

Comparing the performances in Figs. 15 and 16, it is evident that the antenna features inherently excellent antenna-to-body isolation. Further, the relatively flat read range response points out that the antenna's input impedance is not critically affected by the human body. At 900 MHz, the antenna's

DOI 10.1007/978-981-4560-75-7_108-1

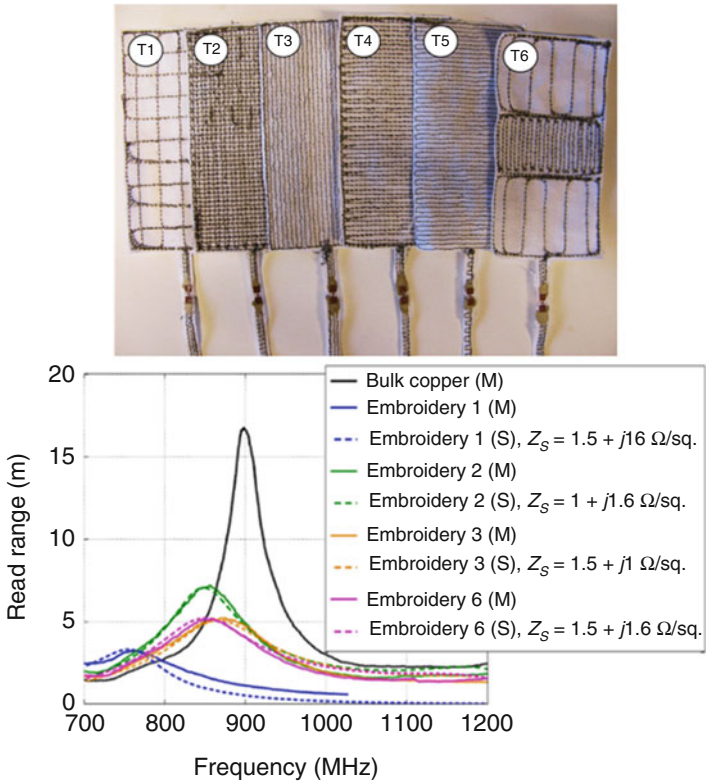


Fig. 11 Measured (*M*) and simulated (*S*) copper tag antenna read range in air in + *z* direction with embroidered top patch. The electro-textile sheet reactance value and consequently the impedance detuning are minimized by optimizing the thread density according to the simulated surface current density in the copper tag antenna top patch. T4 and T5 did not yield any measurable antenna response due to significant cross-polarization level

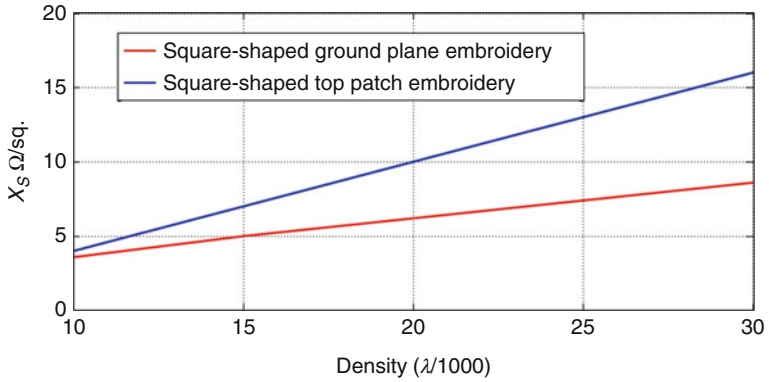


Fig. 12 The relation between square-shaped embroidery sheet impedance imaginary part and density. The curves are attained for the ground planes G1–G3 and top patches T1–T2. The wavelength corresponds to 900 MHz

on-body read range exceeds 3 m, which sets a benchmark for future wearable embroidered antenna designs.

Embroidered On-Body Antenna for Communication with mm-Size Implants

In this section, the feasibility of using flexible electro-textiles in the realization of transmitting loop antennas for wireless brain-machine interface (BMI) systems is explored. For this purpose, both embroidered and commercially available textiles are considered. The neural implant requires a transmitting

DOI 10.1007/978-981-4560-75-7_108-1

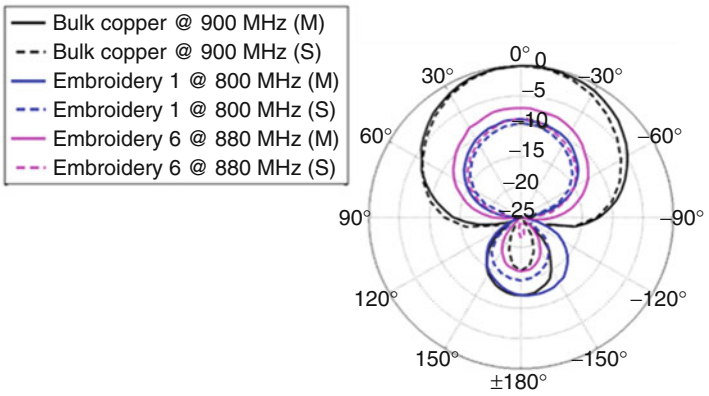


Fig. 13 Measured (*M*) and simulated (*S*) E-plane (*zx*-plane) power patterns normalized to bulk copper antenna for G1 and G6 ground planes. The long current path length in G1 results in increased back radiation

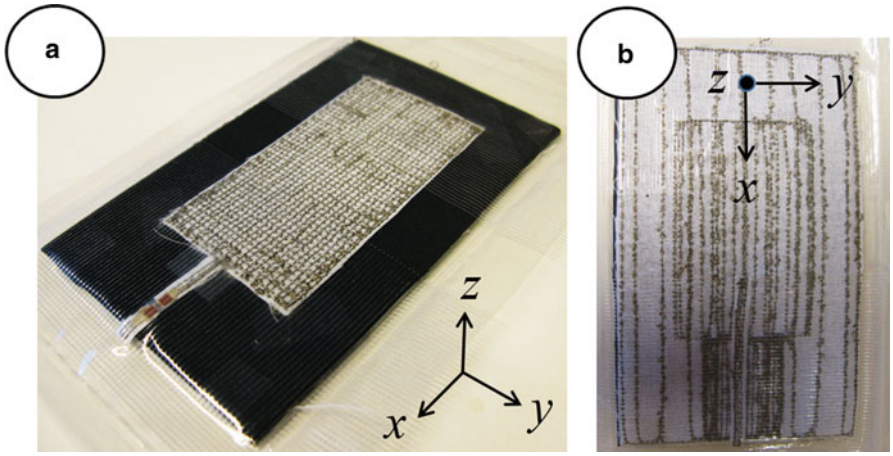


Fig. 14 Wearable embroidered patch tag antenna on EPDM foam substrate; (**a**) *top view*; (**b**) *bottom view*. Size of the substrate and ground plane is 150 × 90 mm, and size of the radiating patch is 90 × 50 mm

antenna outside the human body to establish the wireless body-centric communication link. This antenna should be comfortable to wear, low profile, and optimized for strongest possible coupling to the implant antenna. These requirements are addressed in this section.

Modeling of Antennas and Wireless Link

Powering and communicating with mm-size battery-free cortical implants require transmitting antennas achieving high link power efficiency without violating SAR limits. In addition, the on-body transmit antenna should fulfill the requirements of wearable antennas, having compact size and being flexible, lightweight, and even washable. Electro-textile materials provide compelling means for conformal wearable antennas. Neural implants require a transmitting antenna outside the human body to establish the wireless body-centric communication link. This antenna should be comfortable to wear, low profile, and optimized for strongest possible coupling to the implant antenna. The feasibility of using flexible electro-textiles in the realization of transmitting loop antennas for wireless BMI systems is explored and explained in this section. For this purpose, both embroidered and commercially available conductive textiles will be considered.

DOI 10.1007/978-981-4560-75-7_108-1

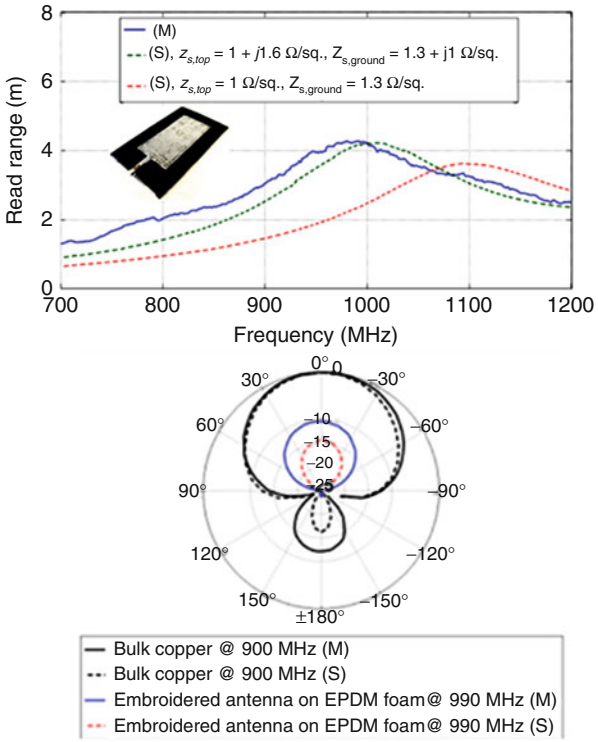


Fig. 15 *Left*: Measured (*M*) and simulated (*S*) read ranges in air in + *z* direction for the embroidered antenna on EPDM foam substrate. A complex sheet impedance is required to properly account for the electromagnetic effects induced by the conductive fibers. *Right*: E-plane power patterns normalized to bulk copper antenna for the embroidered antenna on EPDM foam substrate. Extremely low back lobe level is achieved

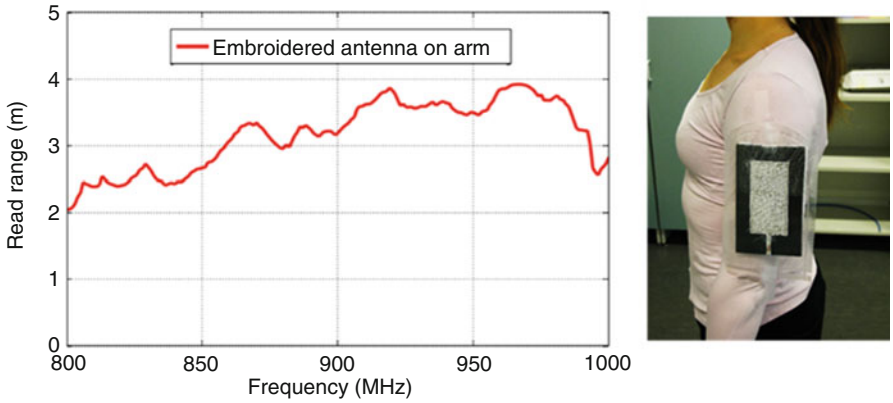


Fig. 16 Read range of the wearable embroidered antenna on the human arm in + *z* direction in polarization-matched configuration when worn on the upper arm

DOI 10.1007/978-981-4560-75-7_108-1

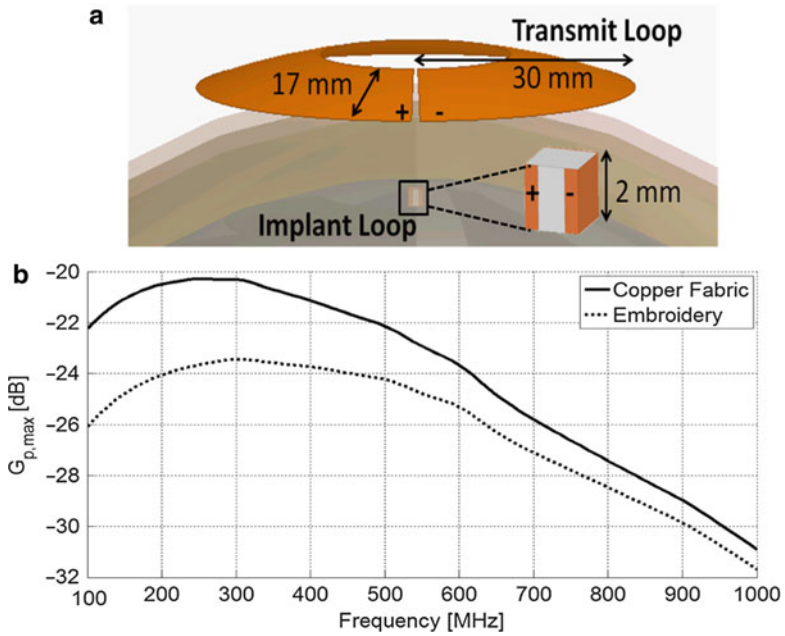


Fig. 17 (**a**) Conformal model of the electro-textile transmit antenna coupled to a mm-size cubic loop. The wireless channel model consists of four spherical layers of tissues. (**b**) Simulated maximum operating power gain of both electro-textile transmit antennas coupled to the SMMDF101 cored $2 \times 2 \times 2$ mm^3 implant antenna

Small mm-sized loop antennas are capable of coupling to an external transmitting loop efficiently through an inductive link in frequencies ranging from hundreds of megahertz to low gigahertz. Analyzing the inductive link as a two-port network allows us to compute the link power efficiency as the maximum operating power gain of the two-port system:

$$G_{p,\,max} = \frac{|Z_{12}|^2}{S + \sqrt{S^2 - |Z_{12}Z_{21}|^2}}, \tag{2}$$

$$S = Re(Z_{11})Re(Z_{22}) - Re(Z_{12}Z_{21}),$$

where $G_{p,max}$ is the maximum operating power gain and Z_{xy} are the two-port system Z-parameters (Moradi et al. 2013). This is the ratio of the power received by the implant IC to the power transmitted by the transmit antenna.

In practice, for the user's comfort, the low-profile transmitting loop needs to be flexible and wearable. Therefore, transmitting loops made of electro-textiles were designed. Copper fabric and an embroidered structure were considered. The electro-textile-transmitting loops were modeled as infinitely thin conductors having the sheet resistance as the only material parameter. The copper fabric and the embroidered pattern were assigned sheet resistance values of $R_S = 0.4\ \Omega/\square$ and $R_S = 1.25\ \Omega/\square$, respectively. These values were determined through wireless reflectometry measurement (Koski et al. 2013b). A flexible cotton fabric ($\varepsilon_r = 2.2$, $\tan\delta = 0.08$) was chosen as the substrate material for both wearable antennas.

Conformal transmitting loops were designed for coupling with a $2 \times 2 \times 2$-mm^3 copper cubic loop with a magnetodieletric core (SMMDF101, $\varepsilon_r = 18$, $\mu_r = 17.5$, $\tan\,\delta_{magn} = 0.15$, $\tan\,\delta_{diel} = 0.02$ at 300 MHz). In simulations, a spherical human-head model (Fig. 17a) with an outer diameter of 9.2 cm was used. The spherical model consisted of 2-mm skin, 2-mm fat, 7-mm bone, and brain layers. Previous studies have shown that small mm-size loop antennas are capable of coupling to an external transmitting

DOI 10.1007/978-981-4560-75-7_108-1

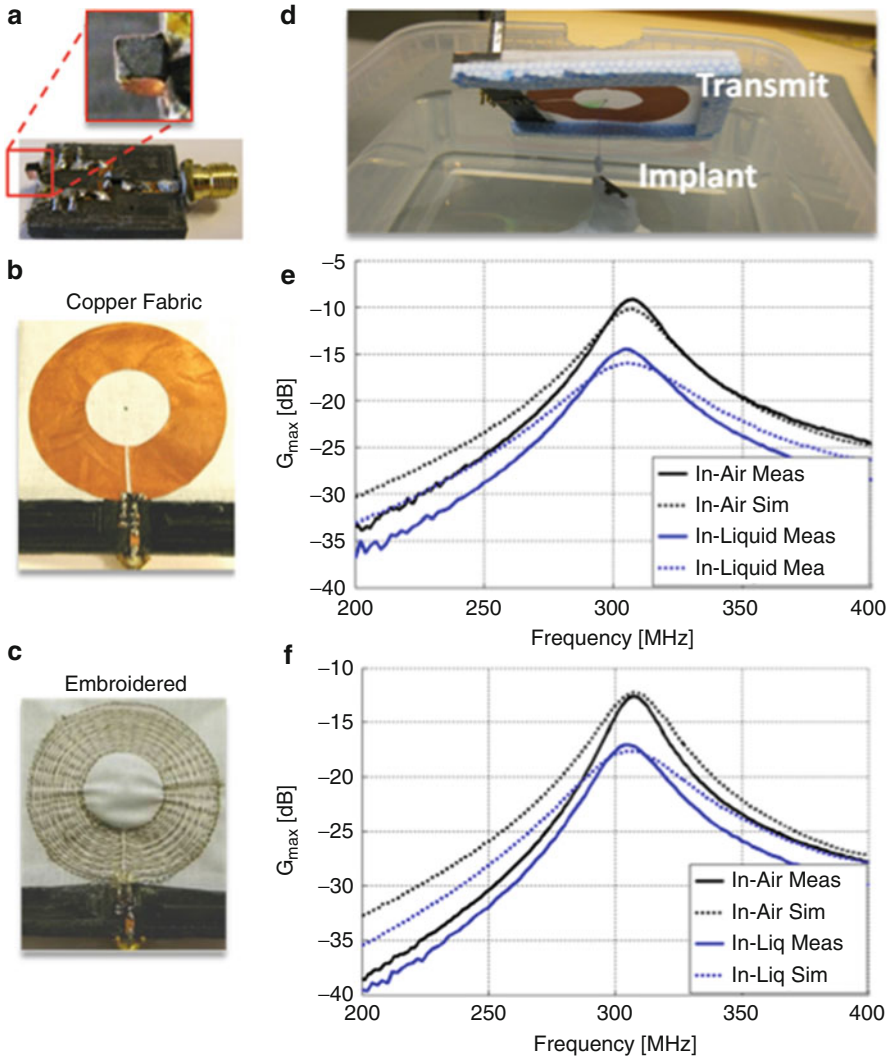


Fig. 18 On the *left side*: the fabricated antenna prototypes with the test boards, the $2 \times 2 \times 2$-mm^3 copper cubic loop with a magnetodieletric core (**a**), the copper fabric (**b**), and the embroidered (**c**) transmit antennas. On the *right side*: two-port measurements setups (**d**), *in-air* and *in-liquid* results of the copper fabric (**e**), and embroidered (**f**) transmit loop antennas. The dashed line refers to the simulation results

loop efficiently through an inductive link in frequencies ranging from hundreds of MHz to low GHz (Yakovlev et al. 2012). Simulations showed that at 300 MHz, the optimized dimensions of the transmit loop with both sheet resistances are the same. The optimized inner radius was 13 mm and the trace width was 17 mm. The simulated wireless link performance is shown in Fig. 17b. At 300 MHz, the peak $G_{p,max}$ of copper fabric (−20 dB) is approximately 3 dB higher compared to that of the embroidered pattern (−23 dB). It is worth noticing that the link power efficiency achieved by the wearable transmit antennas is lower than that achieved by an optimized bulk copper antenna made on solid substrate ($G_{p,max} = -15.6$ dB) (Moradi et al. 2013). However, when using wearable antennas, the received power P_L is in the same order as using a copper transmit antenna. It means that the wearable loop antennas can provide similar powering performance as the bulk copper loop at the expense of a bigger size and higher energy consumption. As presented in Fig. 16, at 300 MHz, the link power efficiency of embroidered transmit antenna is approximately 3 dB lower than that of the copper fabric antenna.

DOI 10.1007/978-981-4560-75-7_108-1

Fabrication and Measurements

The electro-textile antennas were fabricated and experimentally characterized. As the conductive fabric, the commercial conductive textile from LessEMF, pure copper taffeta fabric (35 % copper), was chosen. The embroidered antenna was sewed with HUSQVARNA VIKING computer-aided sewing machine using the conductive thread Shieldex 110f34 dtex 2-ply HC. The embroidered pattern is composed of circular lines of conductive thread sewn along the loop. The stitching density along the loop radius was set equal to 1 line/mm to achieve a sheet resistance of 1.25 Ω/□. The implant antenna was a $2 \times 2 \times 2$ mm^3 SMMDF101 cored cubic loop. It was fabricated using a 50-μm-thick copper tape. The accuracy of VNA is best around 50 Ω, which is the characteristic impedance of its ports. Therefore, additional circuit board fixtures to impedance-match the designed antennas to 50 Ω were fabricated. Figure 18 presents the fabricated components.

Both in-air and in-liquid measurements were conducted. In the latter case, the implant antenna was immersed in human-head equivalent liquid phantom provided by SATIMO. The liquid phantom had similar dielectric properties to those of human head at 300 MHz, following the IEEE standards ($\varepsilon_r = 44.32$, $\sigma = 0.83$ S/m). The implant loop was measured in air with both electro-textile transmit loops at a fixed distances of 16 ± 1 mm. In the *in-liquid* measurements, the implant was placed 11 ± 1 mm beneath the liquid surface, and the transmit loop was 5 ± 1 mm above the liquid surface. These distances were chosen to match with the distances used in the layered human-head model (2-mm skin + 2-mm fat + 7-mm bone = 11 mm). The link power efficiencies between the antennas were computed based on the measured *S*-parameters. Results of both *in-air* and *in-liquid* measurement setup are shown in Fig. 18. In all cases, the peak $G_{p,max}$ is achieved at 305 MHz, near the predicted matched frequency. The link power efficiency of the copper fabric transmit loop at 305 MHz was −9.62 dB and −14.6 dB when the implant was placed in air and in liquid, respectively. Concerning the embroidered prototype, a 3-dB lower peak value of $G_{p,max}$ was measured (−12.3 dB in air, −17.1 dB in liquid), thus confirming the previous simulation results (Fig. 18). Around the matched frequency, simulations are in good agreement with measurements. The maximum difference between the peak values is less than 1.4 dB. This is likely due to the fabrication tolerance and the variability of the discrete components (e.g., capacitors with 5 % tolerance). Overall, the achieved results clearly demonstrate the ability to accurately predict the coupling between the two antennas. The measured peak values of $G_{p,max}$ were significantly higher than those computed by considering only the antennas. This means that the presence of the test boards created an additional coupling mechanism. Hence, the proposed full model enables to evaluate, compensate, and eventually remove the measurement artifacts introduced by the test boards in order to estimate the power that can be delivered to the neural IC in real applications.

Conclusion

Wireless body-centric identification and sensing systems hold an enormous potential to revolutionize wearable intelligence by extending the functionality of advanced garments. Groundbreaking application systems include wireless brain-machine interfaces, which have the potential to revolutionize the treatment of severe neurological conditions, such as paralysis and epilepsy. There are several specific requirements for on-body and implant antennas and their communication: on-body antennas have to be lightweight, conformal, easy to integrate into clothing, and as immune as possible to the performance-degrading effects of human body. When implanted systems are considered, communication link between the on-body and implant antennas has to be as efficient as possible, but at the same time power transmission regulations must be obeyed. Moreover, implants and thereby implant antennas are intended to be as small as possible, and therefore the on-body textile antennas have to be able to efficiently communicate with

DOI 10.1007/978-981-4560-75-7_108-1

mm-size implant antennas. Electro-textiles are becoming widely used in body-centric sensing systems as wearable antenna materials thanks to their suitable RF performance, flexibility, and durability. Electro-textiles typically refer to a group of electrically conductive textiles. These are created either by using conductive sewing threads in a computer-assisted embroidery machine or by plating or interpolating a nonconductive fabric surface with an alloy of metals or with pure metal. In this chapter, the above challenges have been addressed by presenting methods for modeling and characterization of embroidered UHF RFID tag antennas, creating totally embroidered wearable patch-type antennas and co-designing textile-based transmit antenna and mm-size implant antenna for challenging wireless brain-machine interface systems.

Cross-References

- ▶ Advanced Antenna Fabrication Process (MEMS/LTCC/LCP/Printing)
- ▶ Antenna Sensors in Batteryless Wireless Sensing Systems
- ▶ Implanted Antennas in Biomedical Telemetry

References

Bayram Y, Zhou Y, Shim BS, Xu S, Zhu J, Kotov NA, Volakis JL (2010) E -Textile conductors and polymer composites for conformal lightweight antennas. IEEE Trans Antennas Propag 58(8):2732–2736

Carmena JM (2012) Becoming bionic. IEEE Spectr 49(3):24–29

Carmena JM (2013) Advances in neuroprosthetic learning and control. PLoS Biol 11:5

Choi JH, Yeonho K, Kyounghwan L, You CC (2008) Various wearable embroidery RFID tag antenna using electro-thread. In: IEEE Antennas and propagation society international symposium, San Diego, 5–11 July 2008

Collinger JL, Wodlinger B, Downey JE, Wang W, Tyler-Kabara EC, Weber DJ, McMorland AJ, Velliste M, Boninger ML, Schwartz AB (2013) High-performance neuroprosthetic control by an individual with tetraplegia. Lancet 381(9866):557–564

Dobkin DM (2007) The RF in RFID: passive UHF RFID systems in practice. Newnes, Burlington

Foster PR, Burberry RA (1999) Antenna problems in RFID systems. IEE Colloquium on RFID Technology, London, pp 3/1–3/5

Goojo K, Jinseong L, Kyoung HL, You CC, Junho Y, Byung HM, Jeenmo Y, Hee CK (2008) Design of a UHF RFID fiber tag antenna with electric-thread using a sewing machine. In: Asia-Pacific microwave conference, Macau, 16–20 Dec 2008

Hall PS, Hao Y (2006) Antennas and propagation for body centric communications. 1st EuCAP, Nice

Hall PS, Yang Y (2012) Antennas and propagation for body-centric wireless communications. Artech House, Boston

Kellomäki T, Ukkonen L (2010) Design approaches for bodyworn RFID tags. In: Proceeding of the 3rd ISABEL, Rome, Italy, 5 p

Koralek AC, Jin X, Long JD II, Costa RM, Carmena JM (2012) Corticostriatal plasticity is necessary for learning intentional neuroprosthetic skills. Nature 483:331–335

Koski E, Koski K, Björninen T, Babar AA, Sydänheimo L, Ukkonen L, Rahmat-Samii Y (2012) Fabrication of embroidered UHF RFID tags. In: Proceedings of the IEEE AP-S symposium, Chicago, 2 p

DOI 10.1007/978-981-4560-75-7_108-1

Koski K, Vena A, Sydänheimo L, Ukkonen L, Rahmat-Samii Y (2013a) Design and implementation of electro-textile ground planes for wearable UHF RFID patch tag antennas. IEEE Antennas Wirel Propag Lett 12(1):964–967

Koski K, Vena A, Sydänheimo L, Ukkonen L, Rahmat-Samii Y (2013b) Design and implementation of electro-textile ground planes for wearable UHF RFID patch tag antennas. IEEE AWPL 12:964–967

Koski K, Björninen T, Sydänheimo L, Ukkonen L, Rahmat-Samii Y (2014a) A new approach and analysis of modeling the human body in RFID-enabled body-centric wireless systems. Int J Antennas Propag 2014, 368090, 12 pages

Koski K, Lohan ES, Sydänheimo L, Ukkonen L, Rahmat-Samii Y (2014b) Electro-textile UHF RFID Patch antennas for positioning and localization applications. In: IEEE RFID-TA conference, Tampere

Kumar G, Ray KP (2003) Broadband microstrip antennas. Artech House, Boston

Locher I, Klemm M, Kirstein T, Tröster G (2006) Design and characterization of purely textile patch antenna. IEEE Trans Adv Packag 29(4):777–788

Marrocco G (2008) The art of UHF RFID antenna design: impedance-matching and size-reduction techniques. Antennas Propag Mag IEEE 50(1):66–79

Marrocco G (2010) RFID antennas for the UHF remote monitoring of human subjects. IEEE Trans Antennas Propag 55(6):1862–1870

Moradi E, Björninen T, Ukkonen L, Rahmat-Samii Y (2012) Effects of sewing pattern on the performance of embroidered dipole-type RFID tag antennas. IEEE Antennas Wirel Propag Lett 11(1):1482–1485

Moradi E, Björninen T, Sydänheimo L, Ukkonen L, Rabaey JM (2013) Analysis of wireless powering of mm-size neural tags in RFID-inspired wireless brain-machine interface system. In: IEEE RFID conference digest, pp 8–15, April 2013

Nicolelis MAL (2001) Actions from thoughts. Nature 409:403–407

Occhiuzzi C, Cippitelli S, Marrocco G (2010) Modeling, design and experimentation of wearable RFID sensor tag. IEEE Trans Antennas Propag 58(8):2490–2498

Ouyang Y, Chappell WJ (2008) High frequency properties of electro-textiles for wearable antenna applications. IEEE Trans Antennas Propag 56(2):381–389

Polívka M, Švanda M, Hudec P (2009) UHF RFID of people. In: Turcu C (ed) Development and implementation of RFID technology, chapter 4. I-Tech Education and Publishing, Vienna

Rabaey JM (2011) Brain-machine interfaces as the new frontier in extreme miniaturization. ESSDERC, Helsinki

Raumonen P, Sydänheimo L, Ukkonen L, Keskilammi M, Kivikoski M (2003) Folded dipole antenna near metal plate. In: Proceeding of the IEEE AP-S symposium, Columbus, OH. vol 1. pp 848–851

Ukkonen L, Sydanheimo L, Kivikoski M (2005) Effects of metallic plate size on the performance of microstrip patch-type tag antennas for passive RFID. Antennas Wirel Propag Lett IEEE 4:410–413

Ukkonen L, Sydanheimo L, Rahmat-Samii Y (2012) Sewed textile RFID tag and sensor antennas for on-body use. In: Antennas and Propagation (EUCAP), 2012 6th European Conference on Praque. pp 3450–3454, 26-30

Vena A, Moradi E, Koski K, Babar AA, Sydänheimo L, Ukkonen L, Tentzeris MM (2013) Design and realization of stretchable sewn chipless RFID tags and sensors for wearable applications. In: Proceedings of the IEEE international conference RFID, Orlando, FL, pp 176–183

Wang Z, Zhang L, Bayram Y, Volakis JL (2012) Embroidered conductive fibers on polymer composite for conformal antennas. IEEE Trans Antennas Propag 60(9):4141–4147

Wang Z, Zhang L, Volakis JL (2013) Textile antennas for wearable radio frequency applications. Text Light Ind Sci Technol 2(3):105–112

Yakovlev A, Kim S, Poon A (2012) Implantable biomedical devices: wireless powering and communication. IEEE Commun Mag 50(4):152–159

DOI 10.1007/978-981-4560-75-7_108-1

Yang L, Martin LJ, Staiculescu D, Wong CP, Tentzeris MM (2008) Conformal magnetic composite RFID for wearable RF and bio- monitoring applications. IEEE Trans Microwave Theory Tech 56(12):3223–3230

Yeonho K, Kyounghwan L, Yongju K, You CC (2007) Wearable UHF RFID tag antenna design using flexible electro-thread and textile. In: IEEE antennas and propagation society international symposium, Honolulu, 9–15 June 2007

Ziai MA, Batchelor JC (2010) Thin ultra high-frequency platform insensitive radio frequency identification tags. Microwaves Antennas Propag IET 4(3):390–398

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_109-1

Three-Dimensionally Printed/Additive Manufactured Antennas

Min Liang* and Hao Xin
Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ, USA

Abstract

Additive manufacturing (AM), or often referred to as three-dimensional (3D) printing, is an important emerging research area which has received much attention recently. It allows 3D objects with arbitrary geometry to be printed automatically layer by layer from bottom to top. This technology offers several advantages compared to conventional manufacturing techniques including the capability of more flexible design, prototyping time and cost reduction, less human interaction, and faster product development cycle. 3D printing techniques have been applied in many different sectors including mechanical engineering, electrical engineering, biomedical engineering, art, architecture, and landscaping. This chapter reviews state-of-the-art 3D printed antennas from microwave to THz frequencies and offers practical and futuristic perspectives on the challenges and potentials of 3D printed antennas. An overview of various 3D printing techniques relevant to antenna applications is presented first. A number of 3D printed antenna examples categorized by different AM methods are then described. Finally, technical challenges and possible solutions of 3D printing technology specific to antenna application, as well as new and revolutionary antenna design/realization concepts enabled by 3D printing technology, are discussed.

Keywords

Additive manufacturing; 3D printing; Computer-aided design; Automatic fabrication

Introduction

Additive manufacturing (AM), often called “3D printing” or “rapid prototyping”, is an automated fabrication technology to make a three-dimensional physical object directly from digital data. Contrary to the subtractive manufacturing which realizes a product by subtracting a material from a larger piece of material such as cutting out a screw from a piece of metal, it makes a product layer by layer additively.

AM was originated in the United States and was first commercialized in the late 1980s. At that time, it was called “rapid prototyping (RP)” or “generative manufacturing” (Gebhardt 2012), and these terms are still occasionally in use presently. In the early 1990s, several different AM processes including laser sintering (LS) (Agarwala et al. 1995) and fused deposition modeling (FDM) (Griffin and McMillin 1995) were developed and became available commercially. In the mid-1990s, another 3D printing process which creates an object by jetting a liquid binder onto a bed of powder and doing post-processing to solidify the whole structure was invented (Bak 2003). After that, through the rest of the 1990s, further research and development were mainly focused on materials such as various thermoplastics (Kambour 1973) and elastomeric polymers (Kornbluh et al. 2000) in different forms to enable AM techniques to be used in more applications. As the new century begins, the focus was shifted back to improving the AM technology by developing new printing processes. New techniques such as the laser melting (LM) and

*Email: minliang@email.arizona.edu

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_109-1

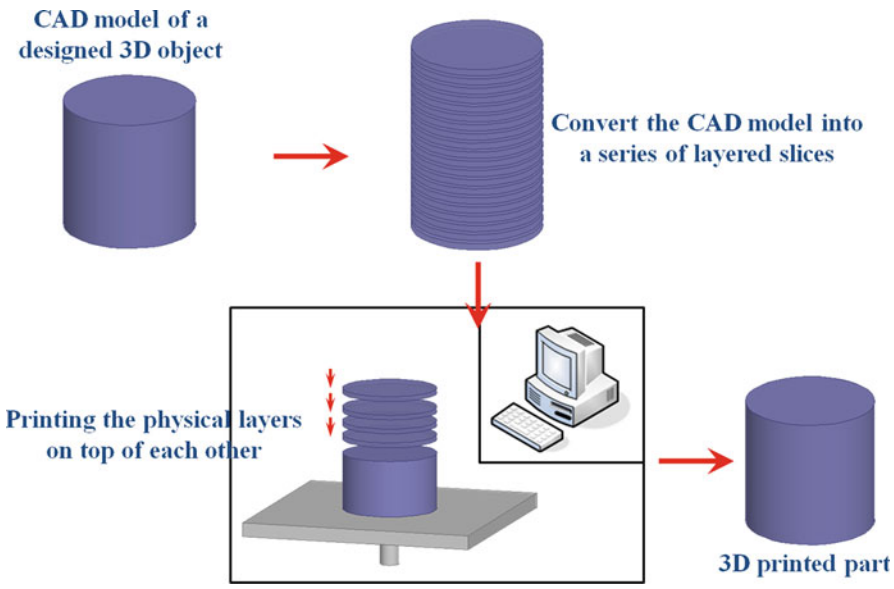


Fig. 1 Schematic illustration of a typical AM process

electron beam melting (EBM) processes were developed. These techniques allow various alloyed materials to become available in the AM process. Over subsequent years, more and more AM companies are founded from all over the world and starting to develop their own printable materials and AM systems. Many new types of materials and systems become available as the demand for AM increases. It is realized that these techniques are not just for rapid prototyping; instead, they can be developed and applied as a new form of manufacturing technology. Therefore, from then on, the name "additive manufacturing" has been coined. Recently, AM has received much attention with impressive demonstrations ranging from musical instruments, to vehicles, to housing components or even entire buildings. Many different structural materials such as metal, polymer, ceramics, concrete, and even biocompatible materials have been incorporated in various 3D printing technologies. Due to its ability to realize desired structures with arbitrarily designed spatial distribution, 3D printing technology has been argued to be the future of manufacturing as it offers huge potentials to revolutionize both the design and manufacturing procedures.

The technical realization of AM is based on layer-by-layer processes, and therefore it is called "layer-based technology" or "layer-oriented technology." The working principle of the layer-based techniques is to create a 3D physical structure from many slices with the same thickness. Each slice is fabricated according to the information from the corresponding 3D model and placed on top of the pervious layer. A schematic illustration of a typical AM procedure is shown in Fig. 1. The process starts with a 3D computer-aided design (CAD) model which represents the 3D object to be printed. This CAD model can be created directly from CAD software or by digital 3D scanning of a real structure. After the CAD model is obtained, specialized software is used to slice the model into layer-by-layer cross sections. As a result, a series of layered slices with equal thickness are generated. The information of these slices including position, layer thickness, and layer number is sent to a machine that could print each layer and bond it to the previous one. The printing and bonding of the layers can be done in many ways based on different physical phenomena. By printing the object layer by layer, the entire structure is built from bottom up.

These basic steps are the same for almost all varieties of AM equipment available today. The differences of different equipment are how they generate the layers, how the adjacent layers are joined together to form the final part, and the corresponding built materials.

Compared to conventional manufacturing methods (such as injection molding, casting, stamping, and machining), the AM approach has the following advantages.

DOI 10.1007/978-981-4560-75-7_109-1

Arbitrary Complexity

AM approach has the ability to create 3D objects with arbitrary shape and complexity. The cost of the 3D printed components is only related to the volume of the parts; there is no additional cost or lead time for making the structure more complex. Also, with multiple printing heads, it is possible to cohesively integrate different materials at the same location simultaneously. Therefore, AM may revolutionize product designs because of the much more flexible object geometry and material property distribution it offers. For example, 3D structures with arbitrary EM property distribution can be printed relatively easily.

Digital Manufacturing

After an object is designed, the whole 3D printing process is accurately controlled by a computer with very little human interaction needed to realize the design. This automatic 3D printing process means that the time between design iterations can be dramatically reduced compared to conventional manufacturing methods.

Waste Reduction

The 3D printed component is created bottom up via layer-by-layer processes so that only materials needed for the design are used. Therefore, material waste in AM process will be much less than conventional subtractive manufacturing techniques.

Various 3D printed antennas have been reported taking advantages of the AM technology. Antennas of different structures such as horn antennas (Huang et al. 2005), patch antennas (Liang et al. 2014a), meander line antennas (Adams et al. 2011), gradient index (GRIN) lens antennas (Liang et al. 2014a), and reflect-array antennas (Nayeri et al. 2014), made of different materials such as all-dielectric antenna (Wu et al. 2012), all-metal antenna (Garcia et al. 2013), and dielectric-metal combined antenna (Liang et al. 2014; Adams et al. 2011; Nayeri et al. 2014), working at different frequencies from GHz to THz, have been realized using different 3D printing techniques. In the next section, an overview of various AM techniques relevant to antenna application is provided, and the pros and cons for each are discussed.

Overview of 3D Printing Techniques

At the present time, there are many kinds of 3D printing techniques, all of which follow the basic steps of AM discussed in the previous section, for example, generating individual physical layers and combining them together. Various materials such as metal, plastic, ceramics, or even biocompatible materials can be used in the generation of the physical layers. According to the methods of generating physical layers and bonding adjacent layers together to form an object, five basic categories of AM processes are commercially available (Gebhardt 2012), including selective sintering and melting, powder binder bonding, polymerization, extrusion, and layer laminate manufacturing (LLM). Key aspects of these five processes are discussed and some commercially available 3D printers as well as printed examples are reviewed.

Selective Sintering and Melting

The 3D printing technique using a laser to selectively sinter or melt powdered material is called selective laser sintering (SLS) (Agarwala et al. 1995) or selective laser melting (SLM) (Kruth et al. 2004). If an electron beam is used instead of laser, the process is called electron beam melting (EBM) (Cline and Anthony 1977).

An SLS printer usually includes a building chamber to be filled with powdered built material and a laser beam on top that can be scanned precisely in the XY (horizontal) plane. The bottom of the chamber is

DOI 10.1007/978-981-4560-75-7_109-1

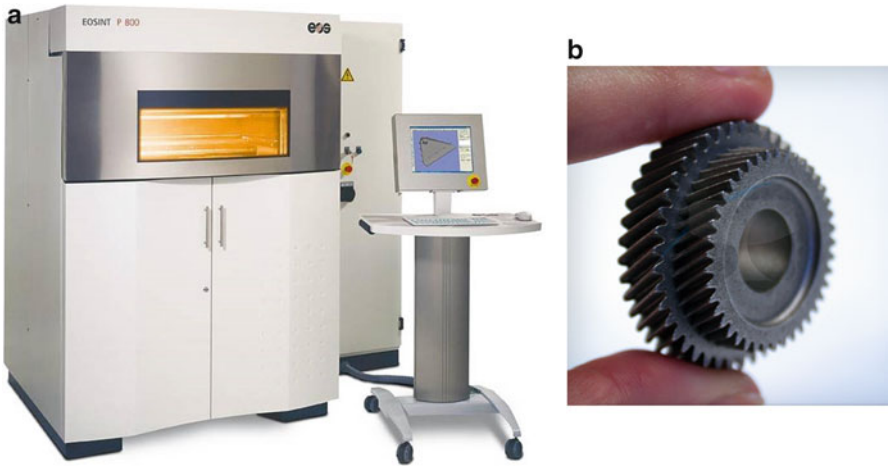


Fig. 2 (**a**) Photo of a selective laser sintering printer (*SLS*) (model EOS P800; size: 2.25 m × 1.55 m × 2.1 m). (**b**) An SLS printed metallic object

moveable in the Z (vertical) direction. During the printing process, the entire chamber is heated to a high temperature close to the melting point of powder so that they are at an optimal temperature for melting. To prevent oxidation, the chamber is often filled with shielding gas (e.g., nitrogen). The scanning laser beam is then used to fuse the powders at designated locations. As the laser beam travels in the XY plane, the melted powders cool down and solidify. After the scanning of an entire layer at designated positions, a solid layer with designed pattern is achieved. After one layer is printed, the powder bed is lowered by the amount of one layer thickness and an automated roller adds a new layer of powdered built material on the top of the previous layer. Then the selective melting process repeats until the entire object is printed. The remaining unsolidified powders are then removed after printing. The SLS technique is quite versatile since it can be used to print several classes of materials, including plastics, metals, and ceramics.

Typically, SLS-fabricated metal parts such as steel and titanium are dense. They may be post-processed by cutting or welding, depending on specific materials involved. Plastic parts such as nylon and polystyrene fabricated using SLS have properties similar to those made by plastic injection molding. As an example, Fig. 2 shows an SLS printer (EOS P800) and a metallic part made by using SLS.

Selective laser melting (SLM) is developed in particular to process metal parts that need to be very dense (>99 %). In this case, the laser beam melts the metal powders completely into liquid phase which results in a close to 100 % density part after resolidification. SLM can be used to print many metals including stainless steel, carbon steel, CoCr, titanium, aluminum, gold, and a large variety of alloys.

EBM is a similar 3D printing process in which metal powders are melted or fused by applying an electron beam under a high voltage (typically in the range of 30 ∼ 60 KV) instead of a laser beam. To avoid oxidation, the process is performed in a high-vacuum chamber. Because the electron beam penetrates much deeper than a laser beam, EBM allows a higher scanning speed. In addition, deeper penetration can be used for powder preheating so that the printing process works at elevated temperatures compared to the laser case. As a result, the mechanical stress and distortion of printed objects are reduced and greater strength can be achieved. Figure 3 shows an example of an EBM 3D printer and a 3D printed object using EBM technique.

Sintering and melting processes are very suitable for applications requiring high strength and/or high temperature. Antennas printed by SLS or EBM can be very dense, void-free, and very strong. The disadvantages of selective sintering and melting techniques are that the printing resolution is limited by the size of the powders (i.e., tens of microns) and a high-vacuum chamber or shielding gas is needed to avoid oxidation (Kruth et al. 2003).

DOI 10.1007/978-981-4560-75-7_109-1

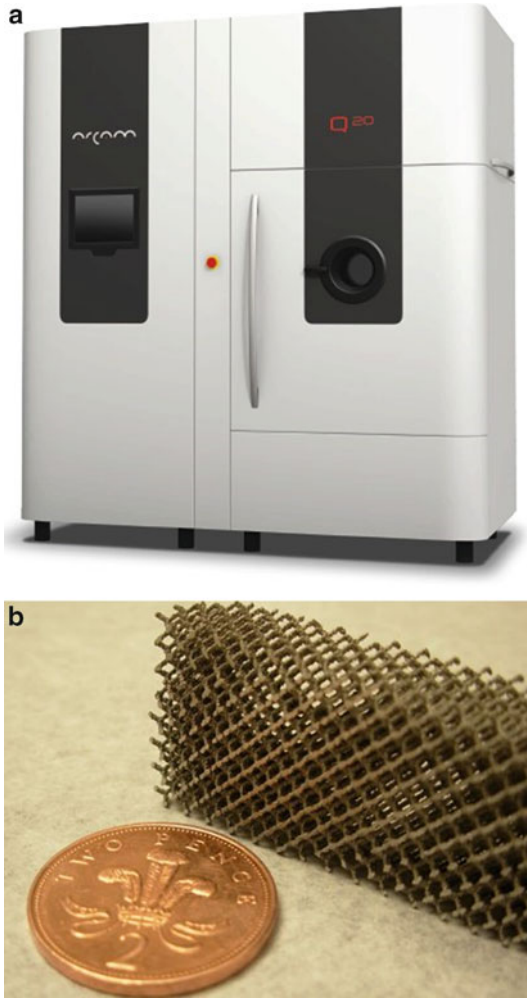


Fig. 3 (**a**) Photo of an electron beam melting (*EBM*) printer (model Arcam Q20; size: 2.3 m × 1.3 m × 2.6 m). (**b**) An EBM-fabricated part

Powder Binder Bonding

Powder binder bonding is another 3D printing technique that implements layer-by-layer bonding of powdered materials by selectively injecting a liquid binder onto the powder bed. This process was first developed in the mid-1990s. Currently, various materials such as plastics, metals, and ceramics can be printed using this technique.

A typical powder binder bonding printer is very similar to a selective laser sintering printer with a piston at the bottom of chamber to adjust the height and a roller to recoat the powders. The printing process starts with depositing small drops of liquid binder onto a layer of built material powders at designated locations. The powders forming the designed structure are bounded together while the surrounding loose powders support the next layer of the structure to be printed. The printing process is then repeated for each layer until the entire structure is completed. Compared to the sintering or melting process, this process is performed at much lower temperature. Therefore, no preheating, shielding gas, or vacuum chamber is needed.

At the end of the printing process, the residual powders are removed and an infiltration process may be performed for enhanced durability. For plastic parts, wax or epoxy resin can be used in the infiltration process. If this technique is used to print a metallic antenna (Lopez et al. 2013), a subsequent high-temperature process is needed to provide strength and durability. For example, to print a bronze object, the

DOI 10.1007/978-981-4560-75-7_109-1

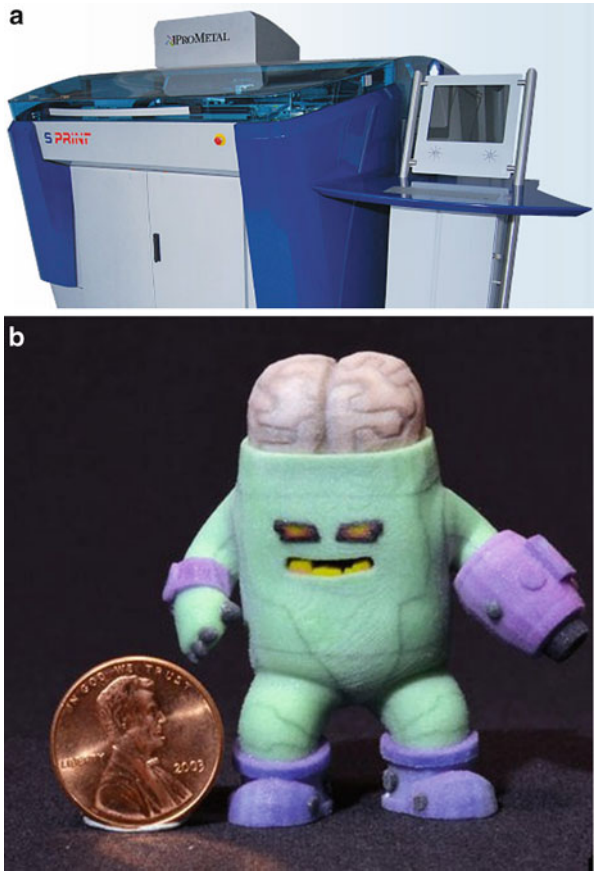


Fig. 4 (**a**) Photo of a powder binder bonding 3D printer (model ProMetal S15; size: 3.1 m × 3.4 m × 2.2 m). (**b**) An example printed using the powder binder bonding technique

printed part needs to be infused into bronze powder and heated up to more than 1,000 °C to replace the binder with bronze (Lipkea et al. 2010). This process can also be used to print alloy materials by changing the sintering temperature and time during the infiltration process (Kruth et al. 2005). Figure 4 shows a powder binder bonding 3D printer and an example printed using this technique. Similar to the sintering and melting processes, the resolution of this technique is also limited by the size of the powders. For the currently available printer on the market, the minimum feature size is 0.1 mm.

Polymerization

Polymerization is a process that selectively solidifies liquid resin using ultraviolet radiation or other power sources. Typically, photosensitive polymers are used as built material. There are several kinds of AM methods based on the polymerization process. Their differences are mainly in how the photon energy is applied and how the layers are created.

Stereolithography is the most accurate polymerization process which employs an ultraviolet laser to solidify a liquid ultraviolet curable polymer. To print each layer, a laser beam scans over the surface of a liquid polymer reservoir to cure the cross section according to the designed pattern. The curing thickness can be adjusted by the laser power and laser scanning speed. After one layer is printed, the building stage descends a distance of one layer thickness. Then, a blade sweeps across the surface of the printed part, recoating it with fresh liquid polymer before the next layer is printed on top. It is possible to incorporate different materials in the printing process, thus achieving multiple material stereolithography (Gebhardt 2012). In this case, the resin needs to be drained and replaced by the new material. After an object is

DOI 10.1007/978-981-4560-75-7_109-1

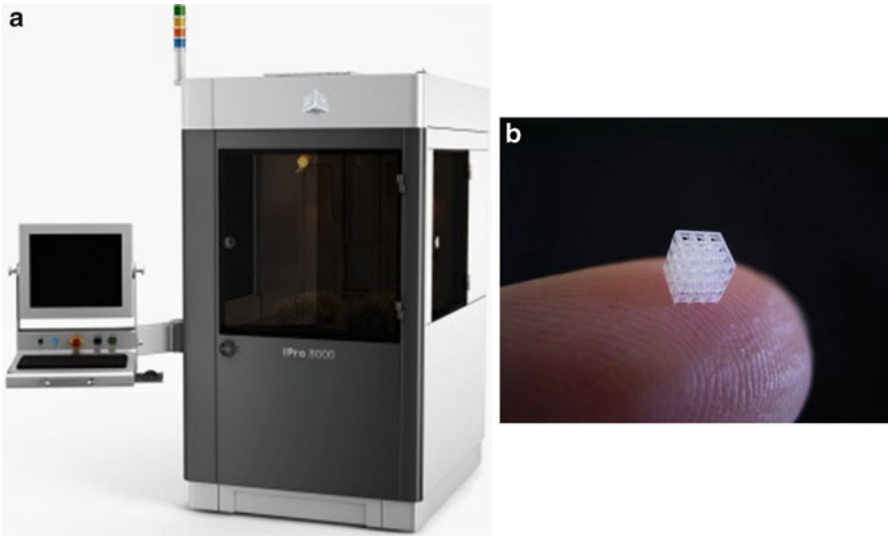


Fig. 5 (**a**) Photo of a laser stereolithography 3D printer (model systems iPro™ 8000; size: 1.26 m × 2.2 m × 2.28 m). (**b**) A sample fabricated using stereolithography technique

printed, it is cleaned and moved into a UV chamber for a final post-curing process to make it more stable. Figure 5 shows a stereolithography 3D printer example and an object realized by stereolithography. Compared to other AM techniques for 3D printing of antennas, the stereolithography process can achieve a very good surface smoothness and finer resolution. In fact, a two-photon stereolithography process has been reported to obtain a submicron printing resolution (http://www.nanoscribe.de/en/). However, the strength of a 3D printed part by stereolithography is weaker than other techniques such as sintering, melting, or powder binder bonding.

If a photosensitive polymer is applied by printer heads, the AM process is called polymer jetting. During printing, the printer head deposits photosensitive polymers onto a stage with designed patterns. Upon jetting, the printed photosensitive polymers are immediately cured by an ultraviolet lamp on the printer head and unlike stereolithography, no post-curing process is needed. The thickness of each layer of this process can be on the order of 20 μm, which provides a very smooth surface. Moreover, multiple types of polymers can be printed simultaneously using multiple printer heads. A gel type of polymer can be used as support material to print overhanging structures and released (e.g., water-soluble support materials can be washed away) after the printing process. A schematic drawing of the polymer jetting procedure is shown in Fig. 6. The polymer jetting method can only be applied to print polymers, limiting its applications to all-dielectric antennas. An additional metallization process would be required to incorporate the conductor part. It has a better resolution than sintering and powder binder bonding techniques. However, similar to stereolithography, parts printed by polymer jetting are not as strong as some of the other AM techniques.

Extrusion

Extrusion, often called fused deposition modeling (FDM), is an AM process that prints an object by extruding thermoplastics through a heated nozzle. A FDM printer includes a feeding roll, a heated extrusion head, and a building platform. The building materials are usually thin thermoplastic filaments which are wound up and stored in a cartridge. The thin thermoplastic filament is guided into the extrusion head by the feeding roll. During the printing process, the heated extrusion head melts down the filament

DOI 10.1007/978-981-4560-75-7_109-1

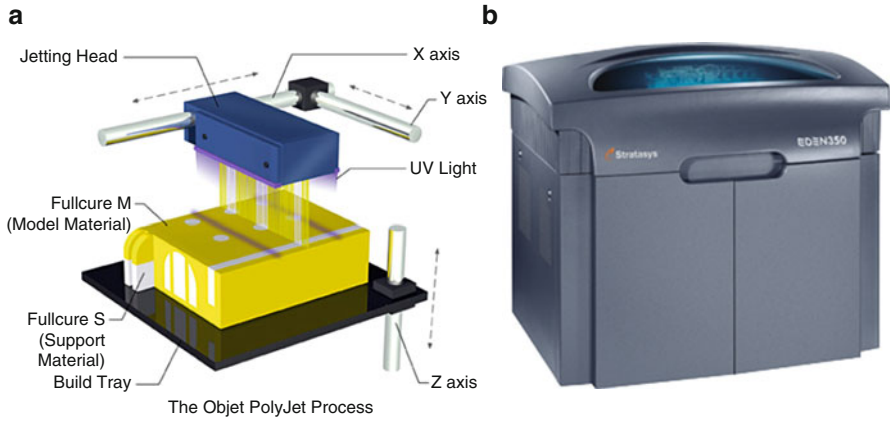


Fig. 6 (**a**) Schematic picture of the polymer jetting technique. (**b**) Photo of a polymer jetting 3D printer (model Stratasys Eden350V; size: 1.3 m × 1 m × 1.2 m)

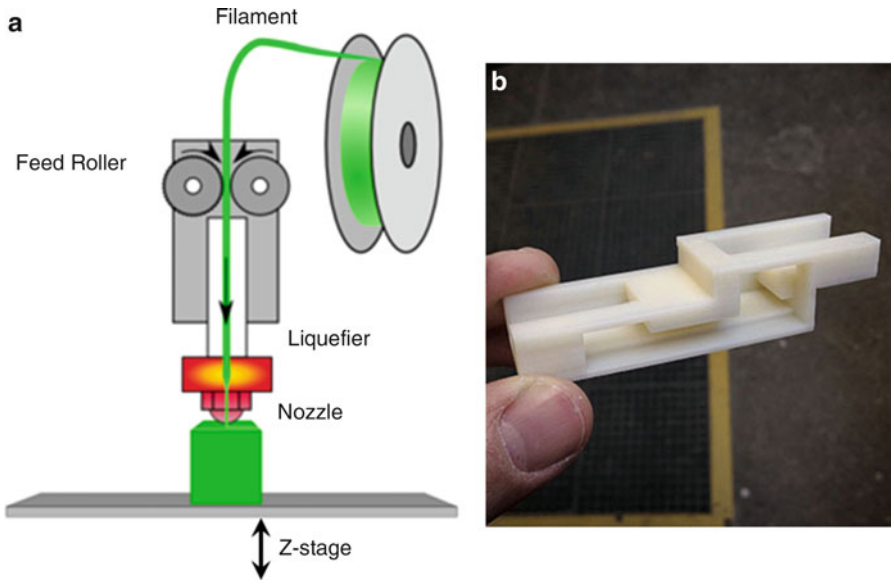


Fig. 7 (**a**) Schematic of FDM. (**b**) An example printed using FDM

and extrudes it through the nozzle at designated locations on the building platform. When the extruded thermoplastic reaches the building platform, it cools down and hardens. After one layer is completed, the platform lowers down by one layer thickness and is ready for printing of the next layer. Figure 7 shows the schematic of an FDM printer and a printed example.

There are a number of available built materials for FDM including polycarbonate (PC), acrylonitrile butadiene styrene (ABS), polyphenylsulfone (PPSF), etc. The advantages of using this technique to print antennas are the relatively simpler processing (i.e., no post-processing needed) and lower printer cost compared to other AM techniques. The disadvantage of FDM is lower resolution (about 0.25 mm (Wong and Hernandez 2012)).

Layer-by-Layer Bonding

Layer-by-layer bonding is an AM technique that creates a 3D structure by cutting a prefabricated sheet or foil into a designed contour and subsequently bonding a number of layers together. It is often called

DOI 10.1007/978-981-4560-75-7_109-1

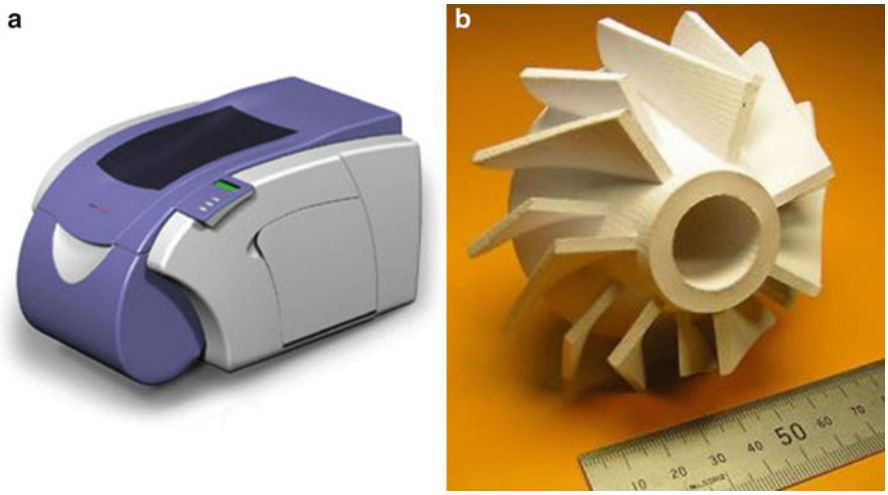


Fig. 8 (**a**) Photo of a LOM 3D printer (model Solidimension SD300; size: 450 mm × 725 mm × 415 mm). (**b**) An object made of paper printed using the LOM method

laminate object manufacturing (LOM). A LOM printer consists of a building platform that can move in the z-direction, a foil supply system to supply and position the foil, and a cutting device to create the contour. The LOM procedure is as follows: First, the foil is positioned and adhered to the building platform by a heated roller. Second, the cutting tool scans on the foil to create the designed contour and perform crosscutting on the non-model area to make it into small pieces for easier removal after printing. After one layer is printed, the platform moves down and the roller positions the next layer of foil on top of the previous layer. Then the platform moves up into position to receive the next layer and the process repeats until the entire 3D object is printed completely. A photo of a LOM 3D printer using paper material is shown in Fig. 8 together with a 3D printed example.

The foil built materials for the LOM technique can be paper, plastic, or metal (Gebhardt 2012). The cutting tool can be a scanning laser, a knife, or a milling machine. To bond adjacent layers, different methods such as gluing, soldering, and ultrasonic or diffusion welding can be used depending on the material properties. Compared to other AM techniques, the advantages for using the LOM in antenna printing include lower material cost and faster building speed for large objects. The disadvantages are less accuracy (e.g., 0.3 mm for the Solidimension SD300 3D printer shown in Fig. 8) and some material waste depending on the geometry.

AM Technique Summary

Most of the AM processes currently available can be classified by the abovementioned five basic categories. Table 1 summarizes the key features of these techniques.

3D Printed Antennas

AM technology enables the flexible and rapid realization of structures with arbitrary shapes and complexity. It has been successfully applied in many scientific and industrial areas such as biomedical, aerospace, and toy industry, architecture, and landscaping (Gebhardt 2012). In the following sections, applications of AM techniques for realizing 3D printed antennas are reviewed. A number of antenna examples printed by different AM techniques including electron beam melting, powder binder bonding, stereolithography, polymer jetting, conductive ink printing, and fused deposition modeling are presented.

DOI 10.1007/978-981-4560-75-7_109-1

Table 1 Summary of key characteristics of the five basic categories of AM processes

Classification	Available material	Process temperature	Resolution	Strength	Can print overhanging structure?	Print multiple materials
Sintering and melting	Polymer, metal and ceramic	High temperature	Low	Strong	Yes	No
Powder binder bonding	Polymer, metal and ceramic	Depending material	Moderate	Moderate	Yes	No
Polymerization	Polymer	Room temperature	High	Moderate	Need support material	Yes
Extrusion	Polymer	200–300 °C	Low	Strong	Need support structure	Yes
Layer by Layer bonding	Paper, plastic and metal	Depending on material	Low	Moderate	Yes	No

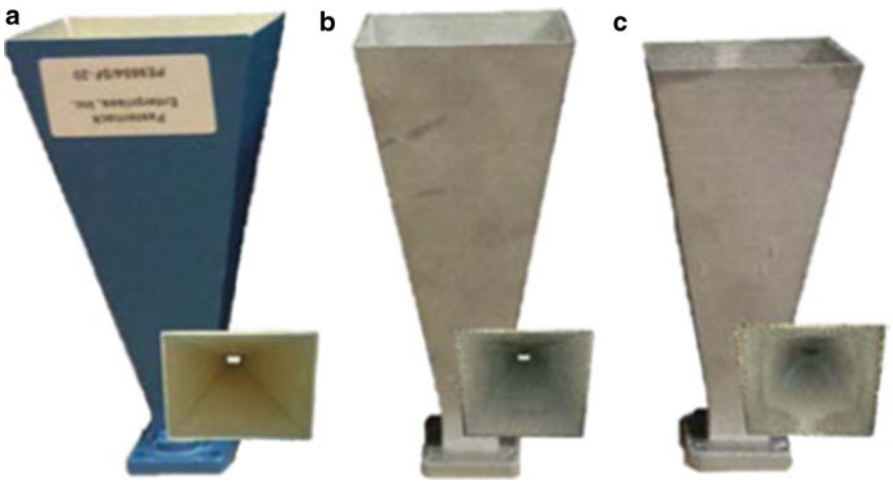


Fig. 9 Front view and side view of three Ku-band horn antennas. (**a**) A commercial standard gain horn as the reference. (**b**) EBM printed horn antenna with a dimension of 145.1 mm × 52.8 mm × 66.2 mm. (**c**) EBM printed horn antenna with a dimension 131.9 mm × 47.3 mm × 55.2 mm (Garcia et al. 2013)

Antenna Printed Using Sintering and Melting

As mentioned in section "Overview of 3D Printing Techniques," sintering and melting is an AM technique which uses laser or electron beam to selectively melt powder material and build 3D structures. In Garcia (2013), two horn antennas operating at Ku-band are printed by employing the EBM technique. Figure 9a is a commercial standard gain horn antenna manufactured conventionally as the reference, while Fig. 9b, c are two 3D printed horns with different dimensions. The RMS (root mean square) surface roughness of the 3D printed horns is measured to be 25.9 μm and 39.7 μm for the antennas in Fig. 9b, c, respectively. The simulated antenna gains without considering surface roughness are compared with measured results at 15 GHz, as shown in Table 2. One can see that the measured antenna gains agree very well with the simulated results for the reference horn antenna and the 3D printed horn antenna in Fig. 9b. For the 3D printed horn antenna in Fig. 9c, the measured antenna gain is about 0.6 dB lower than the simulated result which is attributed to the relatively worse surface roughness.

DOI 10.1007/978-981-4560-75-7_109-1

Table 2 Comparison of simulated and measured gains at 15 GHz for the reference and EBM printed horn antennas shown in Fig. 9 (Garcia et al. 2013)

Name	Simulated gain (dBi)	Measured gain (dBi)
Reference horn	19.56	20.00
3D printed horn 1	18.98	19.02
3D printed horn 2	18.42	17.78

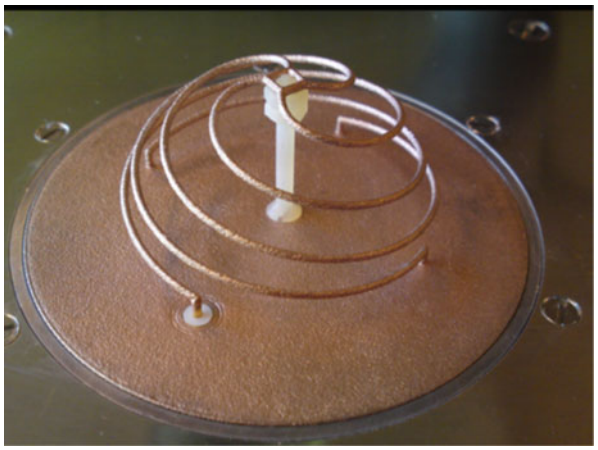


Fig. 10 SLS printed spherical helix antenna with several layers of copper painting (Kim 2013)

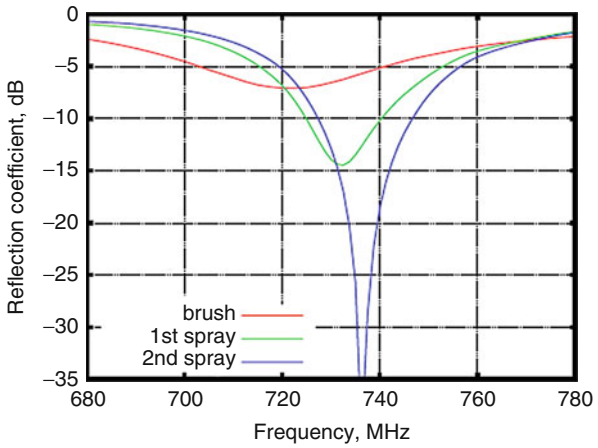


Fig. 11 Measured reflection coefficients of the SLS printed spherical helix antenna during painting process (Kim 2013)

In Kim (2013), an SLS technique is also applied to print an electrically small spherical helix antenna. The antenna is first printed with nylon using an SLS technique and then coated with several layers of conductive paint to make it conductive. A photo of the antenna is shown in Fig. 10. This helix antenna is fed by a SMA connector from the bottom. Figure 11 plots the measured reflection coefficients at different painting steps. The measured resonance frequency is at 736.3 MHz, which is lower than the expected 750 MHz. The reason is attributed to the gravity-induced deformation of the helix arms. The measured

DOI 10.1007/978-981-4560-75-7_109-1

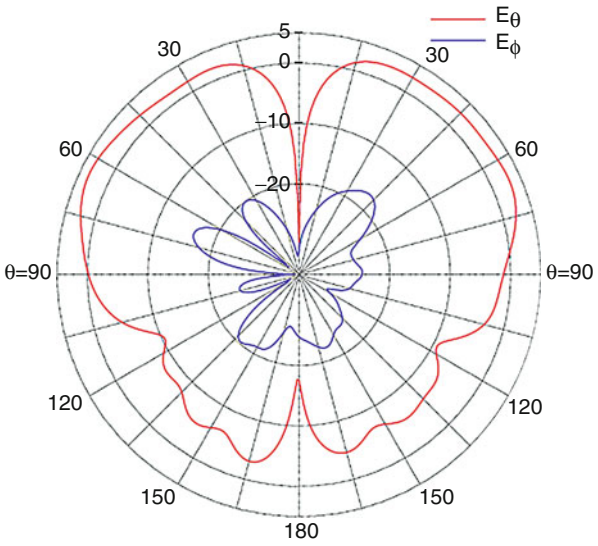


Fig. 12 Measured radiation pattern of the 3D printed spherical helix antenna at 737 MHz (Kim 2013)

radiation efficiency is 80 % compared to the simulated result of 97 % based on solid copper. The discrepancy is from the conductive loss of the copper paint. The measured radiation pattern of the antenna in the elevation plane is plotted in Fig. 12 and it is close to omnidirectional with a null in the broadside direction.

In summary, microwave horn antennas printed by EBM and SLS methods have been successfully demonstrated. It is observed that the surface roughness of 3D printing methods can influence the antenna performance such as gain. This issue will be more severe for higher-frequency bands such as mmW and THz. Moreover, the relatively coarse printing resolution that can be achieved by EBM, SLS, and SLM may also limit their applications for those higher-frequency bands.

Antenna Printed Using Powder Binder Bonding

Another 3D printing technique that is capable of printing pure metallic structures, the powder binder bonding technique, is utilized to realize a 3D volcano smoke antenna (Lopez et al. 2013) for ultrawideband (UWB) applications. The antenna is built using steel material. However, due to the low conductivity of steel, two methods to improve the antenna performance are applied. One method is covering the 3D printed prototype with copper tape. The other method is electroplating the prototype with copper. The antenna geometry and a photo of the 3D printed antenna prototype are shown in Fig. 13. The measured antenna reflection coefficients for both methods are plotted in Fig. 14 together with the simulated result. For the copper tape-covered case, the discrepancy between the measured and simulated results is believed to be due to the glue on the copper tape which would introduce an air gap between the steel and copper. For the electroplating case, the difference between simulation and measurement may come from the insufficient plating of copper layer at some part of the antenna. Nevertheless, the reflection coefficients of the 3D printed antennas show a broadband behavior for both cases.

Antenna Printed Using Stereolithography (SL)

Stereolithography is one of the most accurate AM techniques. It has also been applied in the realization of microwave antennas. An example using stereolithography and electroplating approach to build horn antennas at Ku-band is reported in Huang et al. (2005). Two horn antenna prototypes are first printed using polymer. Then, the stereolithography printed parts are coated by conductive silver ink as a seed layer and electroplated with copper. Figure 15 shows the schematic of the horn and a photo of the antenna after the

DOI 10.1007/978-981-4560-75-7_109-1

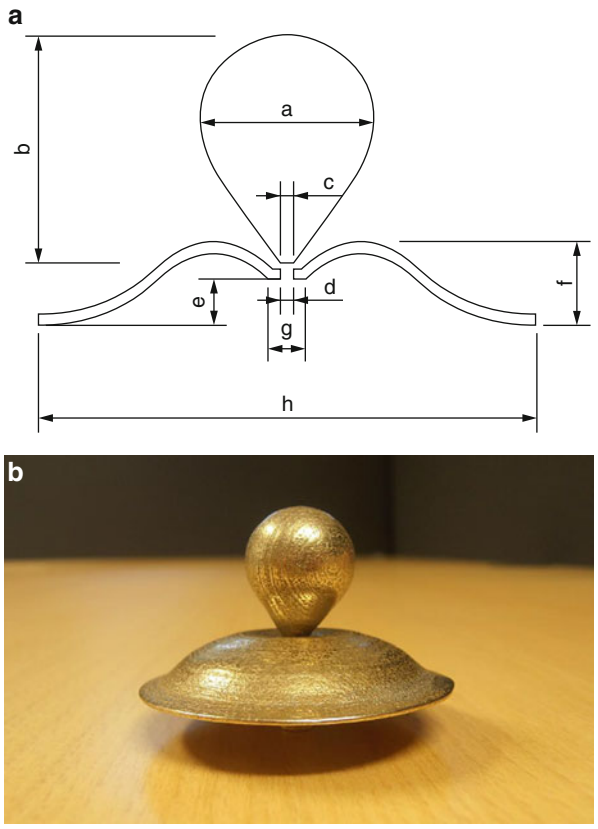


Fig. 13 (**a**) Schematic of a designed 3D volcano smoke UWB antenna. (**b**) A photo of the 3D printed volcano smoke UWB antenna prototype (without copper tape covering or electroplating) by the powder binder bonding technique. The parameters in (**a**) are a = 16.7 mm, b = 21 mm, c = 1.5 mm, d = 2 mm, e = 4.2 mm, f = 8 mm, g = 3.5 mm, and h = 45.6 mm (Lopez et al. 2013)

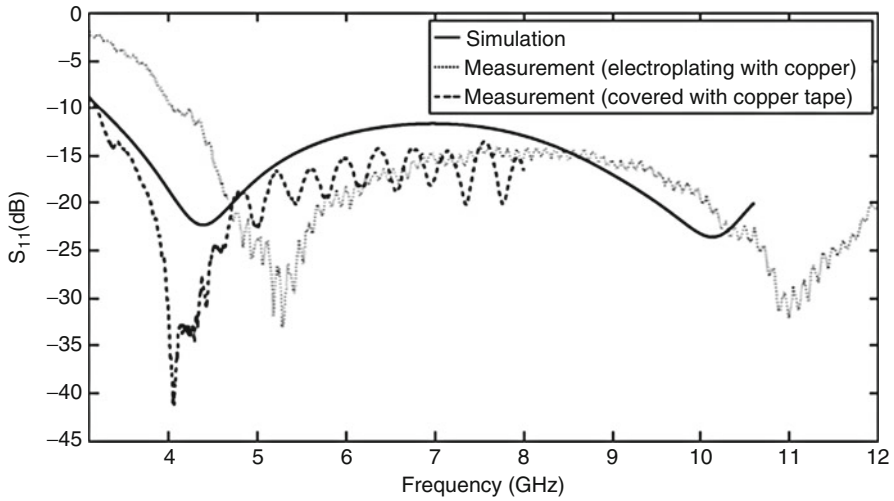


Fig. 14 Measured reflection coefficients of the powder binder printed volcano smoke antennas compared to the simulation result. The *solid line* is the simulation result, the *dotted line* is the measurement result of the electroplated antenna, and the *dashed line* is the measurement result of the copper tape-covered antenna (Lopez et al. 2013)

DOI 10.1007/978-981-4560-75-7_109-1

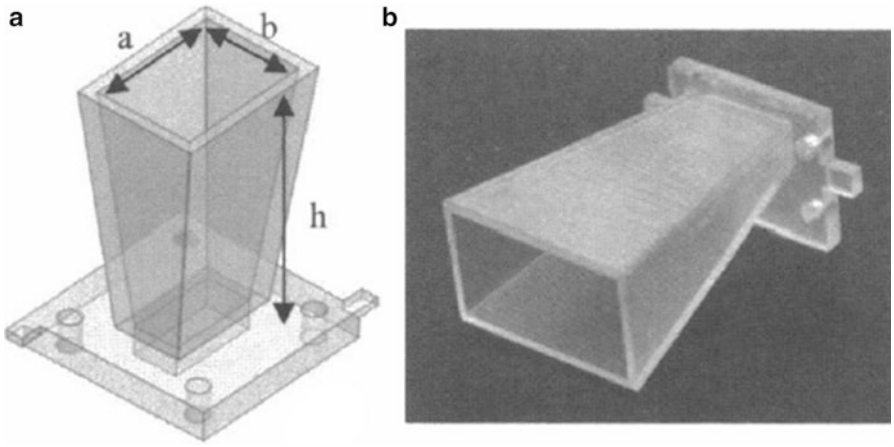


Fig. 15 (**a**) Schematic picture of a stereolithography printed horn antenna. (**b**) A photo of a 3D printed horn using stereolithography technique (Huang et al. 2005)

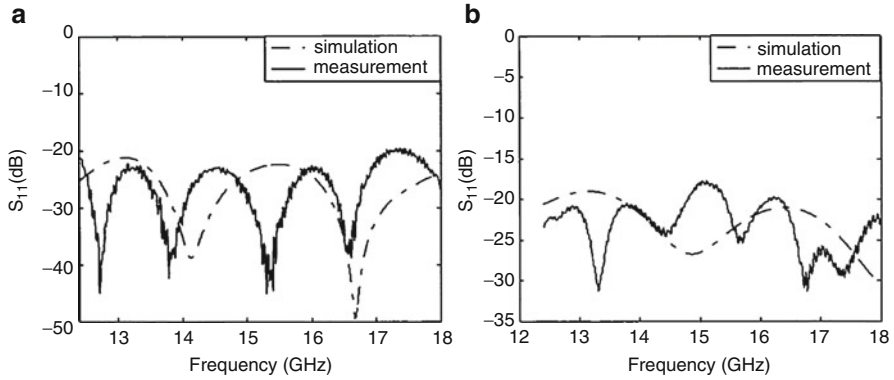


Fig. 16 Measured and simulated reflection coefficients of the stereolithography printed (**a**) horn antenna I and (**b**) horn antenna II

stereolithography printing process. The measured reflection coefficients of two printed horn antennas with different dimensions are plotted in Fig. 16 together with the simulation results. It can be clearly seen that the measured reflection coefficients of both antennas are quite good and they agree reasonably with the simulation results. The measured radiation patterns of the 3D printed horn antennas are plotted in Fig. 17, and the results are also very close to the simulated patterns. The measured and simulated antenna gains of the two printed horns are compared in Table 3.

In Chieh et al. (2014), a Ku-band corrugated conical horn antenna made via the stereolithography technique is also reported. This antenna is printed using acrylonitrile butadiene styrene (ABS) and then coated with conductive aerosol paint. Figure 18 shows the photos of the 3D printed horn antenna before and after the conductive aerosol paint process. The total time in printing the antenna is about 8 h. The surface coating of the antenna is Super Shield 841 which is a conductive spray typically used for EM interference (EMI) reduction. A total DC resistance of 2 Ω is measured across the length of the antenna after final coating. The measured and simulated reflection coefficients are compared in Fig. 19. It can be seen that the measured results agree well with simulation from 11 to 18 GHz. The spike around 18 GHz is believed to be caused by the excitation of the TE_{01} mode from the rectangular-to-circular waveguide transition. The measured co- and cross-polarization patterns at 14 and 16 GHz are plotted in Fig. 20. It is observed

DOI 10.1007/978-981-4560-75-7_109-1

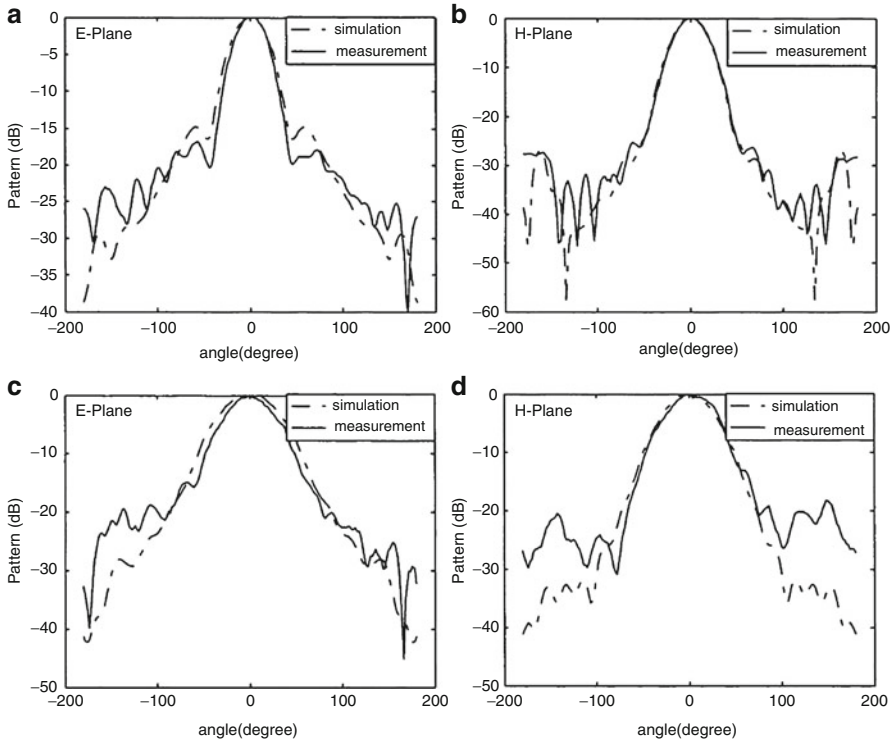


Fig. 17 Simulated and measured radiation patterns of the stereolithography printed (**a**) horn antenna I in E-plane, (**b**) horn antenna I in H-plane, (**c**) horn antenna II in E-plane, and (**d**) horn antenna II in H-plane (Huang et al. 2005)

Table 3 Measured and simulated gain and efficiency of the stereolithography printed horn antennas

	Size($a \times b \times h$)	Measured gain	Simulated gain	Efficiency
Horn antenna I	$40.2 \times 29.3 \times 51.0$ mm^3	14.58 dB	14.52 dB	100 %
Horn antenna II	$23.7 \times 117.3 \times 40.9$ mm^3	10.15 dB	10.35 dB	95.5 %

that the measured main beam agrees with the simulated results well and the cross-polarization level is at least 20 dB lower.

These works show successful examples of using stereolithography printing and metallic coating to fabricate microwave antennas at Ku-band. Because of the high resolution of the stereolithography technique, it can be applied in the realization of a finer structure than other AM techniques and therefore achieved at a higher operating frequency. Nevertheless, the required separate metallization process could be cumbersome and nonideal.

Antenna Printed Using Polymer Jetting

Polymer jetting is another kind of AM technique based on polymerization. Nayeri et al. (2014) report an example using polymer jetting technique to print dielectric reflect arrays as high-gain antennas operating at W-band (75–110 GHz). A reflect-array antenna is a class of antennas that uses one or a number of driving elements in front of a designed flat reflecting surface to produce a high directional beam. In Nayeri et al. (2014), the phase control of the reflect-array elements is achieved by varying the thickness (height) of the dielectric slab in each unit cell, and the designed structure of the reflect-array antenna is printed using a polymer jetting printer (Objet Eden350). There are 400 unit cells in the reflect arrays. After the polymer structure is printed, a thin layer of gold with thickness of about 100 nm is sputtered on the back of

DOI 10.1007/978-981-4560-75-7_109-1

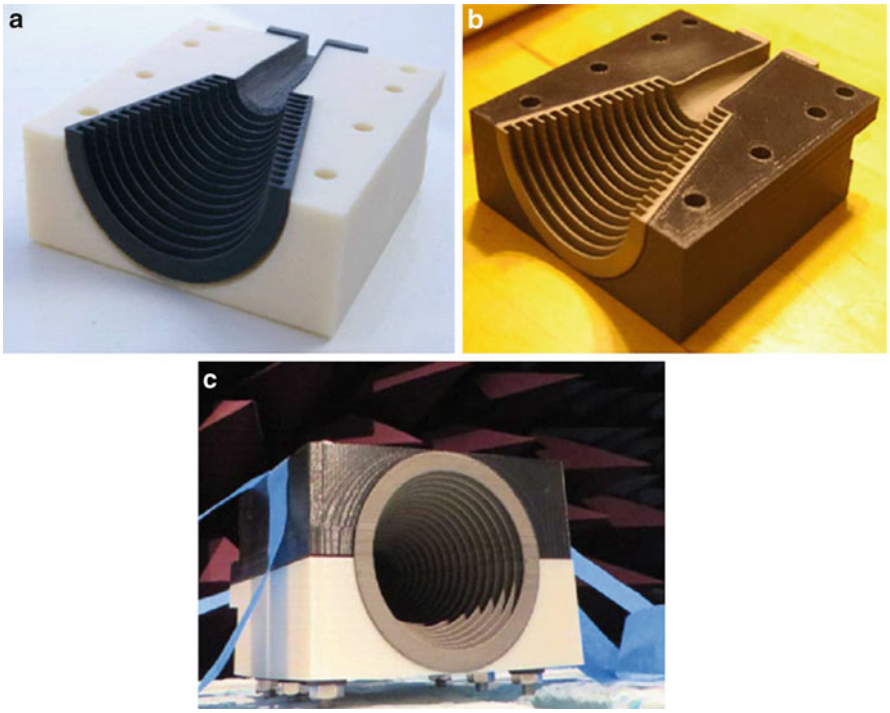


Fig. 18 A 3D printed Ku-band corrugated conical horn antenna using stereolithography before and after conductive aerosol paint (Chieh et al. 2014). (**a**) Half of the printed horn antenna before painting. (**b**) Half of the printed horn antenna after painting. (**c**) The complete final antenna

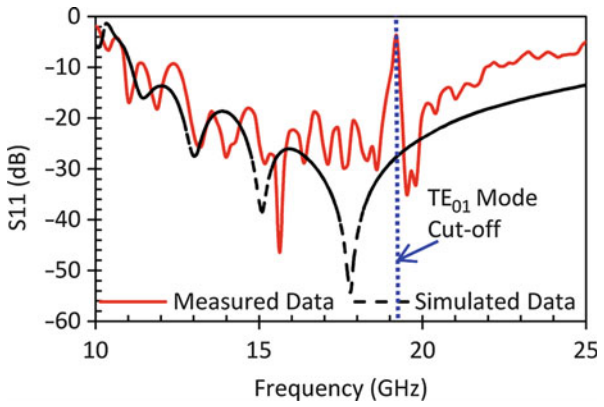


Fig. 19 Measured reflection coefficients of the 3D printed corrugated conical horn antenna compared with simulation (Chieh et al. 2014). The discrepancy around 18 GHz is due to the excitation of the TE_{01} mode from the rectangular-to-circular waveguide transition

the polymer structure as a seed layer. Then, electroplating process is applied to metalize the backside of the reflect array. Photos of the printed reflect-array antennas are shown in Fig. 21. A W-band pyramidal horn placed 22.5 mm away from the reflect-array aperture with a tilted angle of 25° is used as the feed. Figure 22 shows the measured and simulated radiation patterns of the first design in Fig. 21. It can be seen that the measured results agree well with simulation. The measured gain of these three designs at 100 GHz is 22.5 dB, 22.9 dB, and 18.9 dB, respectively. This work shows that the polymer jetting printed dielectric reflect array is a promising approach of realizing a high-gain antenna at a submillimeter frequency.

DOI 10.1007/978-981-4560-75-7_109-1

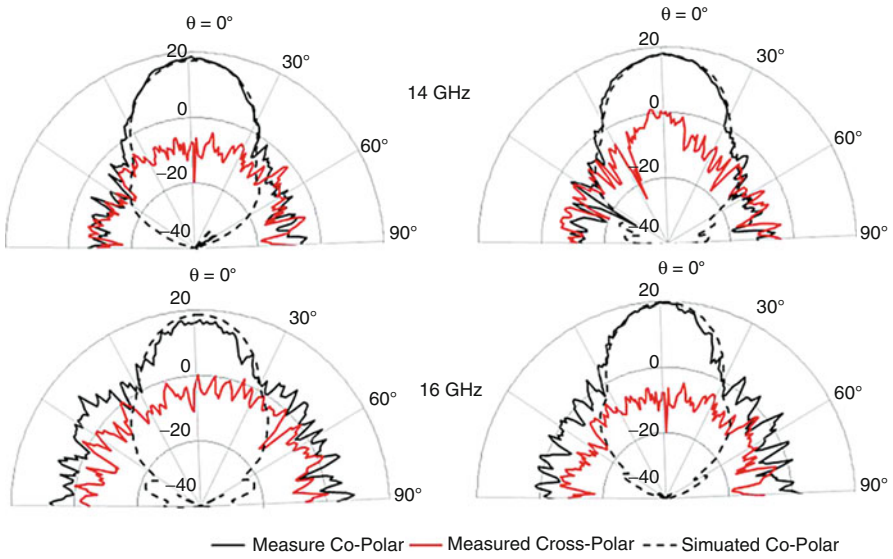


Fig. 20 Measured and simulated co- and cross-polarization radiation patterns of the 3D printed corrugated conical horn antenna at 14 and 16 GHz (Chieh et al. 2014)

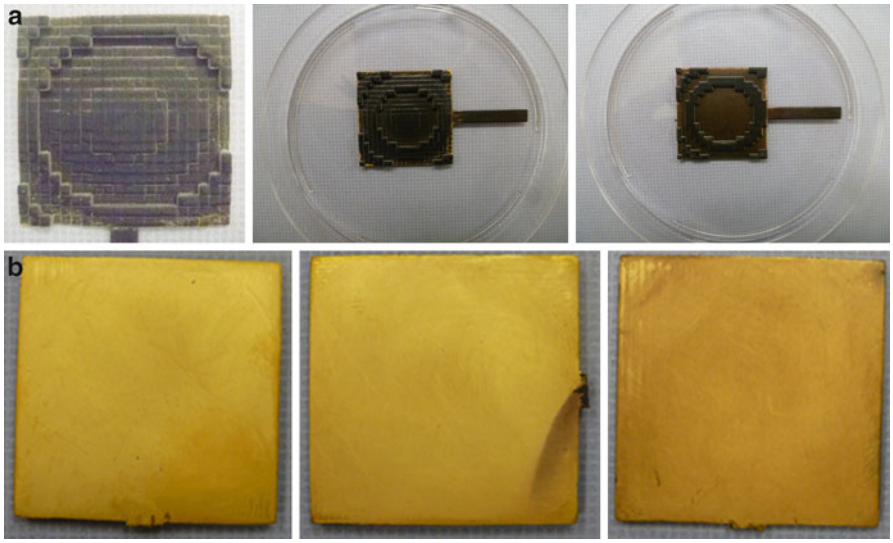


Fig. 21 (**a**) Top view and (**b**) bottom view of the dielectric reflect-array antennas printed using polymer jetting technique and electroplating (Nayeri et al. 2014)

An all-dielectric antenna operating in the mmW/THz frequency range has also been realized by the polymer jetting 3D printing technique. In Wu et al. (2012), a 3D printed electromagnetic crystal (EMXT)-based all-dielectric THz horn antenna is reported. The EMXT structure exhibits electromagnetic band gaps in certain frequency bands due to its periodicity. Within these band gaps, EM wave propagation is prohibited and therefore a hollow channel in the crystal structure will be able to confine and guide wave propagation along the channel. Photo images of the 3D printed EMXT horn antenna are shown in Fig. 23. Because of the band gap structure, in certain frequency bands, the wave propagation will be confined in the horn-shaped hollow core, and the antenna would work similarly like a regular horn antenna. These frequency bands are termed as the passbands of this EMXT horn antenna.

DOI 10.1007/978-981-4560-75-7_109-1

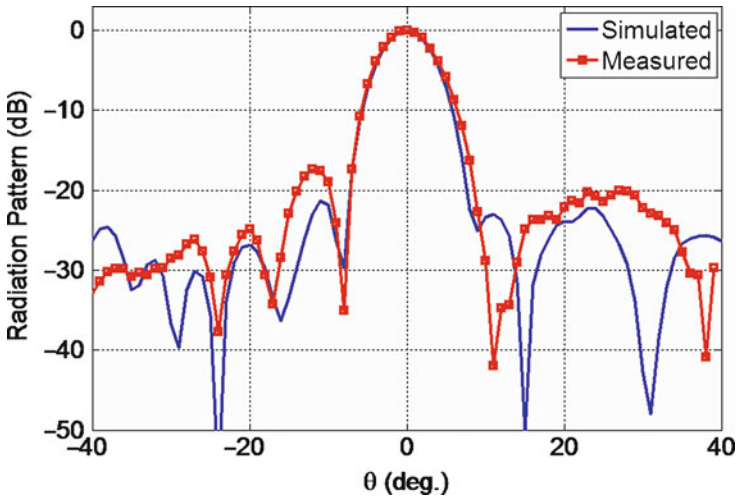


Fig. 22 Simulated and measured radiation patterns of the first reflect array in Fig. 21 at 100 GHz (Nayeri et al. 2014)

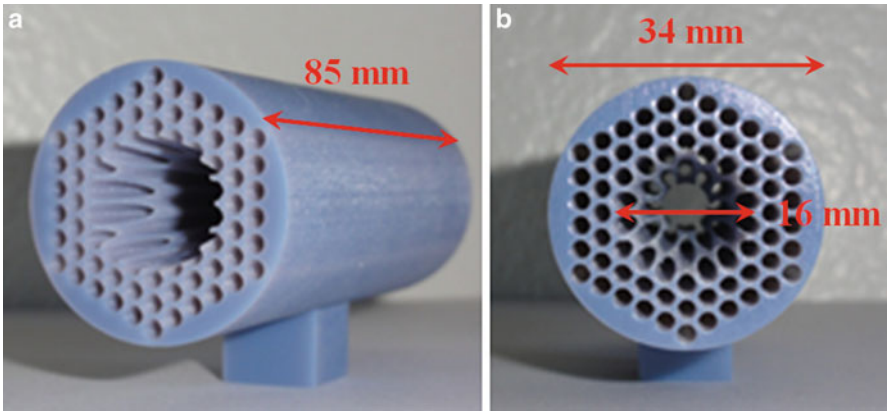


Fig. 23 Photos of the polymer jetting printed THz EMXT horn antenna. (**a**) Side view. (**b**) Front view. The total size of the antenna is 85 mm for length and 34 mm for diameter and the flare angle of the horn is 12.4° (Wu et al. 2012)

This all-dielectric EMXT horn antenna is experimentally characterized by a THz time domain spectroscopy system. Figure 24 plots the measured and simulated radiation patterns of the antenna for the first (105 GHz) and third (146 GHz) passbands. It can be seen that the measured results agree well with simulation. This printed all-dielectric horn antenna demonstrates comparable or better radiation performance compared to a copper horn antenna with the same geometry (Wu et al. 2012). Moreover, as an important free-space coupling component, the demonstration of the all-dielectric antenna may lead to potential integrated THz systems that can be manufactured by the polymer jetting technique.

Another example using polymer jetting technology to print an all-dielectric microwave lens antenna is reported in Liang et al. (2014) in which a broadband 3D Luneburg lens antenna (Luneburg 1964) operating from X to Ku-band is printed by employing the polymer jetting technique. The Luneburg lens is an attractive gradient index device used as an antenna for wide-angle radiation scanning because of its broadband behavior, high gain, and the ability to form multiple beams. Every point on the surface of an ideal Luneburg lens is the focal point of a plane wave incident from the opposite side. The refraction index n distribution of a spherical Luneburg lens is given by equation $n(r)^2 = \varepsilon_r(r) = 2 - (r/R)^2$, where ε_r is the relative permittivity, R is the radius of the lens, and r is the distance from the point to the center of the

DOI 10.1007/978-981-4560-75-7_109-1

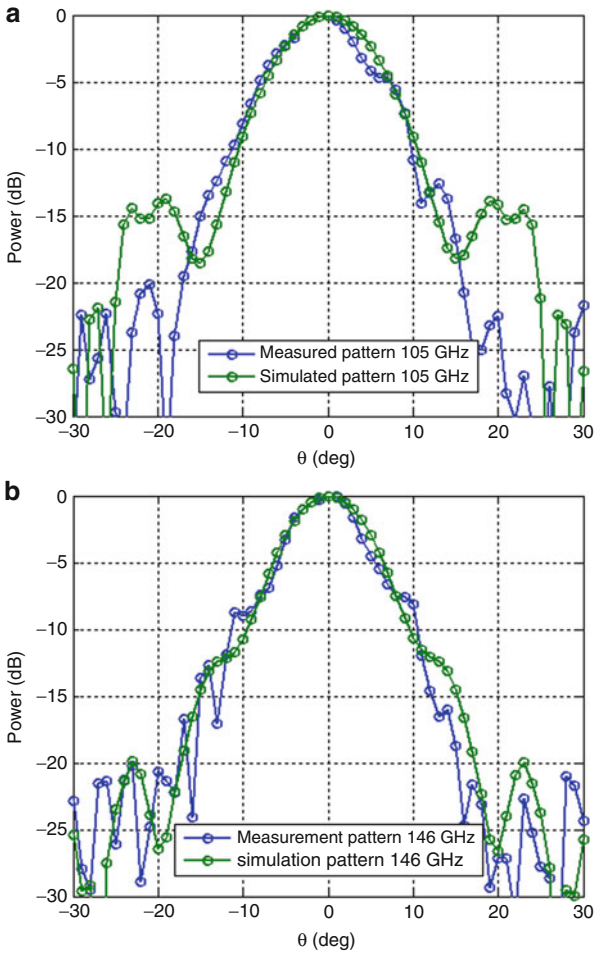


Fig. 24 Measured and simulated radiation patterns of the polymer jetting printed EMXT horn antenna at (**a**) 105 GHz and (**b**) 146 GHz (Wu et al. 2012)

sphere. In Liang et al. (2014), the required gradient index distribution of Luneburg lens is realized by controlling the filling ratio of a polymer-/air-based unit cell and therefore changes the effective permittivity at different locations. For example, at the center part of lens, a larger required effective permittivity will need a larger polymer filling ratio. At the edge of lens, a smaller required effective permittivity will need a smaller polymer filling ratio. One important thing to note is that the unit cell size needs to be small enough to guarantee the effective medium assumption of the design. Figure 25 shows three examples of 3D printed Luneburg lens: the first one has a diameter of 12 cm, the second one is a larger version of the lens with a diameter of 24 cm, and the third one is a Luneburg lens working at Ka- and Q-band with much smaller unit cell size. In the measurement, the feed of the Luneburg lens antenna is either a rectangular waveguide or a coaxial probe mounted on the surface of the lens. Figure 26 plots the simulated and measured radiation patterns of the 12-cm-diameter lens antenna at 10 GHz using a waveguide feed. All the radiation patterns from 8.2 GHz to 20 GHz show a highly directional beam and good agreement between experiment and simulation is obtained. The gain of this lens antenna using a waveguide feed is measured to be from 17.3 dB (at 8.2 GHz) to 24 dB (at 19.8 GHz). The side lobe is measured to be about 25 dB lower than the main beam in the H-plane and about 20 dB lower than the main beam in the E-plane. The difference is due to the asymmetry of the waveguide feed. Compared to traditional Luneburg lens fabrication techniques, the polymer jetting technique has a much lower cost, and the fabrication process is more precise, convenient, and faster.

DOI 10.1007/978-981-4560-75-7_109-1

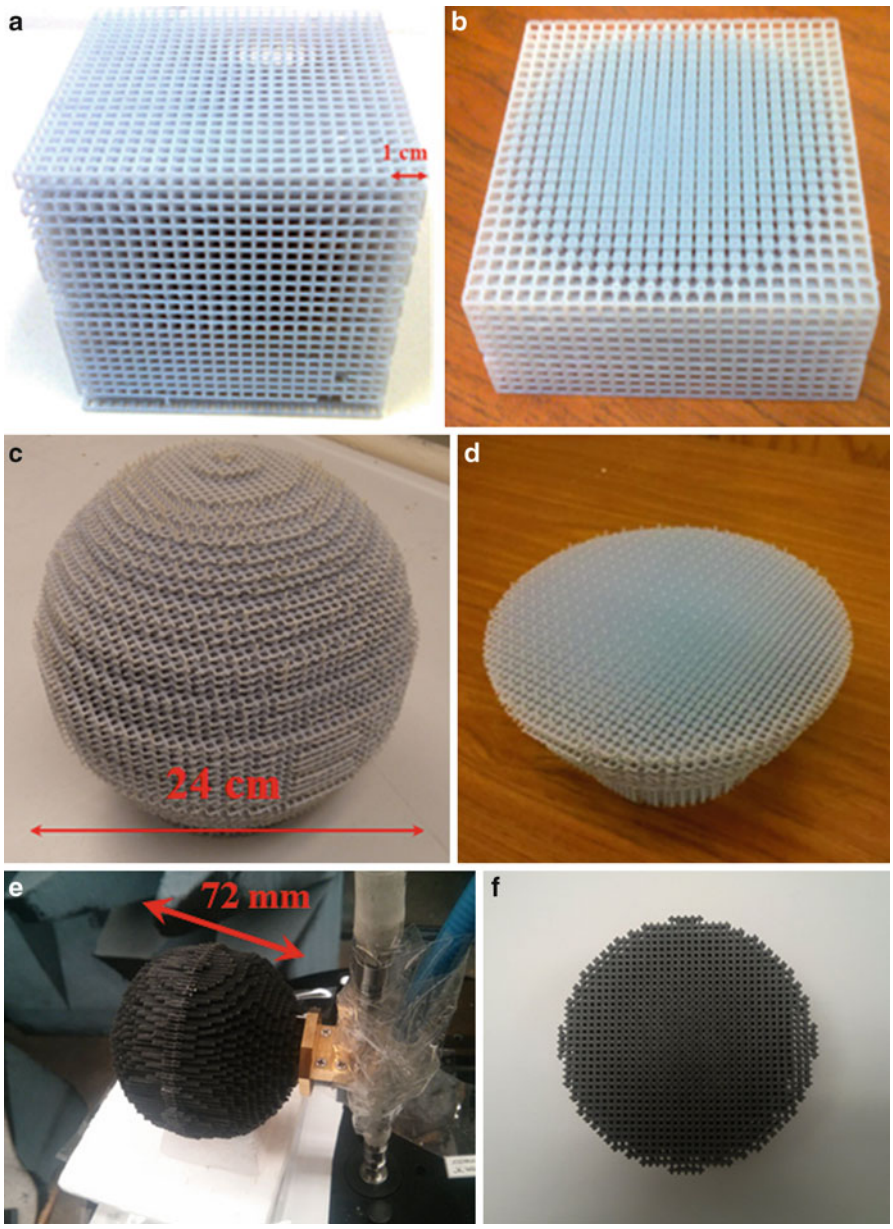


Fig. 25 Photos of the polymer jetting printed Luneburg lenses. (**a**) A lens with a 12 cm diameter. (**b**) The cross-section cut through the center of the 12-cm-diameter lens. (**c**) A larger lens with a 24 cm diameter. (**d**) Cross-section cut of the 24-cm-diameter lens. (**e**) A Luneburg lens working at Ka- and Q-band. (**f**) Cross-section cut of the Ka- and Q-band Luneburg lens (Liang et al. 2014; Gbele et al. 2014)

The above mentioned antenna examples show that the polymer jetting technique is a very good candidate in realizing 3D printed antennas, even up to THz frequency. Some of the complicated structures of these antennas would be very difficult or even impossible to fabricate using conventional manufacturing process. Moreover, compared to the conventional method, the polymer jetting technique will have a lower cost, and the fabrication process is more precise, convenient, and faster.

Antenna Printed Using Direct Printing of Conductive Ink

Direct printing refers to any technology that has the ability to deposit a variety of materials onto a surface in a designed pattern (Lopes et al. 2012). The surface to be printed on can be either flat or curved. Silver

DOI 10.1007/978-981-4560-75-7_109-1

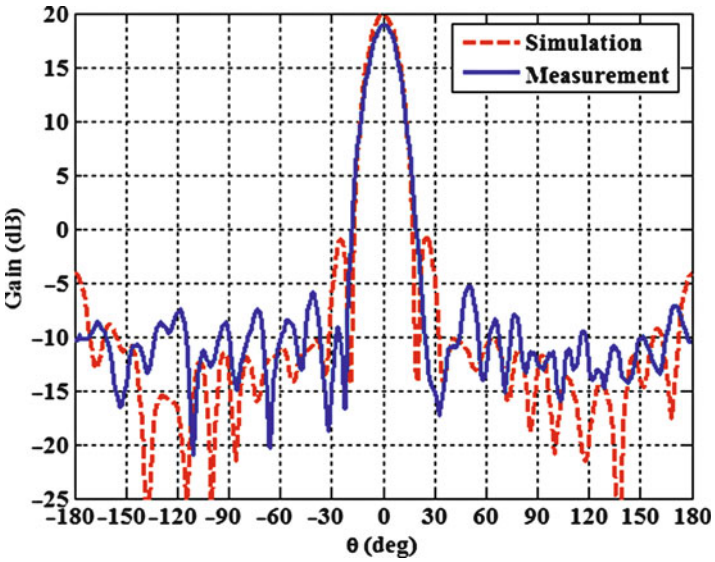


Fig. 26 Simulated and measured H-plane radiation patterns of the polymer jetting printed 12-cm-diameter Luneburg lens antenna at 10 GHz (Liang et al. 2014)

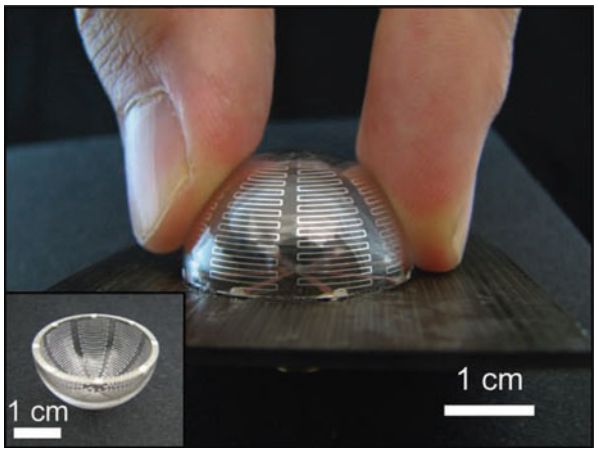


Fig. 27 Photos of the meander line antenna printed using conductive ink before (*inset*) and after connecting to feed lines (Adams et al. 2011)

conductive ink is a commonly used material in direct printing to realize the conductive surface of microwave antennas. After a post-printing annealing process, the conductivity of the printed ink may be on the same order of the conductivity of pure metal.

An example using direct printing technology to implement an electrically small antenna is reported in Adams et al. (2011) in which conductive meander lines are printed through a tapered cylindrical nozzle onto a hemispherical glass substrate to realize the antenna (Ahn et al. 2009). A post-annealing process at 550 °C is applied for 3 h. The printed meander line shows a DC conductivity of 2×10^7 S/m, which is within a factor of 3 compared to the conductivity of pure silver. Photos of the direct printed antenna before and after connecting to the feed lines are shown in Fig. 27. Figure 28 plots the measured VSWR as a function of frequency for one of the printed antennas together with the simulation results. The measured center frequency is at 1.7 GHz with a bandwidth of 12.6 %, indicating a good performance approaching the Chu limit of this electrically small antenna (Chu 1948).

DOI 10.1007/978-981-4560-75-7_109-1

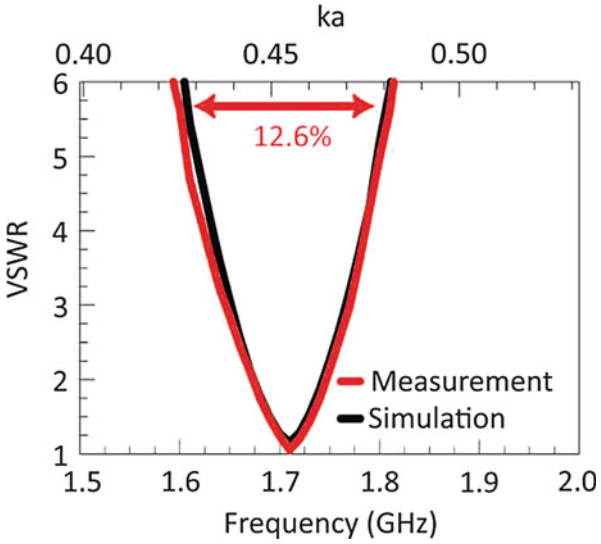


Fig. 28 Measured and simulated VSWR versus frequency of the conductive ink printed meander line antenna (Adams et al. 2011)

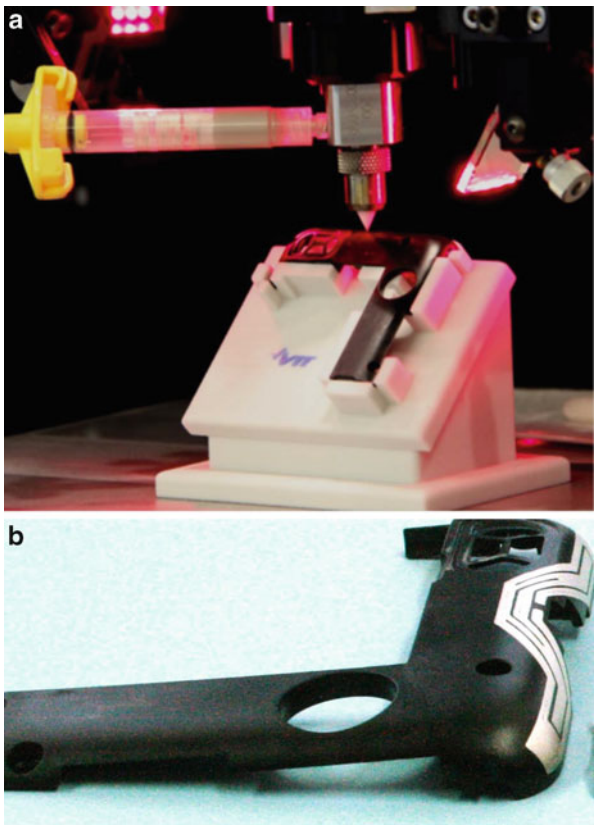


Fig. 29 (**a**) A photo shows a handset antenna during the direct printing process. (**b**) The handset antenna after printing (Salonen et al. 2013)

Salonen et al. (2013) also report a printed antenna using the direct printing technique on a 3D surface. The antenna is realized using the nScrypt@ nozzle dispensing tool (http://www.nscrypt.com/). A photo of the handset antenna during the printing process is illustrated in Fig. 29a, b which shows the antenna after

DOI 10.1007/978-981-4560-75-7_109-1

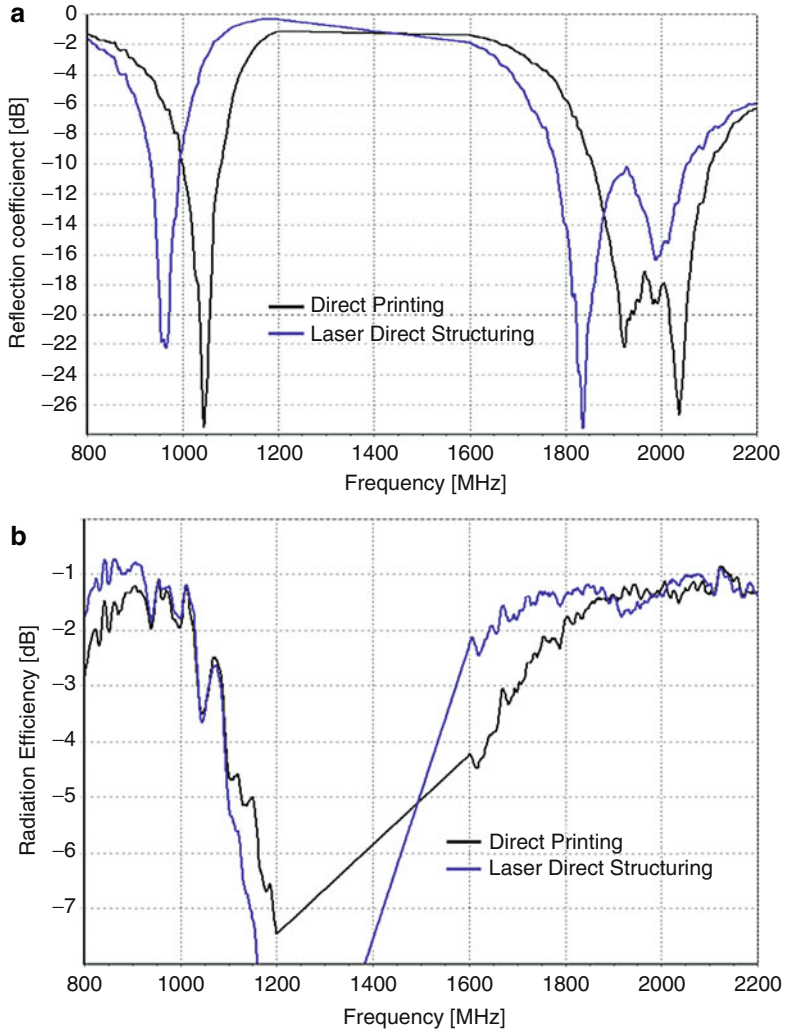


Fig. 30 Measured (**a**) reflection coefficient and (**b**) radiation efficiency of a handset antenna printed using direct printing technique compared to the same antenna made by the laser direct structuring (*LDS*) technique (Salonen et al. 2013)

printing. In the printing process, the conductive silver paste is extruded through a nozzle and deposited onto the substrate. The nozzle position is controlled by a 3-axis motion system and the material flow is adjusted by a precision pump. After printing, the silver paste is heat treated at 120 °C for 10 min to increase the conductivity. 120 °C is selected because a higher temperature causes some deformation of the plastic substrate. Figure 30 plots the measured reflection coefficient and the radiation efficiency of the direct printed antenna compared to those of the standard laser direct structuring (LDS) technique-fabricated antenna. The results show that these two antennas have similar performance in terms of reflection coefficient. The radiation efficiency of the direct printed antenna is 0.5 dB lower at the lower band and nearly equal at the higher band, compared to the LDS antenna. The slightly lower efficiency is believed to be caused by the skin effect which is more dominant at the lower band due to the smaller conductivity of the ink.

Antenna Printed Using Fused Deposition Modeling (FDM)

The FDM technique has the ability to print a large number of thermoplastic materials. In Ahmadloo and Mousavi (2013), a 3D meander line dipole antenna is printed on a V-shaped substrate using the FDM technique. The conductive part of the antenna is realized using a printed conductive ink. The curing

DOI 10.1007/978-981-4560-75-7_109-1

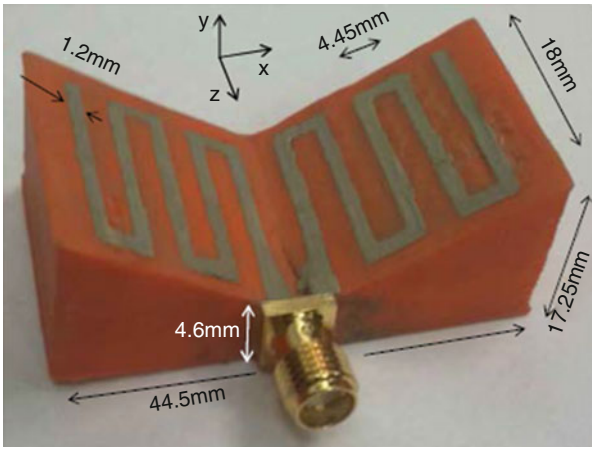


Fig. 31 FDM printed meander line dipole antenna on V-shaped substrate (Ahmadloo and Mousavi 2013)

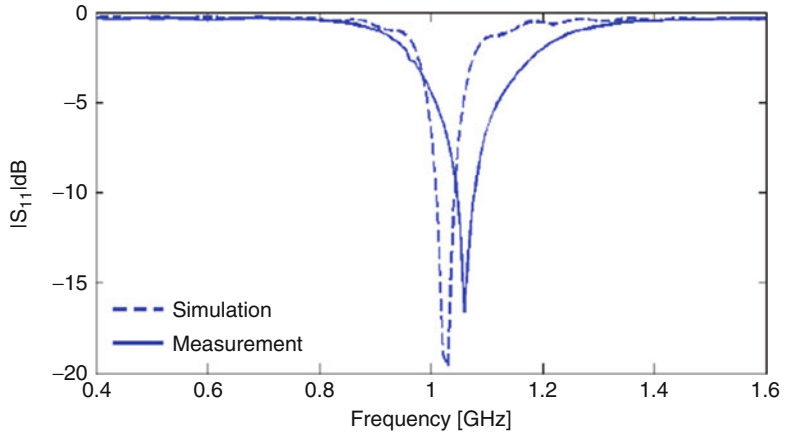


Fig. 32 Measured and simulated reflection coefficients of the FDM printed meander line dipole antenna (Ahmadloo and Mousavi 2013)

process of the printed conductive ink is at 85 °C for 15 min to reduce the resistivity. After curing, the conductivity of the silver ink is measured to be 1.5×10^6 S/m. A higher curing temperature can achieve a higher conductivity. However, to avoid the substrate deformation during the curing time, 85 °C is selected. A photo of the printed antenna is shown in Fig. 31. The meander line dipole is fed by a SMA connector. Figures 32 and 33 compare the measured and simulated reflection coefficients and radiation patterns of this antenna. It can be seen that the measured results agree well with the simulation results.

As a potential solution for integrating dielectric and conductor printing while maintaining high quality for each material, a microwave patch antenna is realized by FDM and an ultrasonic wire mesh embedding process (Liang et al. 2014). The substrate of the patch antenna is created using an FDM 3D printer and the conductive part of the antenna is realized using an ultrasonic machine which has the ability to embed copper wires on a 3D surface. Compared to the conductive ink approach, the ultrasonic wire embedding technique is performed at room temperature and therefore will not influence the thermoplastic substrate. In addition, since pure metal wire is used, the conductivity of the material is much larger than that of the conductive ink. The schematic design of the patch antenna and a photo of the 3D printed patch antenna are shown in Fig. 34. The feed of this antenna is a SMA probe from the back of the patch. The measured reflection coefficient and radiation pattern compared to simulation results using wire mesh and ideal

DOI 10.1007/978-981-4560-75-7_109-1

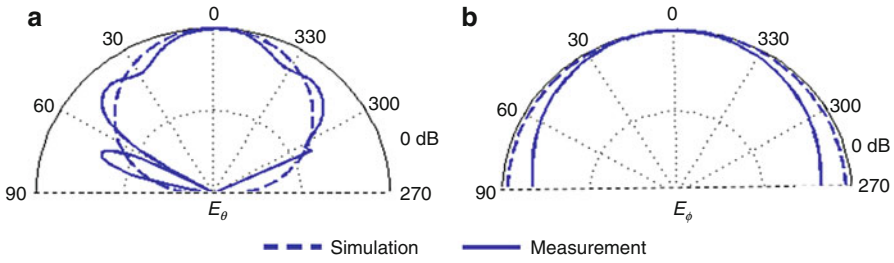


Fig. 33 Measured and simulated radiation patterns of the FDM printed meander line dipole antenna (Ahmadloo and Mousavi 2013)

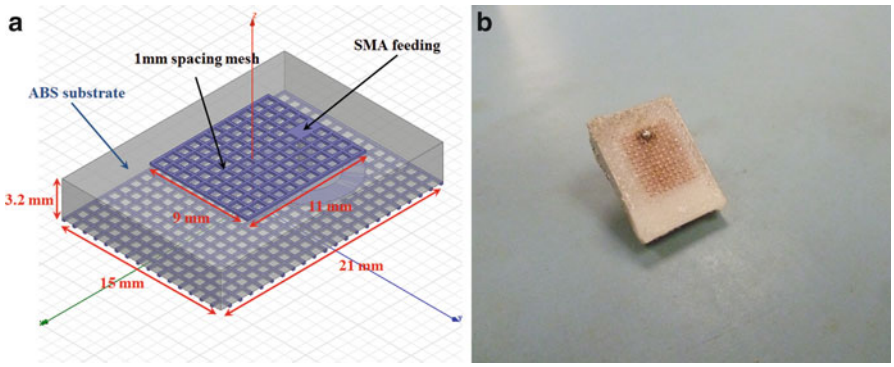


Fig. 34 Schematic of a microwave patch antenna made of FDM printed substrate and ultrasonically embedded wire mesh (*left*) and a photo of the patch antenna printed using FDM and wire mesh embedding technique (Liang et al. 2014)

conductor are illustrated in Fig. 35. It can be seen that the wire mesh structure works as well as a regular conductor sheet at microwave frequency and the measurement agrees well with simulation.

This example successfully demonstrated 3D printing of a microwave patch antenna by combining the FDM method with ultrasonic metal wire mesh embedding. No metal sintering or any other high-temperature conductor printing process is needed, and the printed metal wire mesh works as well as a regular metal sheet at a microwave frequency. The demonstrated 3D printing process of both dielectric and conductor can be applied to the other more sophisticated EM structures such as vertically integrated phased arrays. The disadvantage of this wire mesh embedding technique is the lower resolution compared to other AM techniques due to the minimum diameter of the metal wire (50 μm). This will limit the maximum operating frequency of components printed using the wire mesh embedding technique.

Challenges and Potential Solutions

It has been argued that 3D printing could be the future of manufacturing due to its ability to print structures with more flexible design than conventional methods. Recently, rapid progress has been made in the 3D printing area. However, a number of challenges still need to be resolved before advanced functional antennas can be printed in a 3D fashion robustly.

Surface Roughness

Surface roughness is an important parameter for manufacturing of antennas, especially for higher-frequency ranges such as mmW and THz. A rough surface can seriously degrade the EM performance

DOI 10.1007/978-981-4560-75-7_109-1

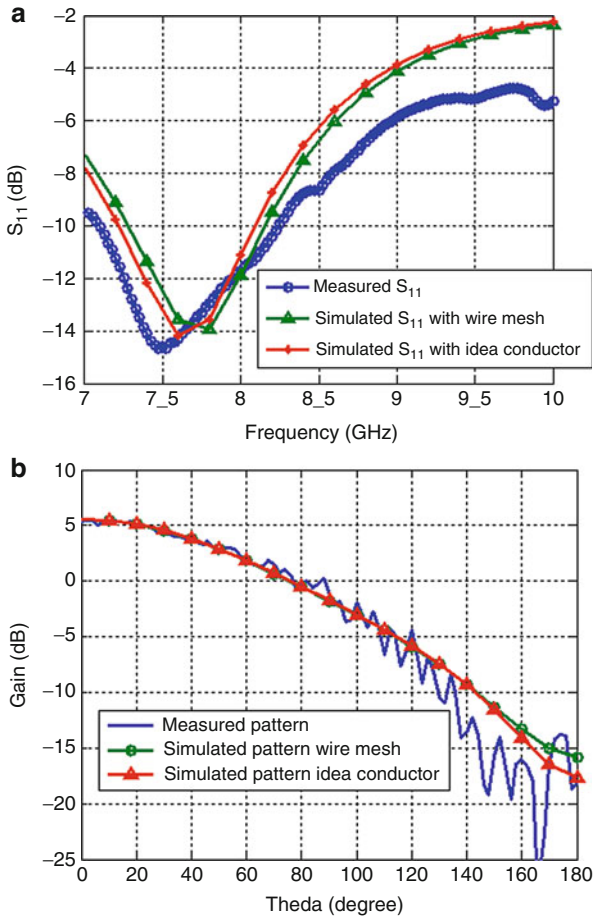


Fig. 35 Measured and simulated (**a**) reflection coefficients and (**b**) radiation patterns of the patch antenna printed using FDM and wire mesh embedding technique (Liang et al. 2014)

in terms of increased conductor loss. Currently, a drawback of many 3D printing techniques is the rough surface finish. For example, the EBM technique has a surface roughness on the order of tens of microns (Garcia et al. 2013). The FDM method has a surface roughness from 10 to 50 μm (Ahn et al. 2008). A post-polishing step often needs to take place after the printing process if fine surface roughness is required. However, for some of the complicated parts, it is inconvenient or even impossible to polish the inner part of the structure. Therefore, improving the surface roughness is one of the challenges in AM for high-frequency antenna applications.

Resolution

The printing resolution is also a challenge for AM technology, especially for those components working at high frequencies (e.g., mmW or THz). A higher resolution means a more accurate structure which is necessary for higher-frequency applications. Currently, most of the 3D printing techniques have a resolution no better than 20 μm with the exception of a two-photon polymerization stereolithography technique which has a sub-micrometer resolution. However, there is always a trade-off between printing resolution, printable size, and printing speed. For example, it is difficult to print large-size components (e.g., >1 cm) using the two-photon polymerization technique because of the printing speed. The long-term solution lies in the integration of multi-scale 3D printing techniques so that high-resolution or high-volume/high-speed techniques can be seamlessly integrated and applied when necessary.

DOI 10.1007/978-981-4560-75-7_109-1

Limited Electromagnetic (EM) Property Range

Another issue for 3D AM of antennas is the lack of available printable materials with desired EM properties (i.e., permittivity ε and permeability μ). At the present time, most of the commercially available materials used in 3D printing technologies are designed or selected with only the mechanical property in mind, and none of the material is designed for the EM applications. The lack of 3D printable materials with desired EM properties restricts their applicability to microwave components. For example, most of the existing printable polymers have a permittivity from 2 to 3. However, a higher dielectric constant is often desired for a microwave circuit and antenna for miniaturization purpose. Therefore, research and development in new printable materials with flexible EM properties would be necessary. A potential solution is using a polymer matrix composite method which mixes a 3D printable polymer with nanoparticles (Liang et al. 2014). By employing this method, novel materials with a wider range of EM properties (i.e., ε and μ) may be realized.

Performance of Printed Conductor

As discussed previously, the incorporation of a high-quality conductor with 3D printed dielectrics is also a challenge. Printing conductive inks on a dielectric substrate is one of the most popular methods to realize 3D spatial control of conducting material. However, to achieve high conductivity, high-temperature curing processes, which are usually not compatible to 3D printing techniques such as thermoplastic extrusion and photopolymerization, are necessary. With the limit of low curing temperature, the conductive inks tend to have relatively poor conductivity which leads to higher loss and could be detrimental to antenna performance, especially at a high frequency. Using the ultrasonic wire embedding technique in Liang et al. (2014) to realize the conductive part of the microwave antenna is an attractive alternative since this method guarantees a high conductivity (σ) at room temperature. Nevertheless, the resolution of the wire mesh embedding technique is lower than other AM techniques. Therefore, developing methods to realize the printing of a high-conductivity material at room temperature with arbitrary geometry will be critical to advance the state-of-the-art 3D printing of electronics including antennas.

Multiple Scales and Multiple Materials

Another desired capability in AM technology is multiple-scale printing. Currently, for most of the commercially available 3D printers, their printer heads only have one resolution which means it would take a long time to print a large object if a small part of the object needs a fine resolution. To address this problem, a multi-scale printer which has the ability to print some part of the object with fine resolution and print other parts of the object with coarse resolution is necessary.

A 3D AM technology capable of printing multiple materials is also highly desirable. Being able to efficiently realize arbitrary 3D spatial distribution of EM (ε, μ, and σ) is the holy grail of 3D printing technology for antenna applications. It will enable many new design possibilities which may lead to revolutionary antenna configurations having unprecedented advantages. An antenna structure with artificially controlled EM properties (ε, μ, and σ) in arbitrary 3D spatial distribution, novel 3D gradient index (GRIN) metamaterial-based lens antenna, and 3D printed vertically integrated phased array system are just some of the examples. If printable materials with time-dependent and/or nonlinear EM properties (i.e., 3D printed semiconductors (Ahn et al. 2006)) become available, AM technology may be used to achieve advanced fully functional systems.

DOI 10.1007/978-981-4560-75-7_109-1

Summary

In this chapter, the working principles and processes for various 3D AM techniques that are relevant to antenna design and implementation are described. Recent progresses and challenges for 3D printing of antennas are reviewed and discussed. The reported experiments demonstrate a number of successful antenna examples printed by various AM techniques. There are still substantial challenges that need to be overcome before complete and fully functional antennas and microwave systems can be truly realized via AM. However, further investigation and development of 3D printing technology in the areas of mechanical engineering, material science and engineering, and electrical engineering will no doubt lead to a new paradigm of 3D printed antennas and other microwave components and systems.

Cross-References

▶ Advanced Antenna Fabrication Process (MEMS/LTCC/LCP/printing)
▶ Lens Antennas
▶ Reflectarray Antennas
▶ RF Material Characterization

References

Adams JJ, Duoss EB, Malkowski TF, Motala MJ, Ahn BY, Nuzzo RG, Bernhard JT, Lewis JA (2011) Conformal printing of electrically small antennas on three-dimensional surfaces. Adv Mater 23:1335–1340

Agarwala M, Bourell D, Beaman J, Marcus H, Barlow J (1995) Direct selective laser sintering of metals. Rapid Prototyp J 1(1):26–36

Ahmadloo M, Mousavi P (2013) A novel integrated dielectric-and-conductive ink 3D printing technique for fabrication of microwave devices. In: IEEE MTT-S international microwave symposium digest (IMS). Seattle, WA

Ahn J, Kim H, Lee K, Jeon S, Kang SJ, Sun Y, Nuzzo RG, Rogers JA (2006) Heterogeneous three-dimensional electronics by use of printed semiconductor nanomaterials. Science 314:1754–1757

Ahn D, Kwon S, Lee S (2008) Expression for surface roughness distribution of FDM processed parts. In: International conference on smart manufacturing application, KINTEX, Gyeonggi-do

Ahn BY, Duoss EB, Motala MJ, Guo X, Park S, Xiong Y, Yoon J, Nuzzo RG, Rogers JA, Lewis JA (2009) Omnidirectional printing of flexible, stretchable, and spanning silver microelectrodes. Science 323:1590–1593

Bak D (2003) Rapid prototyping or rapid production? 3D printing processes move industry towards the latter. Assem Autom 23(4):340–345

Chieh JS, Dick B, Loui S, Rockway JD (2014) Development of a Ku-Band corrugated conical horn using 3-D print technology. IEEE Antennas Wirel Propag Lett 13:201–204

Chu LJ (1948) Physical limitations of omni-directional antennas. J Appl Phys 19:1163

Cline HE, Anthony TR (1977) Heat treating and melting material with a scanning laser or electron beam. J Appl Phys 48:3895–3900

Garcia CR, Rumpf RC, Tsang HH, Barton JH (2013) Effects of extreme surface roughness on 3D printed horn antenna. Electron Lett 49(12):734–736

Gbele K, Liang M, Ng W, Gehm ME, Xin H (2014) Ka and Q band Luneburg lens antenna fabricated by polymer jetting rapid prototyping. Infrared, Millimeter, and Terahertz Waves (IRMMW-THz), Wollongong

Gebhardt A (2012) Understanding additive manufacturing. Hanser publications, Cincinnati

Griffin EA, McMillin S (1995) Selective laser sintering and fused deposition modeling processes for functional ceramic parts. In: Solid freeform fabrication symposium. University of Texas in Austin, Texas, vol 6, pp 25–30

http://www.nanoscribe.de/en/. Last date of accessed 24 Nov 2014

http://www.nscrypt.com/. Last date of accessed 24 Nov 2014

Huang Y, Gong X, Hajela S, Chappell WJ (2005) Layer-by-layer stereolithography of three-Dimensional antennas. In: IEEE Antennas and propagation society international symposium (APSURSI). Washington, DC, pp 276–279

Kambour RP (1973) A review of crazing and fracture in thermoplastics. J Polym Sci Macromol Rev 7(1):1–154

Kim OS (2013) 3D Printing electrically small spherical antennas. In: Antennas and propagation society international symposium (APSURSI). Orlando, Florida, pp 776–777

Kornbluh RD, Pelrine R, Pei Q, Oh S, Joseph J (2000) Ultrahigh strain response of field-actuated elastomeric polymers. In: SPIE's 7th annual international symposium on smart structures and materials. Newport, pp 51–64

Kruth JP, Wang X, Laoui T, Froyen L (2003) Lasers and materials in selective laser sintering. Assem Autom 23(4):357–371

Kruth JP, Froyen L, Van Vaerenbergh J, Mercelis P, Rombouts M, Lauwers B (2004) Selective laser melting of iron-based powder. J Mater Process Technol 149:616–622

Kruth JP, Mercelis P, van Vaerenbergh J, Froyen L, Rombouts M (2005) Binding mechanisms in selective laser sintering and selective laser melting. Rapid Prototyp J 11(1):26–36

Liang M, Shemelya C, MacDonald E, Wicker R, Xin H (2014a) Fabrication of microwave patch antenna using additive manufacturing technique. In: IEEE antennas and propagation society international symposium (APSURSI). Memphis, Tennessee

Liang M, Ng W, Chang K, Gbele K, Gehm ME, Xin H (2014b) A 3-D Luneburg lens antenna fabricated by polymer jetting rapid prototyping. IEEE Trans Antennas Propag 62(4):1799–1807

Liang M, Yu X, Shemelya C, Roberson D, MacDonald E, Wicker R, Xin H (2014c) Electromagnetic materials of artificially controlled properties for 3D printing applications. In: IEEE antennas and propagation society international symposium (APSURSI). Memphis, Tennessee

Lipkea DW, Zhanga Y, Liua Y, Churcha BC, Sandhage KH (2010) Near net-shape/net-dimension ZrC/W-based composites with complex geometries via rapid prototyping and displacive compensation of porosity. J Eur Ceram Soc 30(11):2265–2277

Lopes AJ, MacDonald E, Wicker RB (2012) Integrating stereolithography and direct print technologies for 3D structural electronics fabrication. Rapid Prototyp J 18(2):129–143

Lopez AG, Lopez EEC, Chandra R, Johansson AJ (2013) Optimization and fabrication by 3d printing of a volcano smoke antenna for UWB applications. In: 7th European conference on antennas and propagation (EuCAP). Gothenburg, Sweden, pp 1471–1473

Luneburg RK (1964) Mathematical theory of optics. University of California Press, Los Angeles

DOI 10.1007/978-981-4560-75-7_109-1

Nayeri P, Liang M, Sabory-Garc´ıa RA, Tuo M, Yang F, Gehm M, Xin H, Elsherbeni AZ (2014) 3D Printed dielectric reflectarrays: low-cost high-gain antennas at sub-millimeter waves. IEEE Trans Antennas Propag 62(4):2000–2008

Salonen P, Kupiainen V, Tuohimaa M (2013) Direct printing of a handset antenna on a 3D surface. In: IEEE antennas and propagation society international symposium (APSURSI). Orlando, Florida, pp 504–505

Wong KV, Hernandez A (2012) A review of additive manufacturing. ISRN Mech Eng 2012, Article ID 208760

Wu Z, Liang M, Ng W, Gehm M, Xin H (2012) Terahertz Horn antenna based on hollow-core electromagnetic crystal (EMXT) structure. IEEE Trans Antennas Propag 60(12):5557–5563

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_112-1

Antennas and EM Issues in Medical Diagnosis and Treatment Systems

Kazuyuki Saito* and Koichi Ito
Center for Frontier Medical Engineering, Chiba University, Chiba, Japan

Abstract

In recent years, various types of medical applications of electromagnetic techniques have been investigated. We have also been studying several medical applications of microwave techniques, which can be classified into diagnosis techniques and therapeutic systems. In this chapter, radio frequency techniques in magnetic resonance imaging (MRI) are discussed from the point of view of biomedical electromagnetic compatibility as an example of medical diagnostic devices. Moreover, therapeutic systems that employ the microwave thermal effect of biological tissue are also introduced. These systems are the thermal treatment of cancer and surgical devices using high-power microwave energy.

Keywords

Medical application; Diagnosis; MRI; Thermal therapy; Hyperthermia; Surgical device; SAR; Temperature elevation

Introduction

In recent years, the quality of life of patients has become increasingly important. Therefore, it is an established fact that medical applications of microwave techniques are important (Sterzer 2002) because these applications are effective in reducing the mental and physical burden of patients (Ito 1996). These applications can be classified as follows:

1. Information gathering and diagnostic techniques inside the human body that use microwave, such as computerized tomography (CT) and magnetic resonance imaging (MRI).
2. Thermal treatments, which use microwave energy as a heating source.
3. Transmission of patients' medical information.

Transmission of patients' medical information [point (3) above] is considered to be an extension of communication technologies (Shimizu et al. 1995) and has been described by many papers and books. Therefore, this chapter only focuses on specific aspects of the medical applications of electromagnetic (EM) waves [points (1) and (2) above].

*Email: kazuyuki_saito@faculty.chiba-u.jp

DOI 10.1007/978-981-4560-75-7_112-1

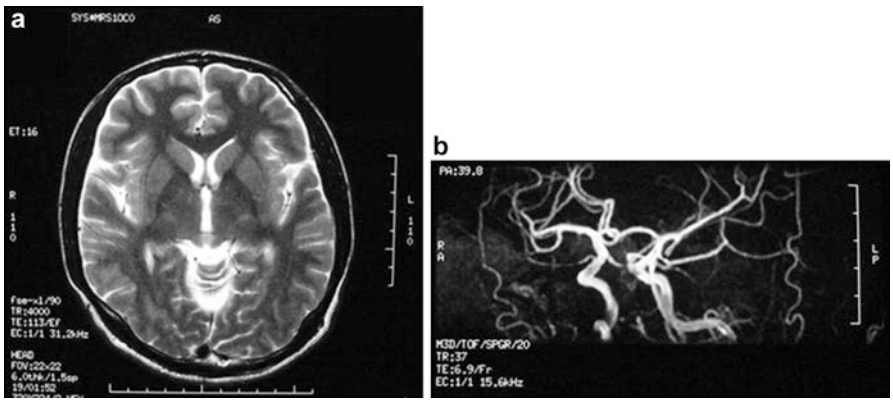


Fig. 1 Magnetic resonance images: (**a**) normal tomographic image of head; and (**b**) magnetic resonance angiography

Microwave Techniques in Magnetic Resonance Imaging (MRI)

Diagnostic Devices Using an Electromagnetic (EM) Field

Diagnoses inside the human body using X-ray, ultrasound, and magnetic fields have been well-studied and widely employed. In recent years, researchers have also tried using microwave techniques, although compared to X-ray CT, microwave CT is less popular. However, microwave CT for detection of breast cancer has been actively studied in recent years (Hagness et al. 1999; Bond et al. 2003). Moreover, MRI might be one of the diagnostic devices that use an EM field with the most success at the moment. There are many papers and books describing the principles and improvement of MRI devices; therefore, in this chapter, after briefly describing the structure of the MRI, the biomedical EM compatibility (EMC) of the device is introduced.

Structure of the MRI

During an MRI , the subject undergoing testing is placed in a high-strength static magnetic field and is radiated with EM wave pulses called radio frequency (RF) pulses. As a result, the spin of the atomic nuclei inside the test subject generate nuclear magnetic resonance (NMR) signals. By receiving the NMR signals and adapting the image reconstruction processing, various anatomical information from inside the object can be obtained. Figure 1 shows a normal tomographic image of the head and an image of blood vessels in the brain (magnetic resonance angiography). Various types of images inside the subject under test can be obtained by changing the parameters of the RF pulses. Figure 2 shows the basic configuration of the MRI device. The main magnet generates a high-strength static magnetic field which brings the spin of the atomic nuclei inside the test subject into line. The gradient coil creates a gradient magnetic field to determine the position of the image. In addition, the RF coil radiates the RF pulses to the test subject and receives the NMR signals from the test subject. Therefore, the RF coil is not a "coil" but rather an "antenna." For a detailed description of the structure of the MRI system, see, for example, Jin (1998). Figure 3 shows the various RF coils for imaging.

Biomedical EM Compatibility During MRI

In this section we describe the biomedical EMC of the human body during MRI, which is caused by the RF pulse radiation from the RF coil. Several kinds of RF coils have now been developed and can be selected according to the imaging portion of the body. During the MRI, the RF coil radiates pulsing EM waves (RF pulses) to the body and in response receives the NMR signals emitted from the nuclei, which constitute the body. Here, according to the shape of the RF pulses, various types of images inside the body

DOI 10.1007/978-981-4560-75-7_112-1

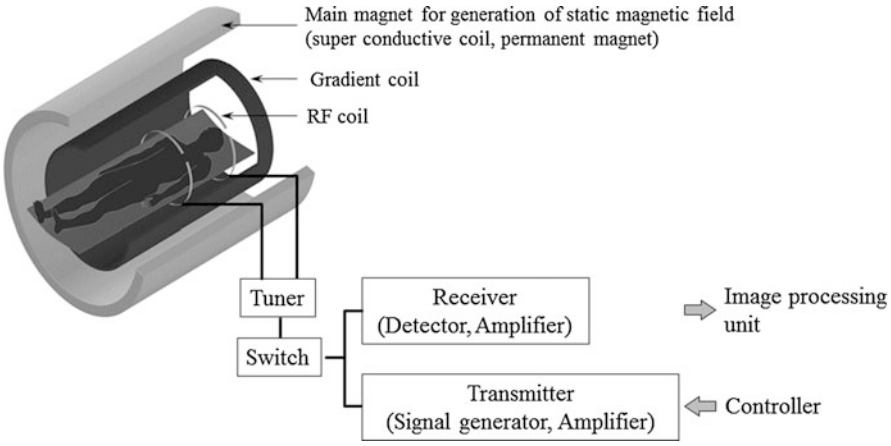


Fig. 2 Basic configuration of a magnetic resonance imaging device

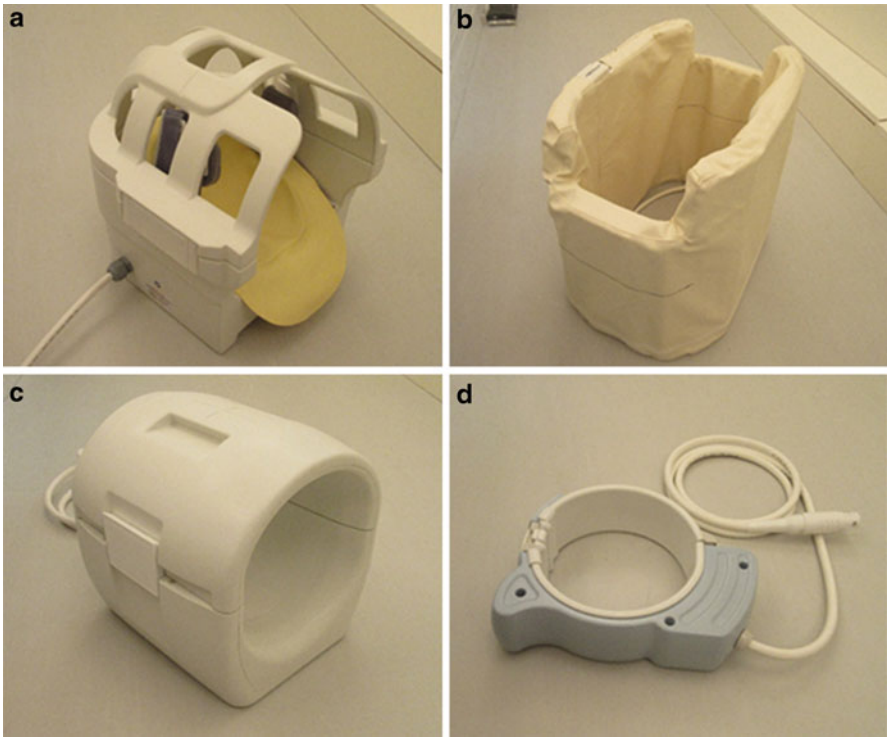


Fig. 3 Various radio frequency coils: (**a**) head coil; (**b**) chest coil; (**c**) knee coil; and (**d**) joint coil

can be obtained, as indicated in Fig. 1. In general, although the widths of the RF pulses are narrow, the amplitudes are not so low. Therefore, it is necessary to estimate the specific absorption rate (SAR) in the human body due to the radiated EM energy from the RF coils.

Until now, evaluation of the SAR in the human head has been studied during MRI (Angelone et al. 2004; Chen et al. 1998; Nguyen et al. 2004; Simunic et al. 1996; Collins et al. 2004), and restriction of the SAR for the safety of patients has been established as the International Electrotechnical Commission (IEC) standard (IEC Standard 2002). However, recently, MRI has been employed not only for the head but also for various other portions of the body. Moreover, MRIs tend to generate high-quality images and reduce the imaging time. In this chapter, the SAR distributions in a pregnant woman and her fetus are

DOI 10.1007/978-981-4560-75-7_112-1

investigated. In general, an ultrasound diagnostic system is used for the medical examination of pregnant women. However, an MRI is sometimes chosen for medical reasons, such as when the diagnosis of a fetal anomaly is unclear via ultrasound (Hubbald et al. 1999; Amin et al. 1999; Levine 2004).

The SAR distributions in a pregnant woman and her fetus exposed to the EM waves from the RF coil were calculated employing the two different types of pregnancy stage high-resolution whole-body pregnant woman models. In addition, two types of birdcage coil were employed for these calculations.

Numerical Calculation Method and Calculation Models

Calculation Method

In the numerical calculation, the electric field in and around the RF coil is analyzed by the finite difference time domain (FDTD) method (Yee 1966) and the EM energy absorption rate (SAR), which is one of the most important indices for evaluation on thermal effect of biological tissue, is calculated from the following equation (Eq. 1):

$$\mathrm{SAR} = \frac{\sigma}{\rho} E^2 [\mathrm{W/kg}] \tag{1}$$

where σ is the conductivity of the tissue [S/m], ρ is the density of the tissue [kg/m^3], and E is the electric field (rms) [V/m]. The SAR takes a value proportional to the square of the electric field around the antennas and is equivalent to the heating source generated by the electric field in the tissue.

Female Models

In this section, the SAR distributions on two types of realistic female models are calculated. Here, the realistic high-resolution whole-body voxel model of an adult Japanese female average figure, developed at the National Institute of Information and Communications Technology, Japan (NICT) (Nagaoka et al. 2004) is used for the calculation of the SAR of a pregnant woman in the early period (Model A). The size of the fetus (embryo) in the early period is very small, and hence it is difficult to confirm pregnancy status. Therefore, it is considered that the person who does not notice her own pregnancy state uses MRI. Moreover, the 26th gestational week pregnant woman model (Model B) (Nagaoka et al. 2004; Saito et al. 2006a; Togashi et al. 2008) is employed as one example of late pregnancy stages. In order to develop this model, the fetus model was first developed from the MR images of a pregnant woman in the 26th gestational week. The fetus model consists of six organs: fetal body, fetal brain, fetal eyes, amniotic fluid, placenta, and uterus wall. Next, the abdomen of the woman in Model A was expanded following the standard shape and structure of the pregnant woman. Finally, these two models were combined and adjusted following comments from medical doctors. More detailed explanation regarding the development of the pregnant woman models are described in Nagaoka et al. (2007).

Figures 4a, 4b show Models A and B, respectively. These high-resolution models are composed of a $2 \times 2 \times 2$ mm cubical voxel. The physical properties of these models were determined from previous publications (Hand et al. 2006; Gabriel 1996; Duck 1990; Van Den Berg et al. 1983). Table 1 summarizes some of the parameters regarding the abdomen of the pregnant woman and her fetus.

Calculation Models for Radio Frequency (RF) Coils

A birdcage coil is employed as one of the most fundamental RF coils for MRI systems. The birdcage coil is often used for both the transmission of the RF pulse and reception of the NMR signal, and is categorized into two types by the different position of the loaded capacitors: a highpass birdcage coil, and a lowpass birdcage coil (Jin 1998). Figures 5a, 5b show the highpass and lowpass birdcage coils, respectively.

DOI 10.1007/978-981-4560-75-7_112-1

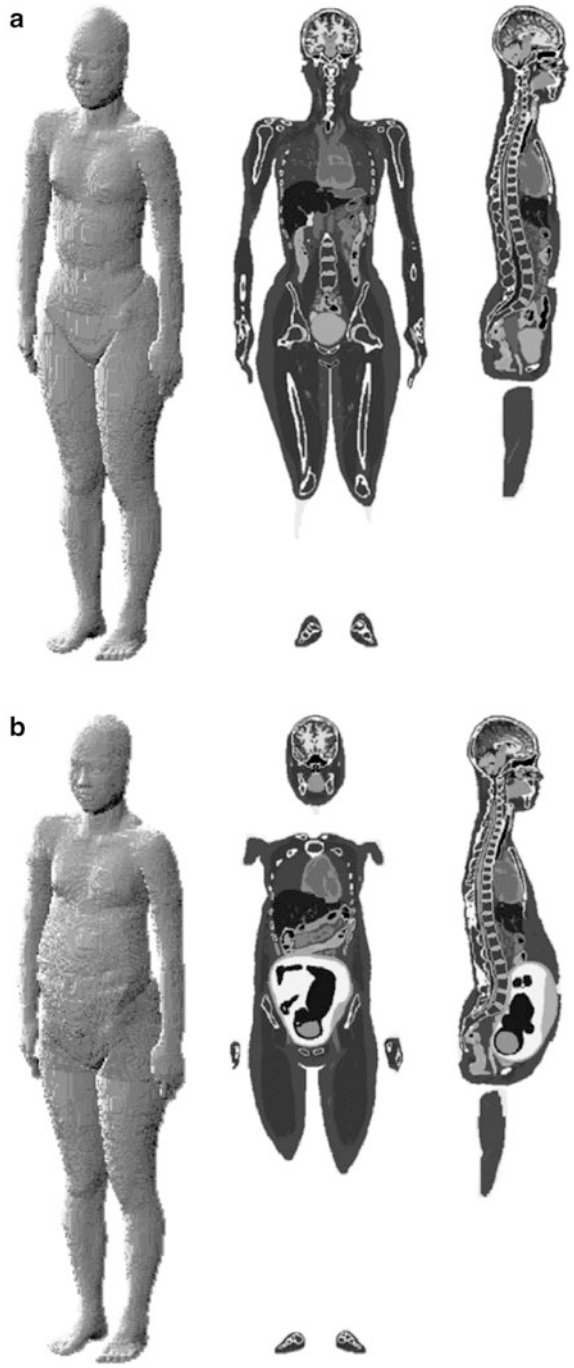


Fig. 4 Realistic female models: (**a**) Model A (pregnant woman model in early period); and (**b**) Model B (pregnant woman model in 26th gestational week)

The operating frequency of the coil is around 64 MHz in both cases, which is used in the generic 1.5 T MRI systems to excite the nuclei in the human body for imaging. These coils consist of two end rings and eight legs, the widths of which are 10 mm, and are modelled as perfect electric conductors for calculations. The diameter and the length of the coil are 600 and 700 mm, respectively, allowing the realistic female model to be used. In order to reduce radiation of RF energy to outside of the coil, the coils are surrounded by an RF shield that is also modeled as a perfect electric conductor. A cylindrical RF shield, with an

DOI 10.1007/978-981-4560-75-7_112-1

Table 1 Example of the physical properties of a pregnant woman and her fetus at 64 MHz

	Density (ρ) [kg/m^3]	Relative permittivity (ε_r)	Conductivity (σ) [s/m]
Maternal body			
Muscle	1040	72.0	0.71
Fat	928	13.6	0.07
Skin	1100	84.4	0.46
Blood	1060	86.5	1.21
Bone – cortical	1990	16.7	0.06
Bone – cancellous	1040	18.2	0.11
Lung	655	56.2	0.41
Liver	1050	80.6	0.45
Stomach	1050	85.8	0.88
Colon	1044	94.7	0.64
Small intestine	1044	118.3	1.59
Ovary	1048	106.8	0.69
Uterus wall	1052	92.1	0.91
Amniotic fluid	1000	97.3	2.07
Placenta	1060	86.5	1.21
Fetal body			
Body	1040	94.2	0.92
Eyes	1009	69.1	1.50
Brain	1030	97.2	0.76

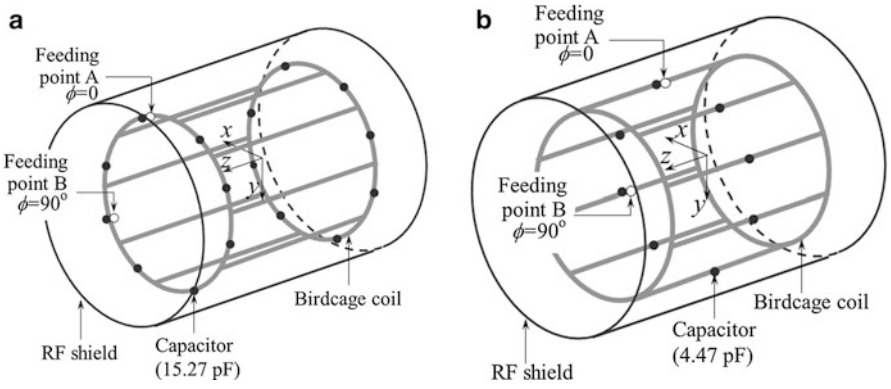


Fig. 5 Two types of birdcage coils: (**a**) highpass type; and (**b**) lowpass type

internal diameter of 740 mm and a length of 1260 mm, was located lateral to the coil. The dimensions of the coil and the RF shield were determined based on previous publications (Hand et al. 2006; Wu et al. 2006; Ochi et al. 1993).

The capacitors were loaded into the end rings on the highpass birdcage coil in Fig. 5a, and into the legs on the lowpass birdcage coil in Fig. 5b. In order to determine the capacitances, a "birdcage builder" (Chin et al. 1998), which calculates the resonance frequency of the birdcage coil using the equivalent electrical circuit model, was employed. Values for capacitors to resonate both coils at 64 MHz were 15.27 pF for the highpass birdcage coil and 4.47 pF for the lowpass birdcage coil. In addition, two feeding points were employed in these calculations and the phase difference between two ports is 90°. This excitation method is called "quadrature excitation" and can generate clear MRI images using a circular polarized field (Jin 1998).

DOI 10.1007/978-981-4560-75-7_112-1

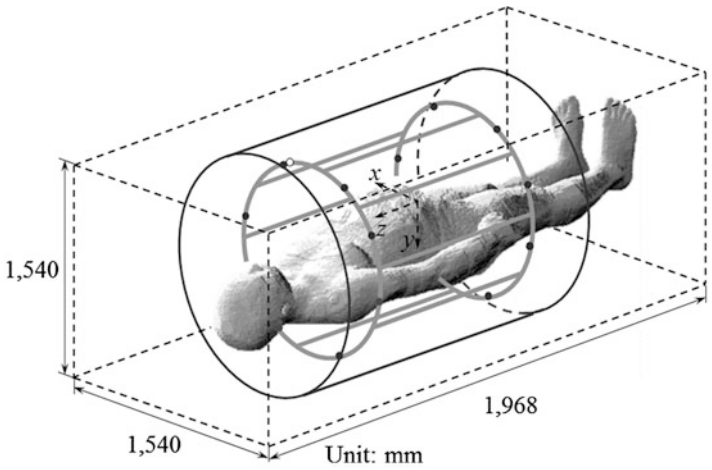


Fig. 6 Finite difference time domain calculation model

Table 2 Parameters for finite difference time domain calculations

Cell size [mm] (minimum)	Δx, Δy	2.0, 2.0
	Δz	2.0 (constant)
Cell size [mm] (maximum)	Δx, Δy	5.6, 5.6
	Δz	2.0 (constant)
Analytical space $x \times y \times z$ (cell)		$534 \times 534 \times 984$
Time step [ps]		3.8
Absorbing boundary condition		Perfect matched layer (PML) (8 layers)

Numerical Calculation Model

Figure 6 shows the FDTD calculation model including the realistic female model. As an example of the model, Model B is inserted into the highpass birdcage coil. There are several possibilities for the coil position, because it is considered that the position of the coil is not the same for each type of imaging. Previously, the relationships of several positions of the RF coil with the SAR were calculated (Kikuchi et al. 2007). However, here, the coil was placed in the center of the uterus in Model A, and placed at the center of the fetal brain in Model B. In order to calculate the whole-body female models, a large analytical region was required and a super technical server (Hitachi SR11000) in the Institute of Media and Information Technology, Chiba University (Chiba, Japan) was employed. In the numerical calculations, the uniform grid size inside the shield including the female model is 2.0 mm. The non-uniform mesh is used for the outside of the shield, because most of the EM waves are not emitted outward by the shield.

The parameters used in the FDTD calculations are listed in Table 2. In addition, steady-state analysis is performed by enforcing a continuous sinusoidal wave of electric field on the feeding gaps to calculate the SAR distribution in the model. Moreover, the coordinate origin is the center of the calculation model. The effectiveness of the calculation technique is confirmed by comparison with an experimental result (Saito et al. 2006b).

SAR Distributions in the Abdomen of Pregnant Woman Models

In the actual MRI system, the various pulse sequences are employed according to the imaging region and type. However, the SARs in the pregnant woman and her fetus were calculated by exciting the continuous sinusoidal wave at the feeding point, because the radiation power of each pulse sequence was not clear. Here, in order to simplify conversion of radiation power, the SAR values are normalized by 1.0 W

DOI 10.1007/978-981-4560-75-7_112-1

radiation power from the coil in all cases. In addition, the radiation power was calculated by a surface integral of the Poynting vector on the closed surface surrounding the coil.

Comparison of the SAR Distributions in Model A with the Two Types of RF Coils

Figures 7a–7h show the calculated SAR distributions inside Model A by employing the highpass and lowpass birdcage coil. The observation planes are the sagittal plane (*yz*-plane) including the uterus, and the coronal plane (*xz*-plane) around the center of the uterus. Moreover, positions of observation line A–A′, B–B′, C–C′, and D–D′ are indicated in the Figs. 7a–7d. Here, Model A (the non-pregnant woman model) is assumed as the early period of pregnancy, when the fetus or embryo is very small. Therefore, the SAR inside the uterus and ovary were observed in place of fetal tissue (Dimbylow 2006).

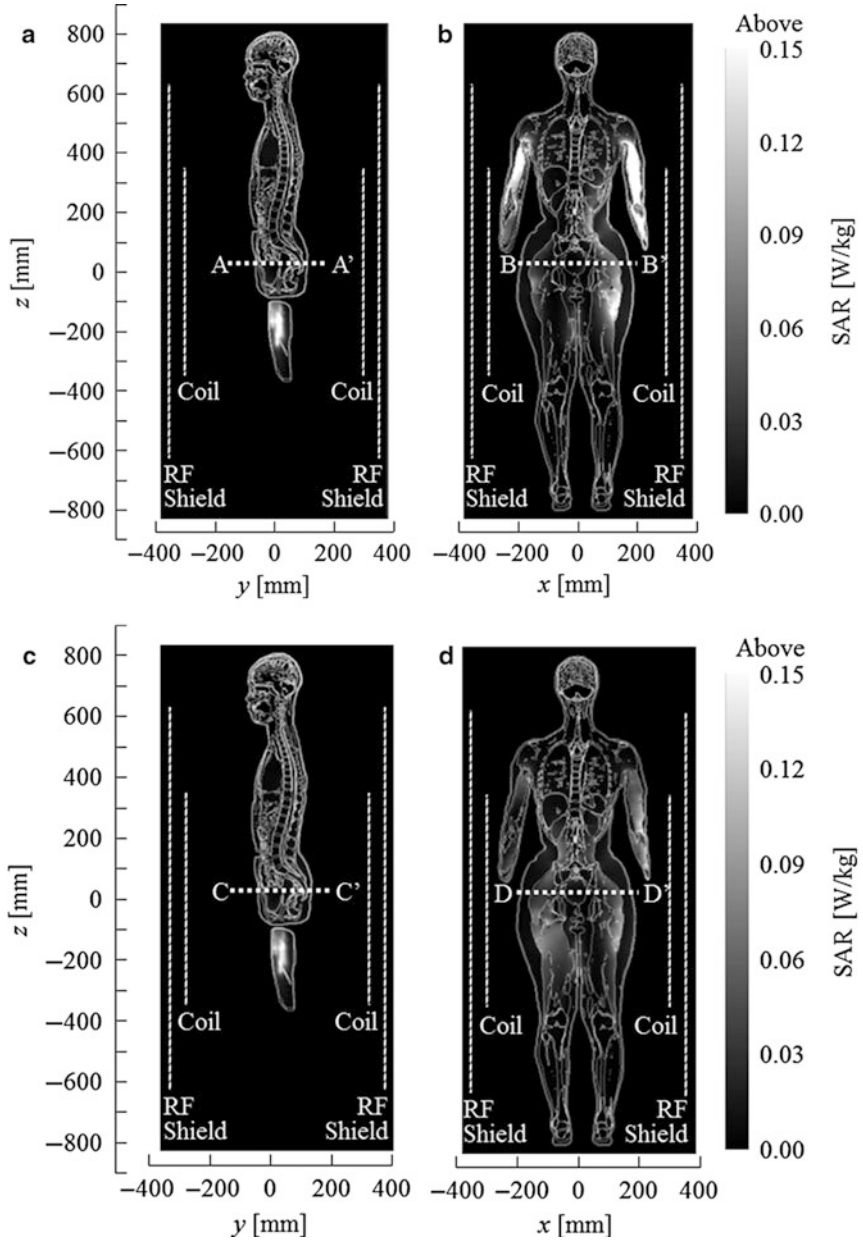


Fig. 7 (continued)

DOI 10.1007/978-981-4560-75-7_112-1

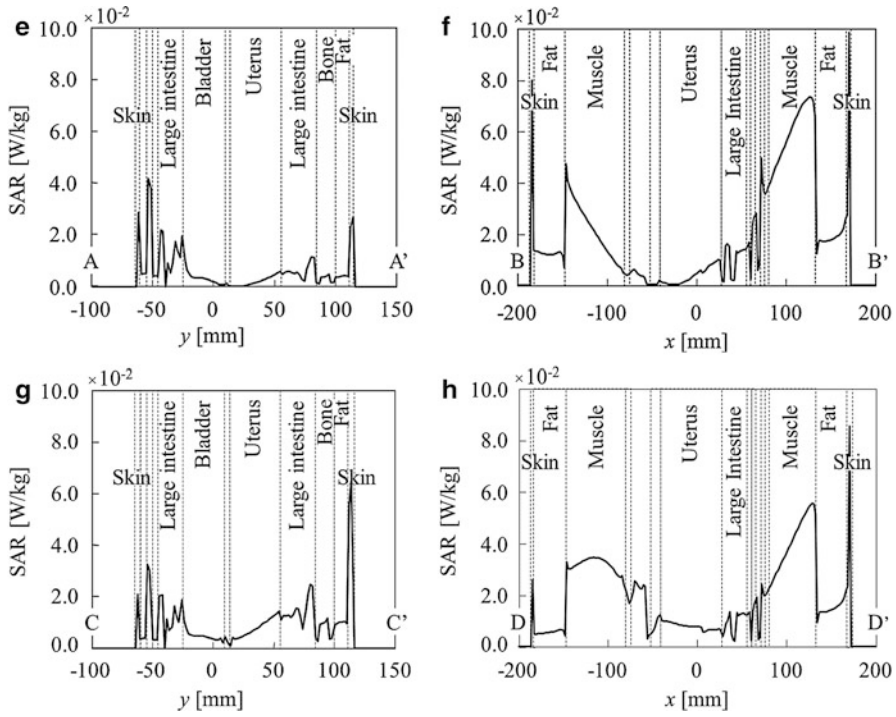


Fig. 7 Calculated specific absorption rate distributions inside Model A due to radiated electromagnetic waves from two types of birdcage coils: (**a**) sagittal plane (highpass birdcage coil); (**b**) coronal plane (highpass birdcage coil); (**c**) sagittal plane (lowpass birdcage coil); (**d**) coronal plane (lowpass birdcage coil); (**e**) A–A′ in sagittal plane (highpass birdcage coil); (**f**) B–B′ in coronal plane (highpass birdcage coil); (**g**) A–A′ in sagittal plane (lowpass birdcage coil); and (**h**) B–B′ in coronal plane (lowpass birdcage coil)

As shown in Figs. 7b, 7d, the SAR distributions on the coronal plane are observed as having an unsymmetrical tendency in comparison with the left side (-225 mm $< x < 0$) and right side ($0 < x <$ 225 mm). Moreover, it was confirmed that the tendency of the SAR distributions depends on their position in the body. Because it was considered that the tendency is influenced by two points of feeding, a symmetry of SAR distributions was confirmed in the case of one point of feeding model. Moreover, from Fig. 7a–7d, relatively high SAR values are observed around the skin, muscle, etc., which have a high electrical conductivity and are located close to the surface of the maternal body. The tendency can be confirmed in Fig. 7e–7h, except for fat tissue which has low electrical conductivity. Moreover, high SAR values are observed inside the thigh ($x = 0$, $-100 < z < -300$ mm in Fig. 7b, 7d) and armpit ($x = \pm 200$ mm, $z = 400$ mm in Fig. 7b, 7d). This is due to the concentrated electric field at those narrow gaps. In comparison with the SAR distributions due to the EM energy from the highpass and lowpass birdcage coil, comparatively high SAR values are observed in the vicinity of each position of feeding points and loaded capacitors. Here, according to the result that confirmed all SAR distributions, it was found that this phenomenon is dependent on whether an electric field concentrates on an existing gap of the human body and a coil.

Meanwhile, the SAR in the uterus and the ovary (around $15 < y < 60$ mm, $-50 < z < 50$ mm in Fig. 7a, 7c, and around $-40 < x < 25$ mm, $-50 < z < 50$ mm in Fig. 7b, 7d) are low compared to the maternal surface of skin and muscle tissue. In addition, a low SAR value is observed at the uterus (around $15 < y <$ 60 mm in Fig. 7e, 7g, and around $-40 < x < 25$ mm in Fig. 7f, 7h) due to the attenuation of EM energy in the muscle tissue. Moreover, it is observed that there is very little difference between the SAR in the uterus and the ovary by different coil types. From these results, it has been confirmed that the SARs in the uterus and the ovary are low, because these areas are located in deep regions of the body.

DOI 10.1007/978-981-4560-75-7_112-1

Comparison of the SAR Distributions in Model B with the Two Types of RF Coils

Figures 8a–8h show the calculated SAR distributions in Model B using the highpass and lowpass birdcage coil. The observation planes are the sagittal plane (*yz*-plane) including the center of the fetal head, and the coronal plane (*xz*-plane) around the center of the fetal body. Positions of the observation line A–A′, B–B′, C–C′, and D–D′ are indicated in Fig. 8a–8d.

As shown in Fig. 8, relatively high SAR values are observed around the maternal surface and unsymmetrical tendencies are observed in the same results as SAR distributions in Model A. Moreover, it is observed that the SAR at the amniotic fluid is also relatively high. In particular, relatively high SAR

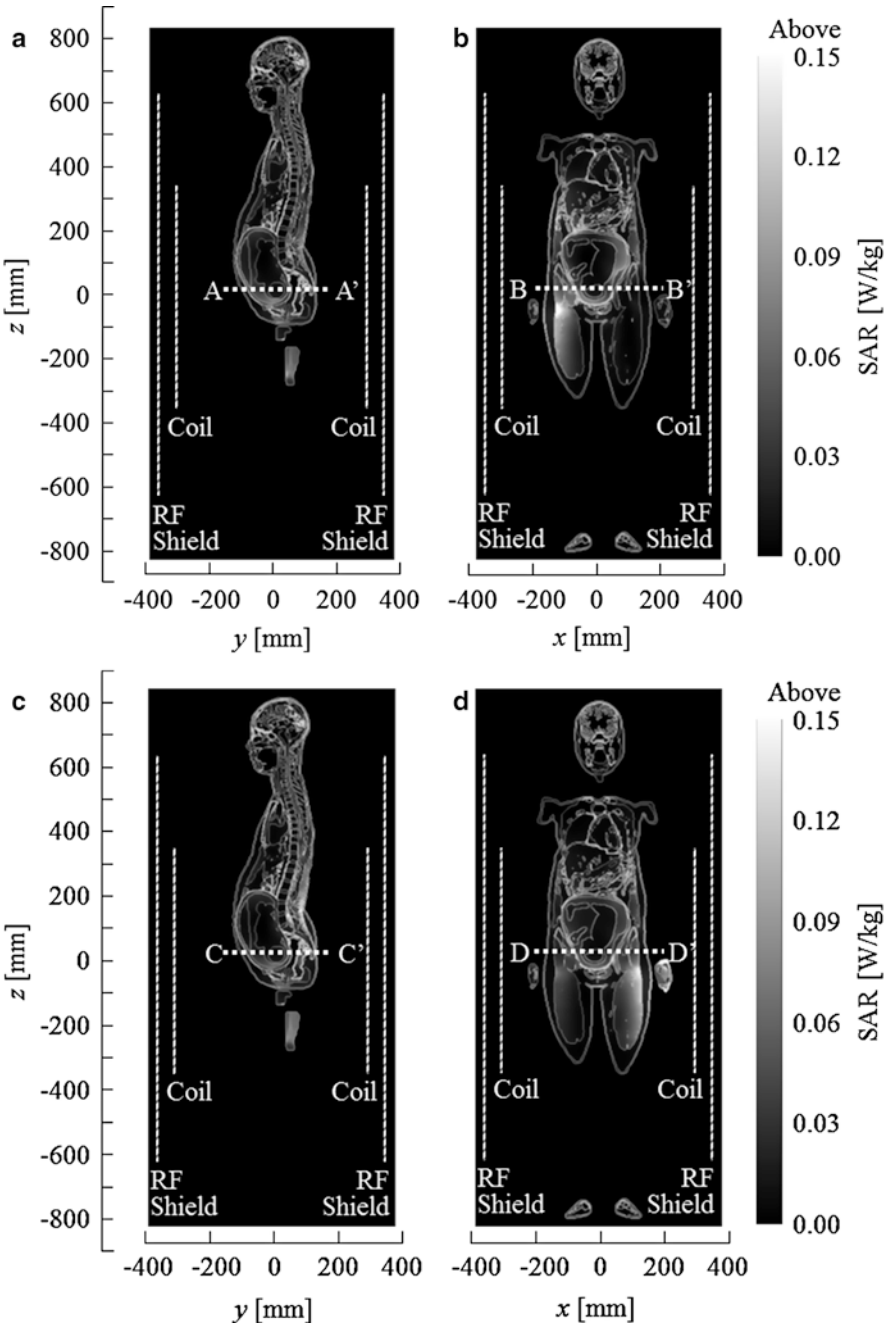


Fig. 8 (continued)

DOI 10.1007/978-981-4560-75-7_112-1

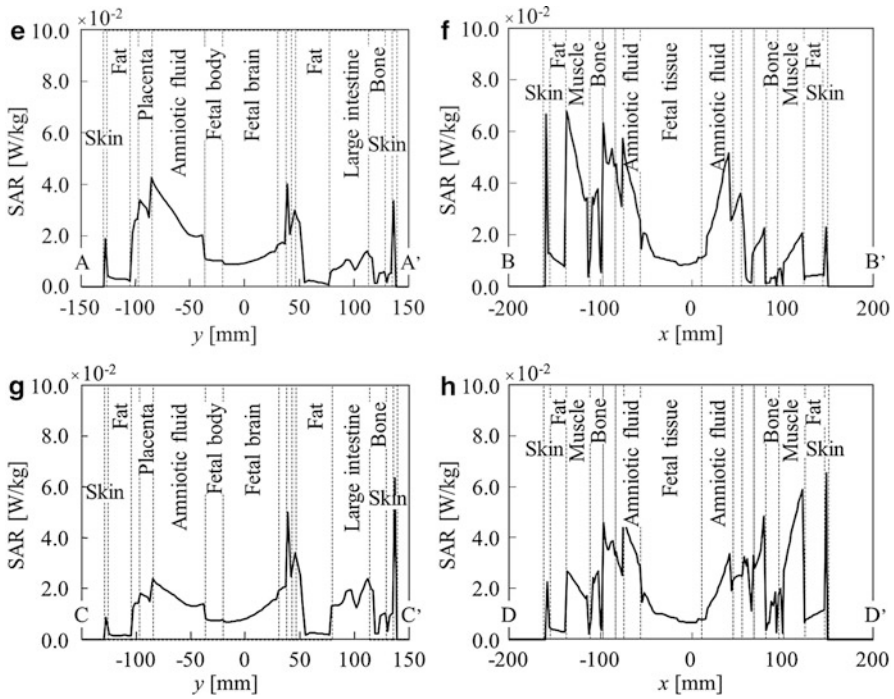


Fig. 8 Calculated specific absorption rate distributions inside Model B due to radiated electromagnetic waves from two types of birdcage coils: (**a**) sagittal plane (highpass birdcage coil); (**b**) coronal plane (highpass birdcage coil); (**c**) sagittal plane (lowpass birdcage coil); (**d**) coronal plane (lowpass birdcage coil); (**e**) A–A′ in sagittal plane (highpass birdcage coil); (**f**) B–B′ in coronal plane (highpass birdcage coil); (**g**) A–A′ in sagittal plane (lowpass birdcage coil); and (**h**) B–B′ in coronal plane (lowpass birdcage coil)

values are observed at the upper and lower portions of amniotic fluid (around $y = -100, z = -50, 150$ mm in Fig. 8a, 8c, and around $x = 0$, $z = -50$ mm and $x = 100$ mm, $50 < z < 200$ mm in Fig. 8b, 8d). In addition, as Fig. 8e–8h indicate, it has been found that the SAR in the amniotic fluid near the fetal tissue is relatively high. This is because the electrical conductivity of amniotic fluid is almost 1.5–2.0 times higher than other tissues, as listed in Table 1. In addition, the SAR distributions around the boundary of each tissue were precipitously varied due to the heterogeneous structure.

However, the SAR in the fetus (around $-20 < y < 50$ mm in Fig. 8e, 8g, and around $-50 < x < 15$ mm in Fig. 8f, 8h) is low compared with the values of the maternal body. Moreover, the maximum SAR value within the fetus is less than 1.0 % of that of the maternal body. From these results, it has been confirmed that the SAR in the fetus is attributed to the attenuation of EM energy in the amniotic fluid which has high conductivity.

Summary

Here, we introduced the SAR calculations of pregnant women during MRI. Deeper analyses of the SAR values are discussed in Kikuchi et al. (2009) and there have been several papers published on SAR analyses for MRI (De Wilde et al. 2005). SAR reduction during MRI via improvement of RF coil structure is also one of the important study themes.

DOI 10.1007/978-981-4560-75-7_112-1

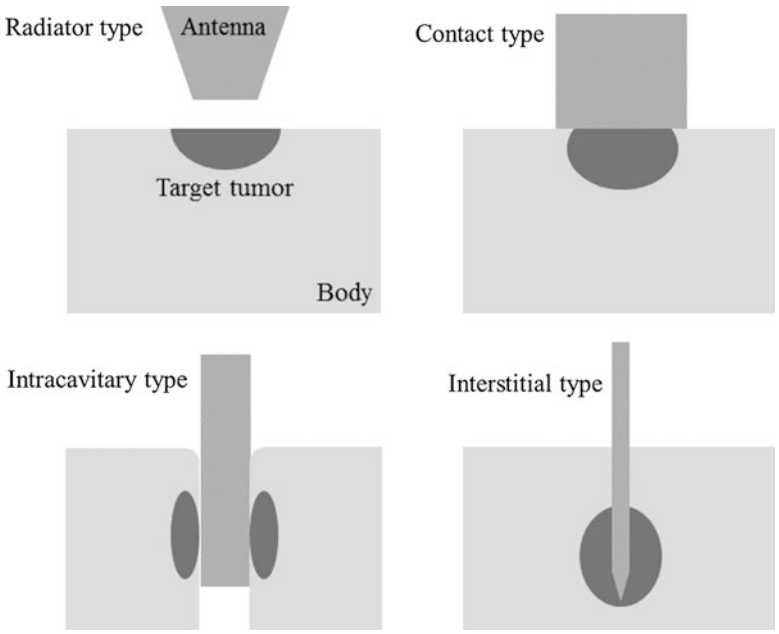


Fig. 9 Various types of antennas for microwave heating

Thermal Treatment of Cancer by Microwave Energy

Classification of Heating Scheme

There are several different methods of cancer treatment, including surgical operations, radiation therapy, chemotherapy, gene therapy, immunotherapy, ablation, and hyperthermia. Two or more different ways are sometimes combined for clinical use. Hyperthermia is one of the promising modalities for cancer treatment, utilizing the difference in the thermal sensitivity between tumors and normal tissue, and is one of the most effective medical applications of microwave techniques. In this treatment, the tumor is heated up to the therapeutic temperature, between 42 °C and 45 °C, without overheating the surrounding normal tissues (Seegenschmiedt et al. 1995). Moreover, the effect of other cancer treatments such as radiotherapy and chemotherapy can be enhanced by using them together with hyperthermia.

There are several energy sources used for heating the tumor, such as hot water, ultrasound, EM wave, etc. Microwave energy is one of the heating sources used for localized hyperthermia (Tsukiyama et al. 1991). The frequency for microwave hyperthermia ranges from several hundred to several thousand MHz (i.e., up to several GHz). For hyperthermia treatments, heating devices using 430 MHz, 915 MHz, and 2.45 GHz have been commercially developed.

The size and shape of the target tumors also vary. Therefore, it is impossible to treat many different types of tumors with a single unique type of antenna (Ito 2006). Figure 9 shows four different types of microwave antennas for hyperthermia treatment. The “radiator type” and the “contact type” are classified into “external antennas,” and the “intracavitary type” and the “interstitial type” are called “internal antennas.” Here, we introduce some topics regarding internal antennas.

Interstitial Microwave Hyperthermia

Interstitial microwave hyperthermia can be applied for a localized tumor. With this technique, thin microwave antennas are inserted into the tumor and radiate microwave energy directly into the target (Tsukiyama et al. 1991; Lin and Wang 1987), permitting effective and reliable heating of the tumor.

DOI 10.1007/978-981-4560-75-7_112-1

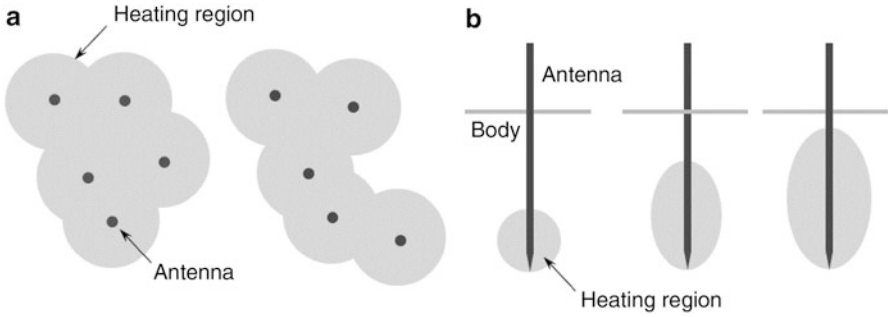


Fig. 10 Heating patterns generated by thin microwave antennas: (**a**) perpendicular to antenna axis; and (**b**) parallel to antenna axis

In this technique, it is possible to change the heating pattern in the perpendicular direction of the antenna axis by varying the number of elements and their insertion points (Fig. 10a). However, control of the heating pattern in the longitudinal direction of the antenna axis is realized by changing the structure of the antenna elements while keeping the thin structure (Fig. 10b). In the following sections, we describe the heating characteristics of the thin microwave antennas, which generate the controllable heating patterns in the longitudinal direction of the antenna axis. Moreover, arbitrary shape heating patterns by an array applicator composed of the developed antennas are also outlined.

Antenna Configuration

In order to generate a controllable heating region in the longitudinal direction, a coaxial-dipole antenna is employed. This antenna was developed for generation of a localized heating region around only the antenna tip (Kikuchi et al. 2006; Saito et al. 2008). Here, we attempt to explain advanced heating pattern control by the coaxial-dipole antenna.

Figure 11 shows the basic configuration of the coaxial-dipole antenna. The operating frequency of the antenna is 2.45 GHz, which is one of the industrial, scientific, and medical (ISM) frequencies in Japan. This antenna is composed of a thin, semi-rigid coaxial cable. A ring slot is cut on the outer conductor of the thin coaxial cable and the tip of the cable is short-circuited. Here, L_{ts} is the length from the shorted point to the center of the slot and is set to 10 mm for impedance matching (Kikuchi et al. 2006). In addition, two conductive sleeves, whose lengths (L_{ld} and L_{ud}) affect the shape of the heating region, are connected to both sides of the slot. The antenna is inserted to a catheter for hygiene. In addition, although the antenna insertion depth depends on the position of the target tumor, it is set 70 mm in this chapter.

SAR Distributions of a Single Antenna

In order to estimate the heating pattern around the antennas, the SAR around the antennas is calculated. Figure 12 and Table 3 show the calculation model of the coaxial-dipole antenna and parameters for the calculations, respectively.

Figure 13 shows the measured and calculated SAR distributions around the antenna. The SAR observation line is parallel to the antenna axis and is 3.0 mm away from the center of the antenna (shown in Fig. 12). Here, the SAR distributions are measured using the thermographic method; details of the measurement are described in Saito et al. (2013a).

The length of the lower sleeve, L_{ld}, is changed from 10.0 to 30.0 mm when L_{ud} is set to 20.0 mm. In these results, the length of the high SAR region in the longitudinal direction is almost the same as the length of $L_{ld} + L_{ud}$ from the tip (the gray region in Fig. 13). Moreover, in all cases, good agreements are observed between the calculations and the measurements except at the slot position. The SAR value at the

DOI 10.1007/978-981-4560-75-7_112-1

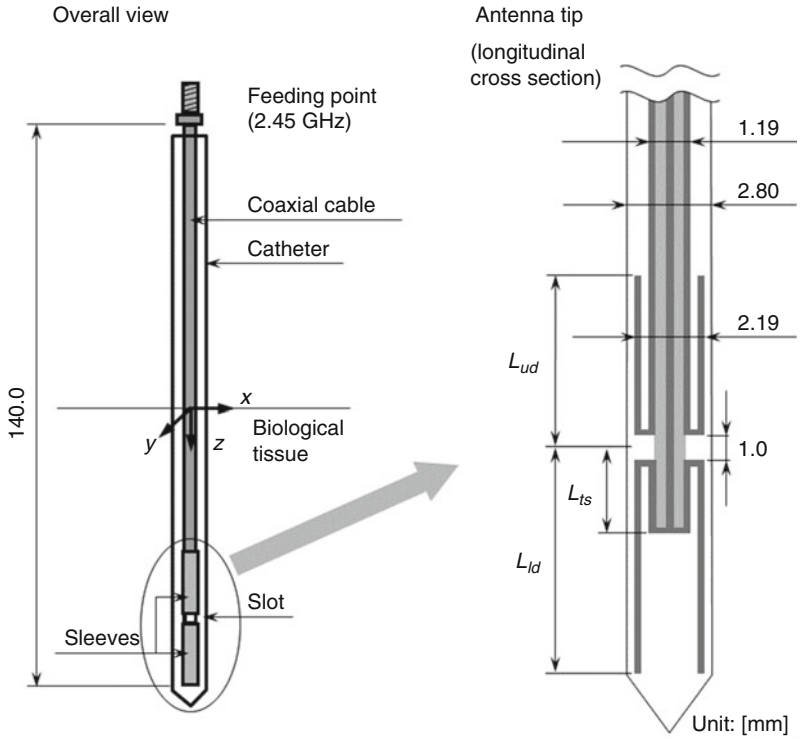


Fig. 11 Basic configuration of the coaxial-dipole antenna

slot is extremely high compared to other positions. In such a case, heat transfer at the peak point cannot be ignored in the thermographic method and the measured SAR value is lower than the calculation.

From these results, it can be said that the coaxial-dipole antennas generate the controllable heating regions in the longitudinal direction. In addition, according to preliminary investigations, the length of the sleeve (L_{ld} or L_{ud}) should be less than 30 mm for generating controllable heating patterns.

Moreover, if an antenna with enough lower sleeve length is prepared, the size of the longitudinal heating region can be adjusted easily by cutting the lower sleeve. In Fig. 13, the minimum length of the lower sleeve L_{ld} is 10 mm, which is the same as L_{ts}.

Generation of Arbitrary Shape Heating Patterns by Array Applicator

The controllable heating patterns in the longitudinal direction of the antenna axis could be generated by use of the coaxial-dipole antennas. Here, generation of three-dimensional arbitrary heating patterns is considered to be by array applicators. (The term "applicator" indicates the part of the heating system that is directly applied on the human body. On this basis, an antenna is an applicator. However, the "applicator" is used only for an array antenna and is not used for a single antenna element conventionally). In this chapter, heating patterns of the array applicators composed of four coaxial-dipole antennas are investigated. Figure 14 shows a calculation model of the array applicator. All of the antenna elements are fed by the same amplitude and in phase.

According to a previous study (Saito et al. 2004), 20 mm array spacing (intervals of antenna elements) was suitable for actual treatment. Therefore, this array spacing is complied with. In addition, antenna insertion depths of all antenna elements of 70 mm are the same.

DOI 10.1007/978-981-4560-75-7_112-1

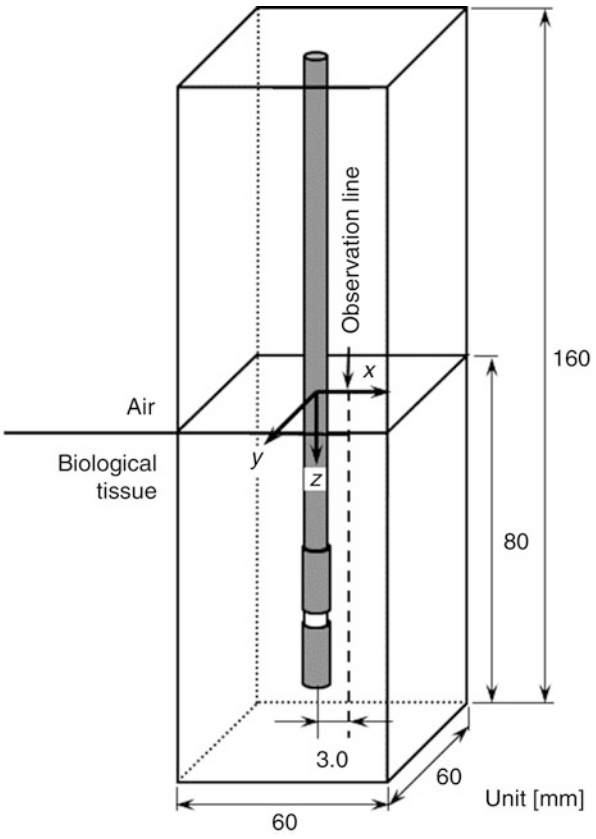


Fig. 12 Finite difference time domain calculation model of coaxial-dipole antenna

Table 3 Parameters for finite difference time domain calculations and material parameters

Parameters for FDTD calculations		
Cell size [mm] (minimum)	Δx, Δy	0.05
	Δz	1.00 (constant)
Cell size [mm] (maximum)	Δx, Δy	1.50
	Δz	1.00 (constant)
Absorbing boundary condition	Mur (first order)	
Material parameters		
	Relative permittivity (ε_r)	Conductivity (σ) [S/m]
Biological tissue (phantom)	42.5	1.5
Catheter	2.6	0.0

Two kinds of array applicators, which are composed of different antenna elements, are considered, as follows:

(a) All antenna elements have the same parameters:
  – Element A, B, C, and D: $L_{ld} = 20.0$ mm, $L_{ud} = 20.0$ mm

(b) Two sleeve parameters:
  – Element A and B: $L_{ld} = 20.0$ mm, $L_{ud} = 20.0$ mm
  – Element C and D: $L_{ld} = 10.0$ mm, $L_{ud} = 10.0$ mm

DOI 10.1007/978-981-4560-75-7_112-1

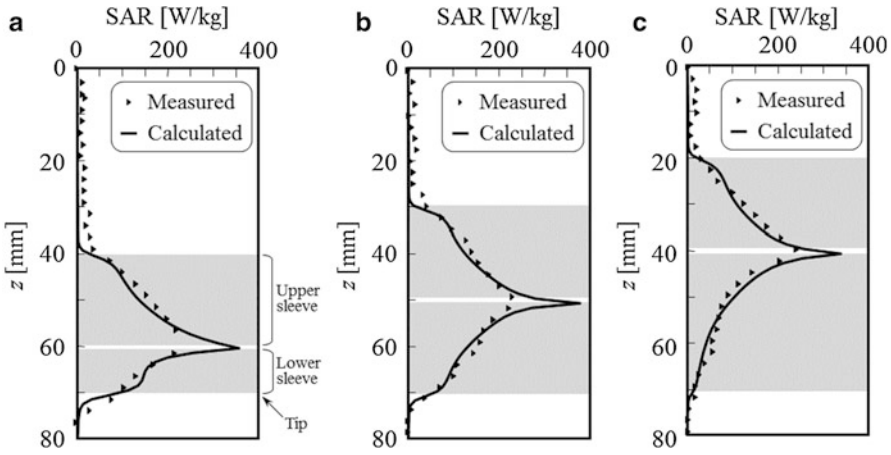


Fig. 13 Measured and calculated specific absorption rate distributions. All results are normalized by 1.0 W radiation power. (**a**) $L_{ld} = 10.0$ mm; (**b**) $L_{ld} = 20.0$ mm; and (**c**) $L_{ld} = 30.0$ mm. L_{ld} length of the lower conductive sleeve

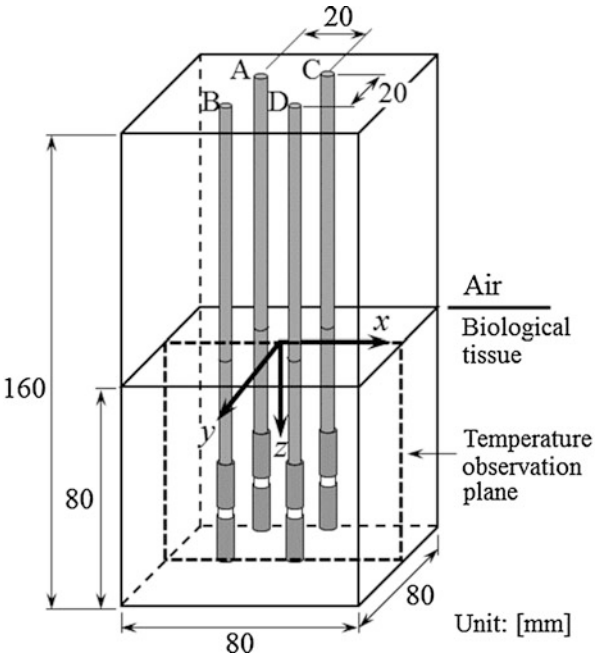


Fig. 14 Calculation model for the array applicator composed of four coaxial-dipole antennas

Here, heating patterns of the array applicators are evaluated by temperature distributions. The temperature distributions inside biological tissue can be calculated by solving a bioheat transfer equation (Penns 1948) (Eq. 2) numerically. The detail of the calculation scheme is the same as Saito et al. (2006c). In addition, electrical and thermal parameters for the calculations are listed in Table 4 (Van Den Berg et al. 1983; Johnson and Guy 1972).

$$\rho c \frac{\partial T}{\partial t} = \kappa \nabla^2 T - \rho \rho_b c_b F (T - T_b) + \rho \cdot \mathrm{SAR} \tag{2}$$

where T is the temperature [°C], t is the time [s], ρ is the density of the tissue [kg/m^3], c is the specific heat of the tissue [J/kg K], κ is the thermal conductivity of the tissue [W/m K], ρ_b is the density of the blood

DOI 10.1007/978-981-4560-75-7_112-1

Table 4 Electrical and thermal parameters for temperature calculations

Electrical properties	
Relative permittivity (biological tissue) (ε_r)	47.0
Conductivity (biological tissue) (σ) [S/m]	2.21
Thermal properties	
Specific heat (biological tissue) (c) [J/kg K]	3500
Specific heat (blood) (c_b) [J/kg K]	3960
Thermal conductivity (κ) [W/m K]	0.60
Density (biological tissue) (ρ) [kg/m^3]	1020
Density (blood) (ρ_b) [kg/m^3]	1060
Blood flow rate (F) [m^3/kg s]	8.33×10^{-6}
Initial temperature [°C]	37.0
Blood temperature [°C]	37.0

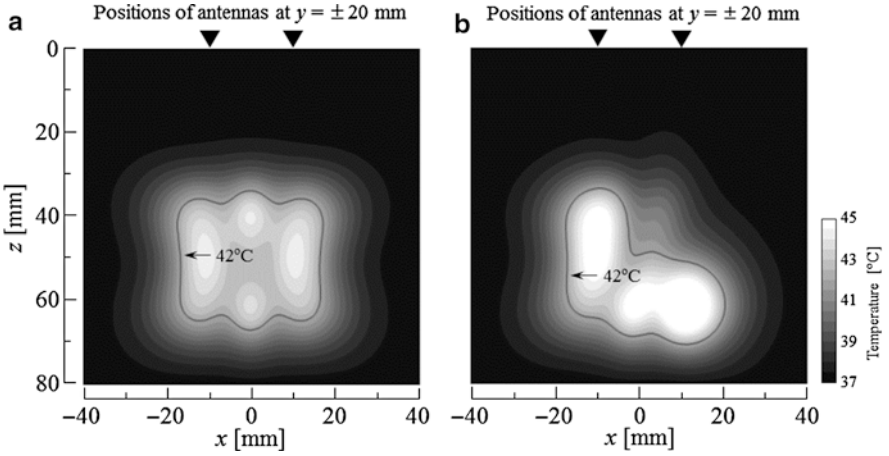


Fig. 15 Calculated temperature distributions by two types of array applicators. (**a**) All antenna elements have the same parameters. (**b**) Two sleeve parameters

[kg/m^3], c_b is the specific heat of the blood [J/kg K], T_b is the temperature of the blood [°C], and F is the blood flow rate [m^3/kg s].

Figure 15 shows the calculated temperature distributions. The temperature observation plane is defined in Fig. 14 and the black solid line indicates the region more than 42 °C, which is the lowest temperature for the treatment (effective heating region). Radiation power from the whole array is 20 W. Even in the array applicator, it is preferable that the size of the effective heating region in z direction is same as the length of $L_{ld} + L_{ud}$ for heating pattern control.

As shown in Fig. 15a, when all antenna elements have the same parameters, the size of the effective heating region in the longitudinal direction (z direction at $x = \pm 10$ mm) is approximately 27 mm. In this case, the length of $L_{ld} + L_{ud}$ is 40 mm. Therefore, the size of effective heating region decreases by 33 %.

On the other hand, as shown in Fig. 15b, in the case of using two kinds of sleeve parameters, the sizes of the longitudinal heating regions depend on the sleeve length. Therefore, the length of the effective heating region in z direction close to the elements A and B (at $x = -10$ mm) is approximately 29 mm, even though the length of $L_{ld} + L_{ud}$ is 40 mm. It decline by 28 %. In addition, the length in z direction around the elements C and D (at $x = +10$ mm) is 19 mm. This is almost the same as $L_{ld} + L_{ud}$.

DOI 10.1007/978-981-4560-75-7_112-1

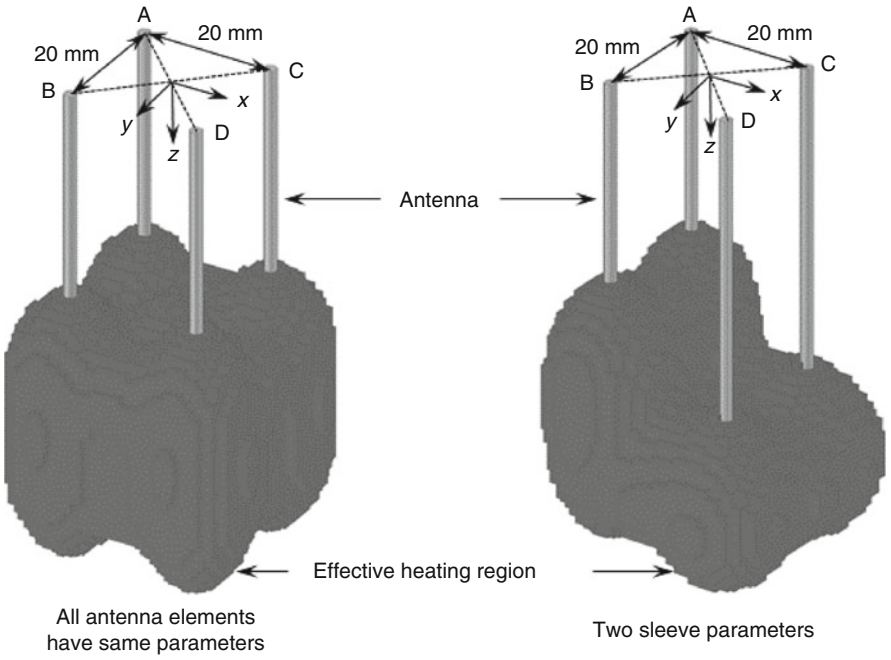


Fig. 16 Three-dimensional calculated temperature distributions

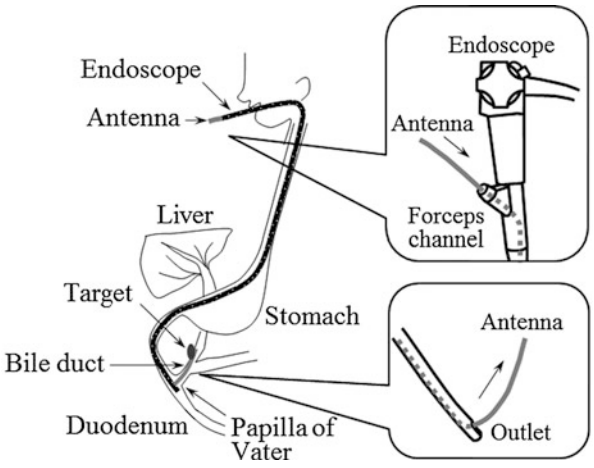


Fig. 17 Scheme of intracavitary microwave hyperthermia for bile duct carcinoma

From these results, it can be said that the heating pattern depends on the sleeve length in the array applicator. This can be clearly seen in Fig. 16, which shows the three-dimension effective heating region.

Intracavitary Microwave Hyperthermia

Microwave antennas can be used for intracavitary microwave hyperthermia if the target tumor is located close to an abdominal cavity, such as the esophagus, large intestine, urethra, etc. We have been studying the microwave antenna with endoscope for intracavitary microwave hyperthermia of bile duct carcinomas, and the scheme of the treatment is shown in Fig. 17. In this treatment, an endoscope is firstly inserted into the duodenum, and a long, thin, flexible microwave antenna is inserted into the forceps channel of the endoscope. Finally, the antenna is guided to the bile duct through the papilla of

DOI 10.1007/978-981-4560-75-7_112-1

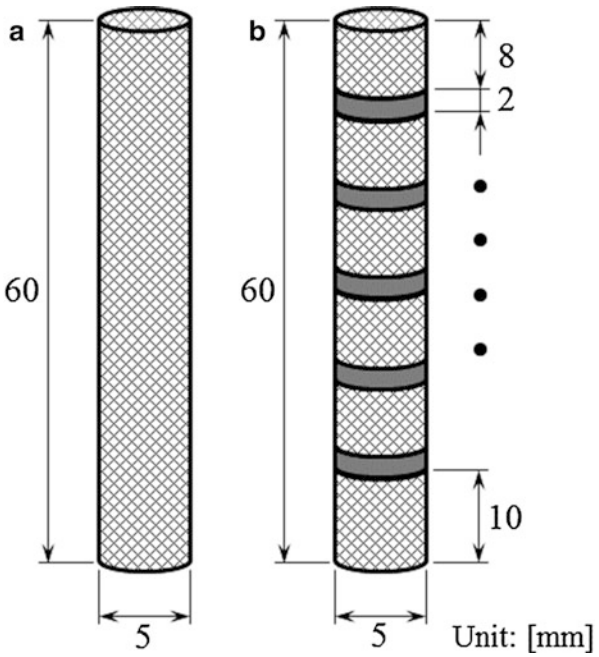


Fig. 18 Structure of biliary stents: (**a**) conventional stent; and (**b**) proposed stent

Vater, which is located in the duodenum. This non-invasive treatment is effective for treating unresectable bile duct carcinoma. Therefore, it is expected to be useful for treating various types and locations of carcinomas.

It should be noted that placement of self-expandable metallic stents is the current standard of care for patients with malignant obstructive jaundice caused by the bile duct carcinoma. However, in less than 50 % of patients with metallic stents, a stent blockage develops within 6–8 months (Loew et al. 2009) by explosive growth of cancer cells. Moreover, according to a previous study, the metallic stent will be shielded from microwave energy from the antenna inserted into it (Saito et al. 2010). Therefore, the authors introduced a new biliary stent compatible with microwave hyperthermia of bile duct carcinoma involving the use of an endoscopic coaxial antenna for intracavitary treatment.

Structure of the Biliary Stent

Figure 18 shows the structure of the two different biliary stent models: the conventional metallic stent (Fig. 18a) and the proposed stent (Fig. 18b) (Itakura et al. 2012). The proposed stent is composed of six metallic stents and five plastic stents of 2 mm in length. Taking the previous study into account, the arrangement of plastic stents at regular intervals could generate extensive heating outside of the stent. Moreover, if the distance between the plastic stents is narrowed to some extent, microwave energy cannot be radiated from plastic stents efficiently. Therefore, plastic stents (relative permittivity: 2.1) are arranged at regular intervals of 8 mm. The length and diameter of the stent, which were determined taking into account a previous study (Püspök et al. 2005), are 60 and 5 mm, respectively. In order to adjust for the length of the stent, a metallic part at the bottom of the proposed stent is set to 10 mm in length. Although stents generally have a mesh structure, a solid cylindrical structure is used in the calculation because the reticulation of a stent mesh is small enough in comparison with the wave length of the feeding EM wave.

Structure of the Coaxial Antenna

Figure 19 shows the structure of the coaxial antenna for feeding. The antenna is composed of a thin, flexible coaxial cable that can be inserted into the endoscope. The diameter of the coaxial antenna is

DOI 10.1007/978-981-4560-75-7_112-1

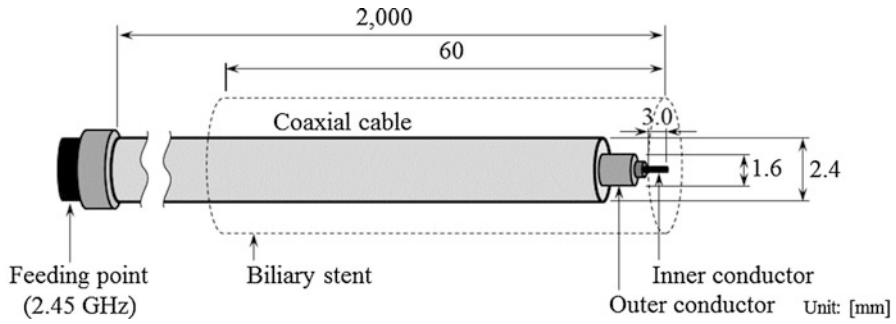


Fig. 19 Structure of feeding coaxial antenna

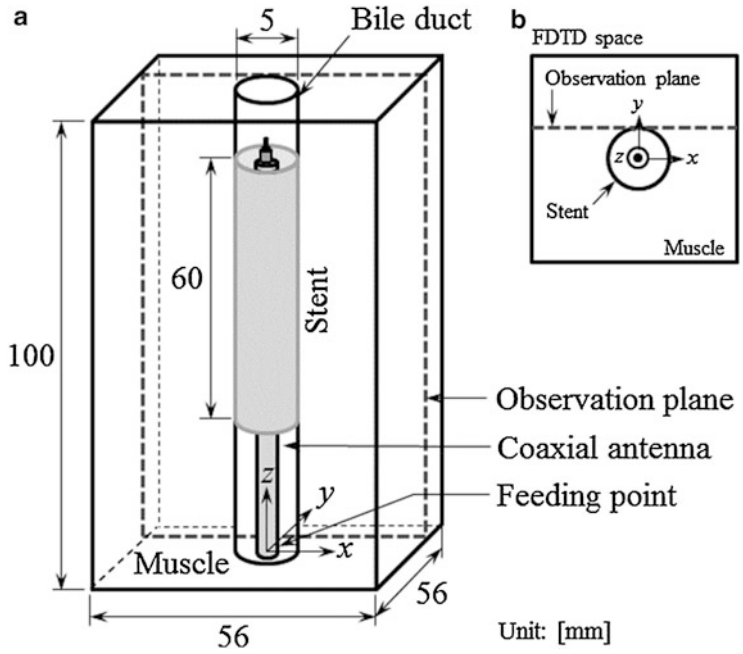


Fig. 20 Finite difference time domain space for the biliary stent and coaxial antenna: (**a**) over view; and (**b**) top view

2.4 mm. An inner conductor of the antenna sticks away 3 mm from the tip of the cable. A feeding point is set at the end of the cable. In this treatment, the whole antenna is placed at the center of the stent, and the antenna passes through the stent end at approximately 1 mm. Therefore, none of the antenna has contact with the stent. Thus, this feeding method utilizing EM coupling is able to be carried out in situations where the biliary stent has been covered with tissues or tumors due to restenosis.

Calculation Model

Figure 20 shows the FDTD calculation model for the biliary stent and the coaxial antenna and observation plane. The stent model is placed in the bile duct filled with the bile, and the bile duct is in the muscle. Moreover, assuming restenosis due to tumors, the inside of the stent is filled with muscle from the viewpoint of their electrical properties. The operating frequency of the antenna is 2.45 GHz. The parameters used in the FDTD calculations and the properties of the biological tissues are listed in Table 5 (Duck 1990; Van Den Berg et al. 1983).

DOI 10.1007/978-981-4560-75-7_112-1

Table 5 Parameters for finite difference time domain calculations and material parameters

Parameters for FDTD calculations		
Cell size [mm] (minimum)	Δx, Δy	0.05
	Δz	1.00 (constant)
Cell size [mm] (maximum)	Δx, Δy	1.50
	Δz	1.00 (constant)
Absorbing boundary condition	Mur (first order)	
Material parameters		
	Relative permittivity (ε_r)	Conductivity (σ) [S/m]
Muscle	47.0	2.21
Bile	67.2	2.77
Catheter	2.6	0.0

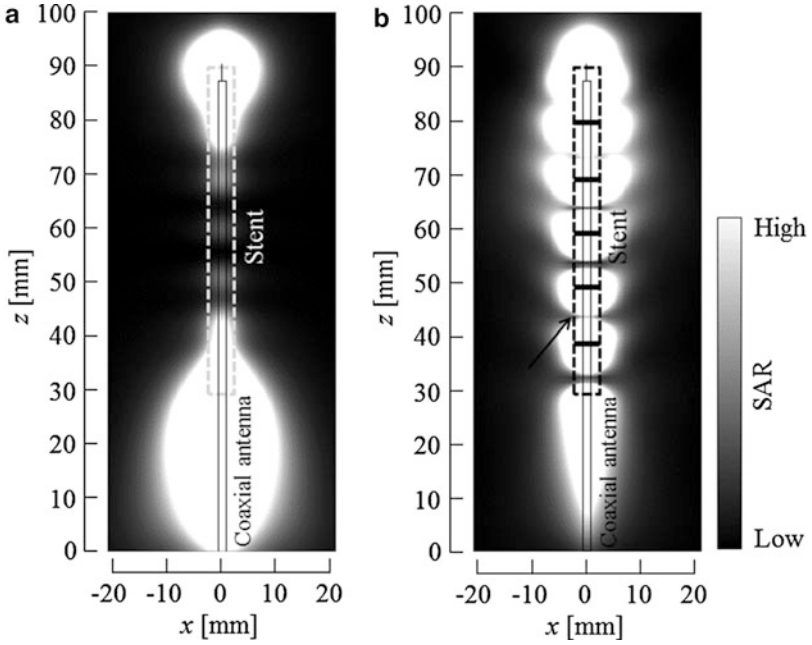


Fig. 21 Calculated specific absorption rate distributions: (**a**) conventional stent; and (**b**) proposed stent

Calculated Results

Figure 21 shows the calculated SAR distributions around the conventional and proposed metallic stent. The observation plane is defined in Fig. 20. In the conventional metallic stent, a lower SAR around the central part of the stent ($40 < z < 70$ mm) than around both edges of the stent ($10 < z < 30$, and $85 < z < 95$ mm) is observed. This is due to the shielding effect by the metallic stent. The target carcinoma should be located at near or around the central part of the stent; thus, this result means that the conventional metallic stent is unsuitable for the treatment. Besides, in the proposed metallic stent, the higher SAR region appears around the whole of the stent ($30 < z < 90$ mm). It is considered that the microwave energy can leak to the outside from the plastic parts of the stent. From the above, it can be said that the proposed stent can have significant heat around it, where the target carcinoma should exist.

DOI 10.1007/978-981-4560-75-7_112-1

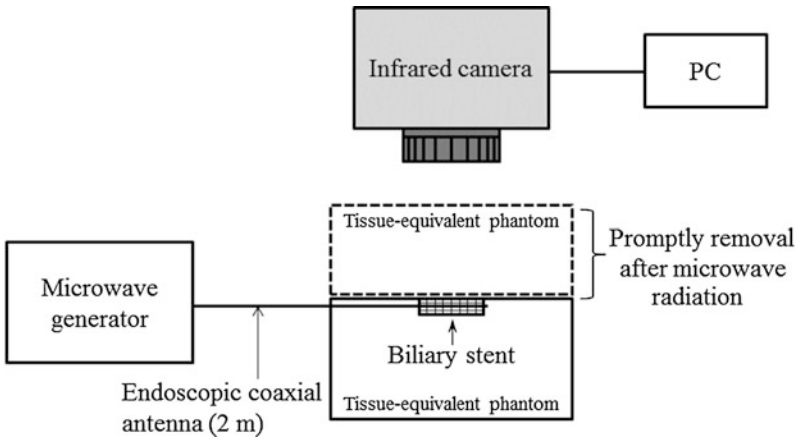


Fig. 22 Specific absorption rate measurement system

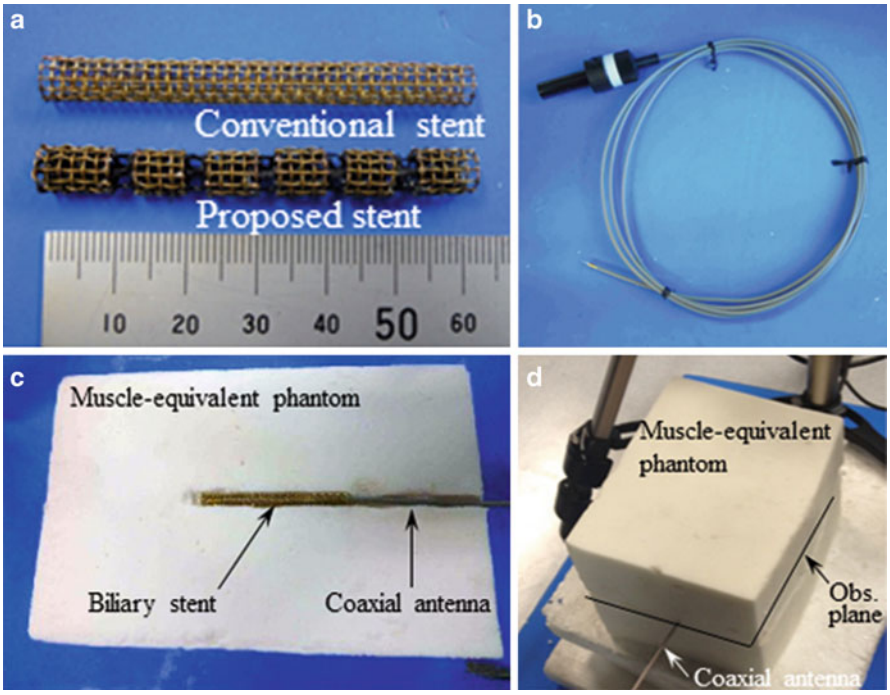


Fig. 23 Specific absorption rate measurement setup: (**a**) trial manufacture of stent; (**b**) endoscopic coaxial antenna; (**c**) cross-section of phantom; and (**d**) measurement setup

Experimental Evaluations

Figure 22 shows the experimental system used in SAR measurements. Figures 23a, 23b show trial manufacture of the biliary stent and the endoscopic coaxial antenna. First, the trial-manufactured proposed and conventional metallic stents are put between the muscle-equivalent solid phantoms (Figs. 23c, 23d). Next, the antenna passes through the stent end at 1 mm and supplies it with microwave energy. The probe is connected to a microwave generator. The input power and the radiation time for the probe are set at 50 W and 60 s, respectively. The measured temperature rise distributions by an infrared camera are converted to the SAR distributions (Okano et al. 2000). The observation plane of the experiment is the same as the calculations shown in Fig. 20.

DOI 10.1007/978-981-4560-75-7_112-1

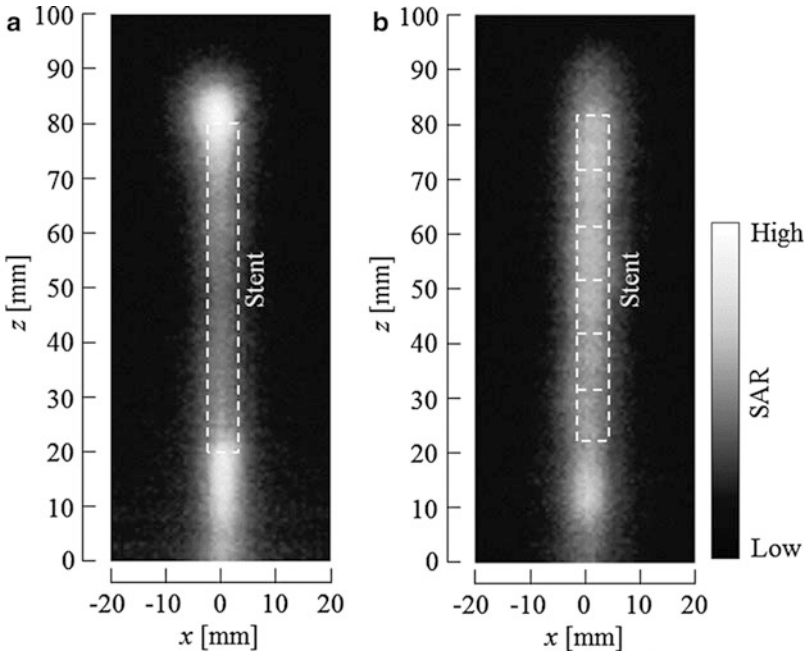


Fig. 24 Measured specific absorption rate distributions: (**a**) conventional stent; and (**b**) proposed stent

Figures 24a, 24b show the SAR distribution around the conventional and proposed metallic stent. In the conventional metallic stent, a lower SAR region around the central part of the stent ($40 < z < 70$ mm) than at both edges of the stent ($20 < z < 30$ mm, and $80 < z < 90$ mm) is observed. Besides, in the proposed stent, the higher SAR region around the whole of the proposed stent can be observed due to the microwave energy leaked to the outside from each plastic part of the stent. From these results, it is confirmed that the proposed stent could generate extensive heating regions around it in the longitudinal direction of up to 60 mm.

Summary

In this section we have introduced the heating characteristics of the microwave antenna for interstitial and intracavitary heating. Not only the development of practical devices but also the treatment protocol should be investigated in further studies.

Surgical Devices Using Microwave Energy

We have also investigated development of surgical devices using the thermal effect of microwave energy. Generally, the electrical scalpel has been widely used for surgical operations. It uses the RF current (from several hundred kHz to several MHz) and can realize both tissue coagulation and dissection. However, this device has some problems. First, it requires an external electrode which sometimes causes burn injuries around it. Moreover, the RF current may flow through an unexpected part, and causing involuntary muscle movement. In addition, since the device generates an extremely high temperature, the biological tissue is carbonized. As a result, fog is be generated. This is a serious problem, especially in laparoscopic surgery. Ultrasonically activated scalpels are also used in surgical operations. In this case, bodily fluid may be splashed by the device, causing the fog that is undesirable in laparoscopic surgery.

DOI 10.1007/978-981-4560-75-7_112-1

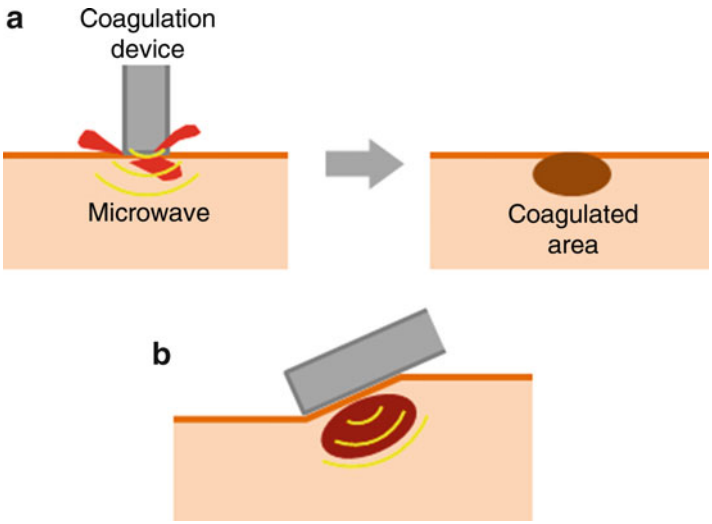


Fig. 25 Microwave tissue coagulation device: (**a**) hemostasis; and (**b**) generation of side coagulated region

It should be noted that massive bleeding during a surgical operation is a danger for the patient. In particular, laparoscopic surgery cannot be continued if there is significant bleeding because the view of the laparoscope is reduced. Therefore, hemostasis is one of the most important treatments (Fig. 25a). Saito et al. (2013b) have studied the tissue coagulation device based on a microwave small-loop antenna and confirmed its effectiveness using numerical calculations and some experiments. However, there is a problem that should be improved: the developed device can generate a coagulated region only around the bottom of it, whereas a coagulated region around the side region of the device is important in some cases (Fig. 25b). Therefore, in the following sections we introduce a tissue coagulation device for hemostasis that can generate coagulated regions around the bottom and the side of it.

Structure of the Device

In our study, a helical structure is employed for the coagulation device shown in Fig. 26. The number of helices is 11 and the pitch of the wire is 2.0 mm. Moreover, the whole helical part is covered by polytetrafluoroethylene (PTFE) material for prevention of tissue adhesion. The operating frequency of the antenna is 2.45 GHz.

Figure 27 shows the FDTD calculation models. The antenna contacts muscle tissue in two ways, and is fed by a coaxial cable. In addition, the entire device is placed in a saline solution that assumes the same properties as bodily fluid.

Heating Characteristics

The heating performance of the device is evaluated by numerical calculations and experiments using a biological tissue. The procedure of the calculation is the same as in section "Thermal Treatment of Cancer by Microwave Energy." First, SAR distribution around the antenna is calculated. After that, temperature distribution in muscle tissue is calculated by solving the bioheat transfer equation numerically. The physical properties of muscle tissue are listed in Table 6.

Figure 28 shows the calculated temperature distributions. The observation planes are a *yz*-plane of $x = -1.7$ mm for Model A and *xy*-plane of $z = -1.0$ mm for Model B. The input power and heating time are 80 W and 20 s, respectively. The maximum sizes of the coagulated region in these planes (a region more than 60 °C) are approximately 13 × 33 mm (*yz*) in Model A and 5 × 4 mm (*xy*) in Model B.

DOI 10.1007/978-981-4560-75-7_112-1

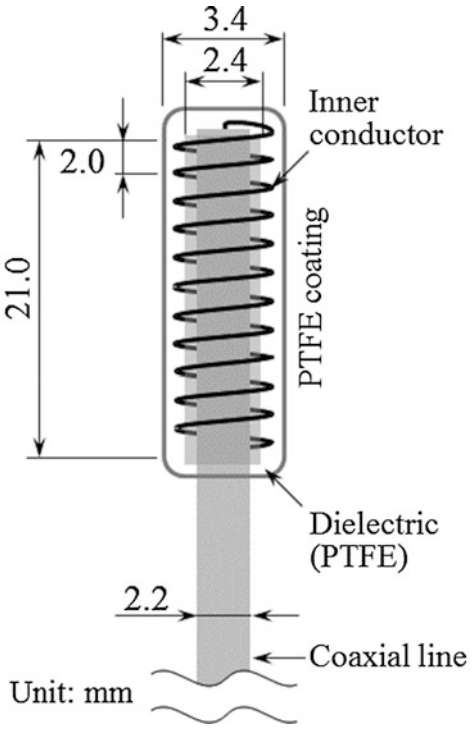


Fig. 26 Developed tissue coagulation device (around tip)

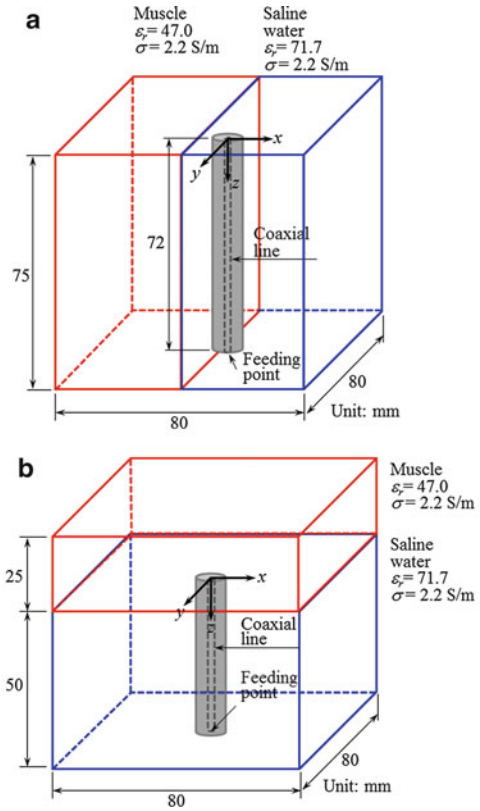


Fig. 27 Calculation models: (**a**) Model A (evaluation of side coagulated region); and (**b**) Model B (evaluation of bottom coagulated region)

DOI 10.1007/978-981-4560-75-7_112-1

Table 6 Physical properties of muscle tissue

Electrical properties (@2.45 GHz)	
Relative permittivity (ε_r)	47.0
Conductivity (σ) [S/m]	2.21
Thermal properties	
Specific heat (c) [J/kg K]	3600
Thermal conductivity (κ) [W/m K]	0.50
Density (ρ) [kg/m^3]	1020
Blood flow rate (F) [m^3/kg s]	8.3×10^{-6}
Initial temperature [°C]	37.0

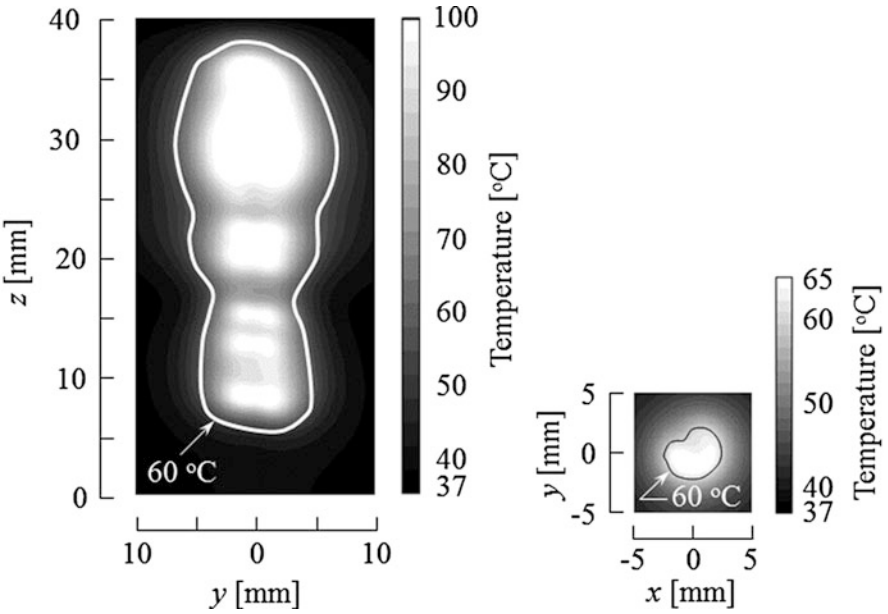


Fig. 28 Calculated temperature distributions

The validity of the device is confirmed by experiments using the removed biological tissue. Figure 29 is photograph of the experiment. Here, the radiation power from the device and radiation time are approximately 25 W and 10 s, respectively. In addition, this experiment was performed without saline water because this was a fundamental investigation.

From Fig. 29, the coagulated region can be confirmed around the side and bottom of the device. Therefore, it can be said that the developed device generates two coagulated regions using one device.

Summary

Prototype devices, which can be used for actual surgical treatment, will be developed in a further study. However, the effectiveness of the device must be confirmed in living animals before its use in actual treatments.

DOI 10.1007/978-981-4560-75-7_112-1

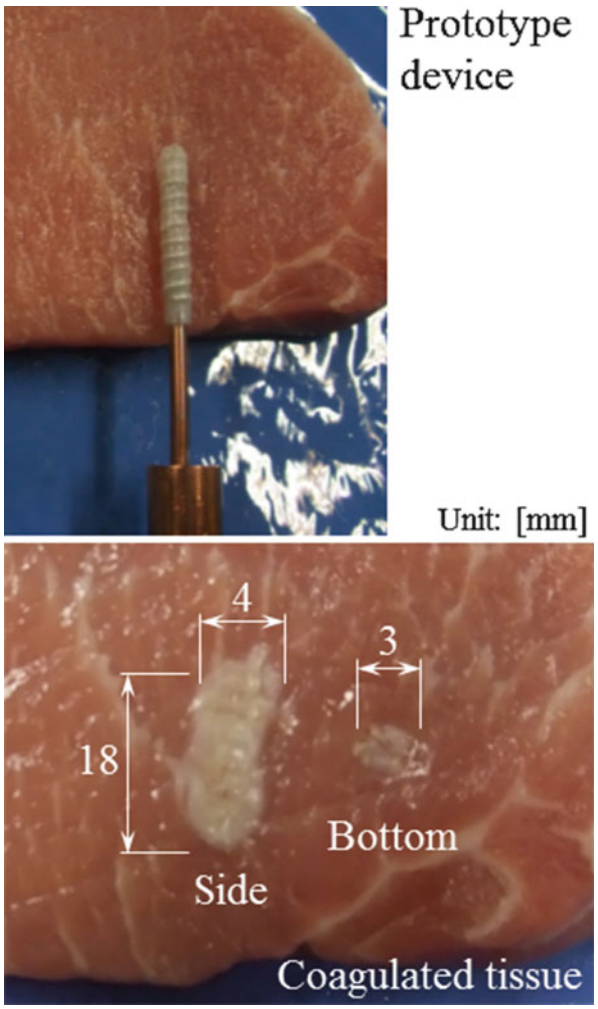


Fig. 29 Experiment on biological tissue

Cross-References

- ▶ Antennas in High-Power High-Voltage Systems and Electromagnetic Pulse Systems
- ▶ Antennas in MRI Systems
- ▶ Evaluation of Wearable and Implantable Antennas with Human Phantoms
- ▶ Implanted Antennas in Biomedical Telemetry

References

Amin RS, Nickolaidis P, Kawashima A, Kramer LA, Ernst RD (1999) Normal anatomy of the fetus at MR imaging. Radiographics 19:S201–S214

Angelone LM, Potthast A, Segonne F, Iwaki S, Belliveau JW, Bonmassar G (2004) Metallic electrodes and leads in simultaneous EEG-MRI: specific absorption rate (SAR) simulation studies. Bioelectromagnetics 25:285–295

Bond EJ, Li X, Hagness SC, Veen DV (2003) Microwave imaging via space-time beamforming for early detection of breast cancer. IEEE Trans Antennas Propag 51:1690–1705

Chen J, Feng Z, Jin JM (1998) Numerical simulation of SAR and B1-field inhomogeneity of shielded RF coils loaded with the human head. IEEE Trans Biomed Eng 45:650–659

Chin CL, Collins CM, Li S, Dardzinski BJ, Smith MB (1998) Bird cage builder, version 1.0. Center for NMR Research, Department of Radiology Pennsylvania. State University College of Medicine, Hershey

Collins CM, Liu W, Wang J, Gruetter R, Vaughan JT, Ugurbil K, Smith MB (2004) Temperature and SAR calculations for a human head within volume and surface coils at 64 and 300 MHz. J Mag Reson Imag 19:650–656

De Wilde JP, Rivers AW, Price DL (2005) A review of the current use of magnetic resonance imaging in pregnancy and safety implications for the fetus. Progress Biophysics and Molecular Biology 87:335–353

Dimbylow P (2006) Development of pregnant female, hybrid voxel-mathematical models and their application to the dosimetry of applied magnetic and electric fields at 50 Hz. Phys Med Biol 51:2383–2394

Duck FA (1990) Physical properties of tissue: a comprehensive reference book. Academic, London

Gabriel C (1996) Compilation of the dielectric properties of body tissues at RF and microwave frequencies. Armstrong Laboratory, Brooks Air Force Technical Report AL/OE-TR-1996-0037. King's College, London, Department of Physics, London

Hagness SC, Taflove A, Bridges JE (1999) Three-dimensional FDTD analysis of a pulsed microwave confocal system for breast cancer detection: design of an antenna-array element. IEEE Trans Antennas Propag 47:783–791

Hand JW, Li Y, Thomas EL, Rutherford MA, Hajnal JV (2006) Prediction of specific absorption rate calculations in mother and fetus associated with MRI examinations during pregnancy. Magn Reson Med 55:883–893

Hubbald AM, Adzick NS, Crombleholme TM, Coleman BG, Howel LJ, Haseigrove JC et al (1999) Congenital chest lesions: diagnosis and characterization with prenatal MR imaging. Radiology 212:43–48

International Electrotechnical Commission (2002) International standard, medical equipment-part 2–33: particular requirements for the safety of magnetic resonance equipment for medical diagnosis, 2nd revision. International Electrotechnical Commission 60601-2-33, Geneva, pp 29–31

Itakura H, Saito K, Takahashi M, Ito K (2012) Development of biliary stent compatible with microwave hyperthermia for bile duct carcinoma. Thermal Medicine 28:43–51

Ito K (1996) Medical applications of microwave. In: Proceedings of 1996 Asia-Pacific microwave conference, vol 1, New Delhi, pp 257–260

Ito K (2006) Antenna technology applied to medicine. Trans IEICE B J89-B:1558–1568 [in Japanese]

Jin J (1998) Electromagnetic analysis and design in magnetic resonance imaging. CRC Press, Boca Raton

Johnson CC, Guy AW (1972) Nonionizing electromagnetic wave effects in biological materials and systems. Proc IEEE 60:692–718

Kikuchi S, Saito K, Takahashi M, Ito K (2006) Control of heating pattern for interstitial microwave hyperthermia by a coaxial-dipole antenna – aiming at treatment of brain tumor. Trans IEICE B J89-B:1486–1492

Kikuchi S, Saito K, Takahashi M, Ito K, Ikehira H (2007) SAR calculations in various parts of pregnant woman during MR imaging. In: Proceedings of the 2007 international symposium on antennas and propagation, Niigata, 20–24 Aug 2007, pp 434–437. http://i-scover.ieice.org/proceedings/ISAP/2007/pdf/2C4-4.pdf. Accessed 28 Jul 2015.

DOI 10.1007/978-981-4560-75-7_112-1

Kikuchi S, Saito K, Takahashi M, Ito K, Ikehira H (2009) SAR computation inside fetus by RF coil during MR imaging employing realistic numerical pregnant woman model. IEICE Trans Commun E92-B:431–439

Levine D (2004) Fetal magnetic resonance imaging. J Matern Fetal Neonatal Med 15:85–94

Lin JC, Wang YJ (1987) Interstitial microwave antennas for thermal therapy. Int J Hyperthermia 3:37–47

Loew BJ, Howell DA, Sanders MK (2009) Comparative performance of uncoated, self expandable metallic biliary stents of different designs in 2 diameters: final results of an international multicenter, randomized controlled trial. Gastrointest Endosc 70:445–453

Nagaoka T, Watanabe S, Sakurai K, Kunieda E, Watanabe S, Taki M et al (2004) Development of realistic high-resolution whole-body voxel models of Japanese adult males and females of average height and weight, and application of models to radio-frequency electromagnetic-field dosimetry. Phys Med Biol 49:1–15

Nagaoka T, Togashi T, Saito K, Takahashi M, Ito K, Watanabe S (2007) An anatomically realistic whole-body pregnant-woman model and specific absorption rates for pregnant-woman exposure to electromagnetic plane wave from 10 MHz to 2 GHz. J Phys Med Biol 52:6731–6745

Nguyen UD, Brown JS, Chang IA, Krycia J, Mirotznik MS (2004) Numerical evaluation of heating of the human head due to magnetic resonance imaging. IEEE Trans Biomed Eng 51:1301–1309

Ochi H, Yamamoto E, Sawaya K, Adachi S (1993) Analysis of MRI antenna inside an RF shield. IEICE Trans Commun J 76-B-II:79–85

Okano Y, Ito K, Ida I, Takahashi M (2000) The SAR evaluation method by a combination of thermographic experiments and biological tissue-equivalent phantom. IEEE Trans Microw Theory Tech 48:2094–2103

Penns HH (1948) Analysis of tissue and arterial blood temperatures in the resting human forearm. J Appl Phys 1:93–122

Püspök A, Lomoschitz F, Dejac C, Hejna M, Sautner T, Gangl A (2005) Endoscopic ultrasound guided therapy of benign and malignant biliary obstruction: a case series. Am J Gastroenterol 100:1743–1747

Saito K, Yoshimura H, Ito K, Aoyagi Y, Horita H (2004) Clinical trials of interstitial microwave hyperthermia by use of coaxial-slot antenna with two slots. IEEE Trans Microw Theory Tech 52:1987–1991

Saito K, Kikuchi S, Takahashi M, Ito K, Ikehira H (2006a) SAR distributions in the abdomen of a pregnant woman generated in a birdcage coil for the MRI system. In: Proceedings of the first European conference on antennas and propagation, 2006 [CD-ROM], Nice, 6–10 Nov 2006

Saito K, Amano T, Takahashi M, Ito K, Ikehira H (2006b) Evaluation of the SAR by two types of RF coils for MRI system [abstract]. In: Abstract Book of the Bioelectromagnetics Society 28th annual meeting, Cancun. Curran Associates, Red Hook, pp 247–250

Saito K, Hiroe A, Kikuchi S, Takahashi M, Ito K (2006c) Estimation of heating performances of a coaxial-slot antenna with endoscope for treatment of bile duct carcinoma. IEEE Trans Microw Theory Tech 54:3443–3449

Saito K, Kawamura T, Takahashi M, Ito K (2008) Generation of controllable heating patterns by two types of thin microwave antennas for interstitial microwave thermal therapy. In: Proceedings of the 2008 international symposium on antennas and propagation (ISAP 2008), Taipei, Oct 2008 [CD-ROM]

Saito K, Tsubouchi K, Takahashi M, Ito K (2010) Practical evaluations on heating characteristics of thin microwave antenna for intracavitary heating. In: Proceedings of the 32nd international conference of the IEEE/EMBS, Buenos Aires, 31 Aug–4 Sep 2010, pp 2755–2758

Saito K, Takahashi M, Ito K (2013a) Generation of controllable heating patterns for interstitial microwave hyperthermia by coaxial-dipole antennas. IEICE Trans Electron E96-C(9):1178–1183

DOI 10.1007/978-981-4560-75-7_112-1

Saito K, Ito K, Ishikawa S, Inoue M, Takahashi M (2013b) Development of surgical devices using microwave energy. In: 2013 7th European conference on antennas and propagation (EuCAP), 8–12 Apr 2013, Gothenberg, pp 2813–2815

Seegenschmiedt MH, Fessenden P, Vernon CC (eds) (1995) Thermoradiotherapy and thermochemotherapy. Springer, Berlin

Shimizu K, Matsuda S, Saito I, Yamamoto K, Hatsuda T (1995) Application of biotelemetry technique for advanced emergency radio system. IEICE Trans Commun E-78B:818–825

Simunic D, Wach P, Renhart W, Stollberger R (1996) Spatial distribution of high-frequency electromagnetic energy in human head during MRI: numerical results and measurements. IEEE Trans Biomed Eng 43:88–94

Sterzer F (2002) Microwave medical devices. IEEE Microwave Mag 3:65–70

Togashi T, Nagaoka T, Kikuchi S, Saito K, Watanabe S, Takahashi M et al (2008) FDTD calculations of specific absorption rate in fetus caused by electromagnetic waves from mobile radio terminal using pregnant woman model. IEEE Trans Microw Theory Tech 56:554–559

Tsukiyama I, Tanaka Y, Yamashita T, Nakazawa M, Terashima H, Uehara S et al (1991) Overview of heating devices by electromagnetic wave. In: Matsuda T (ed) Hyperthermia manual. Magbros, Tokyo, pp 19–42 [in Japanese]

Van Den Berg PM, De Hoop AT, Segal A, Praagmen N (1983) A computational model of the electromagnetic heating of biological tissue with application to hyperthermic cancer therapy. IEEE Trans Biomed Eng BME-30:797–805

Wu D, Shamsi S, Chen J, Kainz W (2006) Evaluations of specific absorption rate and temperature increase within pregnant female models in magnetic resonance imaging birdcage coils. IEEE Trans Microw Theory Tech 54:4472–4478

Yee KS (1966) Numerical solution of initial boundary values problems involving Maxwell's equations in isotropic media. IEEE Trans Antennas Propag AP-14:302–307

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_113-1

Holographic Antennas

Christian Rusch*
Institut für Hochfrequenztechnik und Elektronik, Karlsruhe Institute of Technology, Karlsruhe, Germany

Abstract

The holographic theory known from optics can also be used to describe the functionality of a special kind of leaky-wave antennas. Within the so-called holographic antenna, a hologram builds the radiating aperture, which is fed by surface-wave modes traveling on thin substrates. The hologram can be described as the interference pattern of the superposition of the traveling surface wave and the radiated plane wave. Therefore, it is possible to control the beam direction and beam shape of the holographic antenna by a modification of the hologram form. Compared to other kinds of leaky-wave antennas, the holographic antennas have also advantages in manufacturing and system integration, which make them to be a very promising antenna type for different millimeter-wave applications, e.g., radar systems.

Keywords

Holographic antenna; Leaky-wave antenna; Traveling-wave antenna; Beam scanning; Millimeter-wave antenna; Millimeter-wave radar

Introduction

Mostly known leaky-wave antennas are slotted waveguides or similar antennas realized as substrate-integrated waveguides (SIW). For high frequencies the manufacturing of waveguides is expensive and complex. SIW technology is easier to manufacture, also at low millimeter-wave frequencies, but at sub-THz range the necessary via distance to guide the waveguide mode with low leakage through the sidewalls should not be more than a few micrometers. This recommendation cannot be reached by commercial printed-circuit-board (PCB) processes, which are available today.

Holographic antennas do not need vias and can be realized with different kinds of technologies as etching processes or with thin-film process. Therefore, a manufacturing also for systems working at D or J band is possible. The idea of the holographic antenna was already published in 1968 by P. Checcacci. Due to the low aperture efficiency of that antenna type and the used difficult feeding by an external horn antenna then, the research was nearly stopped for many years. Just in the last century, some groups worked again on this field as they also see the holographic antenna due to the low complexity and the good integration skills in spite of the large aperture as promising candidate for millimeter-wave radars.

Within this chapter the theory of the holographic antenna with its analogy to the optical holography is explained and recommendations for the antenna design are made.

Concrete designs with feeding structures for creating the traveling surface-wave mode are presented. A novel combination with a feeding array also allows the beam scanning within two dimensions. The millimeter-wave radar with integrated holographic antenna is presented at the end of this chapter as an application that extremely benefits from the integration properties and the beam-scanning functionality.

*Email: christian.rusch@kit.edu

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_113-1

Theory

The holographic antennas belong to the group of leaky-wave antennas. Their distinctive feature is that the theory from the optical holography is used to calculate the geometry of the radiating aperture in contrast to other antenna types described in (see chapter on Jackson, "► Traveling-Wave Antennas (Including Leaky Wave and Surface Wave Antennas"), e.g., the slotted waveguide antennas.

After describing the theory of optical holography, this theory is transferred to the antenna design process in section "Using the Holographic Theory for Antenna Designs." Due to the usage of surface-wave modes in the context with holographic antennas, their propagation is theoretically explained in section "Propagating Surface Waves as Reference Waves." Also, possibilities creating these modes are shown in the following section. Due to the fact that holographic antennas are kind of periodic leaky-wave antennas, the relation to this group is pointed out in section "Properties of Leaky-Wave Antennas."

A Short Description of Optical Holography

The optical holography was invented and described the first time in 1948 from the Hungarian engineer Dénes Gábor. This method is used to record the spatial image of an object and allows the three-dimensional reconstruction (Gabor 1948). The term "holography" includes the Greek designations "holos" and "graphein." A translation of these Greek terms means "complete image." This completeness is achieved by recording not only intensity values but also values, which describe the different heights of the object. An explanation follows with a comparison between photography and holography.

During the recording process of photography, an image of the object is built on a photographic film by using an objective. An illumination of the object is necessary. The intensity values of the light, which is reflected by the object, are recorded on the film or today's usually used light sensor. The phase information of the light waves is not recorded and therefore is lost by this method. For the observer results a two-dimensional image of the object.

The method of holography consists of the two-process steps "recording" and "reconstruction." During the recording, the object and the photo film are illuminated with a coherent light source as shown in Fig. 1 on the left. A non-flat object will produce reflected waves with phase variations among each other depending on the positions of the individual reflecting points on the object's surface. These so-called object waves with different phase values interfere with the directly illuminating reference wave (plane wave) on the photographic film. The resulting interference pattern is called "hologram" and contains both the intensity information as brightness contrast (similar to the photography) and the height information (phase) as distance variation between the interference lines.

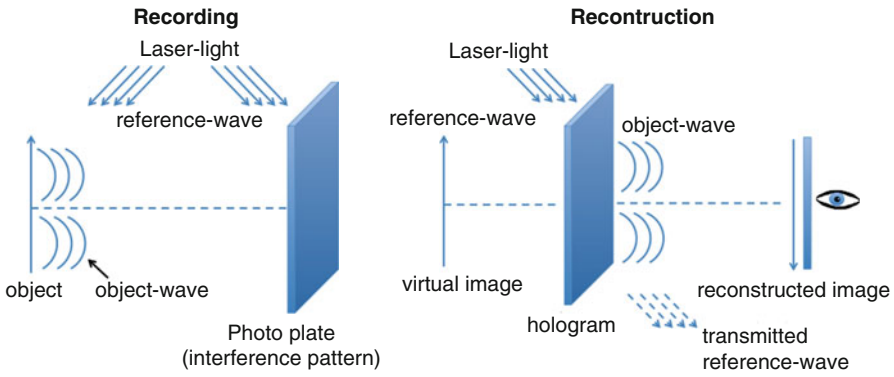


Fig. 1 Process steps of the optical holography

DOI 10.1007/978-981-4560-75-7_113-1

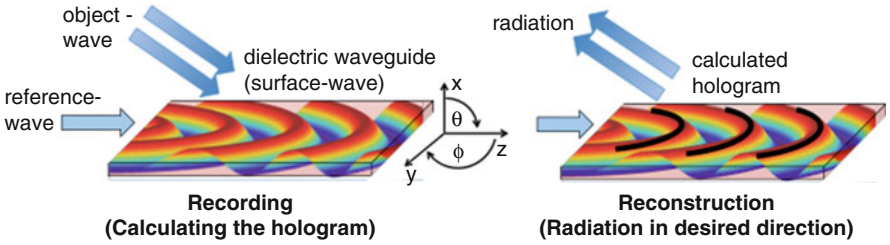


Fig. 2 Schematic illustration of the holographic antenna concept

Illuminating the hologram with the reference light source, which was also used during the recording, does the following reconstruction of the object. The superposition of the light with the interference pattern (hologram) creates a three-dimensional image of the recorded object (see. Fig. 1, right side). This theoretical approach can be used to describe the function of special types of leaky-wave antennas.

Using the Holographic Theory for Antenna Designs

For holographic antennas, the explained theory of the optical holography process is used to calculate the geometry of the radiating aperture. As a photographic film cannot be used as radiating aperture, the recording process is only a theoretical approach to calculate the necessary interference pattern. A schematic representation of the two-process steps transferred to the antenna design is shown in Fig. 2.

The holographic antenna consists of an antenna substrate, which is used as dielectric waveguide for a surface-wave mode propagating in z-direction with its wave number β_0. In relation to the optical holography, this surface wave builds the reference wave. An excitation of this mode type can be done by illuminating the substrate edge with horn antennas (Iizuka et al. 1975; Sooriyadevan et al. 2007; ElSherbiny et al. 2004) or by implementing planar end-fire antennas as surface-wave launchers directly on the substrate surface (Podilchak et al. 2008; Hammad et al. 2003). Within an area with small distance to the source, a traveling wave with circular phase front arises. By using a point source within the substrate, the electric field of the propagating surface wave in a lossless medium can be described as follows (for Cartesian coordinates):

$$\vec{E}_{\mathrm{ref}_{z,\,y}} = E_{\mathrm{ref}_r} \cdot e^{-j\beta_0 r} \tag{1}$$

with

$$z = r \cdot \cos(\phi)$$

$$y = r \cdot \sin(\phi)$$

The surface wave propagates from the source spherically and uniformly within the substrate plane (see. Fig. 2, left side). The parameter r describes the radial distance of the wave to the source. $\vec{E}_{\mathrm{ref}_z}$ and $\vec{E}_{\mathrm{ref}_y}$ are the electric-field components. For ideal circular propagation the amplitude and phase of the electric field are constant at a radial distance to the source. The amplitude at distance r to the surface-wave launcher is called E_{ref_r}.

DOI 10.1007/978-981-4560-75-7_113-1

The radiated TEM wave is defined as object wave, but for calculating the hologram a superposition of this TEM object wave with the reference wave on the substrate surface is assumed. The direction of the incoming object wave, which determines the resulting main-lobe direction of the antenna, is defined by the angles θ_0 and ϕ_0 (see Fig. 2). The electric-field components illuminating the substrate surface ($x = 0$) are described as a function of the angular direction.

$$\vec{E}_{\mathrm{obj}_{z,y}} = E_{\mathrm{obj}_{z,y}} \cdot e^{jk_0 \cdot \sin(\theta_0) \cdot \cos(\phi_0) \cdot z} \cdot e^{jk_0 \cdot \sin(\phi_0) \cdot y} \tag{2}$$

$E_{\mathrm{obj}z,y}$ describes the amplitude of the field components and k_0 is the propagation constant of the TEM wave in free space.

Samples for calculated phase fronts of a reference wave propagating as surface wave within the substrate and of an object wave with $\theta_0 = -45°$ and $\phi_0 = 0°$ at the substrate surface are shown in Fig. 3a, b. For the calculation of the interference pattern, it is assumed that the object wave is not radiated but illuminates the substrate surface and builds a superposition with the circular reference wave. Figure 3c shows the resulting interference pattern. The cylindrical geometry is a result of the spherical phase front of the propagating reference wave close to the surface-wave launcher. To build the hologram on the substrate surface, metallic stripes are used. These stripes shorten the electric-field lines of the surface-wave mode at their positions and therefore build the roots of the interference pattern. Therefore, the created hologram in Fig. 3d builds an interference pattern with reduced phase information. By using this method the remaining phase information is lost.

Nevertheless the feeding of this type of hologram with the reference wave shows a radiation with the previously defined main-lobe direction.

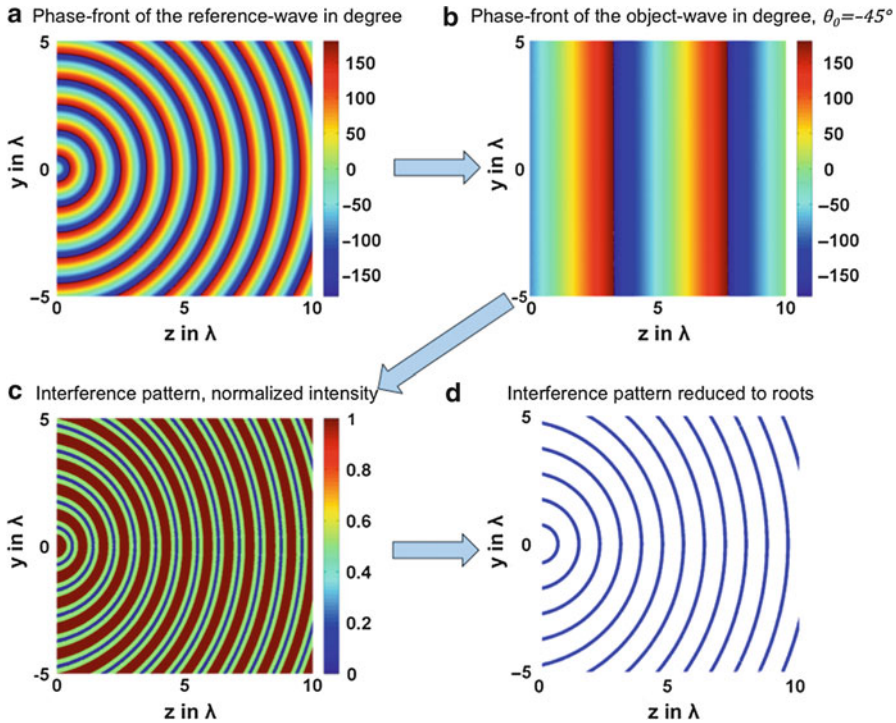


Fig. 3 Creating the hologram geometry by the superposition of reference and object wave

DOI 10.1007/978-981-4560-75-7_113-1

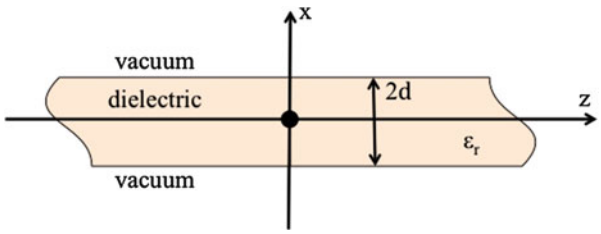


Fig. 4 Definition of the antenna substrate with propagating surface waves

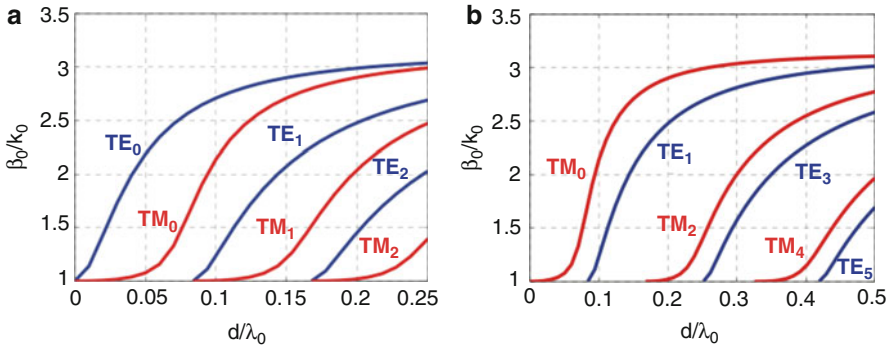


Fig. 5 Mode charts for surface-wave modes on alumina substrates ($\boldsymbol{\varepsilon_r} = 9.9$) without metallization plane (**a**) and with metallization plane (**b**)

Propagating Surface Waves as Reference Waves

The easy way of integrating holographic antennas to millimeter-wave systems is a result of using surface-wave modes to feed the hologram as radiating aperture. For calculating the hologram geometry of a predefined radiation characteristic, the knowledge about the wave number β_0 of the propagating mode is necessary. To calculate β_0 of the different propagable modes on a substrate as shown in Fig. 4 and plotting mode charts as shown in Fig. 5 following assumptions are made.

- The substrate is infinitely extended in y- and z-direction.
- The substrate is lossless.
- The surface-wave propagation direction is $+z \rightarrow e^{-j\beta_0 z}$.
- All electromagnetic fields in y-direction are constant $\left(\frac{\delta}{\delta y} = 0\right)$.
- Continuity of the tangential field components at the air-dielectric transition → no inconstancy of the fields at the boundary.

As described in (Beer 2013), the following four transversal field components for the symmetrical case depending on the longitudinal components exist:

$$E_x\left(\frac{\omega^2}{c^2} - \beta_0^2\right) = -j\beta_0 \frac{\partial E_z}{\partial x} \tag{3}$$

DOI 10.1007/978-981-4560-75-7_113-1

Table 1 Mode equations of the TE and TM surface waves

	Mode equation	Cutoff frequency
Even TM with $n = 0, 2, 4,\ldots$	$\frac{k_c}{k_0}\tan\left(\frac{k_c}{k_0}2\pi\frac{d}{\lambda_0}\right) = \varepsilon_r\sqrt{(\varepsilon_r-1)} - \frac{k_c^2}{k_0^2}$	$f_c = \frac{nc_0}{n4d\sqrt{\varepsilon_r-1}}$
Odd TM with $n = 1, 3, 5,\ldots$	$-\frac{k_c}{k_0}\cot\left(\frac{k_c}{k_0}2\pi\frac{d}{\lambda_0}\right) = \varepsilon_r\sqrt{(\varepsilon_r-1)} - \frac{k_c^2}{k_0^2}$	$f_c = \frac{nc_0}{n4d\sqrt{\varepsilon_r-1}}$
Even TE with $n = 0, 2, 4,\ldots$	$\frac{k_c}{k_0}\tan\left(\frac{k_c}{k_0}2\pi\frac{d}{\lambda_0}\right) = \sqrt{(\varepsilon_r-1)} - \frac{k_c^2}{k_0^2}$	$f_c = \frac{nc_0}{n4d\sqrt{\varepsilon_r-1}}$
Odd TE with $n = 1, 3, 5,\ldots$	$-\frac{k_c}{k_0}\cot\left(\frac{k_c}{k_0}2\pi\frac{d}{\lambda_0}\right) = \sqrt{(\varepsilon_r-1)} - \frac{k_c^2}{k_0^2}$	$f_c = \frac{nc_0}{n4d\sqrt{\varepsilon_r-1}}$

$$E_y\left(\frac{\omega^2}{c^2} - \beta_0^2\right) = j\omega\mu_0\mu_r\frac{\partial H_z}{\partial x} \tag{4}$$

$$H_x\left(\frac{\omega^2}{c^2} - \beta_0^2\right) = -j\beta_0\frac{\partial H_z}{\partial x} \tag{5}$$

$$H_y\left(\frac{\omega^2}{c^2} - \beta_0^2\right) = -j\omega\varepsilon_0\varepsilon_r\frac{\partial E_z}{\partial x} \tag{6}$$

Due to the inhomogeneous medium perpendicular to the z-direction, the propagation of transversal-electromagnetic waves (TEM waves) is not possible. Therefore, the propagable TE and TM modes can be described with the general field components (see Eqs. 3, 4, 5, and 6). The results are summarized in Table 1. The parameters k_c and h build the cutoff wave numbers for the regions within the substrate or vacuum:

$$k_c^2 = \varepsilon_r k_0^2 - \beta_0^2 \text{ for } 0 \le |x| \le d$$

$$h^2 = \beta_0^2 - k_0^2 \text{ for } d \le |x| \le \infty$$

The equations for the single modes in Table 1 can be solved numerically to plot the mode diagrams and to determine the wave number β_0 of the propagating wave. It can be seen that the propagation of the modes TE_0 and TM_0 is already possible on very thin substrates, because there is no cutoff frequency for these two modes. By adding a continuous metallization on one substrate surface (top or bottom layer), the propagation of TE modes with even mode numbers ($n = 0, 2, 4,\ldots$) and TM modes with odd mode numbers ($n = 1, 3, 5,\ldots$) is suppressed. The reason is that the electrical field lines of these modes would be parallel to the metallization.

TE_0 on substrates without metallization plane and TM_0 modes on substrates with metallization are the two modes mostly used as reference waves for holographic antennas, because high mode purity can be guaranteed. The propagation of further modes could create other unwanted radiating main lobes.

Surface-Wave Launchers

During the history of holographic antennas, several ways to create the desired surface-wave mode have been developed. Today especially the planar feeds directly printed on circuit boards are interesting, because they allow the design of compact and easy-to-integrate holographic antennas up to very high frequencies. Since the hologram builds the radiating aperture, the influence of the surface-wave launcher on the radiation pattern is low, as long as the launcher works stable within the desired frequency range and

DOI 10.1007/978-981-4560-75-7_113-1

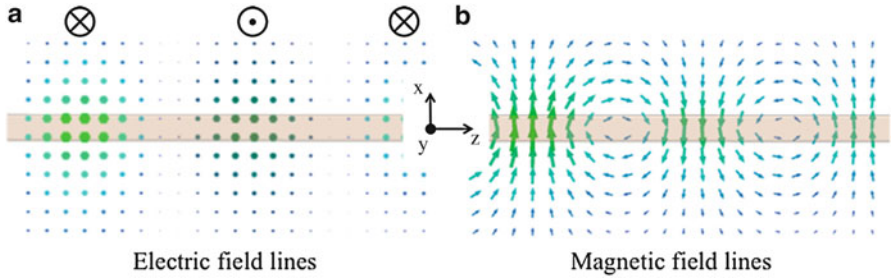


Fig. 6 Field distribution of a TE_0 surface wave propagating in $+z$-direction

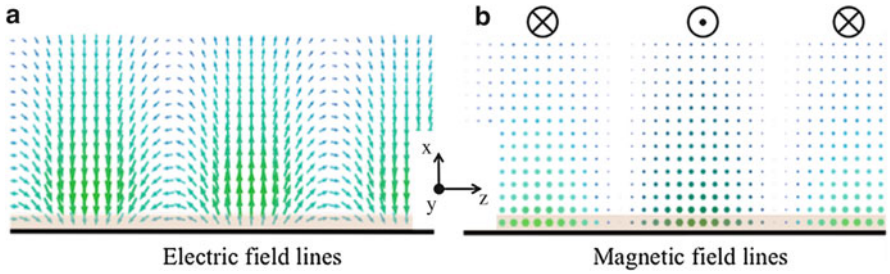


Fig. 7 Field distribution of a TM_0 surface wave propagating in $+z$-direction

illuminates the complete hologram area. There are different possibilities to create the desired mode and because of the low effect to the radiation pattern, the choice can be made on other properties, e.g., the integration level or the feeding line technology of the surface-wave launcher.

A view to the field lines of the two modes, TE_0 on substrates without metallization plane (Fig. 6) and TM_0 on substrates with metallization on one substrate side (Fig. 7), helps to investigate possible surface-wave launchers. With the electric-field perpendicular and the magnetic field lines parallel to the propagation direction, the field distribution of TE_0 waves is similar to a slot-line mode. For this reason tapered slot antennas are advisable to be used as surface-wave launchers for TE_0 modes on substrates without metallization. Different variations of these surface-wave launchers are possible, e.g., the planar Vivaldi antenna with feeding slot line or an antipodal version with microstrip feeding line (Bai et al. 2011).

For TM_0 reference waves Podilchak published a slotted quasi-Yagi-Uda feed (Podilchak et al. 2009). Another possibility is an integrated horn antenna realized in SIW technology. The disadvantage is the need of a high number of vias in small distance to each other, which means high requirements for the manufacturing at high frequencies.

An overview of different surface-wave launchers will follow (see chapter on Rusch, "► Holographic Antennas", section "Propagating Surface Waves as Reference Waves.").

Properties of Leaky-Wave Antennas

After designing the hologram with the calculated β_0 of the propagating reference wave and the desired main-lobe direction at a certain frequency, the antenna's radiation characteristic can be described with leaky-wave antenna theory (see chapter on Jackson, "► Traveling-Wave Antennas (Including Leaky Wave and Surface Wave Antennas"). Figure 8 shows the longitudinal section of the antenna. The surface-wave launcher creates the reference wave, which feeds the periodic single elements of the hologram serially. Due to this point of view, the antenna can be described as periodic leaky-wave antenna. For this

DOI 10.1007/978-981-4560-75-7_113-1

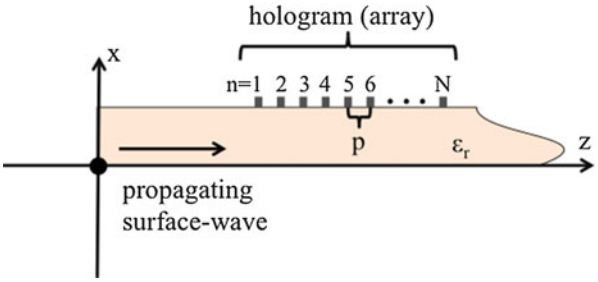


Fig. 8 Side view of the holographic antenna

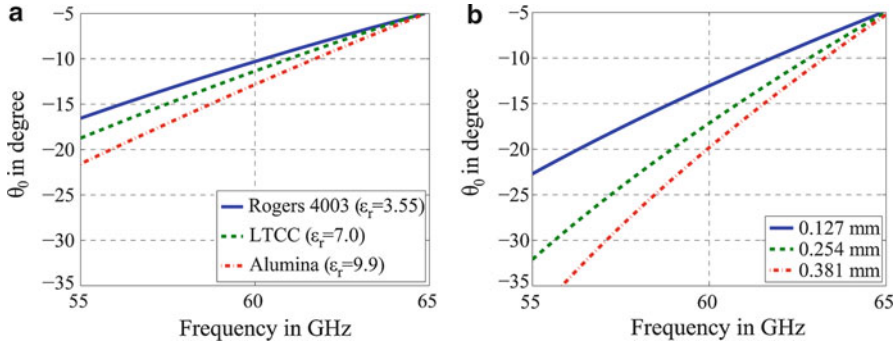


Fig. 9 Influence of material properties on the frequency-dependent scanning range

antenna type, radiation is done by the $n = -1$ spatial harmonic (Oliner 1993), due to the fact that the propagating surface-wave mode with $n = 0$ is a "slow wave" and therefore non-radiating.

With the known β_0 of the slow reference wave and the periodicity p of the hologram, the wave number of the different space harmonics can be calculated with the following equation:

$$\beta_n = \beta_0 + \frac{2\pi}{p} n. \tag{7}$$

The requirement for the radiation of a space harmonic is given by Eq. 8:

$$-k_0 < \beta_n < +k_0. \tag{8}$$

Here k_0 is the propagation constant of a TEM wave in free space. If this condition is met for a space harmonic, this wave is called "fast wave." The main-lobe direction of the antenna using a fast wave with β_n is given by Eq. 9:

$$\sin(\theta_0) \approx \frac{\beta_n}{k_0} \tag{9}$$

With Eq. 9 and the β_0 calculation from (see chapter on Rusch, "▶ Holographic Antennas," section "Propagating Surface Waves as Reference Waves"), the frequency-dependent main-lobe direction can be calculated. Figure 9 shows the scanning range of the antenna for different materials and depending on the substrate thickness. This results show that thick substrates with a high dielectric constant improve the scanning range of the antenna.

DOI 10.1007/978-981-4560-75-7_113-1

Design

With the theory from (see chapter on Rusch, "► Holographic Antennas," section "Theory"), the hologram shape for a holographic antenna radiating in a desired direction can be calculated. Also the effect of the material properties on the frequency-dependent scanning range was explained. Now we will see how the holographic antenna can be integrated within an RF front end. Also the usage of a surface-wave launcher array to build a 2D-scanning holographic antenna will be described. In the end of this chapter, two example designs of 1D- and 2D-scanning holographic antennas are presented.

Multilayer-Substrate Structure

The scanning range of holographic antennas with feeding surface waves is improved by thick antenna substrates with high dielectric constants. To enable interconnects to a front end with transmission-line technology, planar surface-wave launchers are used. Within a front end with signals transmitted via transmission lines (e.g., microstrip lines), the propagation of surface waves can decrease the efficiency of the transmission (Collier and White 1976). Therefore, at this area of the front end, it is advantageous to use thinner substrates to avoid the surface-wave propagation.

To avoid this, a trade-off in (Rusch et al. 2013a) a multilayer substrate is used for the radar front end with integrated holographic antenna (see Fig. 10). The "low-temperature cofired ceramic (LTCC)" has a complete thickness of 0.42 mm and a dielectric constant of $\varepsilon_r = 7.1$. Below the radar circuit, which is realized in microstrip technology, the ground plane (GND) is on the second metal layer. Due to this effective thin substrate (0.127 mm), the propagation of surface waves is avoided within this area and the microstrip lines have a high transmission factor. The GND plane is interrupted close to the surface wave launcher. The effective substrate thickness below the hologram is therefore 0.42 mm. With that value the propagation of surface waves is supported and the frequency-dependent scanning range is improved (see chapter on Rusch, "► Holographic Antennas," section "Properties of Leaky-Wave Antennas"). The simulated gain of this antenna is shown in Fig. 11. Within the frequency range of 55–65 GHz, the main-lobe direction changes by 41°.

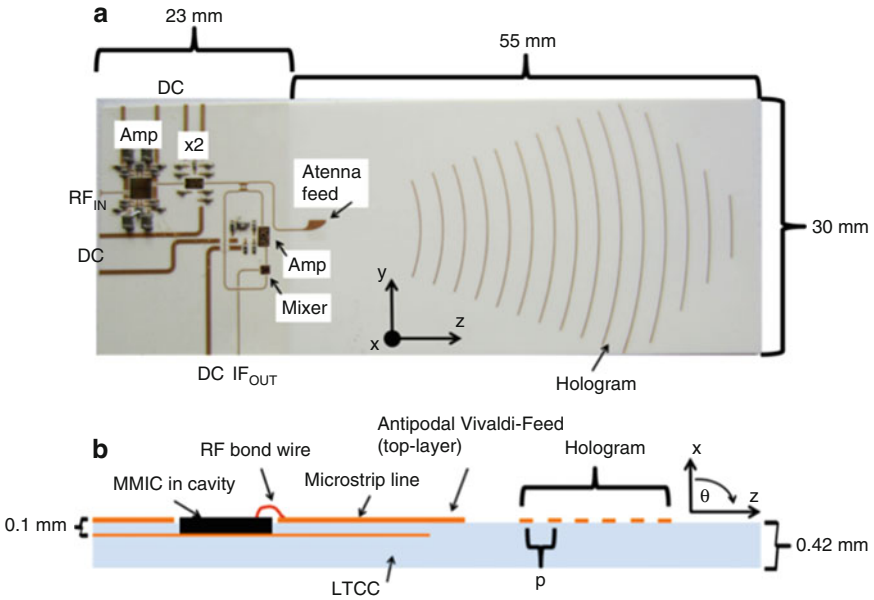


Fig. 10 Radar front end with integrated holographic antenna in LTCC

DOI 10.1007/978-981-4560-75-7_113-1

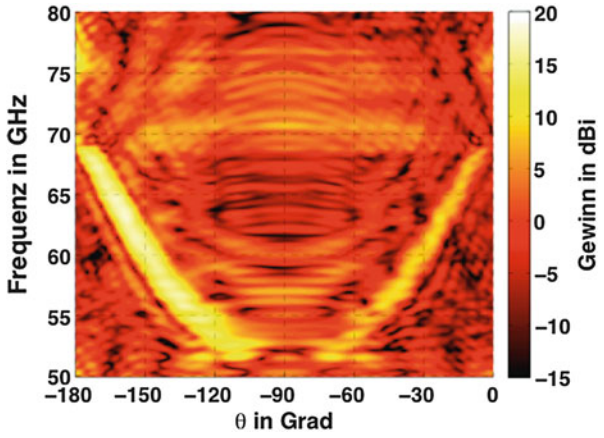


Fig. 11 Simulated gain of the LTCC antenna from Fig. 10

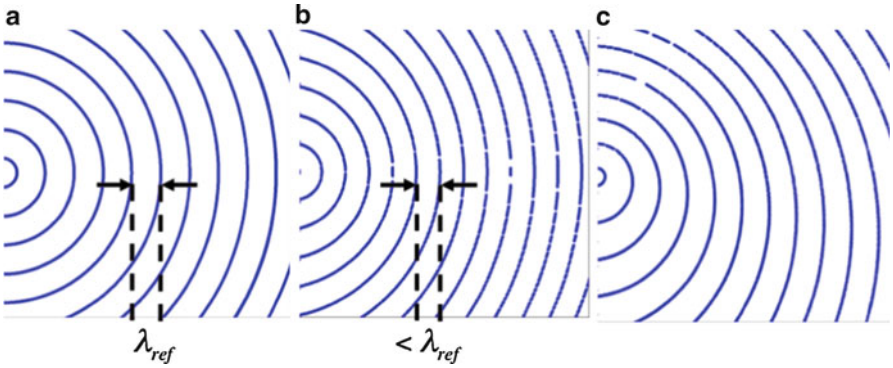


Fig. 12 Calculated holograms for different main-lobe directions. (**a**) $\theta_0 = 0°$, $\phi_0 = 0°$; (**b**) $\theta_0 = -45°$, $\phi_0 = 0°$; (**c**) $\theta_0 = -45°$, $\phi_0 = -40°$

Circular Holograms with Single Surface-Wave Launchers

Due to the holographic theory and the planar phase front of the radiated object wave, the hologram shape represents the phase front of the propagating reference wave. Due to the circular phase front created by a single surface-wave launcher, the single radiating elements are concentric semicircles with distance p to each other. The main lobe radiated within in the plane perpendicular to the substrate and the radiation angle depends on the distance p and the frequency. By changing the periodicity of the hologram, the radiation angle at a desired frequency can be changed (see Fig. 12a, b). Changing the main-lobe direction also within the second plane (xy) results in a hologram shape shown Fig. 12c. The circles are not concentric for this radiation characteristic. Therefore, controlling the main-lobe direction within two planes can only be done by an interchange of the hologram. One possibility is published in Rusch et al. (2013b). Here, the hologram is manufactured on a thin polyamide foil. The planar feed is realized directly on an empty ceramic substrate. By putting the foil on top of the substrate, a feeding of the hologram is possible and radiation in the desired direction is created. Since the foil is not fixed on top of the substrate, an interchange of the hologram allows changing the main-lobe direction of the antenna. The interchange of the hologram may be done by a mechanical solution.

DOI 10.1007/978-981-4560-75-7_113-1

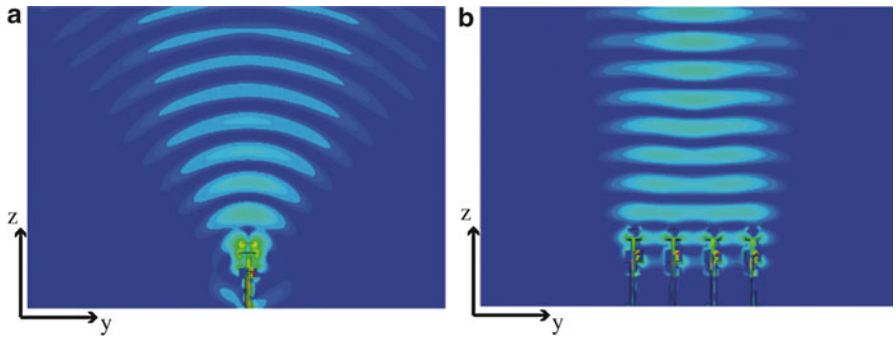


Fig. 13 Comparison of the electric-field distribution of the reference wave by using a single surface-wave launcher or an array

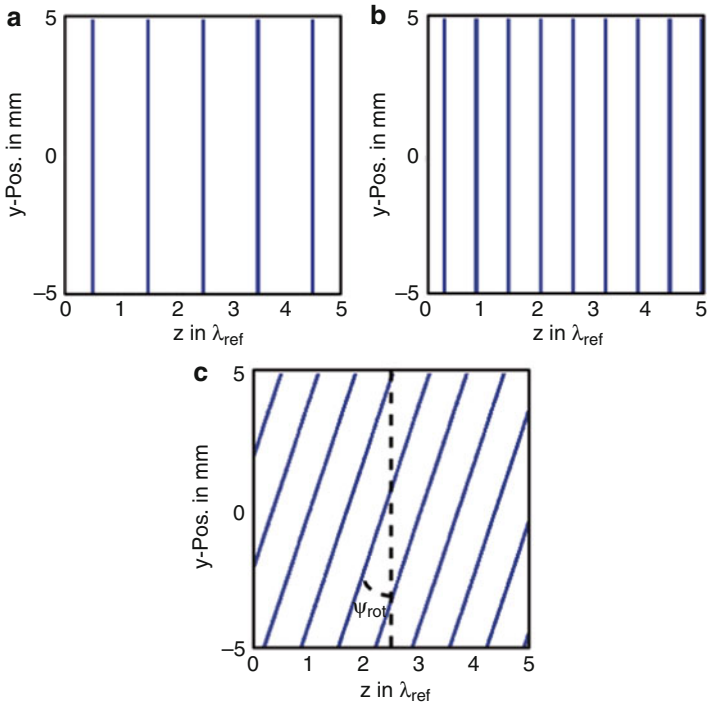


Fig. 14 Calculated holograms for reference waves with linear phase fronts. (**a**) $\theta_0 = 0°$, $\phi_0 = 0°$; (**b**) $\theta_0 = -45°$, $\phi_0 = 0°$; (**c**) $\theta_0 = -45°$, $\phi_0 = 20°$

Linear Holograms with Surface-Wave Launcher Arrays

By using a surface-wave launcher array, the propagating surface wave shows a planar phase front instead of a circular one (see Fig. 13). Due to this fact the single hologram elements also change their shape from circular to linear. As can be seen in Fig. 14, a simple rotation of this hologram with linear stripes means a beam forming of the radiation within the second plane. This effect can be used to build systems with the capability to control the main-lobe direction electronically within two planes.

Such a system is shown in Rusch et al. (2013c). Here, a phased array as surface-wave launcher is combined with the frequency-scanning capability of the holographic antenna. If the tilted propagating reference wave, created by the phase difference between the single launchers, impinges on the hologram, the effect on the main-lobe direction is similar to the rotation of the hologram itself. In Rusch et al. (2013c) a Rotman lens is used to realize the different propagation angles of the reference wave (for more information see chapter on Rusch, "► Holographic Antennas," section "Example I: 2D-Scannable

DOI 10.1007/978-981-4560-75-7_113-1

Fig. 15 Planar quasi-Yagi-Uda surface-wave launcher with balun as slot-line-to-coplanar waveguide transition

Holographic Antenna"). Also other phase-control feeding networks, i.e., the Butler matrix, or the usage of single MMIC phase shifters in front of every surface-wave launcher could be used.

Surface-Wave Launcher Designs

Depending on the desired mode of the reference wave, different types of surface-wave launchers can be used. In general the surface-wave launcher should satisfy the following conditions to get a holographic antenna with high efficiency and low cross polarization:

- Creating a surface-wave mode with high mode purity
- Low antenna mismatch
- Radiation within the substrate plane with high directivity to the hologram direction
- Frequency-independent radiation characteristic

For creating the desired mode, planar end-fire antennas can be used as surface-wave launchers. It has to be noticed that the launcher dimensions have to be matched to the wavelength of the propagating surface wave. Four examples of surface-wave launchers are shown in the following subsections.

Yagi-Uda -Feed

Planar dipoles or folded dipoles are often published surface-wave launchers. Adding a reflector and one or more directors improves the directivity of the launcher and results in a more efficient feeding of the hologram. The relative dimensions and distances of the single antenna elements can be copied from recommended values for Yagi-Uda antennas in air (Rothammel and Krischke 1995) but have to be adapted to the wavelength of the surface-wave mode. A fine adjustment should than be done via full-wave simulation. An integrated balun (see Fig. 15) can be used to feed the launcher with a coplanar waveguide (CPW), which also includes the reflector (Ma et al. 1999). With this balun the holographic antenna can be directly connected to single-ended ICs and also the measurements on a probe-based antenna setup are possible. Due to the resonant launcher type, the antenna matching is narrow band compared to other surface-wave launchers.

Planar Vivaldi Feed

Vivaldi antennas, as shown in Fig. 16, belong to the group of broadband antennas. Due to their broadband antenna matching and frequency-stable radiation characteristic, they are often used for ultra-wideband (UWB) applications (Pancera et al. 2011; Adamiuk et al. 2008). The taper shape and length of slot antennas are analyzed in different publications (Frayne and Leggetter 1991; Yngvesson et al. 1989). Mostly exponential tapers are used, which result in a broadband antenna matching and a high directivity. Due to the strong component of the electric field perpendicular to the propagation direction of the slot-line mode, this antenna type is very well suited to launch a TE_0 surface wave. The layout can be expanded by corrugations to suppress returning currents on the two wings and improve the antenna matching at center

DOI 10.1007/978-981-4560-75-7_113-1

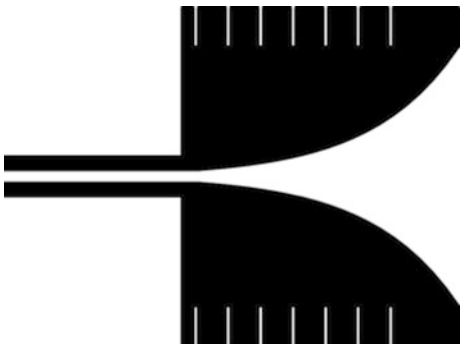


Fig. 16 Vivaldi antenna used as surface-wave launcher with slot-line feed

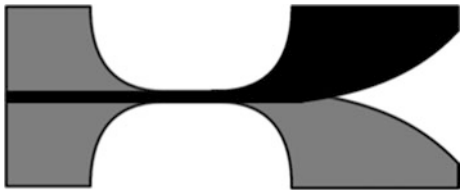


Fig. 17 Antipodal Vivaldi antenna as surface-wave launcher for TE_0 modes

frequency. One disadvantage is the variation of the phase center depending on the frequency. Therefore, it is not guaranteed for a very large bandwidth that the phase center is positioned within the center of the circular hologram. This effect can decrease the antenna efficiency. Due to the feeding slot line, this surface-wave launcher is especially suitable in combination with ICs with differential output ports. To connect with single-ended ICs or for probe-based antenna measurements, the balun from (see chapter on Rusch, "► Holographic Antennas," section "Yagi-Uda-Feed") could also be used.

Antipodal Vivaldi Feed

Also an antipodal version of the Vivaldi feed can be realized. In this case the two antenna wings are positioned on different substrate layers (see Fig. 17; Bai et al. 2011; Hood et al. 2008). The advantage is the feeding with microstrip line technology, which is often used in millimeter-wave front ends. To implement the feed within a microstrip front end, no transition between different transmission-line technologies is necessary and a very high bandwidth can be reached. The effects of taper shape and length are similar to that of the planar Vivaldi antenna. In addition the transition of the feeding microstip line to the strip-line mode between the antipodal wings has to be optimized with respect to the bandwidth and the radiation characteristic, which could have a tilted main-lobe direction, due to the influence of the substrate.

Example I: 2D-Scannable Holographic Antenna

There are different possibilities to control the phase of the surface-wave launcher array (see chapter on Rusch, "► Holographic Antennas," section "Linear Holograms with Surface-Wave Launcher Arrays"). Integrated phase shifters have the advantage to control the phase of every single launcher element stepless and independent of each other (Yun and Chang 2001). The disadvantage is the complex system integration of this kind of phase shifters due to the necessary biasing. Passive phase-control feeding networks like Butler matrix and Rotman lens are easy to implement but have only a limited number of phase shifts dependent on the number of input ports. Compared to the Butler matrix, the Rotman lens

DOI 10.1007/978-981-4560-75-7_113-1

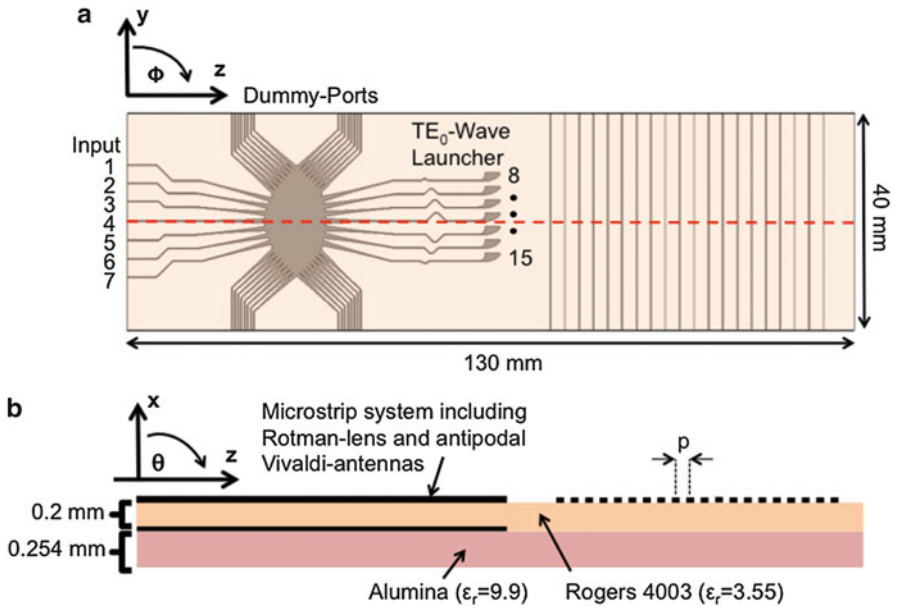


Fig. 18 Layout of the 2D-scannable holographic antenna with Rotman lens (Rusch et al. 2014)

allows more design variances to choose the phase-shift range and the step size within this range. Different kinds of design instructions can be found in the literature, starting with the original paper by Rotman (Rotman and Turner 1963). For the design process in the following example, the description from (Lambrecht et al. 2010) has been used.

Figure 18a shows the top view of the presented antenna. The hologram builds the radiating aperture and consists of linear metal stripes. As described in detail in (see chapter on Rusch, "► Holographic Antennas," section "Using the Holographic Theory for Antenna Designs"), the metal stripes are periodically positioned at the roots of the interference pattern, which results from the superposition of a propagating surface wave (reference wave) and the radiated plane wave (object wave). The surface wave is created by the antipodal Vivaldi array (surface-wave launcher) and propagates as TE_0 mode with planar phase front on the dielectric slab without the ground plane.

The fed input port of the Rotman lens controls the phase difference between the antipodal Vivaldi surface-wave launchers and therefore the propagation direction of the TE_0 surface wave. This results in a change of the main-beam direction in free space within the *xy*-plane (Rusch et al. 2013c). The Rotman lens, the surface-wave launchers, and the hologram are positioned on a dielectric slab of Rogers 4003 ($\varepsilon_r = 3.55$) with a thickness of 0.2 mm. As shown in the layer stack up in Fig. 18b, at the position of the Rotman lens and part of the surface-wave launchers, a GND plane exists at the bottom of the RO layer. This GND plane ends where the surface-wave launchers end. At this point the TE_0 surface wave propagates in the high-permittivity alumina layer below the Rogers substrate. The alumina dielectric sheet with a thickness of 0.254 mm is added to improve the propagation of the surface wave and to increase the frequency-scanning range due to the high ε_r of the material (as described in see chapter on Rusch, "► Holographic Antennas," section "Properties of Leaky-Wave Antennas"). The multilayer stack up allows the suppression of surface waves in the area of the microstrip transmission-line system due to the low dielectric constant of Rogers 4003 and on the other hand increases the frequency-scanning range and promotes the TE_0 surface-mode propagation in the area of the holographic antenna due to the alumina substrate with the high ε_r. This layer stack up can be done with a standard PCB process and is commercially available.

DOI 10.1007/978-981-4560-75-7_113-1

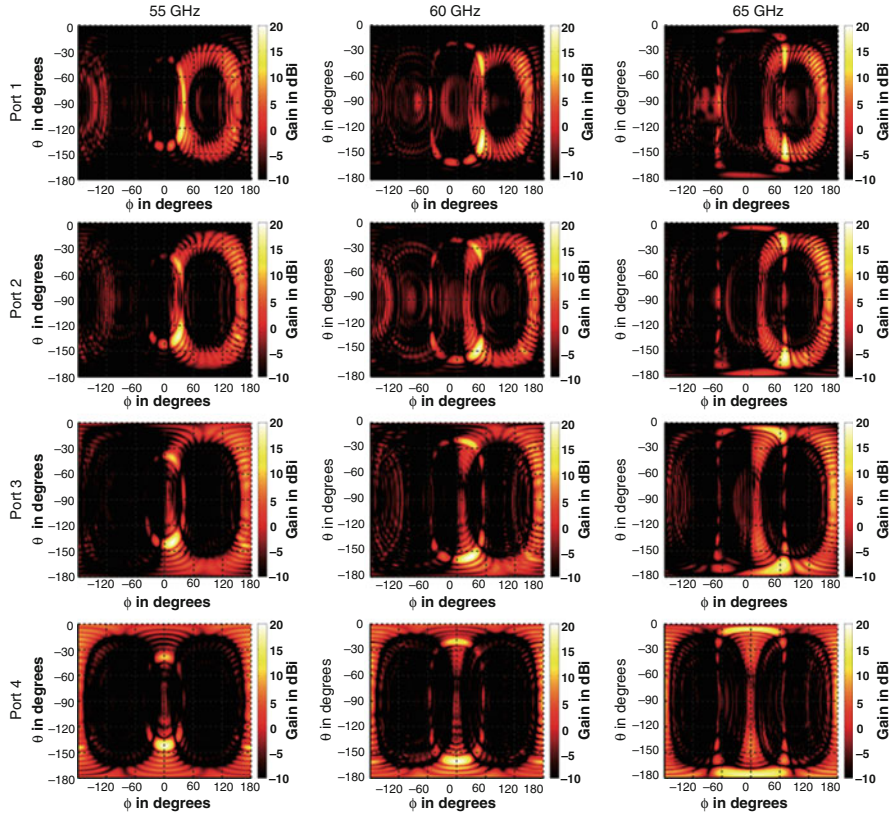


Fig. 19 Simulated radiation patterns of the holographic antenna combined with Rotman lens (Rusch et al. 2014)

Figure 19 shows the simulated radiation patterns for the first four input ports of the Rotman lens. The patterns with feeding at the other ports are not shown, because of the symmetry of the antenna design. The beam-scanning ranges in the two planes meet the expectations. Due to the usage of a dielectric slab without GND plane, there is a second main lobe visible that is mirrored at the symmetry plane. This second beam could be avoided by using reflectors or a volume-type hologram (Iizuka et al. 1975). With the presented concept, a scanning range of $5^\circ < \theta_0 < 40^\circ$ and $-70^\circ < \varphi_0 < +70^\circ$ with a maximum antenna gain of 19.1 dBi was reached.

Example II: Holographic Antenna with TM_0 Surface Wave at 60 GHz

There are limited possibilities of surface-wave launchers for creating the TM_0 mode. A well-working surface-wave launcher for TM_0 modes with good integration properties is a horn antenna built with substrate-integrated waveguide (SIW) technology (Tsao et al. 2012). With its closely spaced vias forming the sidewalls, the manufacturing for mmW frequencies gets difficult. Therefore, this feed can be used only at frequencies up to W band with technology available today. For higher frequency ranges, real planar antennas like the slotted dipole can be used. Podilchak already published a slotted quasi-Yagi-Uda surface-wave launcher (Podilchak et al. 2009) and a scaled version is used here (see Fig. 20). Due to the resonant feed, the bandwidth of the TM_0 antenna is more limited compared to the TE_0 antenna with Vivaldi feed. The feeding line is a coplanar waveguide (CPW), which fits very well to the probe-based measurement system, which was used for measuring the radiation patterns of the holographic antennas (Beer and Zwick 2010).

Since a symmetrical beam forming around broadside is advantageous for most radar applications, a periodicity of the hologram with $p = 3.45$ mm is used to get a main-lobe direction of $\theta_0 = 0^\circ$ at the

DOI 10.1007/978-981-4560-75-7_113-1

Fig. 20 Slotted quasi-Yagi-Uda feed as surface-wave launcher (similar to Podilchak et al. 2009)

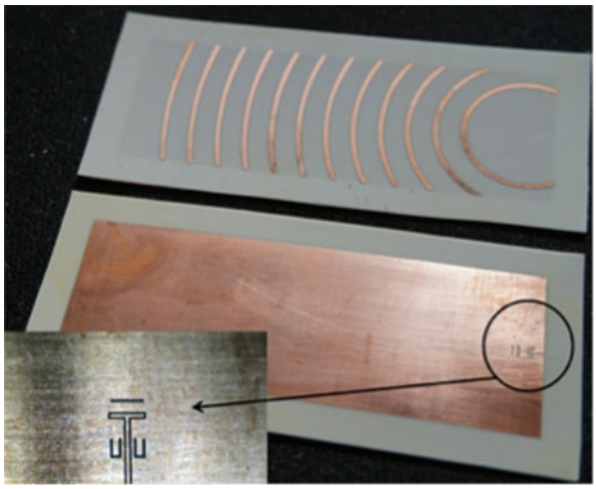


Fig. 21 Manufactured holographic antenna with TM_0 reference wave and slotted quasi-Yagi-Uda surface-wave launcher

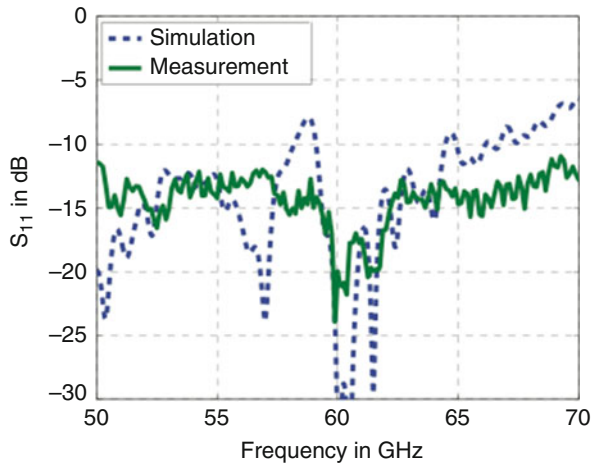


Fig. 22 Reflection coefficient

center frequency $f_c = 60$ GHz. The manufactured structure is shown in Fig. 21. The complete antenna size with its hologram with $N = 12$ elements is 5×2 cm^2. The S parameters for the complete holographic antenna including the feed in Fig. 22 show a good agreement between simulation and measurement. The simulation shows a peak at 58 GHz, which is caused by the stop band close to broadside radiation (Oliner 1993). This is a usual effect of periodic leaky-wave antennas, but its influence

DOI 10.1007/978-981-4560-75-7_113-1

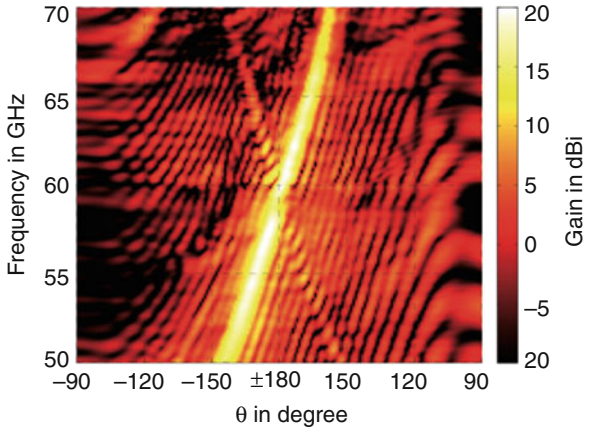


Fig. 23 Simulated antenna gain of the holographic antenna with TM_0 mode and slotted surface-wave launcher

on the radiation depends of the hologram configuration. Although the stop band exists for TM_0 antennas at mmW frequencies as well, the influence on the radiation is lower than for the concepts with TE_0 mode. The reason is that the wave impedance variation caused by the metal stripes with TM_0 mode is decreased. At this frequency and in the range higher than 64 GHz, the agreement between simulation and measurement decreases. The manufactured feed shows a little over-etching which caused a lower directivity than the simulated feed. Therefore, the power of the reference wave, which propagates to the substrate edges perpendicular to the desired propagation direction, does not interact with the hologram and radiates parallel to the substrate plane. This power loss causes a gain decrease and pretends a good antenna matching since it is not coupled back into the surface-wave feed.

With the continuous metal plane there is no bidirectional radiation. The simulated pattern in Fig. 23 shows that the configuration allows the symmetrical beam scanning around broadside. A gain decrease of 3.4 dB due to the stop band is also recognized but is much lower than for similar configurations, i.e., with TE_0-surface waves. The beam-forming range is limited to $\pm 12°$ for a bandwidth of 10 GHz. Due to the asin function in Eq. 9 with the lowest gradient around zero, periodic leaky-wave antennas have the lowest beam-forming range to bandwidth ratio around broadside radiation (Oliner 1993).

Related Planar Periodic Leaky-Wave Antenna Structures

As the holographic antenna is a type of periodic leaky-wave antennas, other radiating structures based on microstrip transmission lines or waveguides with periodic slots have similar radiation characteristics and can be described by the theory approaches explained in (see chapter on Rusch, "► Holographic Antennas," section "Properties of Leaky-Wave Antennas"). In this chapter alternative antenna designs are described that were published in the past and also show the high interest in frequency-scanning antennas.

Most periodic leaky-wave antennas published in literature and used in different applications for radar or communication are realized in waveguide technology or as surface-integrated waveguide to get a planar antenna. Waveguides have the advantage of high robustness against temperature and environmental influences and can therefore be found in space or military applications. The high effort in manufacturing can be reduced by using the SIW technology. The possibilities to feed leaky-wave antennas in SIW technologies with microstrip transmission lines (Martinez-Ros et al. 2011) or coaxial cables (Gandini et al. 2012) facilitate the integration to millimeter-wave RF front ends. By using the fundamental-mode TE_{10}, both slot arrangements transversal (Liu et al. 2011) and parallel to the propagation direction (Gandini et al. 2012) can be used. Transversal-arranged slots influence the currents on the waveguide surface strongly and the resulting stop band prohibits the radiation in broadside direction. Methods have been found by different groups to reduce this stop-band effect and antennas with frequency-scanning

DOI 10.1007/978-981-4560-75-7_113-1

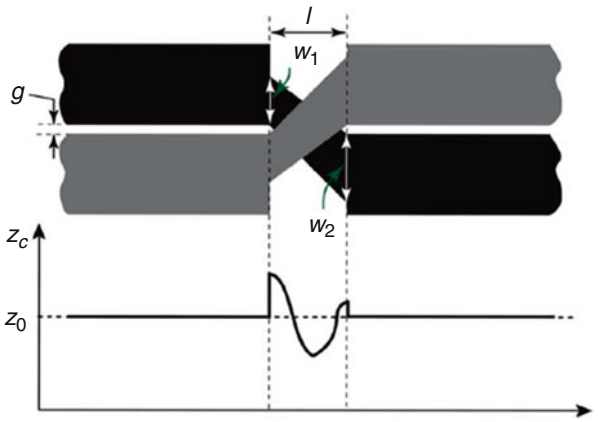


Fig. 24 Unit cell of the periodic leaky-wave antenna based on microstrip transmission-line technology (Yang et al. 2010)

main-lobe directions around the broadband are presented (Dong and Itoh 2011). In Guntupalli and Wu (2012) an easy-to-implement but effective method is presented by using unit cells with transversal-arranged double slots. The slot pairs within a unit cell with a distance of a quarter wavelength at broadside radiation frequency cause a cancelation of the reflected waves and therefore suppress the stop-band effect and allow a continuous frequency-scanning main-beam direction around broadside. Complex systems are published that allow a combination of phased arrays and frequency-scanning antennas and therefore show the possibility of changing the main-beam direction in two dimensions and different spatial directions (Guntupalli and Wu 2013).

By using slots that are arranged parallel to propagation direction, the stop band is not as strong as for the transversal slots due to the lower influence on the traveling wave, but a disadvantage of this configuration is a lower leakage rate of the single unit cells. That means that a higher number of periods are necessary to get an efficient antenna. For a short antenna with a lower number of unit cells, a wave with a high-power level transmits to the end of the antenna which is usually matched to the antenna impedance to avoid the creation of standing waves. In Schuhler et al. (2010) a technique is presented to recycle the power transmitted to the antenna end by guiding it to the antenna feeding port with a meander line to get an in-phase superposition of the waves.

Leaky-wave antennas based on periodic transmission-line structures have the advantage compared to SIW antennas that a high number of vias with very close distance to each other are not necessary. These vias make the SIW antennas expensive in manufacturing and also limit their frequency range of operation. Compared to SIW antennas periodic stubs on a microstrip transmission line are a simple and cheap method to build planar leaky-wave antennas, but also a high number of periods are necessary to get a high antenna efficiency. Also for this type of realization, a method to decrease the stop-band effect at broadside radiation was found by using matching circuits in front of the periodic discontinuities (Paulotto et al. 2009). In Yang et al. (2010), Ke Wu and his group presented the structure shown in Fig. 24. The crossed transmission lines show only a small influence on the traveling wave. Therefore, the stop band at broadside radiation is negligible without using additional techniques and a continuous frequency-scanning main lobe around broadside with an angle range of more than 100° for a frequency sweep from 16 to 39 GHz is presented.

DOI 10.1007/978-981-4560-75-7_113-1

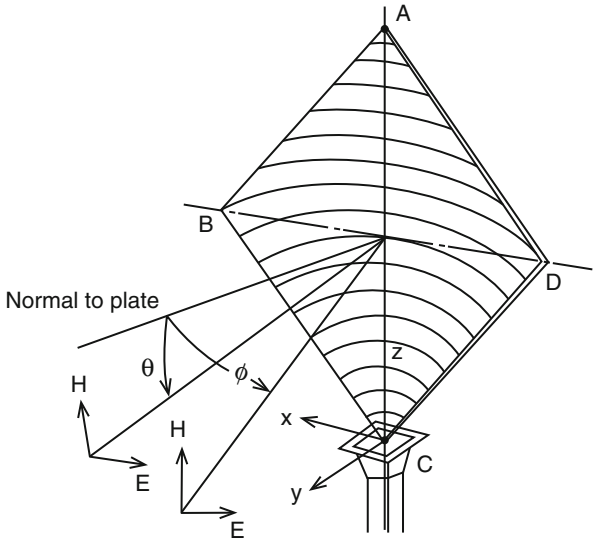


Fig. 25 Structure and geometry of a holographic antenna (Iizuka et al. 1975)

Applications

History of Holographic Antenna Design

The idea to transfer the holographic theory from optics to antenna design was published at first by P. Checcacci et al. (1970). Checcacci created an interference pattern by a superposition of two coherent waves from two different antennas (a horn antenna and a parabolic reflector). In the second step the resulting hologram was illuminated by the horn antenna, which was already used during the recording process. Therefore, the reconstructed antenna pattern of the parabolic reflector was created. Checcacci described his hologram as a pattern consisting of metallic stripes, which represent the roots of the interference pattern. He was the first one who showed that this interference pattern with reduced phase information is sufficient to get the desired main-lobe direction of the holographic antenna. With his antennas Checcacci coined the term holographic antenna.

K. Iizuka (Iizuka et al. 1975) enhanced this technology based on Checcacci's work, which was very important for the practical usage for the millimeter-wave technology. Iizuka rearranged the concept and placed the reference wave and the interference pattern into the same plane, so that both the feed and hologram could be realized on one printed circuit board (see Fig. 25). The radiation perpendicular to this plane built the reconstructed object wave. Iizuka used antenna substrates without a continuous metal plane on top or bottom. The spacing between the metal strips was one wavelength of free space, rather than one wavelength of the surface wave, which should also exist on the interface between the dielectric of the printed circuit board and air. By illuminating the substrate edge with a parallel-arranged rectangular horn antenna, the hologram was fed by the reference wave. Due to the lack of a continuous metallization on one substrate surface, this antenna showed a bidirectional radiation pattern. For that reason another novelty was published the first time in this paper. The authors proposed a way to create a unidirectional radiation by a suitable arrangement of two holographic plates. The two plates were spaced a quarter wavelength apart and the radii of the rings of the second hologram were always a quarter wavelength shorter. Thus, the reconstructed waves interfered constructively in one direction and destructively in the other direction and hereby a unidirectional plane wave was created. They formed the hologram by etching a printed circuit board, which was 12 wavelengths long and wide and used a horn antenna as feed element at the substrate edge. A measured gain was not published, but a unidirectional narrow beam was achieved with a front-to-back ratio of about 20 dB.

DOI 10.1007/978-981-4560-75-7_113-1

In 1980 the idea of dynamically steering an antenna beam based on the concept of holography was mentioned for the first time (Farhat 1980). In this case it was proposed to use semiconductor panels whose charge carrier concentration could be controlled. Thus, changes in the transmission and reflection coefficients of the panel could be realized locally to dynamically create different interference patterns that created a steered beam together with a reference wave. In the presented concept the illuminating antenna did not lie in the same plane as the interference hologram.

Later, the holographic principle was mainly used for antenna measurements. In this case a reference antenna illuminated an interference pattern and synthesized a plane wave that was necessary for the antenna measurements. Especially the measurement of large reflector antennas for satellite communications was eased by the holographic measurement method (Hirvonen et al. 1997; Sazonov 1999).

From 2001 onward, the amount of research on holographic antennas has increased again. Fathy et al. from Sarnoff Corporation published their works performed during their Reconfigurable Aperture Program (ElSherbiny et al. 2004; Fathy et al. 2001, 2003a, (b). Presented is their concept for reconfiguration based on metallic-like conductive plasma islands that are driven by a DC current. These plasma islands could be precisely formed and controlled on silicon substrates. Thus, metallic patterns could dynamically be formed on a silicon wafer (Fathy et al. 2001, 2003b). Holographic antennas were mentioned as one possible application of this concept. Using the DC-driven plasma islands, a dynamically changed interference pattern or hologram could be formed on the silicon wafer that, together with a reference wave, was used for beam steering. To prove their concept, a 35 GHz holographic antenna was realized, however by using simple etching technology instead of the plasma island technology (ElSherbiny et al. 2004; Fathy et al. 2003a). The antenna was realized on a layered substrate of alumina and TMM-4 and consists of the etched concentric rings that form the hologram and a feed antenna in form of an open-ended waveguide. In a distance of a quarter wavelength apart from the substrate, a ground plane was attached that acts as a reflector and creates a unidirectional beam; see Fig. 26. The novelty in their design mainly laid in the fact that a surface wave was used as reference wave rather than a space wave. The surface wave that travels inside the layered material had a shorter wavelength compared to the space wave and thus the hologram could be created denser. They in fact realized several 35 GHz antennas with different holograms to create beams in different angular directions. The hologram dimensions were about 92 $\times$ 92 mm and a gain of about 25 dBi was achieved.

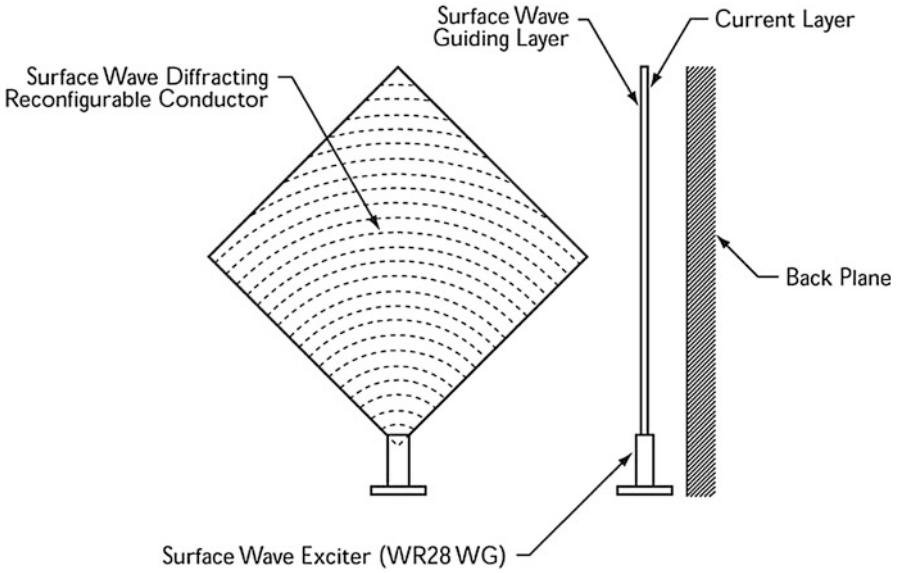


Fig. 26 Holographic surface-wave antenna (ElSherbiny et al. 2004)

DOI 10.1007/978-981-4560-75-7_113-1

The method with rectangular horn antenna as feed was still used in current publications (Sooriyadevan et al. 2007). Although this feeding concept was not suitable for systems with high integration level, the idea using planar end-fire antennas as surface-wave launchers was published a long time after the first mention of holographic antennas (Petosa et al. 2004; Sutinjo et al. 2010). With these types of surface-wave launchers for different kinds of surface-wave modes (Podilchak et al. 2009, 2011), an integration of the holographic antenna in millimeter-wave systems up to very high frequencies (>100 GHz) was possible. In addition to end-fire antennas like Yagi-Uda radiators or tapered slot antennas (Thingvold et al. 2003), also dipoles (Li et al. 2011) or patch antennas (Sutinjo et al. 2010) were published as solutions for creating the surface-wave mode.

With these possibilities holographic antennas had the advantage compared to other types of leaky-wave antennas, e.g., slotted waveguides in "substrate-integrated waveguide (SIW)" technology, that a manufacturing without vias became possible. Therefore, thin, ceramic substrates with high dielectric constants could be used which are favored for thin-film technology to build antennas at very high frequency ranges.

Example: The Frequency-Scanning Millimeter-Wave Radar

Using leaky-wave antennas due to their frequency-scanning properties in frequency-modulated continuous-wave (FMCW) radars can be a cheap solution with low complexity to measure distance, and the angle of a target position has been shown by different groups. A large system bandwidth is necessary to get sufficient resolution for both measurement dimensions. That is the reason why the usage of this antenna type is especially interesting for integrated millimeter-wave radars, and at high center frequencies, the antennas become small enough to make packaging solutions similar to the one presented in Beer et al. (2013) possible.

General Principle of Operation

Radar systems with frequency-scanning antennas have the advantage that range and cross range of targets in one plane can be measured at the same time. Therefore, a combination with frequency-modulated continuous-wave (FMCW) radars is useful. This approach is discussed in Mayer et al. (2003) and Alvarez-Lopez et al. (2012) and allows a system with lower complexity compared to radars with phased-array or mechanically scanning antennas. In the case of frequency scanning, no mechanical movement of components or of the entire antenna is needed, which reduces the complexity of the design and reduces the risk of failure due to material fatigue. By sending a frequency chirp with a desired bandwidth B, the main-lobe direction of the holographic antenna scans over the measurable angle range. The back-reflected wave from a target is mixed with the transmitted chirp signal, as it has also been done in traditional FMCW radars, so that an intermediate frequency is generated which includes the range information of the illuminated target.

$$R = \frac{c_0}{2B} f_{IF} t_{\mathrm{mod}} \tag{10}$$

Smaller targets at different angular positions to the radar will reflect only parts of the frequency chirp back to the radar depending on the time of illumination. This results in a modulation of the received signal power. After mixing transmit and receive signal, this power modulation is still included within the intermediate baseband signal. By analyzing this power modulation, the cross range of targets can be measured.

DOI 10.1007/978-981-4560-75-7_113-1

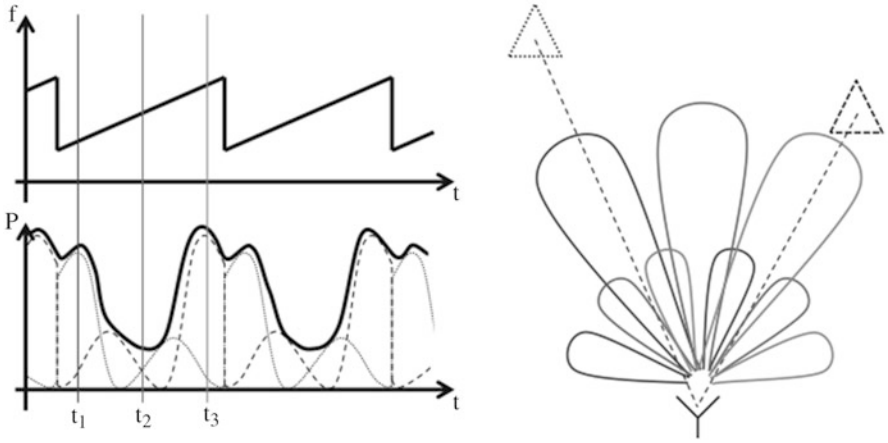


Fig. 27 Scheme of the radar principle with frequency-scanning antenna

Figure 27 shows this in a schematic diagram. The curve on the top-left corner shows the transmitted frequency-modulated signal. The example scenario with two targets is shown on the right side, in which also the antenna characteristics for different time points are indicated. The second graph shows three curves. The two dashed curves represent the echo signals created by the individual targets. The bold line is the superposition of these echoes, which is received by the radar. Depending on the configuration of the transmitted frequency chirp, the main-lobe direction at different times varies. At the beginning of the sending process at the time t_1, the main beam illuminates the first object; thereby a strong echo signal is received (short-dashed line). At the time t_2 the main lobe points to an empty range without target, which results in a low receiver signal. At the third time t_3, the received power increases again because the antenna now illuminates the second target (long dashed).

Range Measurement

The maximum range resolution $\Delta R_{\max}$ of an FMCW radar is described by Eq. 11 and therefore is inversely proportional to the system bandwidth B and proportional to the propagation velocity of the wave c.

$$\Delta R_{\max} = \frac{c}{2B} \tag{11}$$

If a leaky-wave antenna with frequency-scanning main-lobe direction is combined with this radar type, not the complete bandwidth B is available at the receiver. Due to the fact that the antenna illuminates the individual target just within a shorter time range, the new bandwidth B' that was transmitted during this time determines the resolution. If a constant beam width θ_{3dB} and a linear slope of the frequency-dependent main-lobe direction $\frac{\delta\theta}{\delta f}$ are assumed, the decreased distance resolution can be calculated by the relation of scanning range θ_S and beam width θ_{3dB} (Mayer et al. 2003):

$$\Delta R = \frac{c}{2B'} = \frac{c}{2B}\frac{B}{B'} = \frac{c}{2B}\frac{\theta_S}{\theta_{3dB}}. \tag{12}$$

In general the range information is given by the proportionality of the intermediate frequency f_{IF} and the target distance as known from the traditional FMCW radar. Due to the different angle ranges, illuminated by the antenna at different times during the modulation time t_{mod}, the maximum measurable target range becomes dependent from the target's cross range. This is shown in Fig. 28. For an FMCW radar the echo

DOI 10.1007/978-981-4560-75-7_113-1

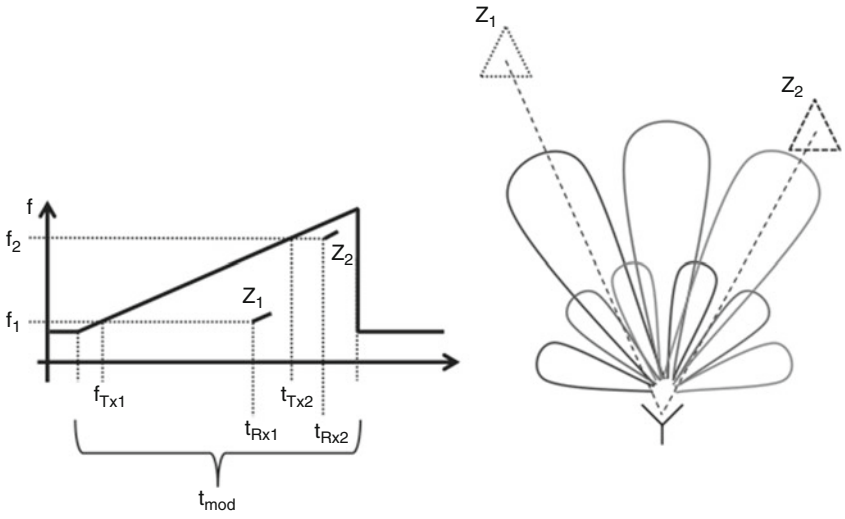


Fig. 28 Decreasing of range resolution and maximum range by the usage of frequency-scanning antennas

of a faraway positioned target should be received within the modulation time t_{mod} due to the mixing process of T_x and R_x chirp (for a real FMCW radar: $t_{\mathrm{Rxn}} << t_{\mathrm{mod}}$).

For the following example the simplified approach $t_{\mathrm{Rxn}} \leq t_{\mathrm{mod}}$ applies. If there is a target Z_1 within the angle range, which is illuminated at the beginning of the transmitted chirp, the transmitted and reflected signal can travel a longer way within the remaining modulation time without violating the condition than for the target Z_2, which is illuminated later. A simple calculation will illustrate this circumference. For the frequency-scanning radar, a modulation with bandwidth $B = 10$ GHz and modulation time $t_{\mathrm{mod}} = 10$ ms is used. Target Z_1 is illuminated by a signal with frequency f_1, which was transmitted at the time $t_{Tx1} = 1$ ms and received as echo at t_{Rx1}. Accordingly target Z_2 is illuminated by a signal with frequency f_2, which was transmitted later at $t_{Tx2} = 9$ ms and received due to the propagation time at t_{Rx2}. To satisfy the condition that the last received echo is within the modulation time, the maximum distance results:

$$R_{\max} = \frac{c_0}{2}(t_{\mathrm{mod}} - t_{\mathrm{Txn}}), \tag{13}$$

For the described scenario this means maximum distances of the two targets with $R_{\max 1} = 1,350$ km and $R_{\max 2} = 150$ km. The values show that this effect is only important for very fast frequency sweeps or large target distances.

Cross Range Measurement

For measuring the cross range of a target, the amplitude modulation of the received power signal is analyzed. The echo signal of a point target at angle position θ_Z includes the frequency-dependent antenna radiation characteristic of the radar. Due to the fact that the relation between time and transmitted signal frequency is known, the cross range information can be extracted. This requires synchronization between the transmitter and receiver for an exact knowledge of the transmitted signal frequency at the time, when the echo was received (see Fig. 27). To get a high-quality convergence of the antenna characteristic, the IF frequency should not be too low. If there are only a small number of periods within the IF signal, there is an undersampling of envelope power and results in a lack of measurement accuracy (Chan and Judah 1998). This limit depends on the scenario and on the desired resolution. A practical approach to move this limitation is the usage of meander lines to create higher IF frequencies for equal target distances

DOI 10.1007/978-981-4560-75-7_113-1

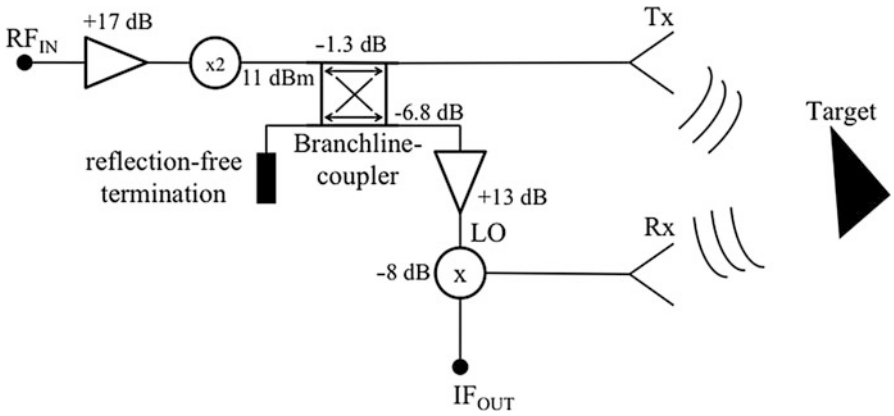


Fig. 29 Equivalent circuit of the radar front end in LTCC

(Mayer et al. 2003). To get a high accuracy of the cross range measurements, the time-shift caused by the propagating velocity of the signal has to be noticed. This time-shift causes a measurement inaccuracy due to the lack of synchronization between transmit chirp and power modulation at the receiver. The inaccuracy can be corrected by measuring the range with Eq. 10 in a first step. If the range is known, the resulting time-shift can be calculated and therefore used for the calibration of the cross range measurement.

The cross range resolution is defined by the 3 dB beam width θ_{3dB} of the antenna main lobe (Skolnik 1990). If the targets are in a closer distance to each other, a reliable separation is not guaranteed.

Cross Range/Distance Trade-Off

As described in (see chapter on Rusch, "► Holographic Antennas," section "Range Measurement"), a wide antenna beam width is necessary to get a high-range resolution. This is in contrast to the condition of a small beam width for a high cross range resolution as described in (see chapter on Rusch, "► Holographic Antennas," section "Cross Range Measurement"). This trade-off is a principal characteristic of the combination of FMCW radar and frequency-scanning antenna. With Eq. 14 the uncertainty relation can be defined.

$$\Delta R \cdot \Delta\theta = \frac{c}{2B}\theta_S \tag{14}$$

With this equation, the limit of one parameter can be calculated if the other parameter is known (Solbach and Schneider 1999).

Implementation: Radar Front End in LTCC with Integrated Holographic Antenna at 60 GHz

This chapter presents a radar front-end prototype with integrated holographic antenna. Test-measurement results in laboratory environment are shown to clarify the operating principle and the requirements of FMCW radars with frequency-scanning antennas.

Measurement Setup

For test measurements the radar front end is positioned on the antenna measurement setup described in Beer and Zwick (2010). The movable arm, which holds the reference horn antenna during antenna gain measurements, is used to align radar and target. Figures 29 and 30 show an equivalent circuit of the front end and the measurement setup with all used instruments.

DOI 10.1007/978-981-4560-75-7_113-1

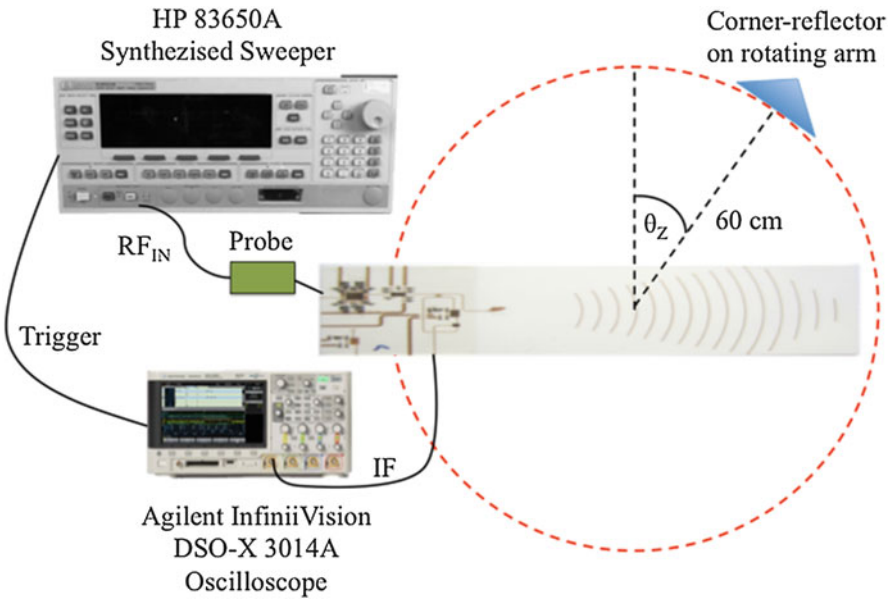


Fig. 30 Test-measurement setup of the radar front end with integrated holographic antenna

The radar input RF_{in} is fed by ground-signal-ground(GSG) probes. With the signal generator HP 83650A, the frequency chirp with a bandwidth of 6 GHz is created around the center frequency 31.5 GHz. With the implemented frequency multiplier ($x2$), the transmit signal has a bandwidth of 12 GHz with a center frequency of 63 GHz. The used frequency range is adapted to the measured characteristics of the manufactured antenna, which are slightly different to the simulation results, due to tolerances of the material parameters and manufacturing process. The minimum available sweep time $t_{\mathrm{mod}} = 20$ ms has been chosen, to get IF frequencies as high as possible. The movable arm rotates in a distance of approximately 70 cm around the radar. With Eq. 15 an IF frequency of $f_{IF} = 2.8$ kHZ is expected, if the signal propagation time within the front end is neglected.

$$f_{IF} = \Delta f \frac{\Delta t}{t_{\mathrm{mod}}} = \frac{\Delta f^2 r / c_0}{t_{\mathrm{mod}}} \tag{15}$$

(with r as distance between radar and target in m, c_0 as wave velocity in air, and t_{mod} as sweep time)

The sampling of the IF signal is done with the oscilloscope Agilent InfiniiVision DSO-X 3014A. The oscilloscope and signal generator are synchronized by a trigger signal. This synchronization is necessary to get the relation between the transmitted frequency at the time when the target was illuminated and the main-lobe direction of the antenna. A corner reflector is used as target to get a high radar cross section. In Fig. 30, the target is placed above the radar. This has been done within the illustration to have the same angle definition for θ and θ_Z as in previous figures. For the measurements the target is placed beyond the substrate ($\theta_Z > 180°$), because of the higher radiation of the antenna within the lower hemisphere. The approximate target distance of 60 cm, which results by the measurement setup, meets the far-field condition from (Balanis 1997).

Short-Time Fourier Transform

The short-time Fourier transform (STFT) is one of the mostly used time-frequency analyzers. The function is based on the discrete Fourier transform (DFT). Instead of windowing the complete signal, a short window is used and moved from sample to sample. Therefore, a time resolution of the analyzed

DOI 10.1007/978-981-4560-75-7_113-1

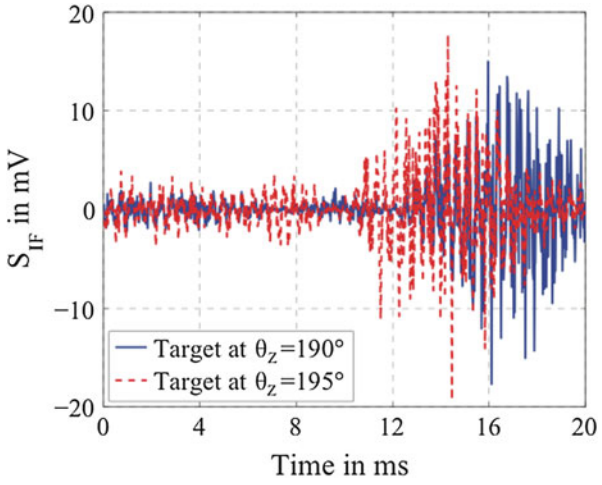


Fig. 31 Measured IF signals for two different angular positions of the corner reflector

signal results and a higher number of individual DFTs are created, which include the information about the signal spectrum at different times. A sequence of these single DFTs results in a STFT. The generated spectrogram as square of the absolute STFT shows the power of the analyzed signal as a function of time and frequency. To get a high temporal resolution, a short window function should be used. Accordingly a short window decreases the frequency resolution. Therefore, an accurate choice of the window parameters is important to get high-quality measurement results. By using the STFT the following four parameters should be considered with reference to the resolution and the computing time:

- The number of samples in frequency domain determines the length of the DFT (possible increase by using "zero padding" (Skolnik 1990)).
- The number of samples in time domain determines the number of DFTs, which must be calculated.
- The shape of the window influences the intenseness of the spectral leakage.
- The length of the window determines the trade-off between time and frequency resolution.

Depending on the significance of the two parameters' time and frequency resolution, the window length should be adapted.

Measurement Results

Two measurement examples with a corner-reflector target at two different angular positions within the described measurement scenario are shown in Fig. 31. To eliminate effects of antenna mismatch or coupling between transmit and receive path, the signals are corrected via subtraction with a reference measurement within the empty scenario. With a constant distance the change of the angular position is clearly visible, due to the amplitude maximum at different times.

Due to the synchronization between FMCW sweep generator and oscilloscope, the transmitted signal frequency at every time step is known. To get the main-lobe direction of the antenna at the time of the maximum received amplitude, the measured antenna gain should be used for the signal processing. The range and angular position resolution can be configured with the definition of the window length (here, Hamming window) for the STFT. Figure 32 clarifies this trade-off.

This figure shows the STFT of two different measurements with the same corner reflector. The corner reflector is mounted to the movable arm. The distance to the radar is reduced to approximately 60 cm. The two defined angular positions are $\theta_Z = 190°$ and $\theta_Z = 200°$. The STFT implementation in Matlab

DOI 10.1007/978-981-4560-75-7_113-1

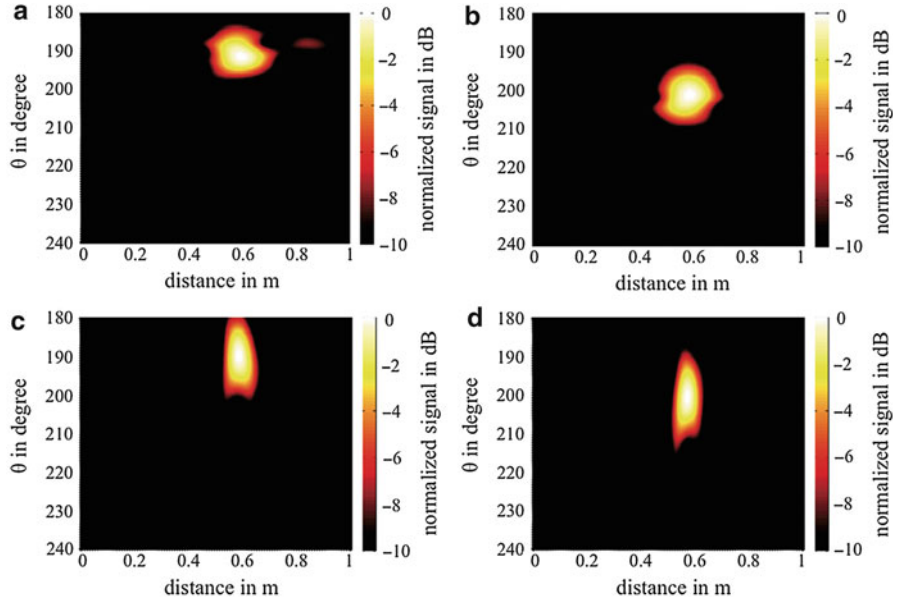


Fig. 32 Measurement results in a 1-target scenario; (**a**) $\boldsymbol{\theta_Z} = 190°$, $\boldsymbol{N_W} =$ 101 samples; (**b**) $\boldsymbol{\theta_Z} = 200°$, $\boldsymbol{N_W} =$ 101 samples; (**c**) $\boldsymbol{\theta_Z} = 190°$, $\boldsymbol{N_W} =$ 401 samples; (**d**) $\boldsymbol{\theta_Z} = 200°$, $\boldsymbol{N_W} =$ 401 samples

allows defining the window function with its length and window type. Due to the knowledge of the measured antenna pattern and the wave velocity, the time and frequency axes can be replaced with range and angular position. The complete signal has a length of $N_S = 2,001$ samples.

A longer window ($N_W = 401$ samples) narrows the measured signal within the range axes, which results in a higher range resolution. The same effect can be seen on the angular position axes by using a shorter window ($N_W = 101$ samples). A maximum resolution should be reached by using a Dirac for maximum angular position resolution or a rectangular window with a length according to the complete number of samples for the maximum range resolution. The intermediate frequency is measured within the complete sampled time range. This effect is caused by the nonideal antenna pattern, which illuminates the target also with its side lobes. Therefore, a longer window increases the signal-to-noise ratio as well. To decrease the effect of spectral leakage, a Hamming window with length N_W was used for the shown measurements. During the post-processing, two different window lengths are applied to every measurement to get high accurate results of both measured variables. Furthermore, the sample number is increased by zero padding to $N_S = 4,096$. In the first processing step, the Hamming window with length $N_W = N_S$ is applied. From the calculated spectrum the intermediate frequency is determined. No temporal resolution is available during this first step, but the range is determined with the maximum accuracy. To get the angular position of the target, the STFT is applied to the data with a shorter Hamming window. A length of $N_W = 1$ should result in a maximum angular position resolution, but it means the determination of the maximum amplitude of the signal in time domain. This amplitude is strongly influenced by environmental effects or noise. Therefore, a short window length with $N_W = 101$ samples has been selected to get a good temporal resolution on the one hand and to decrease the effects from noise on the amplitude on the other hand.

The corner reflector is positioned on the movable arm and rotated with $-5°$ steps started at $\theta_Z = 220°$ around the radar. The center of rotation is aligned with the hologram. Figure 33 shows the measurement results. The distance values are calculated with the determined intermediate frequency and the wave velocities within the dielectric waveguide and air. The second y-axis shows the transmit frequency at the time of the maximum measured echo signal. A linear FMCW chirp after triggering the sweep generator is

DOI 10.1007/978-981-4560-75-7_113-1

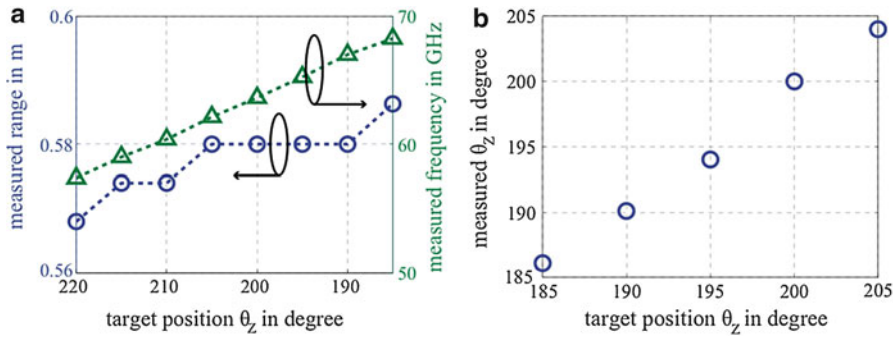


Fig. 33 Radar measurement results for the 1-target scenario

assumed. With the available measured antenna pattern, these axes can be interchanged with the measured angular positions of the target (Fig. 33b). The measurement error is only $\pm 1°$ and the limitation in this case is given by the resolution of the measured antenna pattern. For the antenna pattern measurement, a step width of 2° has been used to move the reference horn around the AUT.

The deviation of the distance measurement at smaller position angles is explained by a nonconstant phase-center position of the antenna, due to its large longitudinal expansion and the nonconstant amplitudes of the single hologram elements.

Conclusion

It has been shown that the usage of planar holographic antennas on dielectric substrates within has several advantages, which makes them to be promising candidates for millimeter-wave systems. Due to the usage of surface waves as feeding modes, the manufacturing of holographic antennas is possible up to very high frequencies (e.g., D and J band) also with standard PCB technology.

In the previous chapter, the theory of holographic antennas and their analogy to the optical holography were explained. Concrete relations between material parameters and antenna properties as the frequency-dependent beam-scanning angle were shown.

The deviation of the propagation properties of different surface-wave modes on grounded and ungrounded dielectric slabs served as background for the followed investigations. Three feeding structures, in literature often called surface-wave launchers, realized as planar feeds similar to the Vivaldi antenna and the Yagi-Uda antenna in two different layouts were presented.

Building an array of these surface-wave launchers causes the propagation of surface waves with linear phase fronts and therefore influences the hologram shape. This novel structure allows the realization of systems with beam-scanning functionalities in two dimensions.

An FMCW radar as one of the most interesting applications of holographic antennas was explained. A possible implementation with LTCC technology allows a compact RF front end, which measures range and angular position of targets within a scenario simultaneously.

DOI 10.1007/978-981-4560-75-7_113-1

Cross-References

- ▶ Antennas in Automobile Radar
- ▶ Beam-Scanning Leaky Wave Antennas
- ▶ Linear Wire Antennas (Including Yagi-Uda Antennas)
- ▶ Millimeter-Wave Antennas and Arrays
- ▶ Phased Arrays
- ▶ Traveling-Wave Antennas (Including Leaky Wave and Surface Wave Antennas)
- ▶ Waveguide Slot Antennas and Arrays
- ▶ Waveguide Slot Array Antennas

References

Adamiuk G, Zwick T, Wiesbeck W (2008) Dual-orthogonal polarized Vivaldi antenna for ultra wideband applications. In: 17th international conference on microwaves, radar and wireless communications, Wroclaw, pp 1–4

Alvarez-Lopez Y, Garcia-Gonzalez C, Vazquez-Antuna C, Ver-Hoeye S, Andres FL (2012) Frequency scanning based radar system. Prog Electromagn Res 132(9):275–296

Bai J, Shi S, Prather D (2011) Modified compact antipodal Vivaldi antenna for 4–50-GHz UWB application. IEEE Trans Microwave Theory Tech 59(4):1051–1057

Balanis CA (1997) Antenna theory: analysis and design. John Wiley & Sons, Inc., Hoboken, New Jersey

Beer S (2013) Methoden und Techniken zur Integration von 122 GHz Antennen in miniaturisierte Radarsensoren, Karlsruher Forschungsberichte aus dem Institut für Hochfrequenztechnik und Elektronik; 70, KIT Scientific Publishing, Karlsruhe, ISBN 9783731500513

Beer S, Zwick T (2010) Probe based radiation pattern measurements for highly integrated millimeter-wave antennas. In: Proceedings of the fourth European conference on antennas and propagation (EuCAP), Barcelona

Beer S, Rusch C, Gulan H, Gottel B, Girma M, Hasch J, . . . Zwick T (2013) An integrated 122-GHz antenna array with wire bond compensation for SMT radar sensors. IEEE Trans Antennas Propag 61(12):5976–5983

Chan K-L, Judah SR (1998) A beam scanning frequency modu- lated continuous wave radar. IEEE Trans Instrum Meas 47(5):1223–1227

Checcacci P, Russo V, Scheggi A (1970) Holographic antennas. IEEE Trans Antennas Propag 18(6):811–813

Collier RJ, White PD (1976) Surface waves in microstrip circuits. In: 6th European microwave conference, Rome, pp 632–636

Dong Y, Itoh T (2011) Substrate integrated composite right-/left-handed leaky-wave structure for polarization-flexible antenna application. IEEE Trans Antennas Propag 60:760–771

ElSherbiny M, Fathy A, Rosen A, Ayers G, Perlow SM (2004) Holographic antenna concept, analysis, and parameters. IEEE Trans Antennas Propag 52(3):830–839

Farhat N (1980) Holographically steered millimeter wave antennas. IEEE Trans Antennas Propag 28(4):476–480

Fathy A, ElSherbiny M, Rosen A, Ayers G, Perlow S (2003a) Experimental demonstration of a 35 GHz holographic antenna. IEEE MTT-S Int Microwave Symp Dig 3:1833–1836

Fathy A, Rosen A, OwenH, Kanamaluru S, McGinty F, McGee D, . . . ElSherbiny M (2001) Silicon based reconfigurable antennas. IEEE MTT-S Int Microwave Symp Dig 1:377–380

Fathy A, Rosen A, Owen H, McGinty F, McGee D, Taylor G, . . . ElSherbiny M (2003b) Silicon-based reconfigurable antennas – concepts, analysis, implementation, and feasibility. IEEE Trans Microwave Theory Tech 51(6):1650–1661

Frayne PG, Leggetter AJ (1991) Wideband measurements on Vivaldi travelling wave antennas. IEEE Colloq Multi-Octave Microwave Circuits 5/1–5/6

Gabor D (1948) A new microscopic principle. Nature 161:777–778

Gandini E, Ettorre M, Casaletti M, Tekkouk K, Le Coq L, Sauleau R (2012) SIW slotted waveguide array with pillbox transition for mechanical beam scanning. IEEE Antennas Wirel Propag Lett 11:1572–1575

Guntupalli AB, Wu K (2013) Full-space scanning phased array system for future integrated high data rate communication over E-band and beyond. In: European microwave conference (EuMC), Nuremberg, pp 1607–1610

Guntupalli A, Wu K (2012) Multi-dimensional scanning multi-beam array antenna fed by integrated waveguide Butler matrix. IEEE MTT-S Int Microwave Symp Dig 1–3

Hammad H, Antar YM, Freundorfer A, Mahmoud S (2003) Uni-planar CPW-fed slot launchers for efficient TM_0 surface-wave excitation. IEEE Trans Microwave Theory Tech 51(4):1234–1240

Hirvonen T, Ala-Laurinaho J, Tuovinen J, Raisanen A (1997) A compact antenna test range based on a hologram. IEEE Trans Antennas Propag 45(8):1270–1276

Hood AZ, Karacolak T, Topsakal E (2008) A small antipodal Vivaldi antenna for ultrawide-band applications. IEEE Antennas Wirel Propag Lett 7:656–660

Iizuka K, Urasaki S, Ushigome H (1975) Volume- type holographic antenna. IEEE Trans Antennas Propag 23(6):807–810

Lambrecht A, Beer S, Zwick T (2010) True-time-delay beamforming with a Rotman-lens for ultrawideband antenna systems. IEEE Trans Antennas Propag 58(10):3189–3195

Li Y, Zhu Q, Mo R (2011) Studies on the holographic antenna: theories and experiments. In: Asia-pacific microwave conference proceedings, Melbourne, pp 654–657

Liu J, Jackson D, Long Y (2011) Substrate integrated waveguide (SIW) leaky-wave antenna with transverse slots. IEEE Trans Antennas Propag 60:20–29

Ma K-P, Qian Y, Itoh T (1999) Analysis and applications of a new CPW-slotline transition. IEEE Trans Microwave Theory Tech 47(4):426–432

Martinez-Ros A, Gomez-Tornero J, Goussetis G (2011) Planar leaky-wave antenna with flexible control of the complex propagation constant. IEEE Trans Antennas Propag 60:1625–1630

Mayer W, Wetzel M, Menzel W (2003) A novel direct-imaging radar sensor with frequency scanned antenna. IEEE MTT-S Int Microwave Symp Dig 3:1941–1944

Oliner AA (1993) Leaky-wave antennas. In: Johnson R (ed) Antenna engineering handbook. McGraw-Hill, New York

Pancera E, Zwick T, Wiesbeck W (2011) Spherical fidelity patterns of UWB antennas. IEEE Trans Antennas Propag 59(6):2111–2119

Paulotto S, Baccarelli P, Frezza F, Jackson D (2009) A novel technique for open-stopband suppression in 1-D periodic printed leaky-wave antennas. IEEE Trans Antennas Propag 57:1894–1906

Petosa A, Thirakoune S, Levis K, Ittipiboon A (2004) Micro- wave holographic antenna with integrated printed dipole feed. Electron Lett 40(19):1162–1163

Podilchak SK, Freundorfer AP, Antar YM (2008) Broadside radiation from a planar 2-D leaky-wave antenna by practical surface-wave launching. IEEE Antennas Wirel Propag Lett 7:517–520

Podilchak SK, Freundorfer AP, Antar YM (2009) Surface-wave launchers for beam steering and application to planar leaky-wave antennas. IEEE Trans Antennas Propag 57(2):355–363

Podilchak SK, Freundorfer AP, Antar YM (2011) A new leaky-wave antenna design using simple surface-wave power routing techniques. IEEE Int Symp Antennas Propag 3052–3054

Rothammel K, Krischke A (1995) Antennenbuch. Franckh-Kosmos, Stuttgart

Rotman W, Turner R (1963) Wide-angle microwave lens for line source applications. IEEE Transa Antennas Propag 11(6):623–632

Rusch C, Beer S, Gulan H, Zwick T (2013a) Holographic antenna with antipodal feed for frequency-scanning radar. In: 2013 I.E. antennas and propagation society international symposium (APSURSI), Orlando, pp 234–235

Rusch C, Beer S, Pahl P, Zwick T (2013b) Multilayer holographic antenna with beam scanning in two dimensions at W-band. In: 7th European conference on antennas and propagation (EuCAP), Gothenburg, pp 2625–2628

Rusch C, Beer S, Pahl P, Gulan H, Zwick T (2013c). Electronic beam scanning in two dimensions with holographic phased array antenna. In: International workshop on antenna technology (iWAT), Karlsruhe, pp 23–26

Rusch C, Schäfer J, Gulan H, Zwick T (2014) 2D-scanning holographic antenna system with Rotman-lens at 60 GHz. In: European conference on antennas and propagation (EuCAP), The Hague

Sazonov D (1999) Computer aided design of holographic antennas. IEEE Antennas Propag Soc Int Symp 2:738–741

Schuhler M, Wansch R, Hein M (2010) On strongly truncated leaky-wave antennas based on periodically loaded transmission lines. IEEE Trans Antennas Propag 58:3505–3514

Skolnik M (1990) Radar handbook, 2nd edn. McGraw-Hill, New York

Solbach K, Schneider R (1999) Antenna technology for milli-meter wave automotive sensors. In: 29th European microwave conference, Munich, pp 139–142

Sooriyadevan P, McNamara DA, Petosa A, Ittipboon A (2007) Electromagnetic modelling and optimisation of a planar holographic antenna. IET Microwaves Antennas Propag 1(3):693–699

Sutinjo A, Okoniewski M, Johnston RH (2010) A holographic antenna approach for surface wave control in microstrip antenna applications. IEEE Trans Antennas Propag 58(3):675–682

Thingvold SR, Ittipiboon A, Sebak A, Petosa A (2003) Holographic antenna efficiency. IEEE Antennas Propag Soc Int Symp 3:721–724

Tsao H-Y, Yang D-H, Cheng J-C, Fu J, Lin W-P (2012) W-band SIW H-plane horn antenna development. In: 4th international high speed intelligent communication forum (HSIC), Nanjing, pp 1–3

Yang N, Caloz C, Wu K (2010) Full-space scanning periodic phase-reversal leaky-wave antenna. IEEE Trans Microwave Theory Tech 58:2619–2632

Yngvesson KS, Korzeniowski TL, Kim Y-S, Kollberg EL, Johansson JF (1989) The tapered slot antenna – a new integrated element for millimeter-wave applications. IEEE Trans Microwave Theory Tech 37(2):365–374

Yun T-Y, Chang K (2001) A low-cost 8 to 26.5 GHz phased array antenna using a piezoelectric transducer controlled phase shifter. IEEE Trans Antennas Propag 49(9):1290–1298

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_116-1

Antennas in Microwave Wireless Power Transmission

Xue-Xia Yang*
School of Communication and Information Engineering, Shanghai University, Shanghai, China

Abstract

The transmitting and receiving antenna designs for the microwave wireless power transmission (MPT) systems are illustrated in this chapter. After introducing a brief history, this chapter describes the components of a MPT system and the operation band selections. Several rectennas and arrays with different structures and performances are designed after analyzing the special considerations of the receiving antennas. To obtain the maximum DC-DC (direct current to direct current) system efficiency, the aperture level distributions of the transmitting antenna are investigated with consideration of the receiving rectennas. A MPT system design operating at C-band is explained based on the above discussion and the summary is made finally.

Keywords

Microwave wireless power transmission (MPT); Rectennas; Antennas; Rectifying circuits; Conversion efficiency; Aperture level distribution; Optimization

Introduction

Among several kinds of wireless power transmission (WPT) technologies, the microwave power transmission (MPT) is noticeable for the features of being applied in long-distance and high power transmission. It had been suggested to power helicopter, transfer power to the far-reached areas, and transfer the solar power to the earth by the solar power satellite (SPS). Now the WPT technology has explored its applications in many fields, such as charging portable devices and automobile, powering wireless sensors, telemetry and radio frequency identification (RFID), microwave-driven smart material actuators, and recycling ambient microwave energy. With the development of the wireless communications, radars, and radio astronomy, the techniques of the microwave generators and transmitting antennas have been much developed and could be adopted by the MPT systems. However, the rectennas, which are composed of receiving antennas and rectifying circuits, are unique for the MPT systems. After a brief history and operation principle of MPT technology, the emphasis is put on the designing of the receiving rectennas and the transmitting antennas for the MPT systems.

Section "Introduction" presents the brief history of MPT technology. Section "MPT Technology and System" describes the components of a MPT system and the operation frequency selections. A receiving antenna should be designed to be easily integrated with a rectifying circuit and receives the maximum microwave power, which is detailed in section "Receiving Rectennas Design." Section "Transmitting Antennas Design and *BCE*" shows the transmitting antenna design considering the receiving rectenna to obtain the high system efficiency. Finally, several MPT systems are designed in section "Efficiency Evaluation of C-Band MPT System" and the conclusion is in section "Summary."

*Email: xxyang@staff.shu.edu.cn
*Email: xuexiay@hotmail.com

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_116-1

Brief History of Wireless Power Transmission and Applications

The gap spark experiment using metal coils made by Heinrich Hertz in 1887 demonstrated the electromagnetic wave propagation in free space. Over 120 years passed, electromagnetic waves have explored many kinds of communications applications including telegraph, broadcast, television, radar, radio astronomy, satellite navigation, and all sorts of mobile communications. Even so, early in 1899, Nikola Tesla suggested transferring power by the electromagnetic wave and did effort to verify the idea. It was said that experiment of transferring power by the resonant radio frequency of 150 kHz using Tesla coil had been carried out at Colorado Springs in 1899. It is a pity that the detailed data had not been recorded. The second famous WPT experiment was performed in the Westinghouse Laboratory by H. V. Noble and was exhibited at the Chicago World's Fair of 1933–1934. Several hundreds of power were transferred by 100 MHz electromagnetic wave. The transmitting and the receiving antennas were dipoles and the distance was 25 ft.

It was not until the end of the 1950s that the microwave WPT technology was noticed again because of the emerging and maturity of the key microwave elements, such as the klystrons and magnetrons that generate the microwave and transistor diodes that rectify efficiently. The microwave generators at the higher radio frequencies and the phased arrays and high-gain antennas in radar systems rapidly developed during World War II make it possible to propagate EM wave by a narrow beam.

The first MPT technology application was microwave platform put forward by Raytheon Company in 1959. The microwave platforms would provide long-endurance operations for communication relay, environment surveillance, space detection, remote sensing, etc. at a high altitude space. Although the microwave platforms were not developed practically, it activated the MPT research work from the beginning of the 1960s. The first MPT experiment in laboratory was performed by Dr. Brown at JPL in 1963 (Brown 1964), in which 400 W microwave power was transmitted and 104 W DC (direct current) power was obtained at the receiver from 7.4 m distance away. It was also reported that a microwave-powered helicopter flew about 10 h at 10 ft height (Brown 1984). In the 1990s and the early 2000s, the concept of microwave platform was investigated further in many plans, such as SHARP plan of Canada (East 1992); MILIX, MILAX, and ETHER plans of Japan (Shinohara 2011; Fujino et al. 1996; Shinohara and Matsumoto 1998); KERI plan of Korea (Youn et al. 1999); and SERT plan of the USA (Strassner and Chang 2003). To verify these plans, some MPT experiments were carried out on the earth from one place to another one or from one point on the earth to the air.

The solar power satellite (SPS) concept introduced by Dr. Peter Glaser in 1968 is the second important MPT technology application (Glaser 1992), which possesses perpetual value as a kind of green energy until now and in the future. The SPS is a power converting station, on which the solar power is transferred to the earth by the microwave beam using its low loss performance in the atmosphere, regardless of day or night. Three steps are included. Firstly, the SPS collects solar power and converts it into DC (direct current) power by the optic-electric converters. Secondly, the microwaves are produced from the DC power by the generators and transmitted to the earth by the transmitting antennas on the SPS. Thirdly, the rectenna arrays on the earth, which are composed of the receiving antennas and the rectifying circuits, receive the microwaves and convert them back to DC power. The conversion efficiency of DC-DC is main parameter concerned by the researchers and the users. As one of the key technologies, MPT research has been strongly inspired by the SPS concept. Some relative areas have been researched, for example, the influences of the microwave beams on the birds, on the soil for growing crops, on the ionosphere, etc. The fractionated spacecraft concept presented by DARPA in 2007 is a good combination of the microwave platforms and SPS. The fractionated spacecraft is a satellite system composed of some modules, one of which obtains energy from the SPS and powered other modules by MPT technology.

MPT technology was also proposed to be applied in powering unmanned aircrafts and transfer power to the faraway districts (Celeste et al. 2004). The microwave beam strength and steering direction can be

DOI 10.1007/978-981-4560-75-7_116-1

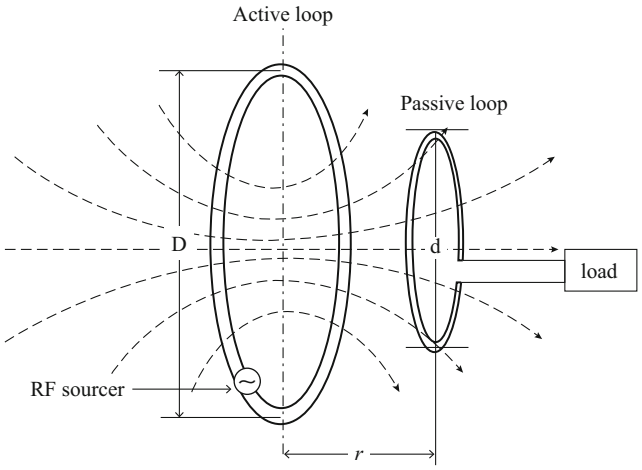


Fig. 1 Near-field magnetic induction

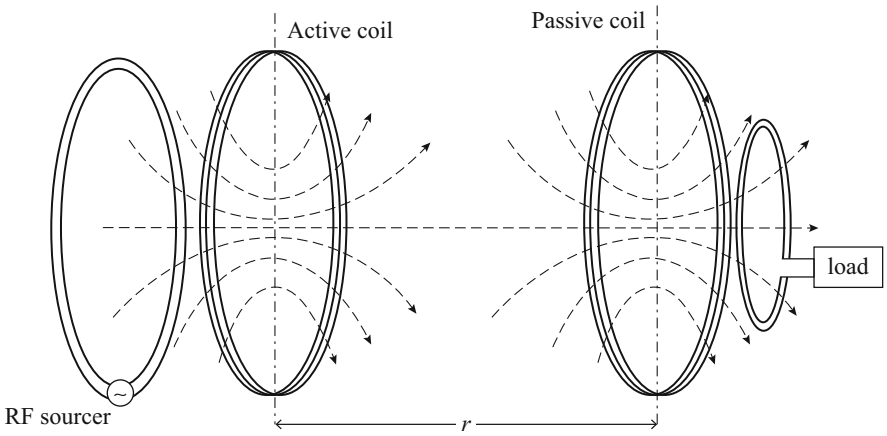


Fig. 2 Strongly coupled magnetic resonance

controlled easily to meet the requirement of safe standards. Now it has explored its applications in low power conditions, such as wireless sensors embedded in concrete bridge or bodies, radio frequency identification (RFID), recycling ambient microwave energy (Zoya et al. 2014), etc.

Remarkably, the planar printed technology developed in the mid-1950s made MPT more practical because the rectennas that receive the microwave beam and convert it into DC power could be very thin and conformal to the DC demanders.

Since the beginning of this century, consumer electronic products have been developed rapidly. It is attractive to recharge electric products wirelessly. Tailless TVs and wireless recharge mobile devices have been shown by the main companies although they have not been widespread in the market. In 2008, the Wireless Power Consortium developed an interface standard for the near-field inductive electrical power transfer, which is named as Qi. The Qi standard is renewed constantly since then (2014). This near-field magnetic induction can wirelessly transfer power to only a centimeter range (Fig. 1). In 2007, strongly coupled magnetic resonances were suggested to transfer much power wirelessly to a little long distance (Kurs et al. 2007). Sixty watts power was collected over 2 m at the efficiency of about 40 % in MIT (Fig. 2).

There are three kinds of wireless power transmission technologies based on different operation principles: magnetic induction, magnetic resonance, and MPT by radiation beam. Generally, the magnetic

DOI 10.1007/978-981-4560-75-7_116-1

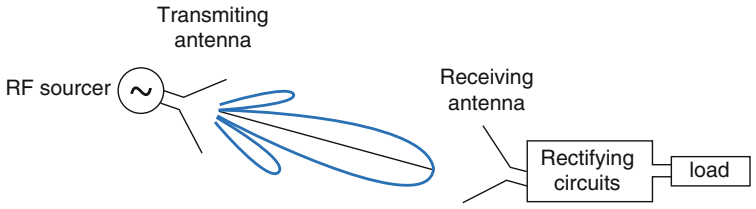


Fig. 3 MPT by radiation beam

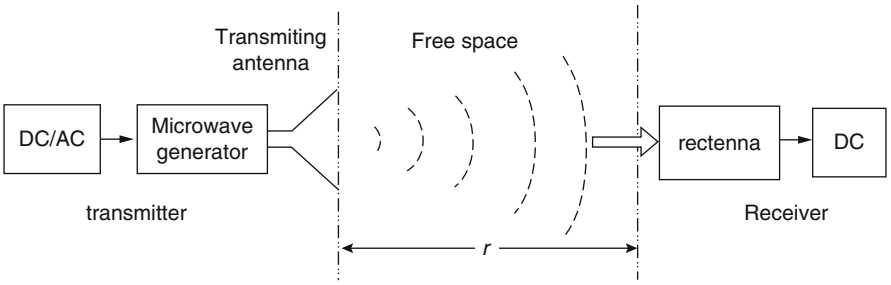


Fig. 4 MPT system

induction and the magnetic resonance manners are based on the near-field theories and the MPT system should be designed by the far-field theory. This chapter put the emphasis on designing the transmitting and receiving antennas for the MPT systems.

MPT by Radiation Beam

It is well known that the losses of microwaves below X-band are low in the atmosphere. The strength and the direction of the microwave beam could be controlled by the transmitting and the receiving antennas as shown in Fig. 3. So it is a feasible manner to transmit power to a far distance on the stage of hundreds of kilometers.

MPT Technology and System

MPT technology has been proposed firstly on the applications of high power and long distance, such as microwave powering high altitude airships and unmanned vehicles, SPS, and transmitting power to the far-reached places. Now MPT has explored the applications to low power density situations of harvesting the ambient electromagnetic wave and recharging the portable electronic devices. Meanwhile, the power density must meet the safety standards for the living environments of human, animals, and plants (IEEE Std 2006).

System Components and Efficiency

A MPT system consists of a transmitter and a receiver as shown in Fig. 4. The microwave transmitted by the transmitter is received and rectified by the receiver after propagating an R distance. The transmitter is composed of a microwave generator and a transmitting antenna or array. The microwave could be excited from a DC or an AC (alternating current) resource. The receiver has a receiving antenna and a rectifying circuit, which is named as a rectenna.

DOI 10.1007/978-981-4560-75-7_116-1

The DC-DC (or AC-DC) efficiency is the key parameter of a MPT system, which is contributed by several factors. The transmitter includes two factors: the efficiency of the microwave generator from DC to mw (microwave) η_g and the transmitting antenna efficiency η_t. The microwave propagation efficiency is indicated as η_p, in which the attenuation in the atmosphere should be considered. The receiver of the rectenna has two factors: the microwave capture efficiency η and the mw-DC conversion efficiency η_r of the rectenna. So the total efficiency of a MPT system is

$$\eta_{sys} = \eta_g \cdot \eta_t \cdot \eta_p \cdot \eta_c \cdot \eta_r \tag{1}$$

To design a MPT system with a high DC-DC efficiency, the transmitting and the receiving antenna apertures must be optimized jointly, which will be detailed in section "Transmitting Antennas Design and *BCE*" where the microwave beam capture efficiency (*BEC*), i.e., η_c, will be investigated. The five efficiency factors in formula (1) are all related to the operation frequency.

Operation Frequency of MPT

The study and applications of MPT technology mainly focus on the ISM (industrial, scientific, and medical) bands of L (915 MHz), S (2.45 GHz), and C (5.8 GHz). With the development of microwave devices, millimeter waves have been applied in MPT technology, such as K- (24.125 GHz), E- (61.25 GHz), and W-bands. The operation frequencies are mainly determined by the application environments.

For the applications of transferring power between a long distance across the atmosphere, such as powering an unmanned vehicle from the earth or an aerostat in the near space, transferring power from one place to another on the ground, and transferring power from the space to the ground (SPS), the losses in the atmosphere must be considered so L-, S- and C- bands should be used. For the applications in the space, where there are nearly no atmosphere losses, the millimeter waves are preferable because of the relatively compact and light generators, transmitting antennas, and rectennas. The lightness and compactness of microwave devices are very important because of the actual load limitation in the space. The millimeter MPT technology is recommended to power the fractionated satellites.

The generator efficiencies of η_g are relative to the operation frequencies. Generally, the generator technique is mature and the efficiencies η_g is higher at the low microwave bands. The propagation efficiency η_p is mainly decreased by the attenuations of the oxygen and water vapor in the atmosphere. Generally, the path losses are very low at the bandwidths below around 10 GHz and can be neglected. However, the losses must be considered when the operation frequencies come to higher. The oxygen molecules resonate at 60 and 118.75 GHz, while the water molecules do at 22.24, 183.31, and 325.5 GHz, which cause serious attenuations. There exist atmospheric windows among these absorbing frequencies, such as 35, 94, 140, and 220 GHz, which can be used to MPT.

The rectennas are generally attached to the surface of a DC power demander so the planar printed process is preferable. As far as the present PCB process level, the printed rectennas at microwave bands are acceptable, while those at millimeter wave bands have high losses and the mw-DC efficiency is relatively low.

With the development of the radar, wireless, and satellite communications, many kinds of high-power microwave antennas and generators have been proposed, which could be utilized in the MPT systems. In this chapter, the emphasis will be put on two items. One is the receiving rectennas and array designs with the high mw-DC efficiency, i.e., η_r. Another is the optimization of the transmitting and the receiving apertures for the high microwave capturing efficiency of *BEC*.

DOI 10.1007/978-981-4560-75-7_116-1

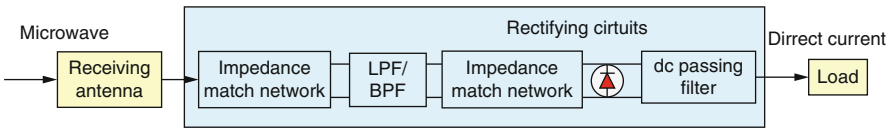


Fig. 5 Components of a rectenna

Receiving Rectenna Design

A rectenna is the integration of a receiving antenna with a rectifying circuit as shown in Fig. 5. The rectifying circuit is composed of a rectifying diode, an input matching network, a bandpass or lowpass filter, a DC-pass filter, and a load. The research on rectenna elements shows that the mw-DC conversion efficiency could reach 80 % when a Schottky diode receives an input power about 100 mW (Strassner and Chang 2003). The common mw-DC efficiencies of rectenna elements are about 75 % at the operation frequencies not higher than X-band (Strassner and Chang 2003; Yang et al. 2008). The mw-DC efficiencies of rectenna arrays would decrease apparently because of the transmission line loss of the circuits, the feedline loss of the antennas, and the nonuniform of the power density on the rectenna elements (Shinohara and Matsumoto 1998).

The antennas should receive enough microwave power to keep the rectifying circuit have a high and constant mw-DC efficiency. Thus, to design a receiving antenna for a rectenna, some factors should be considered, including a high gain, easiness to be integrated with the rectifying circuit, the polarization calibration for the transmitting and the receiving antennas, etc.

Evaluation of mw-DC Conversion Efficiency of Rectennas

The mw-DC conversion efficiency is the most concerned parameter for a rectenna. Supposing the DC voltage on the load R_L is V_{dc}, the mw-DC can be calculated by Eq. 2:

$$\eta_r = \frac{P_{dc}}{P_{in}} = \frac{V_{dc}^2}{R_L P_{in}} \tag{2}$$

where P_{out} and P_{in} are the output DC power and the incident microwave power, respectively. The incident microwave power of the rectenna is calculated by the Friis transmission formula:

$$P_{in} = \left(\frac{\lambda}{4\pi r}\right)^2 P_t G_t G_r \tag{3}$$

where P_t and G_t are the power and the gain of the transmitting antenna, G_r is the receiving antenna gain, and λ is the free space wavelength at the center frequency. Figure 6 illustrates the rectenna efficiency measurement system. The transmitting power P_t can be monitored by a power meter with a little microwave being coupled by a directional coupler.

Considerations for Designing Receiving Antennas

The most commonly used antennas for rectenna systems are the planar antennas, which could be integrated with the rectifying circuits with a minimum loss and be conformal to the electric equipment.

To meet with the rectenna demand of high mw-DC efficiency and engineering applications, the selection of the receiving antenna for a particular application should be based on these considerations: high gain, the same feedline with the transmission lines of the rectifying circuit, necessary bandwidth, and the polarization state.

DOI 10.1007/978-981-4560-75-7_116-1

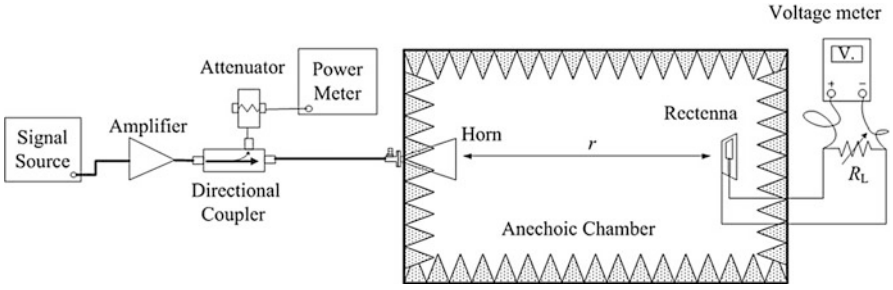


Fig. 6 mw-DC conversion efficiency measurement system of rectennas

High Gain

High-gain demand of the receiving antennas lies on two points. Firstly, on the safety condition of a limited power density, the high-gain antennas can easily obtain enough input microwave power for the rectifying circuits to maintain the high mw-DC conversion efficiency. Secondly, antennas with high gain can reduce the total number of rectenna elements on a given area. Planar printed antennas are preferred to keep low profile and conformal with the DC power demanders.

Matching and Integrating with Rectifying Circuits

Impedance match between the receiving antenna and the rectifying circuit is necessary to transfer the maximum microwave rectifying diode. To minimize the energy lost in the impedance match networks, the feedlines of receiving antennas are the same as the transmission lines of the rectifying circuits. The commonly used planar transmission lines have CPS (coplanar striplines), microstrip lines, and CPW (coplanar waveguides). Generally, to simplify the design process, the input impedances of the antenna and the circuit are all 50 Ω. However, it is better to design the input impedance of the antenna equivalent to that of the diode to make the rectenna compact if the LPF/BPF can be omitted. The design examples of the three different planar transmission lines, including CPS, microstrip line, and CPW, will be introduced in this section.

Bandwidth

Theoretically, MPT can operate on single frequency. Considering the errors of design and manufacture and the frequency uniformity between the receiving antenna and the rectifying circuit, a certain bandwidth is essential. To harvest the ambient electromagnetic energy, a wideband or a multi- band receiving antenna, which covers the microwave bands, is required.

Polarizations

As all know, a linear polarized antenna can receive only half the power of a circularly polarized wave and vice versa. So for the applications of transfer power from one point to another one, the polarization alignment is critical for the receiving and the transmitting antennas with linear polarization operations while only same direction of turning is demanded for the circularly polarized waves. For the MPT applications of one point to multi-point or inverse, circularly polarized rectenna can receive constant microwave at any polarization directions when the transmitted microwaves are linear polarization or the same circular one.

Other Considerations

The functions of the LPF/BPFs in Fig. 9 are used to restrain the high-order harmonic waves excited by the nonlinear diode from being radiated again because re-radiation will decrease the rectenna efficiency.

DOI 10.1007/978-981-4560-75-7_116-1

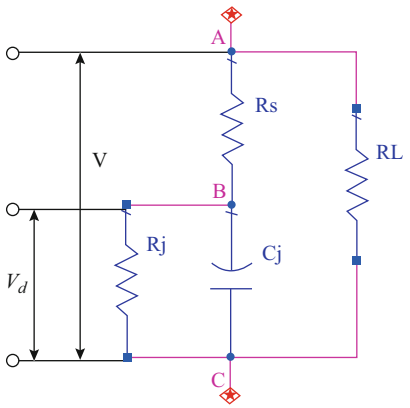


Fig. 7 Equivalent circuit of unpackaged Schottky diode

A receiving antenna with harmonic suppression can avoid the need of the LPF/BPFs and allows for miniaturizing the rectenna.

For the WSN and RFID applications, the electrical devices usually transmit data and obtain power simultaneously. The dual-polarized receiving antenna is a strategy with one polarization for communication and another for MPT. Similarly, a dual-frequency antenna is also an option.

Rectenna Design Examples

CPS Circularly Polarized Rectenna

Dr. Strassner and Chang proposed a kind of CPS circularly polarized rectenna, whose prominent virtues are the high gain of the receiving antenna and the broadband of the antenna and the rectifying circuit (Strassner and Chang 2003). A dual-rhombic loop with a reflecting plane was utilized as the receiving antenna with 10.7 dB gain and 10 % 2:1 VSWR being obtained. The rectifying circuit included a CPS band-reject filter that suppresses the re-radiated harmonics by 20 dB. The rectifying diode was an unpackaged MA4E1317 Schottky diode, whose equivalent circuit is illustrated in Fig. 7. The equivalent circuit parameters are as follows: series resistance $R_S = 4\ \Omega$, zero-bias junction capacitor $C_{j0} = 0.02$ pF, built-in turn-on voltage $V_{bi} = 0.7$ V, and breakdown voltage $V_B = 7$ V. The mw-DC conversion efficiency of approximately 80 % was measured for an input power level of 100 mW on the load of 250 Ω.

A similar rectenna was investigated as shown in Fig. 8, in which a low pass filter (LPF) was used to inhibit the high-order harmonics. According to the formula derived in Strassner and Chang (2003), the input impedance of the MA4E1317 diode was 172 Ω. The input impedance of the antenna and the characteristic impedance of the CPS transmission line could be designed as 172 Ω to omit the impedance match networks in the rectifying circuit. The rectenna was etched on the substrate of 2.55 relative permittivity and 0.8 mm thickness. So the CPS has trace width of 0.6 mm and the slot width of 0.4 mm.

The LPF structure and the S parameter responses are shown in Fig. 9. The main sizes are $L_1 = 5.6$ mm, $L_2 = 3.7$ mm, $L_3 = 3.5$ mm, and $L_4 = 4.3$ mm. $|S_{21}|$ parameters at 5.8, 11.6, and 17.4 GHz are −0.08, −29.45, and −22.9 dB, respectively. The higher-order harmonics have been inhibited effectively.

A dual-rhombic-loop antenna was designed at 5.8 GHz with the reflect plate at the distance of $0.21\lambda_0$. The measured gain is 10 dB and the input impedance was 170 Ω. The antennas and the LPF were simulated together by Designer software and the frequency response of $|S_{11}|$ was shown in Fig. 10. $|S_{11}|$ parameters at the fundamental frequency of 5.8 GHz and the second and third harmonics were −28.8, −1.05, and −1.02 dB, respectively, which illustrated that the harmonics were suppressed successfully. The distance between the diode and the capacitor in Fig. 8 was 9.5 mm and the optimized load was 250 Ω.

DOI 10.1007/978-981-4560-75-7_116-1

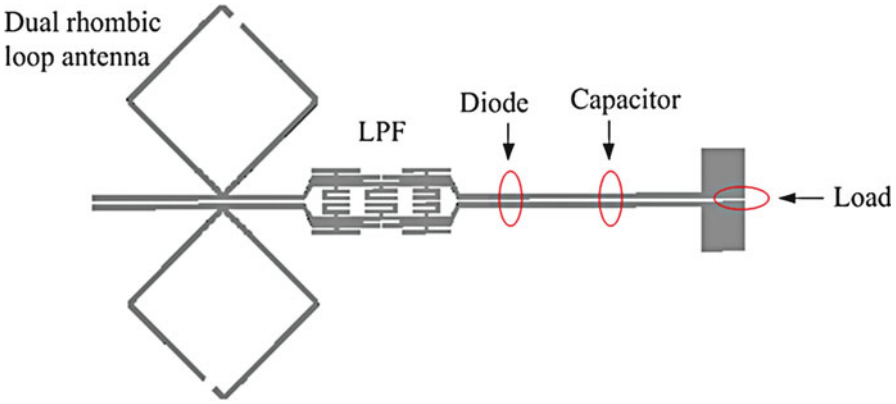


Fig. 8 CPS circularly polarized rectenna

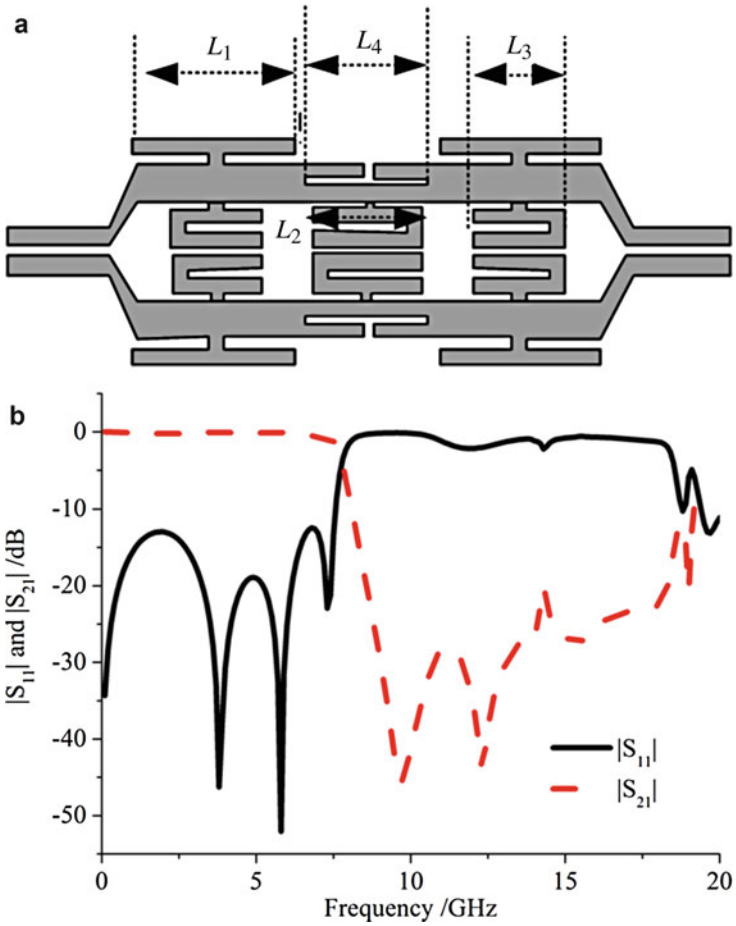


Fig. 9 LPF structure and the S parameter responses. (**a**) LPF structure and (**b**) frequency responses of $|S_{11}|$ and $|S_{21}|$

The experiments had been carried out in a chamber. The maximum mw-DC conversion efficiency was 74 % at 5.95 GHz operation efficiency when the input power was 100 mW. The center frequency shift induced the decrease of the efficiency.

DOI 10.1007/978-981-4560-75-7_116-1

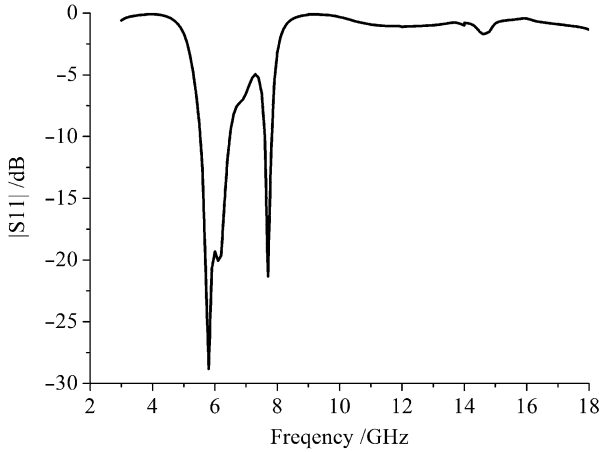


Fig. 10 Simulated $|S_{11}|$ versus frequency of the antenna integrated with the LPF

Circularly Polarized Rectenna with No LPF/BRF

The LPF (or BRF) could be omitted if the receiving antenna has harmonic inhibition function. Itoh et al. proposed a circular-sector antenna with a sector angle of 240° and an inset feeding point at 30° from the edge suppressing effectively the radiations of the second and the third harmonics (Park et al. 2004). Gao et al. designed a circularly polarized patch antenna with DGS (defect ground structure) under the microstrip feedline to avoid the harmonics (Gao et al. 2010). The structure, the simulation, and the measurement performances are illustrated in Fig. 11. The simulation software of HFSS has been used (Ansoft 2011).

The truncated-corner square patch excites the circular polarization wave and the DGS of dual dumbbell below the microstrip line on the ground suppressed the harmonics. The antenna has been etched on a substrate with the thickness of 1.5 mm, the relative dielectric constant ε_r of 2.55, and the tanδ of 0.002. From Fig. 11b, it is obvious that the high-order harmonics are inhibited effectively. The reflection coefficient is −30 dB at 5.8 GHz and about −3 dB at 11.6 GHz. The measured minimum AR is 1.0 dB and the bandwidth of AR less than 3 dB is 60 MHz from 5.78 to 5.84 GHz on the main direction from Fig. 11c. The measured gain is 6.5 dB at 5.8 GHz operating frequency.

The schematic diagram of the rectifying circuit is shown in Fig. 12. The diode used in the circuit was also MA4E1317 and the substrate is the same as the rectenna in Fig. 8. Simulated by ADS software (Agilent 2005), the input impedance of the diode was (230 + j18)Ω. The input impedance match network consists of two $\lambda/4$ microstrip lines and an open stub which is sensible to the operating frequency so the effect of the high harmonics deduced by the diode is low. The output match network is composed of a chip capacitor and a $\lambda/4$ microstrip line, which smoothes the DC voltage and reutilizes harmonics energy. The DC power is collected by a resistive load. This rectifying circuit had the highest simulated mw-DC efficiency of 81.4 % on the load of 320 Ω at 5.8 GHz when the input power was 100 mW.

The receiving antenna and the rectifying circuit were connected by a SMA connector as shown in Fig. 13a. The overall RF-DC conversion efficiency of rectenna versus frequency is illustrated in Fig. 13b. The voltage of 4.34 V was obtained and the highest conversion efficiency reached to 68.4 % on the load of 298 Ω at 5.86 GHz, which was exactly the value of the rectifying circuits. In applications, the antennas and the rectifying can be integrated directly on one substrate by omitting SMA connector. Without the loss of SMAs, the efficiency would be higher.

DOI 10.1007/978-981-4560-75-7_116-1

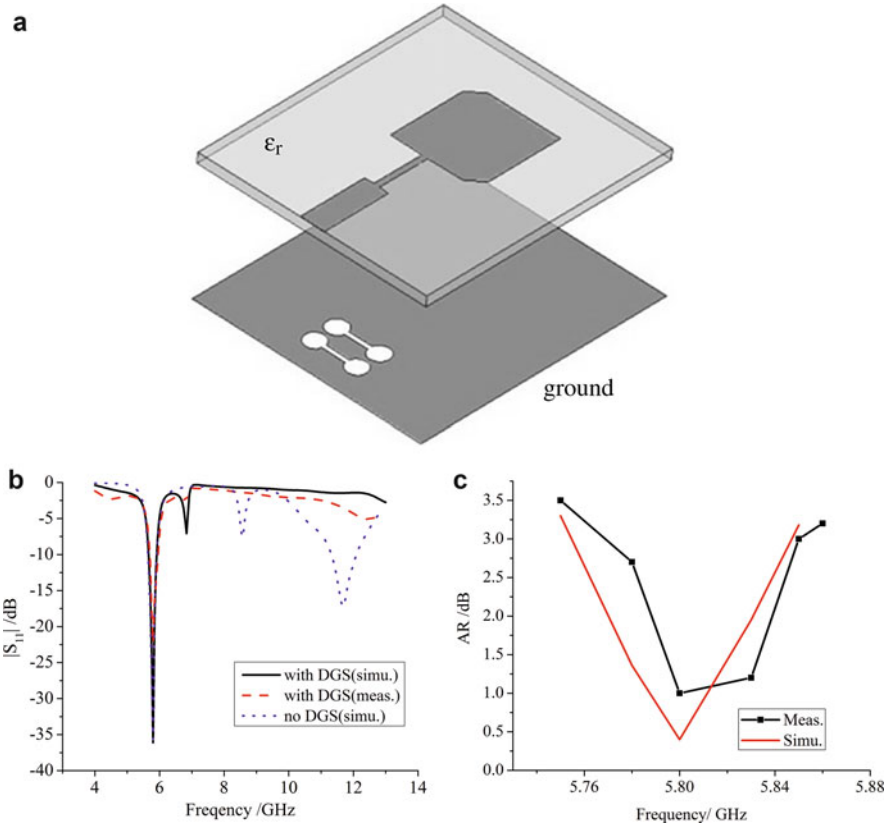


Fig. 11 Circularly polarized truncated-corner patch antenna. (**a**) Antenna structure, (**b**) simulated and measured S_{11} versus frequency, and (**c**) simulated and measured AR versus frequency

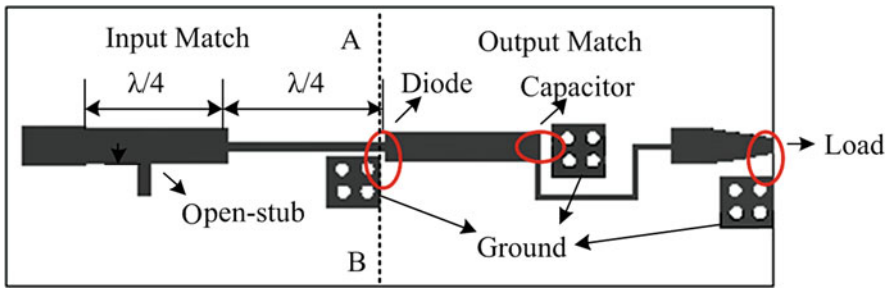


Fig. 12 Schematic diagram of the rectifying circuit

Rectenna with Data Communication Function

Usually, electrical devices obtaining DC power from the microwave also need transmitting and receiving information. A general scheme is designing two antennas, one for communication and another for the rectenna. However, a rectenna with communication function would have two good advantages. First, the RF front end would be compact, which is a critical demand for the most electronic devices. Second, the electromagnetic compatibility could be considered and optimized when designing the antenna for rectenna and communication simultaneously. A triple-band implantable antenna with a rectenna was designed for biotelemetry communication in Huang et al. (2011), in which data telemetry was at 402 MHz, wireless powering transmission at 433 MHz, and wake-up controller at 2.45 GHz. Although the rectifying circuits had a mw-DC conversion efficiency of 86 % at 433 MHz, the rectenna's efficiency

DOI 10.1007/978-981-4560-75-7_116-1

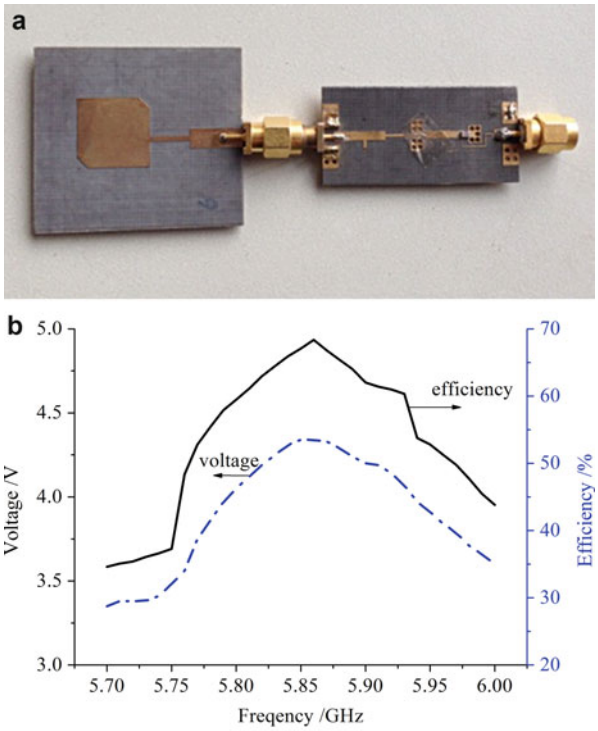


Fig. 13 Rectenna picture and the measured efficiency. (**a**) Picture of rectenna and (**b**) measured voltage and mw-DC efficiency versus frequency

was very low because of the low receiving power capacity of the antenna, which shows that the impedance match is very important for a rectenna design. The example as follow is a rectenna with communication capability (Yang et al. 2013).

Dual-Polarized Antenna Design

The receiving antenna is a dual-polarized square-patch antenna fed by aperture coupling as shown in Fig. 14a. The antenna has two layers of substrate with depths of h_1 and h_2. The square patch is on the top of the upper substrate and the microstrip feedlines are on the bottom plane of the lower substrate. The two orthogonal H-shaped apertures, located in the perfectly conducting plane placed between the two substrates, were used to improve the isolation between the two feed ports. Two orthogonal feed ports V- and H-port excite vertical and horizontal polarization waves, which are used as power supply and data communication, respectively.

The operating frequency f_0 is mainly determined by the side length of the square patch with less influence from the aperture position. This double-layer and aperture coupling fed patch antenna provides higher gain than the single layer and good suppression of harmonic waves so the filter in the rectifying circuit is omitted and the rectenna is compact. The detailed geometrical parameters and the substrate characteristics could be referred to Yang et al. (2013).

The simulated and measured reflection coefficients in terms of $|S_{11}|$ of the V-port, S_{22} of the H-port, and the isolation $|S_{21}|$ versus frequency are displayed in Fig. 14b. For the two polarizations, the measured lowest values of $|S_{11}|$ are -30 and -18 dB at their corresponding center frequencies of 5.8 and 6.1 GHz, respectively, which is mainly caused by the different feed positions and structure. The isolation $|S_{21}|$ is above 25 dB within a broad bandwidth. The V-port, operating at 5.8 GHz, is used as power transformer, while the H-port, operating at 6.1 GHz, is used for communication systems. A wideband simulation shows that the reflection coefficient of the second-order harmonic wave at 11.6 GHz is less than 2 dB. This

DOI 10.1007/978-981-4560-75-7_116-1

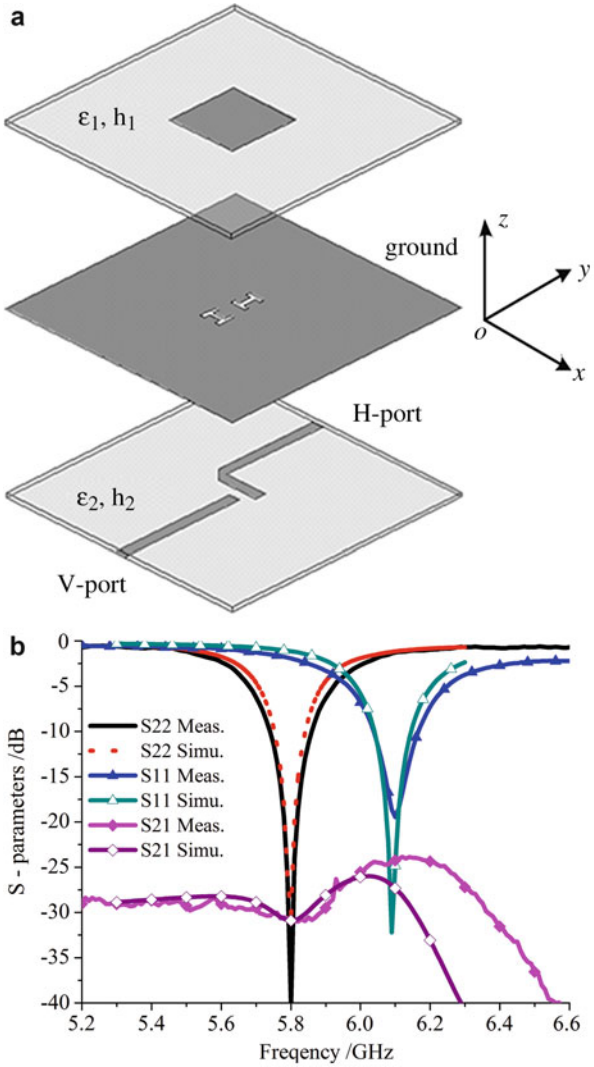


Fig. 14 Configuration of the receiving antenna and S parameter versus frequency. (**a**) Perspective view of the receiving antenna and (**b**) $|S_{11}|$, $|S_{22}|$, and $|S_{21}|$ versus frequency

feature mainly attributes to the antenna symmetry and the aperture coupling feed structure and a special LPF/BRF will not be required for the rectenna. The maximum gains of the H-port for the communication and V-port for the rectifying circuit at their corresponding resonant frequencies are 7.6 and 7.2 dBi, respectively.

Rectifying Circuit Design

The input matching network and the DC-pass filter of the rectifying circuit are designed using microstrip lines as shown in Fig. 15. The substrate is the same as that for the receiving antenna. The open stub of microstrip line 2 and the short stub of microstrip line 3 are used to cancel out the imaginary part of the diode's input impedance Z_{in}, while microstrip line 1 is a $\lambda/4$ impedance transformer to match the impedance to 50 Ω at the input port. Three fan-shaped stubs of different sizes act as a DC-pass filter to smooth the output DC power.

The diode used in the circuit is a packaged Schottky diode of HSMS2860. Apart from the parameters in Fig. 7, the equivalent circuit includes the parasitic inductance L_p and the parasitic capacitance L_c. The equivalent parameters are listed in Table 1. According to ADS simulation, the input impedance Z_{in} of the

DOI 10.1007/978-981-4560-75-7_116-1

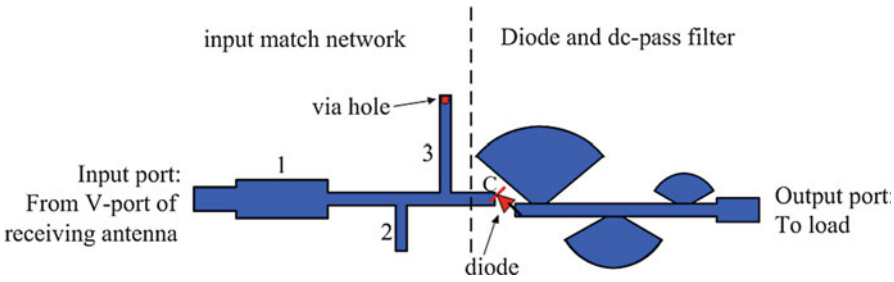


Fig. 15 Structure of the rectifying circuit

Table 1 Equivalent parameters of packaged Schottky diode HSMS2860

V_f Turn-on voltage	V_B Breakdown voltage	L_p Parasitic inductance	C_p Parasitic capacitance	R_s Junction resistance	C_{j0} Junction capacitance
0.3 V	7 V	0.5 nH	0.08 pF	6 Ω	0.18 pF

diode is (22-j55) Ω with 10 mW input power and 1200 Ω load at 5.8 GHz operating frequency to obtain the maximum mw-DC efficiency. Microstrip lines 1, 2, and 3 match the diode input impedance to 50 Ω.

The simulated reflection coefficient at the operation frequency of 5.8 GHz and the second harmonic frequency of 11.6 GHz are −40.2 and −0.13 dB, respectively. So the input port is matched well at the operating frequency and the higher-order harmonics are effectively suppressed. The simulated $|S_{21}|$ of the direct pass filter for the fundamental wave and the second and third harmonics are 50, 25, and 30 dB, respectively, which shows that the fundamental wave and the harmonics are inhibited successfully.

Figure 16 shows the simulated and measured mw-DC efficiency of the whole rectifying circuit. The simulation results show that this rectifying circuit had maximum efficiency of 81 % on the condition of 1200 Ω load and 10 mW input power at 5.8 GHz operating frequency. Experiments showed that the operation frequency with the highest efficiency was 5.78 GHz. The highest conversion efficiencies of 68.7 %, 69.3 %, and 69.9 % were measured on the loads of 1200, 900, and 600 Ω when the input power was 20, 25, and 35 mW, respectively. The differences between the simulations and the experiments lie mainly on several factors: the fabrication errors, the diode model in ADS, and the measurement errors.

The measured optimized condition of 900 Ω load and 5.78 GHz operation frequency would be preferred in the measurements of rectenna in the following section.

Rectenna Performances

The antenna and the rectifying circuit are frequency dependent; hence, a little mismatch would cause large losses. The co-simulation of the receiving antenna and rectifying circuit using HFSS and ADS becomes very important for a high mw-DC conversion efficiency design. To investigate the rectenna performances, the co-simulation by ADS has been carried out. The characteristic of the antenna's input impedance versus frequency could be embedded in the rectifying circuit's input port by S1P file with identical format. The simulated $|S_{11}|$ of the rectenna, receiving antenna, and rectifying circuit versus frequency are then displayed in Fig. 17. Comparing the dash line for the rectifying circuits and the dash dotted line for the receiving antenna reveals that the rectifying circuit has a wider bandwidth than the receiving antenna. From Fig. 17, the simulated power reflection coefficients at 5.8 GHz for the receiving antenna, the rectifying circuit, and the rectenna are −28.1, −40.2, and −27.7 dB, respectively. Thus, a close to perfect impedance match could be assumed and the receiving power of the antenna is considered equal to the

DOI 10.1007/978-981-4560-75-7_116-1

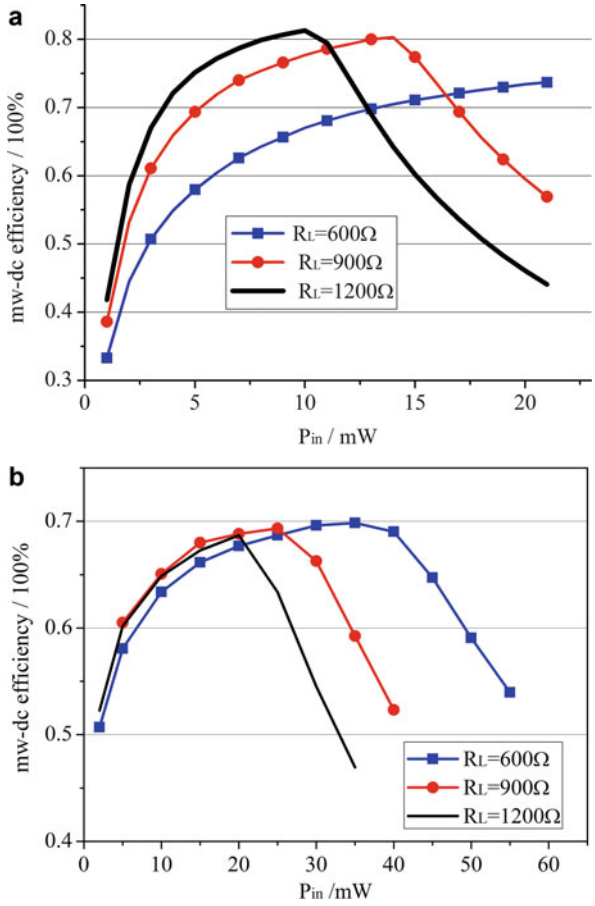


Fig. 16 Efficiency versus input power of the rectifying circuit: (**a**) ADS simulation at 5.8 GHz and (**b**) measurement at 5.78 GHz

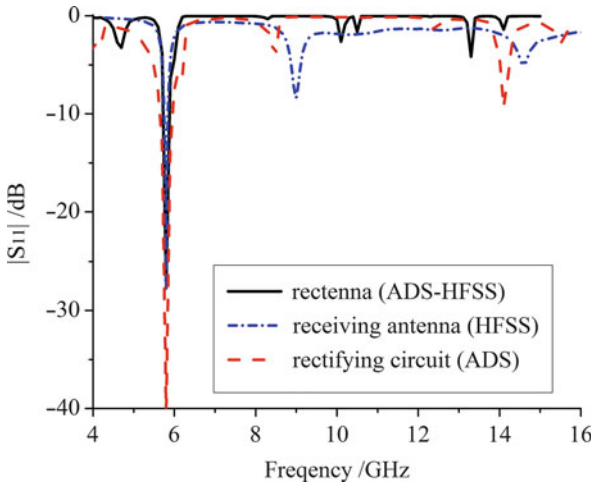


Fig. 17 Comparison of reflection coefficients

DOI 10.1007/978-981-4560-75-7_116-1

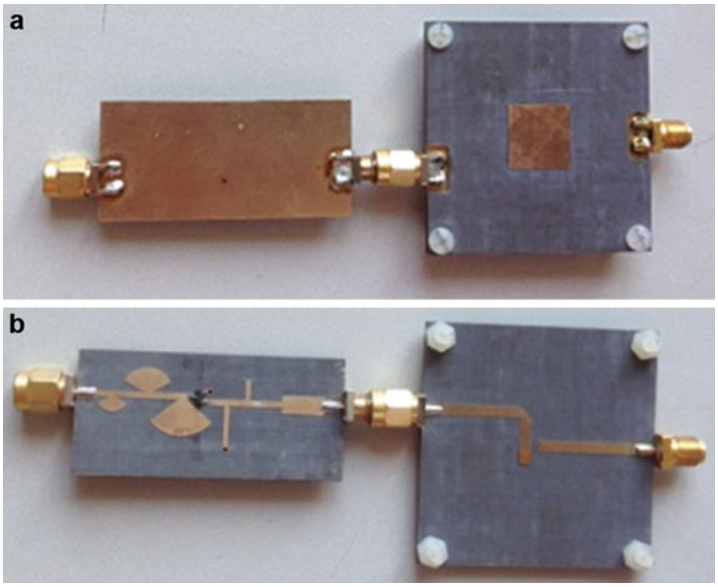


Fig. 18 Photograph of the rectenna: (**a**) front view and (**b**) back view

input power of the diode. The antenna bandwidth is narrower than that of the circuit so the rectenna bandwidth will be mainly decided by that of the antenna.

The receiving antenna and the rectifying circuit were connected by an SMA connector with 50 Ω characteristic impedance as shown in Fig. 18. The mw-DC conversion efficiency was measured by the measurement system shown in Fig. 6. The transmitting horn had a gain of 15.6 dBi and a size of 13.5 × 10 cm. The rectenna was fixed at the distance of 80 cm from the horn, which is in the far-field region. The measured gain of the receiving antenna was 7.0 dBi, which is used to calculate the received RF power of the rectifying circuits.

The output voltage and conversion efficiencies of different loads with the receiving power at 5.78 GHz operation frequency are illustrated in Fig. 19. The rectenna with a higher load achieved the maximum output voltage and conversion efficiency at a relatively low receiving power. The output voltages of the rectenna with various loads approached 4.5 V, which was the breakdown voltage of the diode. The highest conversion efficiency of 63 % was measured on the load of 900 Ω when the receiving power was 25 mW.

Communication Performances

To investigate the communication characteristics of the H-port, the rectenna model shown in Fig. 20 is simulated in HFSS software. The power reflection coefficient $|S_{11}|$ has nearly no change so it is not provided here. The gain patterns for the communication port deteriorate a little, as shown in Fig. 21. Nevertheless, the cross-polarization is still low and is approximately −20 dB in the main radiation direction. The gain decreases to 7.2 dBi from the expected 7.6 dBi because of the rectifying circuit.

When the antenna is connected to the rectifying circuit, the measured reflection coefficient of the communication H-port has nearly no change and is similar to the experimental $|S_{11}|$ curve of Fig. 14b. Figure 22 is the measured patterns of the H-port with and without the circuit. Compared to the simulation results of Fig. 21, the E-plane characteristic is good with −15 dB cross-polarization in the main radiation direction, even though there is an obvious sidelobe in the 120° direction on the H-plane. The measured maximum gain is 7.0 dBi.

The compact size, the ease of fabrication, and the compliance with safety standards are the key advantages of this rectenna.

DOI 10.1007/978-981-4560-75-7_116-1

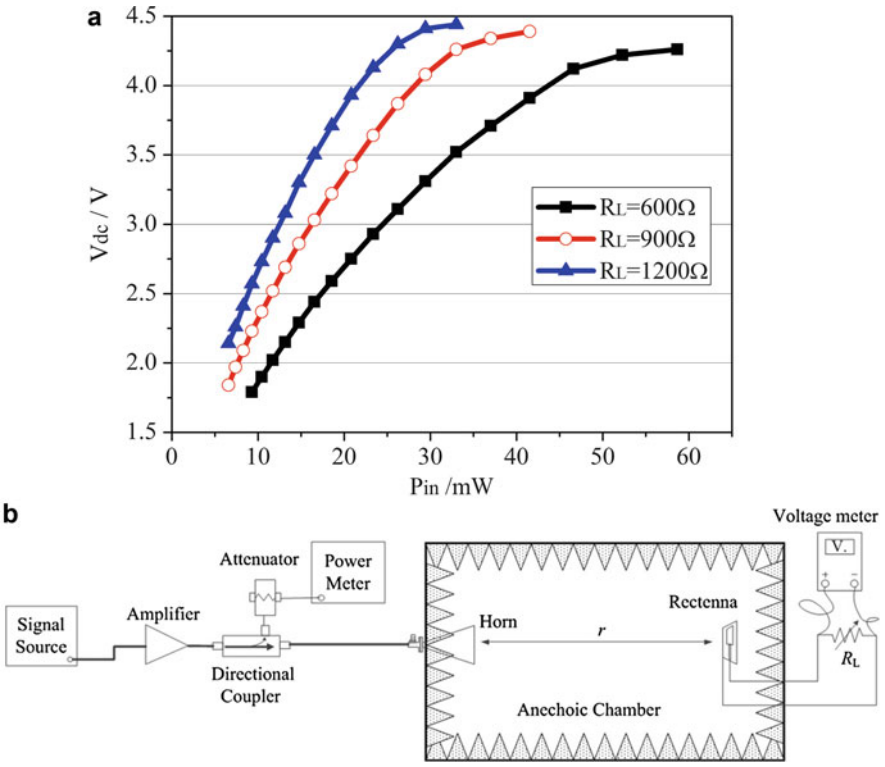


Fig. 19 Measured output voltage and efficiency of the rectenna versus the input power at 5.78 GHz. (**a**) Output voltage and (**b**) mw-DC conversion efficiency

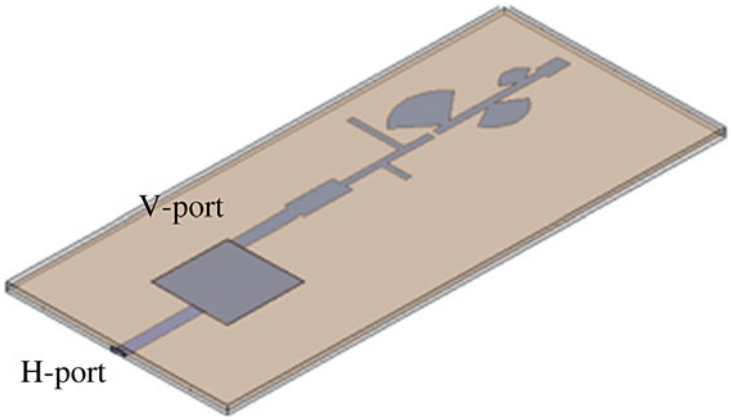


Fig. 20 Model of rectenna in HFSS

CPW Rectenna with Broad Operation Band

CPW transmission lines have several advantages, such as the easiness of integration with active and passive elements, high circuit density, low dispersion, low radiation loss, and avoidance of ground vias, which are preferable for rectifying circuits design. Grounded coplanar waveguide (GCPW) is a modification of CPW. It has an additional ground plane on the back of the substrate. Combining CPW and GCPW transmission lines, a broadband CPW rectenna was designed (Nie et al. 2015) and Fig. 23 plots the CPW rectenna.

DOI 10.1007/978-981-4560-75-7_116-1

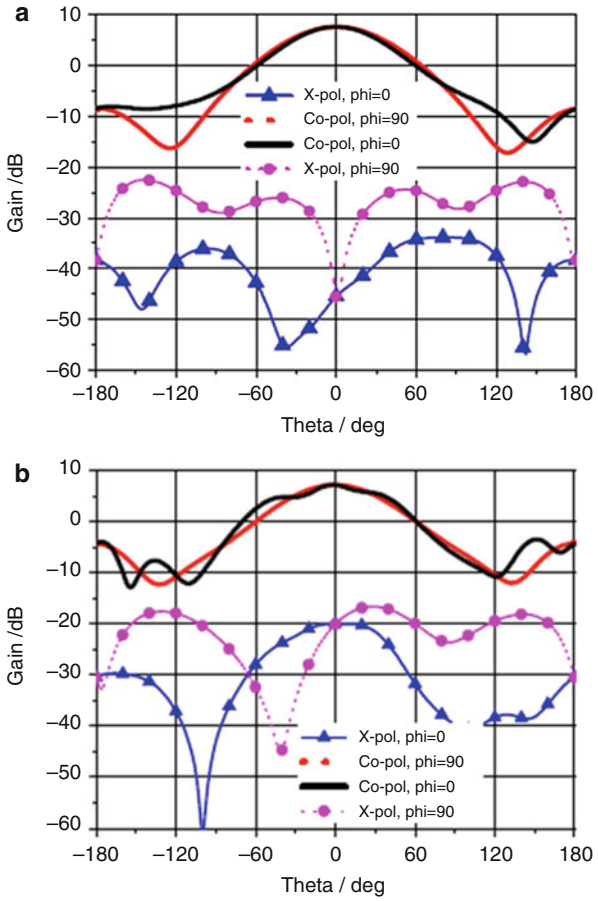


Fig. 21 Pattern comparison with and without rectifying circuit simulated by HFSS. —— E-plane/Co-P, ····· H-plane/Co-P, —+— E-Plane/X-Pol, ··▲·· H-Plane/X-Pol. (**a**) Without rectifying circuit. (**b**) With rectifying circuit

The rectifying circuit consists of a capacitor, a Schottky diode pair, and a DC-pass filter, which are connected in shunt or series on the same front side of the substrate of the GCPW transmission line. The characteristic impedance of the GCPW transmission line is 50 Ω. The input impedance Z_{in} of the diode is (64.2 + j9.8) Ω at 2.45 GHz when the input power is 15 dBm from ADS simulation. Z_{in} is matched to 50 Ω at the input port of the rectifying circuit by designing the input and the output impedance match networks. The input impedance match network is realized by the GCPW sections of L_3 and L_4. The output impedance match network consists of the capacitor pair of C_1 and C_2 and the GCPW section L_5.

In order to match the GCPW rectifying circuit directly, the antenna is fed by the CPW with the length of L_2, while the beginning section is GCPW structure, which could be recognized from Fig. 23b. In order to increase the antenna gain by directing its broadside beam in one direction, a reflecting plane is placed behind the antenna at the distance of $D = 0.18\lambda_0$.

The center frequency is 2.45 GHz when the slot length L_1 is $0.408\lambda_0$. The input impedance of the antenna is well matched to 50 Ω by carefully tuning the length L_1 of the CPW stub. Both the simulated and measured $|S_{11}|$ are found to be better than −30 dB at 2.45 GHz as plotted in Fig. 24a. The measured −10 dB bandwidth is more than 28.6 % (from 2.0 to 2.7 GHz). A good impedance match performance within a broad bandwidth has been achieved. The measured peak gain is 10 dBi at the center frequency of 2.45 GHz, while the simulated one is 10.3 dBi from Fig. 24b. The measured gain is stable and varies between 9.4 and 10 dBi in the bandwidth from 2.0 to 2.5 GHz.

DOI 10.1007/978-981-4560-75-7_116-1

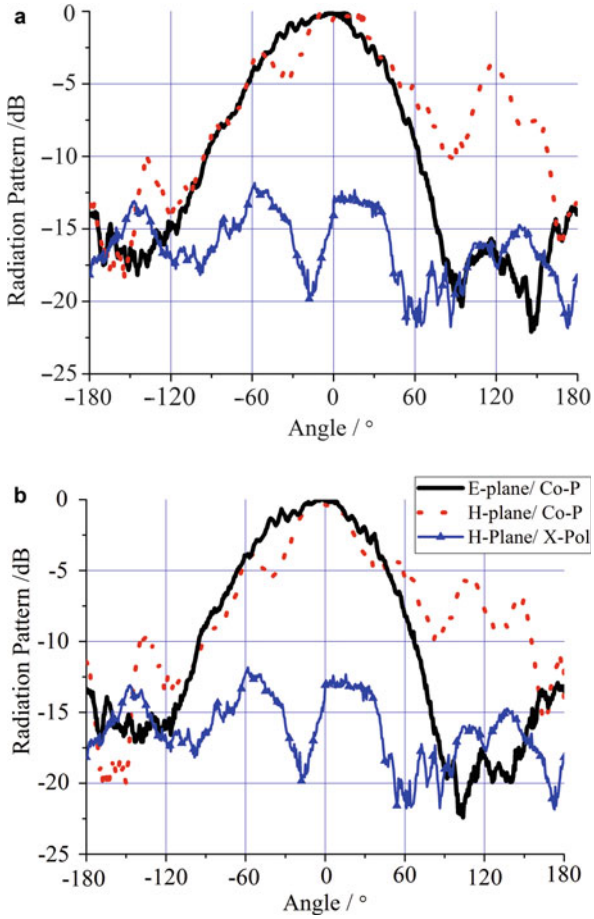


Fig. 22 Measured gain patterns with and without rectifying circuit. —— E-plane/Co-P, ····· H-plane/Co-P, ···▲··· H-Plane/X-Pol. (**a**) Without rectifying circuit. (**b**) With rectifying circuit

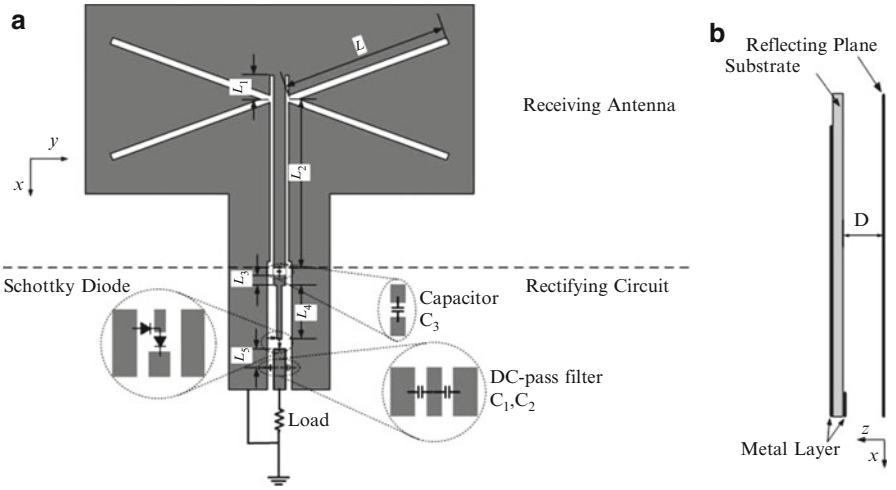


Fig. 23 Configuration of the CPW rectenna. (**a**) Top view. (**b**) Side view

Figure 25 illustrates the simulated and measured gain patterns in E-plane and H-plane at 2.45 GHz. Its half-power beamwidth (HPBW) is about 60°, which alleviates the demand of the rectenna alignment.

DOI 10.1007/978-981-4560-75-7_116-1

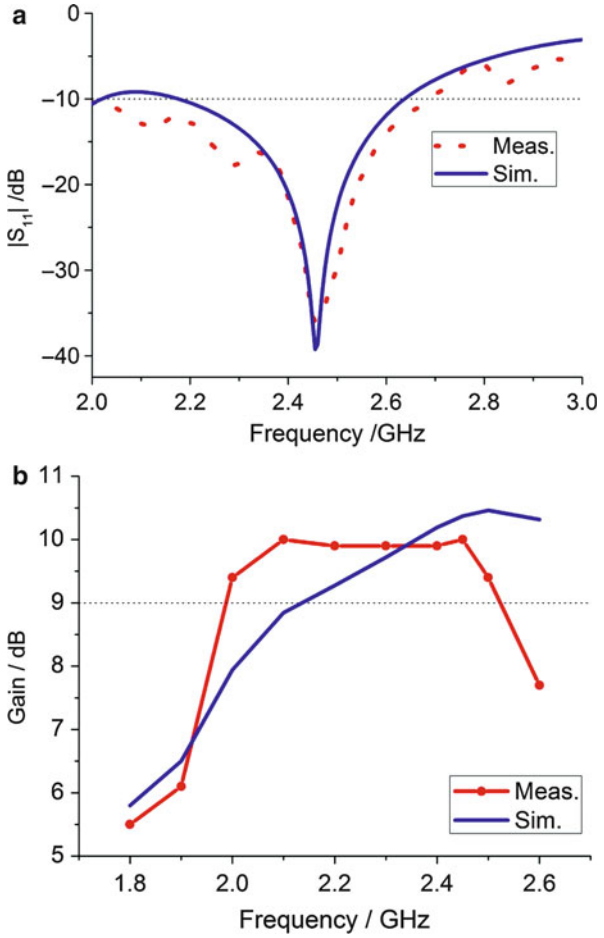


Fig. 24 Simulated and measured $|S_{11}|$ and gain versus frequency of the antenna. (**a**) $|S_{11}|$ and (**b**) gain

The measured conversion efficiencies of the rectenna are shown in Fig. 26. From Fig. 26a, the peak conversion efficiencies for both the rectenna and the rectifying circuit are about 62 % on the same load of 900 Ω, which show good impedance match between the receiving antenna and the rectifying circuit. From Fig. 26b, it can be found that the bandwidth of the efficiencies higher than 50 % is from 2.2 to 2.6 GHz (16.3 %) at 13 dBm received power. This broadband performance benefits from the wide operation bandwidths of the antenna and the rectifying circuit.

This CPW rectenna has a wide operation bandwidth. The receiving antenna has a high gain and a wide half-power beamwidth, which lessens the demand of the strict direction alignment. The rectifying circuit is designed by the GCPW transmission line, which avoids the use of ground vias and it is easy for the diodes and capacitors to be connected.

Rectenna Arrays

The power capacity of a Schottky diode is limited, so rectenna array design is necessary to obtain higher DC output power for most applications. A rectenna element can be equivalent to the series connection of a voltage source E and a load R_L. R_L, which is relative to the input mw power to the diode, is the optimized load with the aim of obtaining the highest mw-DC conversion efficiency. Rectenna arrays have three kinds of connections: series, parallel, and series/parallel mixed connections.

The rectenna element with communication function in section "Rectenna with Data Communication Function" will be referred to investigate the operation laws of rectenna arrays. The Schottky diodes used

DOI 10.1007/978-981-4560-75-7_116-1

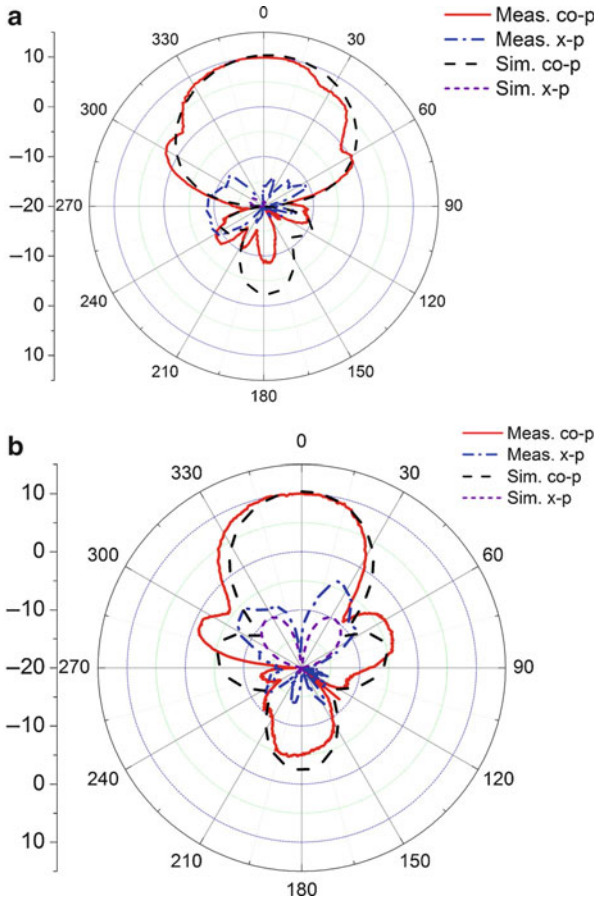


Fig. 25 Simulated and measured gain pattern at 2.45 GHz of the antenna in (**a**) E-plane (*xoz*-plane) and (**b**) H-plane (*yoz*-plane)

in the circuits are also the packaged Schottky diode of HSMS2860, whose equivalent circuit parameters are listed in Table 1.

Parallel Connection of Rectenna Arrays

Figure 27 schematically shows the parallel connection of n rectenna elements, which has the output DC power of nE^2/R_L. Based on the equivalent circuit principle, this parallel connection can be simplified as the series connection of a source of E and a resistor of $R_L./n$, whose equivalent load R_p and equivalent current I_p are

$$R_p = R_L/n \tag{4}$$

$$I_p = nE/R_L \tag{5}$$

Two-element parallel array is simulated by ADS software. The output DC power and mw-DC conversion efficiency versus the load are shown in Fig. 28. Apparently, the output current is two times of that of the element when the load is lower than 600 Ω, while the output voltage moves to the saturation of 3.2 V when the load becomes higher. The optimized loads of the element and the two-element array are 1200 Ω and 600 Ω, respectively, which meet with formulas (4) and (5) of the equivalent model. It can be seen that the maximum mw-DC efficiencies for the element and the two-element array are all about 81 %.

DOI 10.1007/978-981-4560-75-7_116-1

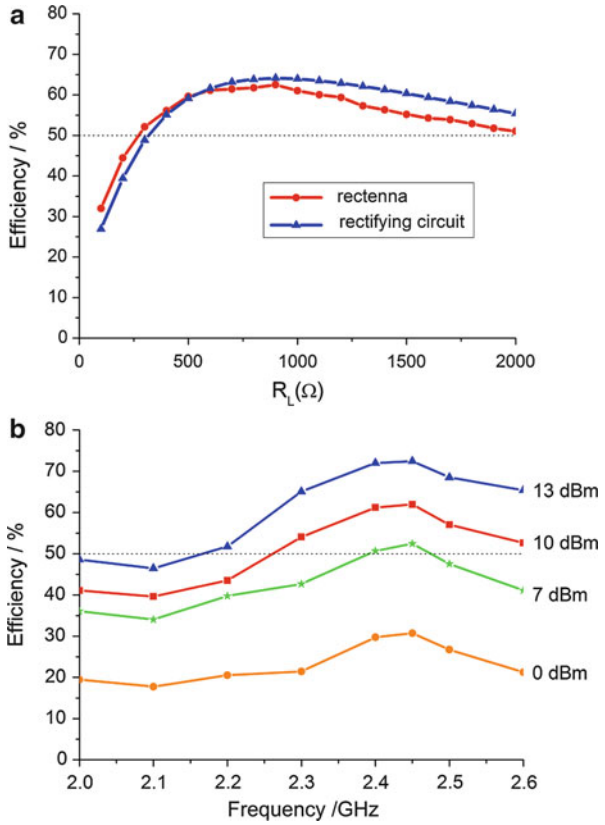


Fig. 26 Measured efficiency of the rectenna. (**a**) Efficiency versus load. (**b**) Efficiency versus frequency

Figure 29 gives the picture of the two-element array, the measured output voltage, and the mw-DC

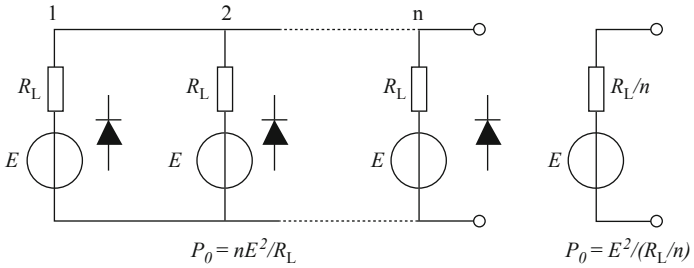


Fig. 27 Equivalent circuit model and its simplified model of parallel connection of n rectenna elements

efficiency versus the load with the element results being compared. The trends also accord with formulas (4) and (5). The maximum mw-DC efficiency of two-element array is about 60 %, which is lower than that of the element of 65 %. The difference between the simulated and measured results has been analyzed previously.

Series Connection of Rectenna Arrays

The series connection of n rectenna elements and the corresponding simplified equivalent circuits are illustrated in Fig. 30. The output DC power is also nE^2/r. The equivalent load R_S and equivalent voltage U_S are as follows:

DOI 10.1007/978-981-4560-75-7_116-1

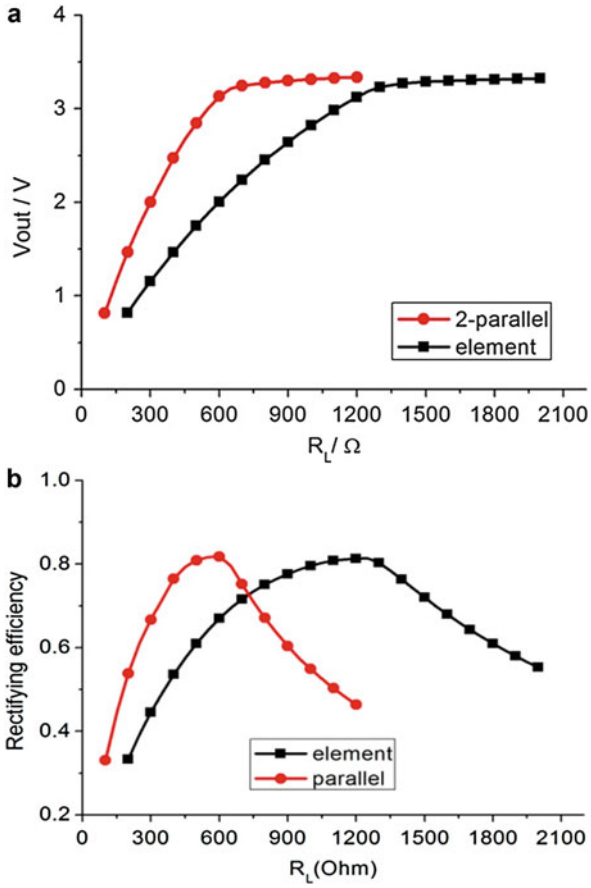


Fig. 28 ADS simulated output voltage and efficiency of element and two-element parallel array. (**a**) Output voltage versus load. (**b**) mw-DC efficiency versus load

$$R_s = nR_L \tag{6}$$

$$U_s = nE \tag{7}$$

The picture of the two-element series array and the measured output voltage and mw-DC efficiency are illustrated in Fig. 31 with the measured results of the element being plotted for comparing. When the load is less than 500 Ω, the diodes are not saturated, so the current on the loads of the element and the array are the same. The optimized loads of the element and the series array are 800 Ω and 1600 Ω, with the corresponding mw-DC efficiencies of 64.3 % and 62.1 %, respectively. Formulas (6) and (7) are validated.

So the conclusions can be drawn as follows. The array design accords with the general circuit theory of series and parallel connections when a rectenna element is equivalent to the series connection of a voltage source E and a load R_L. The optimized load of a series array is n times that of the element, while the current is the same if the diodes are not saturated. The optimized load of a parallel array is 1/n times that of the element and the output voltage is the same. At the same time, the minimum distance between two elements must comply with the principle of antenna array design. Extra narrow distance between two rectenna elements will decrease the receiving efficiency of the antennas, while extra wide one will enlarge unnecessarily the rectenna array area.

DOI 10.1007/978-981-4560-75-7_116-1

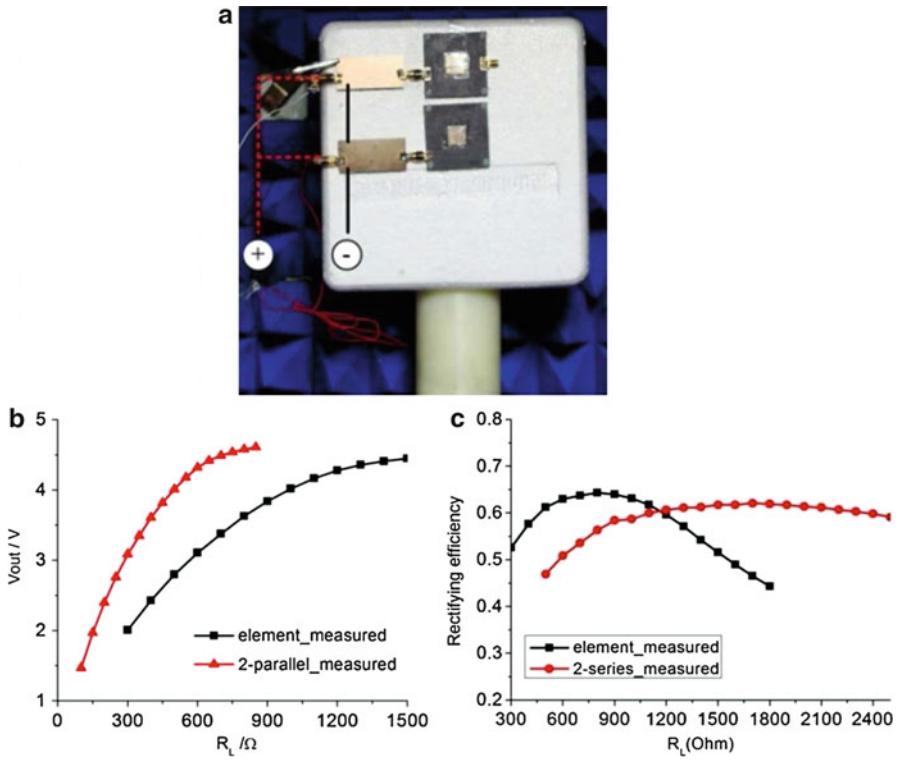


Fig. 29 Two-element parallel array and measured output voltage and efficiency. (**a**) Array picture. (**b**) Output voltage versus load. (**c**) mw-DC efficiency versus load

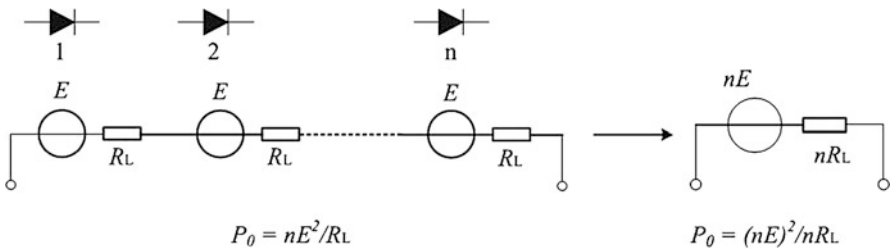


Fig. 30 Equivalent circuit model and its simplified model of series connection of *n* rectenna elements

Series/Parallel Mixed Connections of Rectenna Arrays

Generally, the load is given for an actual application so the array design should be carried out to get the maximum mw-DC efficiency after a rectenna element has been optimized. Two kinds of six-element rectenna arrays are shown in Fig. 32. Figure 32a is two branches in parallel connection after three elements in series, whose load is $3R_L/2$. Figure 32b is three branches in parallel after two elements in series, whose load is $2R_L/3$. Two arrays have the same output DC power of $P = 6E^2/R_L$ and different optimal loads.

Thus, if an array has n branches in parallel and every branch has n elements in series connection, the optimized load of the array is the same as that of the element of R_L.

A circularly polarized rectenna element in Fig. 33 is used as the array design. The quasi-square patch fed by aperture coupling is used as the circularly polarized receiving antenna. The rectifying circuit is on the same layer with the microstrip feedline of the antenna. This double-layer structure not only minimizes

DOI 10.1007/978-981-4560-75-7_116-1

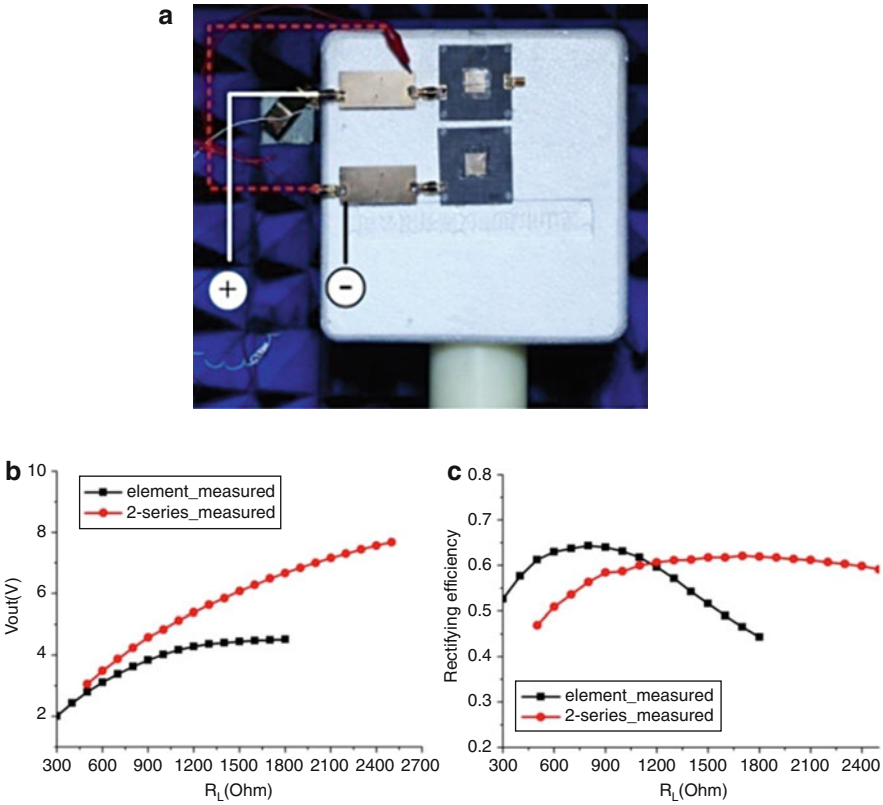


Fig. 31 Picture of two-element series array and measured output voltage and efficiency. (**a**) Picture. (**b**) Output voltage versus load. (**c**) mw-DC efficiency versus load

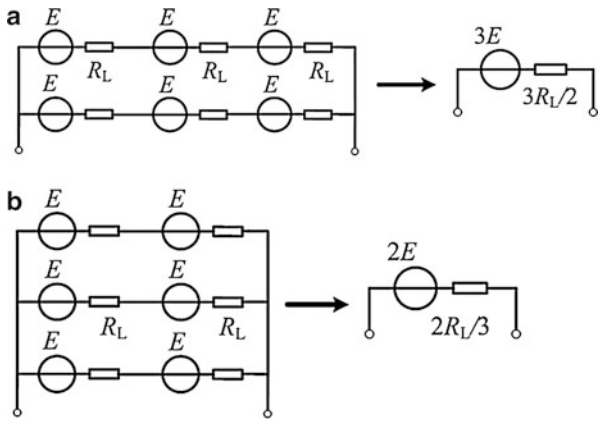


Fig. 32 Equivalent models of two six-element array with different loads and same DC output power. (**a**) Two branches with three elements in series. (**b**) Three branches with two elements in series

the size of the rectenna but also decreases the effects of the circuits on the antenna. To make the rectenna array compact, several microstrip lines are bended, so the rectifying circuit is just below the antenna.

The rectenna operates at C-band. The measured mw-DC efficiencies versus the loads on different input microwave powers are plotted in Fig. 34. From Fig. 34a, a maximum mw-DC efficiency of 63 % is measured when the input power is about 85 mW and the optimized load is 250 Ω. From Fig. 34b, the mw-DC efficiency on the 45° direction is only lower 8 % than that of the 0° direction, which shows a preferable circular polarization performance.

DOI 10.1007/978-981-4560-75-7_116-1

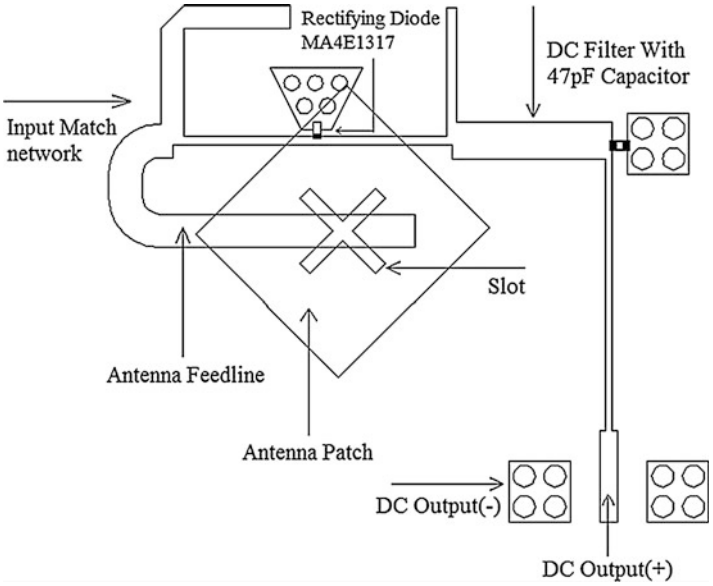


Fig. 33 Circularly polarized rectenna element

A rectenna array of 3 × 3-element with three branches in parallel and every branch having three elements in series connection is designed as illustrated in Fig. 35. The measured output DC voltage and the mw-DC efficiency versus loads are shown in Fig. 36 with the element results being plotted for comparing. When the input mw power P_{in} for every rectenna element is 81.1 mW, the measured maximum mw-DC efficiency for the array is 63 %, while the optimal loads for the array and the element are the same with the values of 250 Ω from Fig. 36a. The center frequency is 6.04 GHz and the maximum efficiency direction is at 45°. The output DC voltages of the array are about nine times by those of the element from Fig. 36b. The output DC power is about 461 mW on the load of 250 Ω with the receiving aperture of 12 × 10 cm.

The measured maximum efficiencies of the rectenna element and the 3 × 3-element array both occurred at 6.04 GHz while the designed center frequency had been 5.8 GHz, which was the main factor that decreased the efficiency. Another factor was that the input microwave power was 84.9 mW in the experiment. The mw-DC efficiency could be improved in some degree until the breakdown voltage of 4.5 V is reached, on which the input mw power is about 100 mW. So the mw-DC conversion efficiency of this sub-array could be expected to improve to 75 % after a careful design. Considering the rectenna array losses, the mw-DC efficiency is estimated to be 65 % for a large array. Under this assumption of 100 mW input mw power and 65 % conversion efficiency, 585 mW DC power could be expected on the rectenna array size of 12 × 10 cm.

Transmitting Antennas Design and *BCE*

The most concerned performance for a point-to-point MPT system is the DC-DC efficiency. Early in 1964, Degenford et al. studied level distribution of Gaussian function on the transmitting antenna aperture and verified experimentally on a reflecting beam waveguide with near 100 % *BCE* (beam capture efficiency) being obtained (Degenford et al. 1964). However, the reflecting beam waveguide propagation path model is impractical for the applications of long-distance microwave transmission. 10 dB Gauss amplitude tapered beam has been verified to be efficiently transferring the power from the transmitting

DOI 10.1007/978-981-4560-75-7_116-1

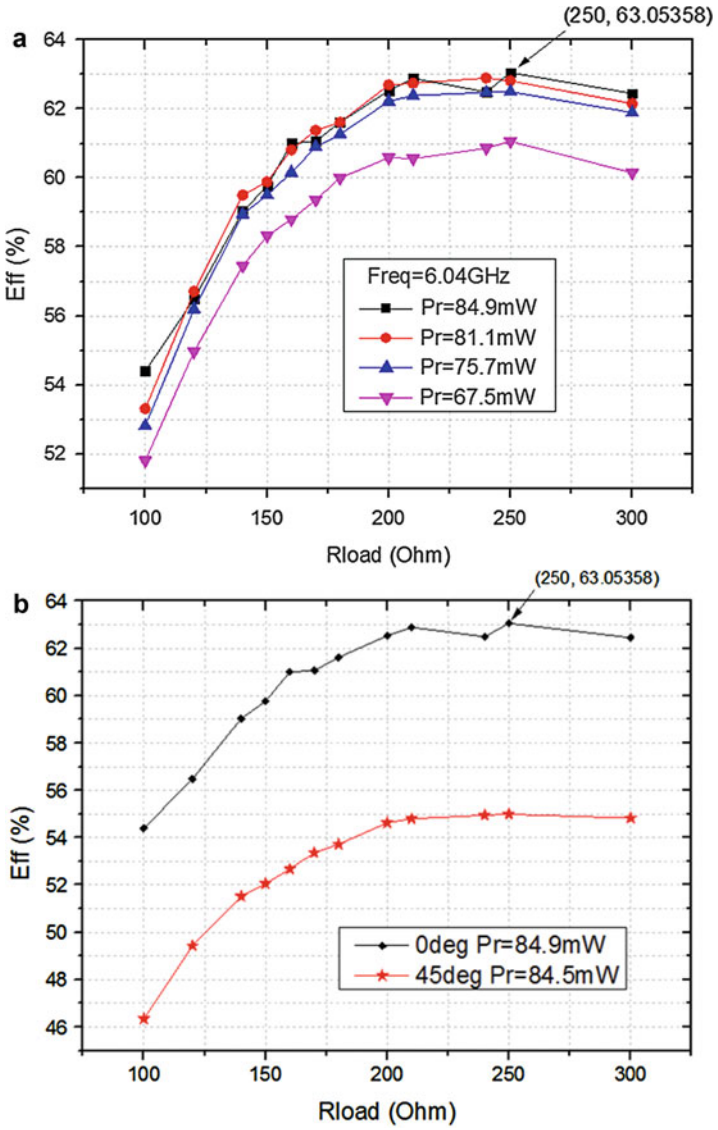


Fig. 34 Measured mw-DC efficiency of the rectenna element versus load. (**a**) Different input mw power. (**b**) Different receiving directions

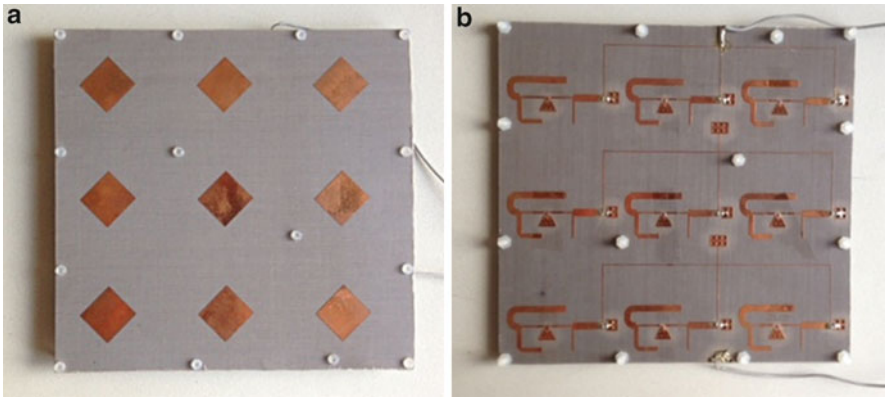


Fig. 35 Pictures of the 3 × 3-element array: (**a**) front view and (**b**) back view

DOI 10.1007/978-981-4560-75-7_116-1

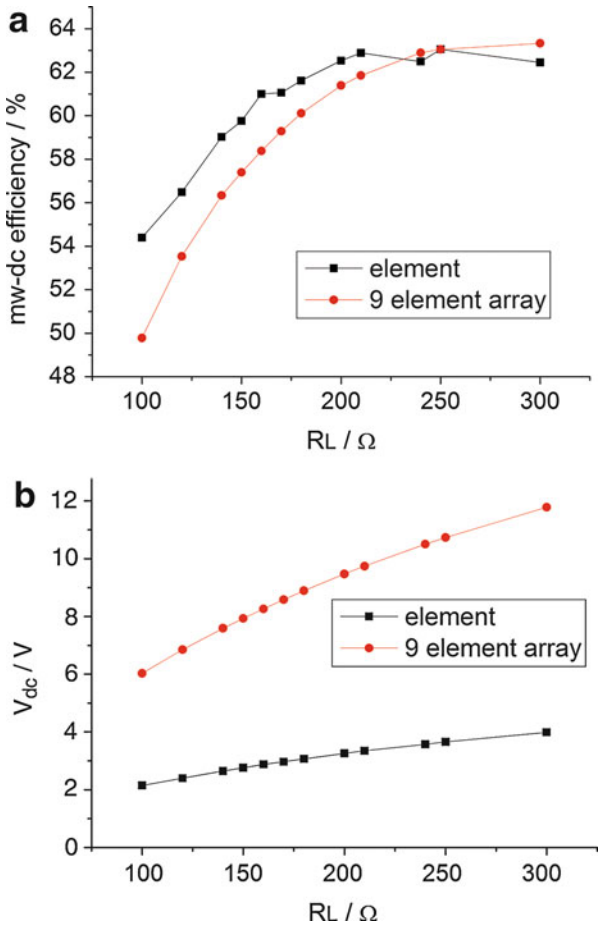


Fig. 36 Measured mw-DC efficiency and DC voltages versus loads of 3 × 3 array and element. (**a**) mw-DC efficiency and (**b**) DC voltage

aperture to the receiving rectenna array (Murao 2000). The high level at the center of the transmitting and the receiving apertures easily damages the center microwave devices so the 10 dB Gauss tapered beam is limited in some practical applications. Some kinds of truncated taper beams were studied in Zepeda et al. (2003). In 2013, Oliveri et al. investigated the maximum *BCE* optimization by using eigenvalue method (Oliveri et al. 2013). It was concluded that near 100 % *BCE* for any shape of transmitting and receiving apertures could be obtained only if the apertures are large enough. Normally, the transmitting antenna and the receiving rectenna have square apertures, which are convenient to be manufactured and fixed. The circular apertures of the transmitting and the receiving antennas can capture more microwave beam at a low budget on the concept that every element of the transmitting antenna operates efficiently (Zhou and Yang 2014).

This section mainly discusses two schemes for the maximum *BCE* of MPT systems. One is the topological models of square, truncated square, and circle for the transmitting apertures. Another is the various level distributions on the transmitting apertures.

Topological Models of Transmitting Antenna

The transmitting antenna aperture is composed of N radiating elements, numbered as 0, 1, 2. . .$N-1$ as shown in Fig. 37. The nth element is located at (x_n, y_n), $n = 0, 1, 2. . .N-1$, with the normalized element weights w_n. Considering the directivity of the transmitting antenna, the receiving rectenna in the far-field

DOI 10.1007/978-981-4560-75-7_116-1

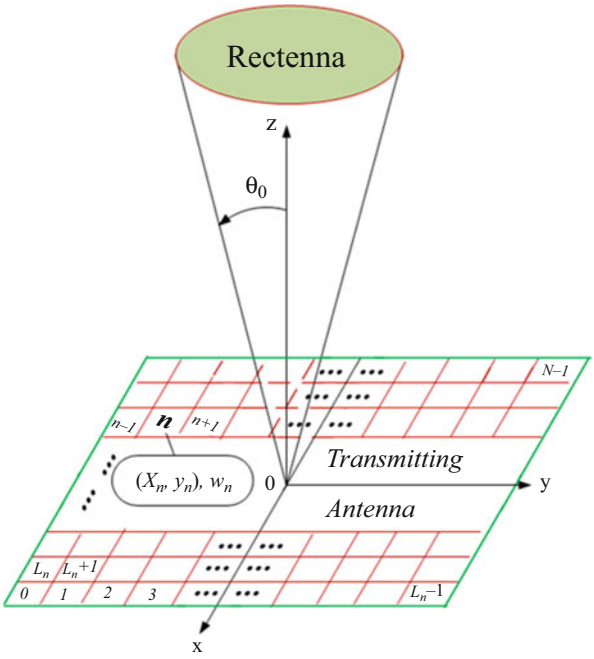


Fig. 37 Transmitting and receiving apertures of a MPT

region is supposed to have a circular aperture. The distance between the transmitter and the receiver is r and the receiving angle of the rectenna aperture is θ_0.

Every element of the transmitting aperture is equivalent to a square Huygens element and is a source with identical amplitude E_n and inphase angle. The electrical field strength in the radiation field is

$$E = \frac{C_0}{r} f_e \sum_{n=0}^{N-1} w_n e^{jk(ux_n + vy_n)} \tag{8}$$

where parameters u and v are $u = \sin\theta\cos\varphi$ and $v = \sin\theta\sin\varphi$, respectively. f_e is the normalized pattern of the Huygens element and C_0 is the coefficient as follows:

$$f_e = \frac{1 + \cos\theta}{2} \operatorname{sinc}\left(\frac{kau}{2}\right) \operatorname{sinc}\left(\frac{kbv}{2}\right) \tag{9}$$

$$C_0 = \frac{jk}{2\pi} e^{-jkr} E_n ab \tag{10}$$

Supposing the level distribution is symmetrical to the center, E of the radiation field could be calculated as

$$E = \frac{C_0}{R} f_e \left[\sum_{\substack{x_n>0 \\ y_n>0}} 4w_n \cos(kux_n) + \sum_{\substack{x_n=0 \\ y_n\geq 0}} 2w_n \cos(kvy_n) + \sum_{\substack{x_n\geq 0 \\ y_n=0}} 2w_n \cos(kux_n) + w_n = y_n = 0 \right] \tag{11}$$

DOI 10.1007/978-981-4560-75-7_116-1

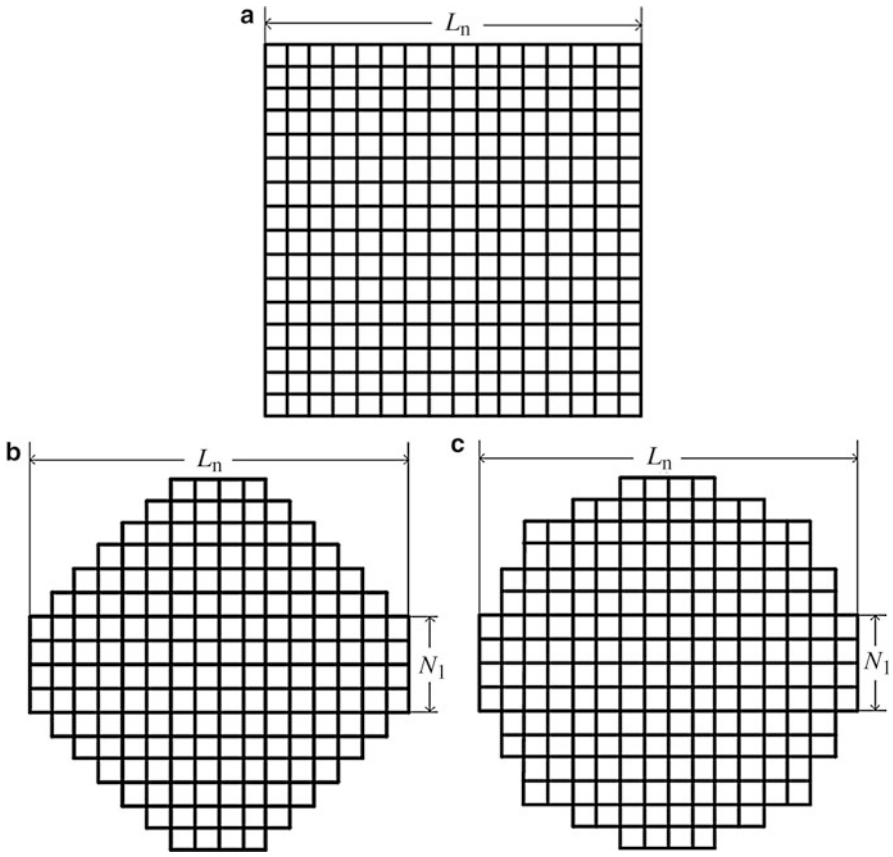


Fig. 38 Typical examples of three topological models: (**a**) SQ, (**b**) TQR, and (**c**) CR

The power density at the radiation field is

$$S = \frac{|C_0|^2}{2\eta R^2} f_e^2 \left| \sum_{n=0}^{N-1} w_n e^{jk(ux_n + vy_n)} \right|^2 \tag{12}$$

The total transmitting power is the integral of power density S from $\theta = 0$ to $\theta = \pi$ and the receiving power is from $\theta = 0$ to $\theta = \theta_0$. So the beam capture efficiency can be calculated as

$$BCE = \frac{P_r}{P_t} = \frac{\int_0^{2\pi} \int_0^{\theta_0} f_e^2 \left| \sum_{n=0}^{N-1} w_n e^{jk(ux_n + vy_n)} \right|^2 \sin\theta d\theta d\phi}{\int_0^{2\pi} \int_0^{\pi} f_e^2 \left| \sum_{n=0}^{N-1} w_n e^{jk(ux_n + vy_n)} \right|^2 \sin\theta d\theta d\phi} \tag{13}$$

The square model (SQ) for the transmitting antenna aperture plotted in Fig. 38a is the most used one, which has the total element number N, the column and row number L_n, and the element number N_1 of the first column. The topological models of truncated square (TSQ) and circle (CR) have the same L_n as SQ model, as shown in Fig. 38b, c. The total N numbers of the three topological models are different when the row (or column) L_n numbers are the same.

DOI 10.1007/978-981-4560-75-7_116-1

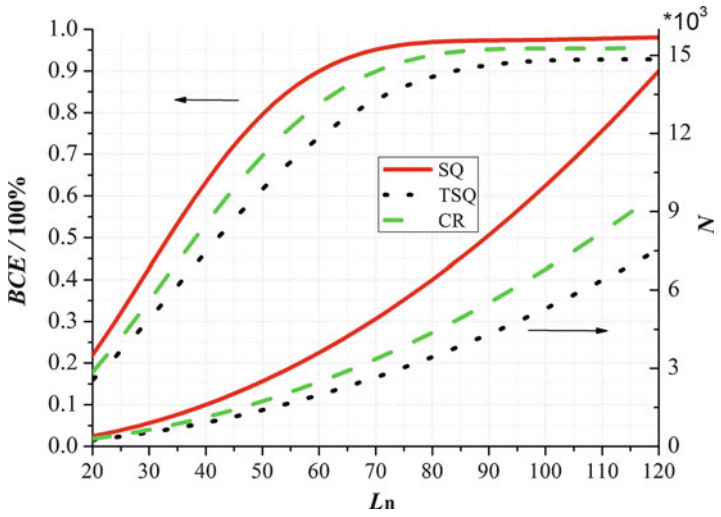


Fig. 39 Total element number N and BCE of three models versus column number L_n

Assuming element aperture a is $\lambda/2$ in Eq. 10, and the transmitting antenna aperture being normalized to wavelength λ, the weight of every element for 10 dB Gaussian taper distribution on the transmitting aperture can be calculated by

$$w_n = e^{-1.1513\rho_n^2} \tag{14}$$

$$\rho_n = \begin{cases} \dfrac{4\sqrt{(|x_n| - 0.25)^2 + (|y_n| - 0.25)^2}}{L_n}, \text{ for even } L_n; \\ \dfrac{4\sqrt{x_n^2 + y_n^2}}{L_n}, \text{ for odd } L_n. \end{cases} \tag{15}$$

For long-distance MPT systems, given the receiving angle θ_0 of the rectenna as 0.03, the three topological models are compared subsequently.

Figure 39 shows the total number N and BCE versus column number L_n for the three topological models. When L_n is 80, BCEs of SQ, TSQ, and CR models are 96.9 %, 88.6 %, and 93.8 %, and the corresponding total element N are 6400, 3436, and 4372, respectively. CR model saves transmitting elements 31.7 % at the cost of losing 3.1 % BCE compared with the SQ model.

Figure 40 shows the transmitting power P_t and BCE of the three models versus total element number N. It should be noticed that the actual transmitting power P_t is the product of the value in Fig. 41 by the factor of $|C_0|^2/2\eta$.

BCEs of three models are very near to each other if the total transmitting elements are the same. When N is less than about 4500, BCEs of the CR and SQ models are higher than that of the TSQ model. The transmitting power P_t of the TSQ and CR models is obviously higher than that of the SQ model.

Several conclusions can be drawn from above analysis. For the same column number L_n, CR and TSQ models can save much more array elements than SQ one with a little BEC being lost. For the same total element number N, CR and TSQ models can obtain higher BCE than SQ when N is lesser than a certain number and transfer much more power. CR model has a relative higher BEC than other two at any total element number N. So CR topological model for the transmitting antenna aperture is efficient for any scale of MPT systems.

DOI 10.1007/978-981-4560-75-7_116-1

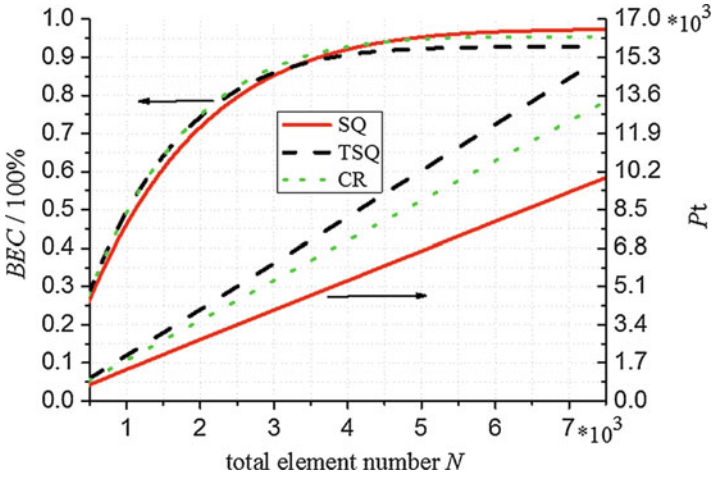


Fig. 40 Transmitting power P_t and *BCE* of three models versus total element number N

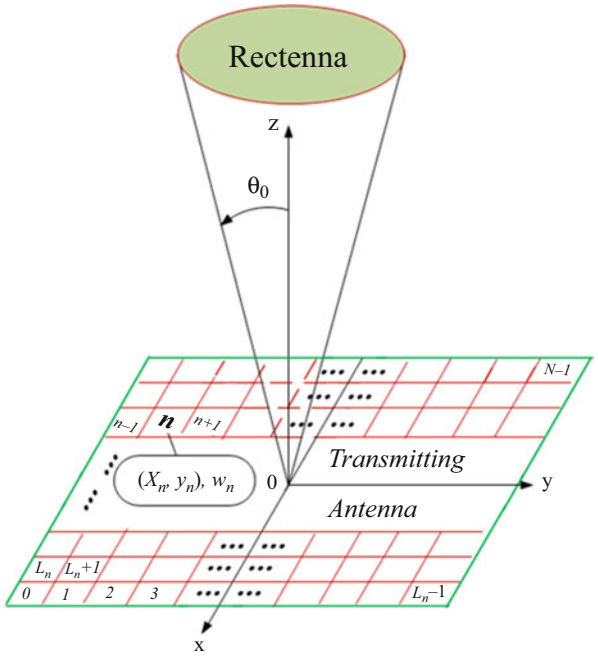


Fig. 41 Transmitting and receiving apertures of a MPT

Level Distributions on Transmitting Antenna Apertures

Considering trade-off between *BCE* and the operation efficiency of all the transmitting elements, the MPT system has circular transmitting and receiving apertures as shown in Fig. 41. The transmission distance is R, while the radii of the transmitting and the receiving aperture are R_t and R_r, respectively. The receiving angle of the rectenna aperture is θ_0.

The aperture level distribution of 10 dB Gauss pulse is verified having high *BCE* but the high power density at the center makes the microwave devices easy to be damaged. The *BCE* and sidelobe level of the truncated taper function of SC (Split Circle) will be investigated and compared with those of the 10 dB Gauss in formulas (14) and (15). SC function expression is as follows:

DOI 10.1007/978-981-4560-75-7_116-1

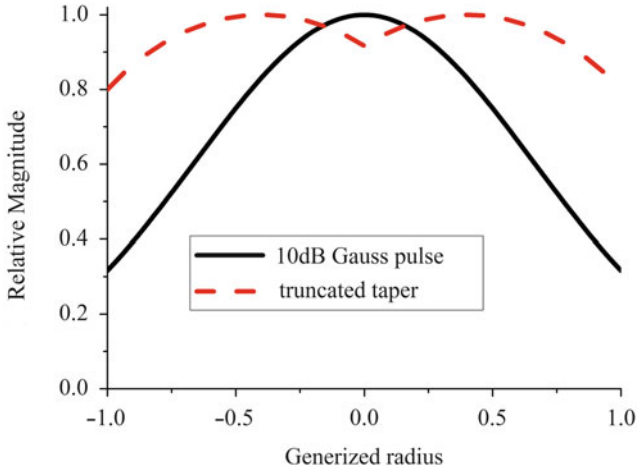


Fig. 42 10dB Gauss pulse and truncated taper functions

$$f(x) = \begin{cases} \sqrt{1-(x-0.4)^2} & x > 0 \\ \sqrt{1-(x+0.4)^2} & x < 0 \end{cases} \tag{16}$$

The curves of 10 dB Gauss pulse and the SC truncated taper functions are shown in Fig. 42 Resembling the 10 dB Gauss pulse and using the topological model of CR in Fig. 38, the weight of every element for SC function distribution on the transmitting aperture can be calculated by

$$w_n = \sqrt{1-[a(\rho_n - b)]^2} \tag{17}$$

where $a = 1.0$, $b = 0.4$, and

$$\rho_n = \begin{cases} \dfrac{4\sqrt{(|x_n|-0.25)^2+(|y_n|-0.25)^2}}{L_n}, \text{ for even } L_n; \\ \dfrac{4\sqrt{x_n^2+y_n^2}}{L_n} \qquad\qquad \text{, for odd } L_n. \end{cases}$$

Given the receiving angle θ_0 of the rectenna as 0.03 for a long-distance MPT system, Fig. 43 shows the total number N and microwave beam capture efficiency BEC versus column number L_n for SC level distribution. It could be seen that BEC increases fast with the number of L_n until L_n is greater than 60 while the total number N increases continuously. Considering the cost of MPT system, L_n is chosen as 74. On this condition, the total element number N is 3748, BEC is 85.2 %.

The weight coefficient distribution for $L_n = 74$ and $N_1 = 4$ is illustrated in Fig. 44. A relative constant level distribution from 0.8 to 1.0 could be found. Figure 45 shows the power density distribution in the radiation field of the transmitting antenna. It could be seen that the first sidelobe is lower −25 dB than the main beam and 85.2 % power could be captured by the receiving aperture with $\theta_0 = 0.03$ rad.

DOI 10.1007/978-981-4560-75-7_116-1

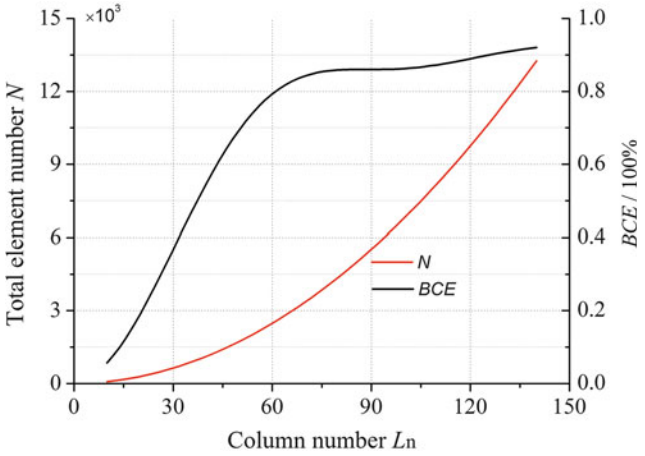


Fig. 43 Total number N and BEC versus column number L_n for SC level distribution

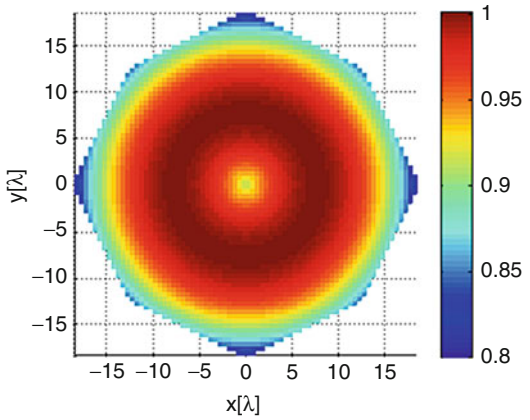


Fig. 44 Weight coefficients of the model for $L_n = 74$, $N_1 = 4$

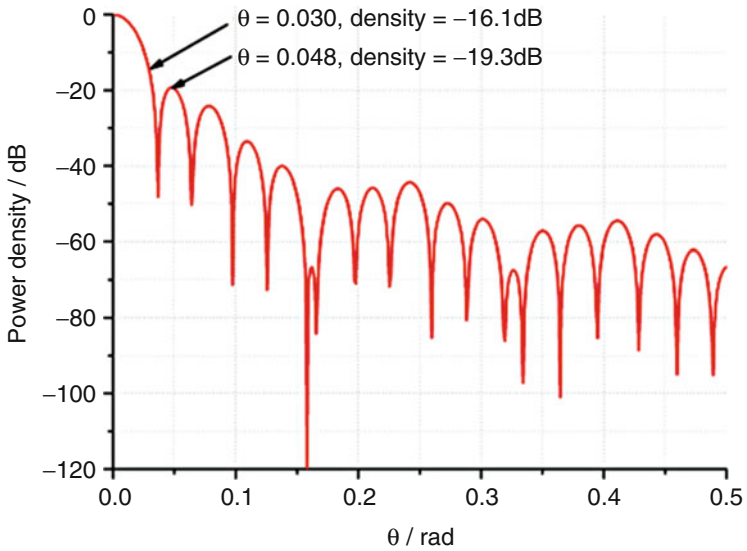


Fig. 45 Power density distribution of the transmitting antenna in the radiation field

DOI 10.1007/978-981-4560-75-7_116-1

Efficiency Evaluation of C-Band MPT System

For a MPT system in Fig. 4, the available microwave generators mainly include the magnetic amplifiers, the klystron, and the solid state amplifiers. Magnetrons have high power handling capability and have the maximum DC-mw efficiency η_g of 70 ~ 80 % at the microwave bands of S-, C-, and X- bands.

A magnetron with 80 % of DC-mw conversion efficiency η_g at 5.8 GHz is supposed to be applied as the microwave generator. A parabolic antenna or a waveguide slot array antenna has high transmitting efficiency η_t of about 80 %, which can be used as the transmitting antenna element. Considering the array losses for the generators and the transmitting antennas composed of large array to obtain the demanded level distribution, the efficiencies are evaluated as 70 % for both η_t and η_t.

The microwave power capture efficiency *BEC* has been optimized to be 85.2 % in the above section. In section "Series/Parallel Mixed Connections of Rectenna Array," based on the experiment of a 3 × 3-element rectenna array and the reasonable analysis of the results, 585 mW DC power could be obtained on the receiving aperture of 12 × 10 cm and a large rectenna array is estimated to have 65 % mw-DC efficiency at 5.8 GHz. From Eq. 1, the DC-DC system efficiency is evaluated about 27 %.

From Fig. 38c, the diameter of the transmitting aperture is 1.91 m at 5.8 GHz when the column number L_n is 74. The diameter of the receiving aperture is 8.4 m at the far-field distance of 140 m when the receiving angle is 0.03 radial. About 2700 W DC power will be obtained on this receiving rectenna array aperture.

It should be noticed that the results are obtained in an inaccurate evaluation instead of a strict calculation. The power density at the 140 m far field may not be uniform, which will decrease the mw-DC conversion efficiency obviously. The rectifying circuit losses for a large rectenna array may be high. So the optimization method of the transmitting antenna and the receiving rectenna array is the breakthrough that should be investigated further in the near future.

Summary

MPT is a challenging technology in long-distance applications. At the present level with the transmitter of DC-mw and transmitting antenna efficiencies of both 70 %, the beam capture efficiency of 85.2 %, and the rectenna array of 65 % mw-DC efficiency, the DC-DC system efficiency of 27 % can be probably achieved. To further improve the MPT efficiency, some key techniques should be overcome. Firstly, the rectifying Schottky diodes should be specially developed to increase the mw-DC efficiency of the rectifying circuits. The diodes should have lower series resistance R_S, low zero-bias junction capacitor C_{j0}, and higher breakdown voltage V_B. On the applications of low power density, the lower built-in turn-on voltages V_{bi} are required. Secondly, the beam capture efficiency can be increased by optimizing the level distribution on the transmitting aperture. Thirdly, to maintain 70 % DC-mw of the generators and 70 % transmitting efficiency of the transmitting antenna array for the optimized level distribution, the transmitter design is a challenging topic.

Cross-References

- ▶ Antenna Sensors in Batteryless Wireless Sensing Systems
- ▶ Antennas in Wireless Charge Systems
- ▶ Circularly Polarized Antennas
- ▶ Microstrip Patch Antennas
- ▶ Millimeter-Wave Antennas and Arrays

DOI 10.1007/978-981-4560-75-7_116-1

References

Ansoft High Frequency Structure Simulator (HFSS) (2011) Ver. 12. Ansoft Corp.

Brown WC (1964) Experiments in the transportation of energy by microwave beam. IRE Int Conv Rec 12(2):8–17

Brown WC (1984) The history of power transmission by radio waves. IEEE Trans MTT 32(9):1230–1242

Celeste A, Jeanty P, Pignolet G (2004) Case study in Reunion island. Acta Astronaut 54(4):253–258

Degenford JE, Sirkis MD et al (1964) The reflecting beam waveguide. IEEE Trans MTT 12(4):445–453

East T (1992) A self-steering array for the SHARP microwave-powered aircraft. IEEE Trans Antennas Propag 40(12):1565–1567

Fujino Y, Fujita M et al (1996) A dual polarization patch rectenna for high power application. IEEE Antennas Propag Soc Int Symp Dig 3:1560–1563

Gao Y, Yang XX et al (2010) A circularly polarized rectenna with low profile for wireless power transmission. Prog Electromagn Res Lett 13:41–49

Glaser PE (1992) An overview of the solar power satellite option. IEEE Trans MTT 40(6):1230–1238

Huang FJ, Lee CM et al (2011) Rectenna application of miniaturized implantable antenna design for triple-band biotelemetry communication. IEEE Trans Antennas Propag 59(7):2646–2653

IEEE standard for safety levels with respect to human exposure to radio frequency electromagnetic fields, 3 kHz to 300 GHz (2006) IEEE Std C95.1-2005 (Revision of IEEE Std C95.1-1991):20−30

Kurs A, Karalis A, Moffatt R et al (2007) Wireless power transfer via strongly coupled magnetic resonances. Science 317(83):83–86

Murao Y (2000) An investigation on the design of a transmission antenna and a rectenna with arrayed apertures for microwave power transmission. Electron Commun Jpn, Part 1 83(2):1–9

Nie MJ, Yang XX et al (2015) A compact 2.45-GHz broadband rectenna using grounded coplanar waveguide. IEEE Antennas Wirel Propag Lett 14:986–989

Oliveri G, Poli L, Massa A (2013) Maximum efficiency beam synthesis of radiating planar arrays for wireless power transmission. IEEE Trans Antennas Propag 61(5):2490–2499

Park J-Y, Han S-M et al (2004) A rectenna design with harmonic-rejecting circular-sector antenna. IEEE Antennas Wirel Propag Lett 3:52–54

Shinohara N (2011) Power without wires. IEEE Microw Mag 12(7):S64-S73

Shinohara N, Matsumoto H (1998) Experimental study of large rectenna array for microwave energy transmission. IEEE Trans MTT 46(3):261–268

Strassner B, Chang K (2003) Highly efficient c-band circularly polarized rectifying antenna array for wireless microwave power transmission. IEEE Trans Antennas Propag 51(6):1347–1356

System description, wireless power transfer, vol. I: low power, part 1: interface definition (2012)

The advanced design system (ADS) (2005) Agilent Corp.

Yang XX, Xu JS et al (2008) X-band circularly polarized rectennas for microwave power transmission applications. J Electron 25(3):389–393

Yang XX, Jiang C et al (2013) A novel compact printed rectenna for data communication systems. IEEE Trans Antennas Propag 61(5):2532–2539

Youn DG, Park YH et al (1999) A study on the fundamental transmission experiment for wireless power transmission system. In: Proceedings of the IEEE Region 10 conference in Cheju Island, vol. 2, pp 1419−1422

DOI 10.1007/978-981-4560-75-7_116-1

Zepeda P (2003) Modeling and design of compact microwave components and systems for wireless communications and power transmission. Ph.D. dissertation, Texas A&M University

Zhou H-W, Yang X-X (2014) Aperture optimization of transmitting antennas for microwave power transmission systems. IEEE AP-S Digest 1357–1358

Zoya P, Sean K, Steven D et al (2014) Scalable RF energy harvesting. IEEE Trans MTT 62(4):1046–1056

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_117-1

Near-Field Antenna Measurement Techniques

Flaminio Ferrara, Claudio Gennarelli* and Rocco Guerriero
Dipartimento di Ingegneria Industriale, Università di Salerno, Fisciano (Salerno), Italy

Abstract

A complete description of the near-field antenna measurement techniques is provided in this chapter. After a discussion of the state of the art, the key steps of the classical near-field–far-field (NF-FF) transformations with plane-rectangular, cylindrical, and spherical scannings, in their probe-uncompensated and probe-compensated versions, are summarized, by also providing some analytical details on the wave expansions commonly adopted to represent the antenna radiated field. The nonredundant sampling representations of electromagnetic field are then introduced and applied to drastically reduce the number of required NF data and related measurement time with respect to the classical NF-FF transformations. At last, the NF-FF transformations with innovative spiral scannings, allowing a further measurement time saving, are described.

Keywords

Antenna measurements; Near-field – far-field transformation techniques; Plane wave expansion; Cylindrical wave expansion; Spherical wave expansion; Probe compensation; Nonredundant sampling representations of electromagnetic fields; Spiral scannings

Introduction

When dealing with electrically large antennas, far-field (FF) range size limitations, transportation, and mounting problems make it absolutely impractical or impossible to measure their radiation patterns on a conventional FF range. On the other hand, the increasing use of high-performance antennas, as those employed in radar and satellite systems, requires an accurate measure of their radiating characteristics. For instance, satellite antennas are designed to transmit or receive over long distances, and accordingly the requirements to beam pointing may be severe; thus a high measurement accuracy is mandatory to verify that the antenna fulfills all the specifications. As a consequence, the problem of the determination of the antenna FF pattern from near-field (NF) measurements has attracted considerable attention in the last 50 years (Appel-Hansen et al. 1982; Yaghjian 1986; Gillespie 1988; Hald et al. 1988; Gennarelli et al. 2004; Gregson et al. 2007; Francis and Wittmann 2008; Gennarelli et al. 2012; Francis 2012).

NF measurements may be performed in a controlled environment, as an indoor shielded anechoic chamber, which allows one to overcome those drawbacks that, due to weather conditions (rain, snow, etc.), electromagnetic (EM) interference, and other, cannot be eliminated in FF measurements. In addition, NF scanning techniques are the better choice when complete pattern and polarization measurements are required. Moreover, they provide the necessary information to determine the field at

*Email: cgennarelli@unisa.it

DOI 10.1007/978-981-4560-75-7_117-1

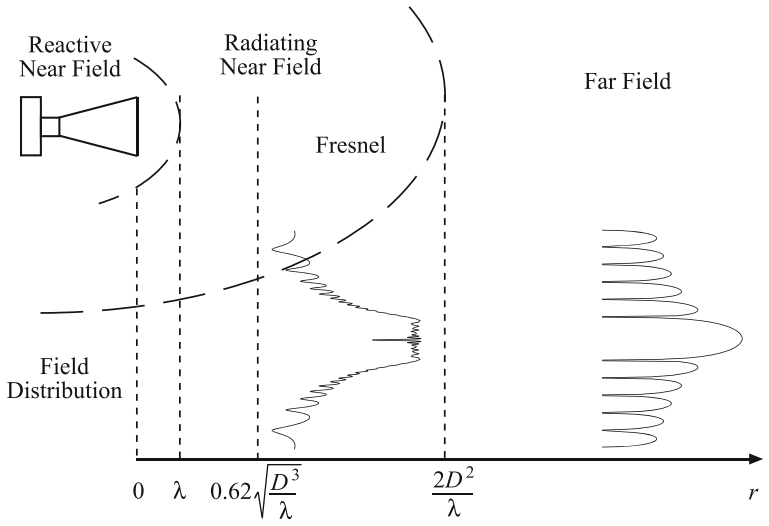


Fig. 1 Field regions of an electrically large antenna

the surface of the antenna. Such an information can be properly employed for the diagnostics of surface errors in a reflector antenna or of faulty elements in an array (microwave holographic diagnostics (Yaccarino et al. 1994)).

Another advantage of NF measurements stems from the fact that the reflected signal will be weaker than in a FF measurement facility because it is transmitted and received through far-out sidelobes. Moreover, the absorbers will work more efficiently than in a FF range, where they are employed close to the grazing incidence condition.

It is the authors' opinion that a comprehensive chapter on the NF-FF transformation techniques cannot begin without a preliminary discussion which highlights the field behavior when the distance from the antenna increases. The free space surrounding an antenna is usually subdivided into three regions: the reactive near-field, the radiating near-field, and the far-field region (see Fig. 1). Although no abrupt change in the field behavior can be observed when their boundaries are crossed, the field configuration is quite different in them. For an antenna focused at infinity, the optical term *Fraunhofer region* can be used synonymously with far-field region. For such an antenna, the optical term *Fresnel region* can be also employed to denote a subregion of the radiating near-field zone. The reactive near-field region is the zone immediately surrounding the antenna wherein the reactive field predominates. It extends up to a distance of about $\lambda/2\pi$ from the antenna surface, λ being the wavelength. However, experience with NF measurements indicates that λ is a more reasonable limiting distance for such a region. Outside this zone the reactive field decays rapidly and can be neglected at a distance of a few wavelengths from the antenna surface. The radiating near-field region is the intermediate zone between the reactive near-field and the far-field regions. In such a region the radiation fields predominate, but the angular distribution of the field is dependent on the distance from the antenna, and the field does not exhibit the dependence $e^{-j\beta r}/r$ typical of the antenna far field, β being the wavenumber. It is worthy to note that a time dependence $e^{j\omega t}$ has been implicitly assumed. It will be assumed and suppressed throughout the chapter. The Fresnel region is the radiating near-field subregion wherein a quadratic phase approximation can be used in the vector potential integral. The far-field region is the zone of the free space where the relative angular field distribution is independent of the distance from the antenna and the electric and magnetic fields vary according to the $e^{-j\beta r}/r$ dependence. Commonly, for electrically large antennas, the inner boundaries of the Fraunhofer and Fresnel regions are set at $2D^2/\lambda$ and $0.62\sqrt{D^3/\lambda}$, respectively, where D is the maximum dimension of

DOI 10.1007/978-981-4560-75-7_117-1

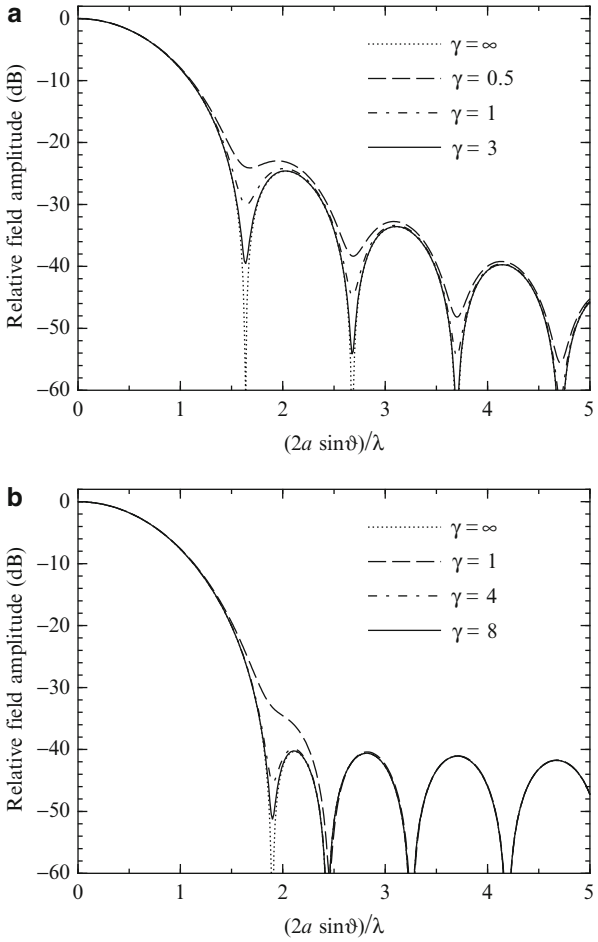


Fig. 2 Pattern behavior referred to the normalized distance $\gamma = r/(2D^2/\lambda)$: (**a**) for a circular aperture with $f(\rho) = (1 - \rho^2)$, (**b**) for a circular Taylor aperture distribution

the antenna. These boundaries are determined by assuming acceptable a maximum phase error of $\pi/8$, when a linear or quadratic phase approximation is used in the expression of the vector potential integral.

It has been recognized that distance requirements depend both on first sidelobe level of the antenna and on the desired accuracy (Silver 1984; Hollis et al. 1972). The widely used Rayleigh $2D^2/\lambda$ distance criterion gives rise to negligible pattern errors only for antennas with moderate sidelobe level (−25 dB). When measuring antennas having low (−30 to −40 dB) and ultralow (below −40 dB) sidelobe levels, a distance far larger than $2D^2/\lambda$ is needed (Hansen 1984; Corona et al. 1989). As shown in Fig. 2, particularly for antennas having low sidelobes, a significant increase in the sidelobe level and a more considerable raising of the null between it and the main lobe occur. As a conclusion, the distance requirements for measuring in a FF range the near-in sidelobes, which are below −30 dB, are very severe (Hansen 1984; Corona et al. 1989). For example, a distance of at least $6D^2/\lambda$ is required to measure, within a 1 dB accuracy, the first sidelobe of a Taylor antenna with sidelobe ratio (SLR) = 50 dB.

Usually, the NF antenna characterization can exploit complex field data or phaseless field data, the first choice being the most commonly used in practice.

In a NF facility processing complex field data, the probe, located near the antenna under test (AUT), is moved through a surface (scanning surface) which can be planar, cylindrical, or spherical (see Fig. 3). It collects complex voltage samples which are stored together with their positions. From these amplitude and phase data, measured for two different orientations of the probe (the probe is rotated by 90° around its

DOI 10.1007/978-981-4560-75-7_117-1

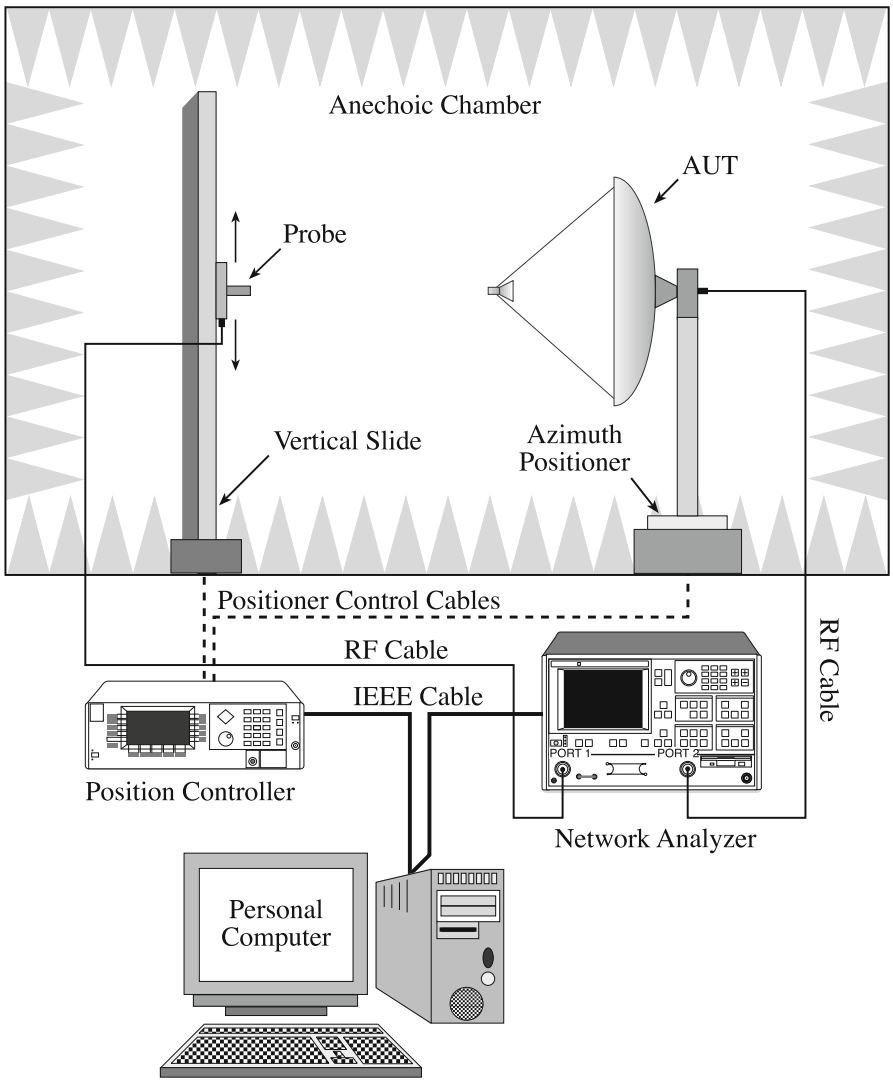


Fig. 3 NF measurement system: cylindrical scanning

axis in the second set), and taking into account the probe effects, one can compute the antenna FF pattern. It is worth noting that both the copolar and cross-polar components of the antenna far field can be recovered. Moreover, during the NF data acquisition, no information on the AUT polarization is needed. Nevertheless, such an information can be accurately determined by properly processing the reconstructed far field. The probe and its mounting structure should introduce minimum disturbance into the field to be measured, and its characteristics must be stable with time, environmental conditions, and orientations (Francis 2012). Commonly, the measured NF data are transformed into FF patterns by using an expansion of the AUT field in terms of modes, namely, a complete set of solutions of the vector wave equation in the region outside the antenna. Plane, cylindrical, or spherical waves are generally used. The type of modal expansion employed for representing the field determines the kind of the NF scanning surface, which accordingly will be a plane, a cylinder, or a sphere. The orthogonality properties of the modes on such surfaces are then exploited to obtain the modal expansion coefficients, which allow the reconstruction of the AUT far field. The development and the spreading of NF-FF transformation techniques employing planar, cylindrical, or spherical scanning systems is justified from the fact that each approach has its own particular advantages, depending on the AUT and the measurement requirements. The complexity of the

DOI 10.1007/978-981-4560-75-7_117-1

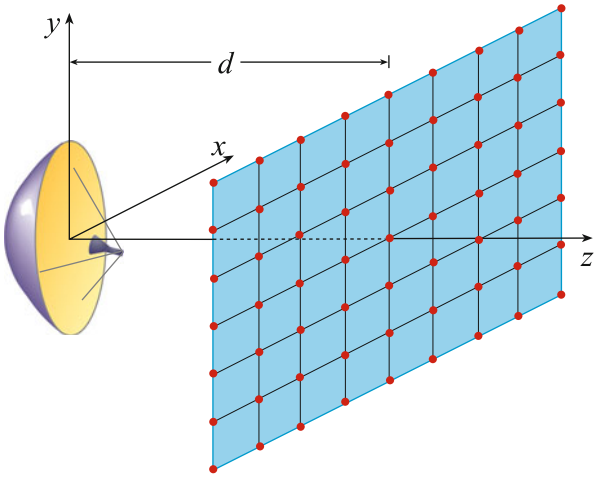


Fig. 4 Plane-rectangular scanning

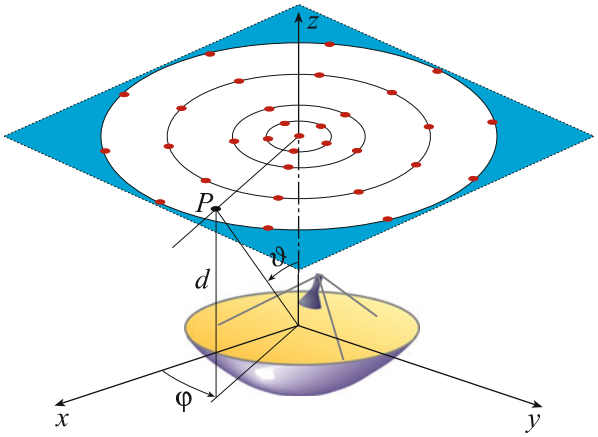


Fig. 5 Plane-polar scanning

analytical transformation increases from the planar to the cylindrical and from the cylindrical to the spherical surfaces.

The NF-FF transformations using a planar scanning are the most simple and efficient ones from the analytical and computational viewpoint. Their main disadvantage is that the pattern can be reconstructed only in a cone with an apex angle less than 180° without repeating the measurements. Accordingly, they are particularly suitable for highly directive antennas which radiate pencil beam patterns. There are quite different ways to realize a planar scanning: plane-rectangular (Kerns 1970, 1981; Joy and Paris 1972; Paris et al. 1978; Joy et al. 1978) (see Fig. 4), plane-polar (Rahmat-Samii et al. 1980; Gatti and Rahmat-Samii 1988; Yaghjian and Woodworth 1996; Bucci et al. 1991a, 1998a, 2000; Fig. 5), and bipolar (Yaccarino et al. 1994; Williams et al. 1994; D'Agostino et al. 2003; Fig. 6).

At the cost of a modest increase in the analytical and computational complication with respect to the planar scannings, the NF-FF transformation with cylindrical scanning (Joy et al. 1978; Leach and Paris 1973; Yaghjian 1977; Appel-Hansen 1980; Bucci et al. 1998b; D'Agostino et al. 2002, 2012a; Qureshi et al. 2013) (see Fig. 7) allows one to reconstruct, from a single set of NF measurements, the AUT complete radiation pattern save for the zones surrounding the spherical polar angles. Such a scanning is particularly attractive when considering antennas that concentrate the EM radiation in an angular region centered on the horizontal plane, as the radiating systems for radio base stations.

DOI 10.1007/978-981-4560-75-7_117-1

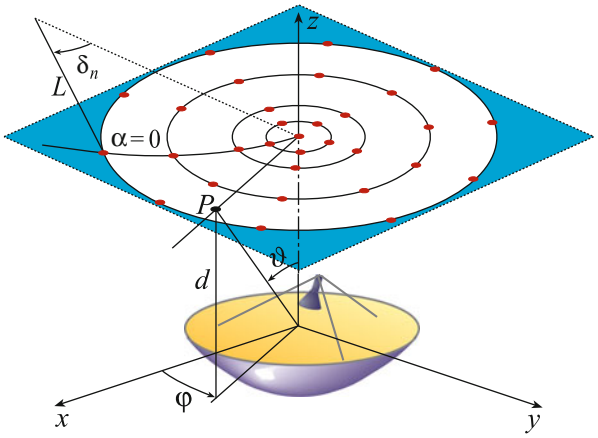


Fig. 6 Bipolar scanning

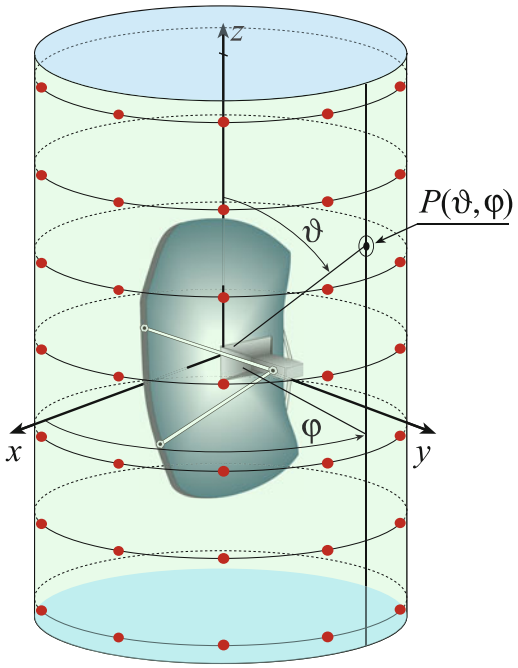


Fig. 7 Cylindrical scanning

The NF-FF transformation with spherical scanning (Hald et al. 1988; Qureshi et al. 2013; Wacker 1975; Larsen 1980; Yaghjian and Wittmann 1985; Hansen 2011; Bucci et al. 2001a; D'Agostino et al. 2011, 2013a) (see Fig. 8) allows one to reconstruct, from a single set of NF measurements, the complete radiation pattern of the AUT. However, the data processing to get the far field is considerably more complicated than that needed by planar and cylindrical facilities. The NF spherical scanning is, obviously, particularly tailored to measure low-gain and omnidirectional antennas.

In recent years, NF-FF transformations techniques based on the reconstruction of a proper set of equivalent currents have been developed (Petre and Sarkar 1992; Taaghol and Sarkar 1996; Sarkar and Taaghol 1999; Las-Heras and Sarkar 2002; Las-Heras et al. 2006; Alvarez et al. 2008). These (unknown) equivalent currents, lying on a selected surface enclosing the antenna, are evaluated by solving a set of integral equations relating them to the NF data acquired on the scanning surface. Once these equivalent

DOI 10.1007/978-981-4560-75-7_117-1

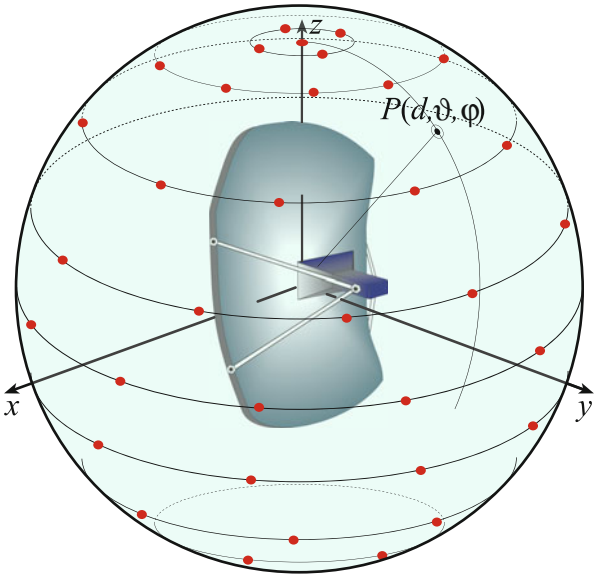


Fig. 8 Spherical scanning

currents have been determined, according to Love's equivalence theorem, it is possible to obtain the field at any point outside the equivalent source domain and, then, to evaluate the FF pattern. Unlike the NF-FF transformations using the modal expansion approach, they do not require the use of canonical scanning surfaces (plane, cylinder, sphere) but can be applied to an arbitrary measurement surface. On the other hand, their main drawback is the remarkably increased computational cost. As a matter of fact, the solution of a system equation (eventually ill-conditioned) with several thousands of unknowns is usually required. It is worthy to note that, for these NF-FF transformations too, it is possible to correct the distortion due to the nonisotropic radiation pattern of the employed probe. The probe correction is obtained by considering the antenna pattern of the probe as a weighting function in the integral equations relating the fields and the equivalent sources (Alvarez et al. 2008).

As already stressed, near-field techniques usually require the measurement of both the amplitude and phase of the antenna near field. Nevertheless, phaseless NF measurements have recently attracted a considerable interest, due to the less expensive measurement facilities required and to the increasing difficulty to perform accurate phase measurements in the millimeter and submillimeter frequency range. Over the years, several techniques have been proposed to determine the antenna far field from only amplitude NF measurements. A first possibility is the use of interferometric techniques (Bennett et al. 1976), which require an additional reference antenna whose transmitted signal, interfering with that transmitted by the AUT, allows to recover the lacking phase information. Other approaches exploit the functional relationship existing between two sets of only amplitude NF data collected by a probe on two scanning surfaces (Bucci et al. 1990, 1999; Isernia et al. 1996; Yaccarino and Rahmat-Samii 1999) or by two probes on the same scanning surface (Pierri et al. 1999) to retrieve the phase. A basically interferometric approach, avoiding the use of a reference antenna and using two identical probes and a simple microstrip circuit, has been also proposed to retrieve the phase information (Costanzo and Di Massa 2002; Costanzo et al. 2005).

Each member of the antenna measurement techniques community can profit today by about 50 years of research activity on NF data acquisition and related NF-FF transformations. Over these years, many solutions have been proposed to meet the demands of the various applications. In this framework, significant improvements in the performance of NF measurements have been recently achieved.

DOI 10.1007/978-981-4560-75-7_117-1

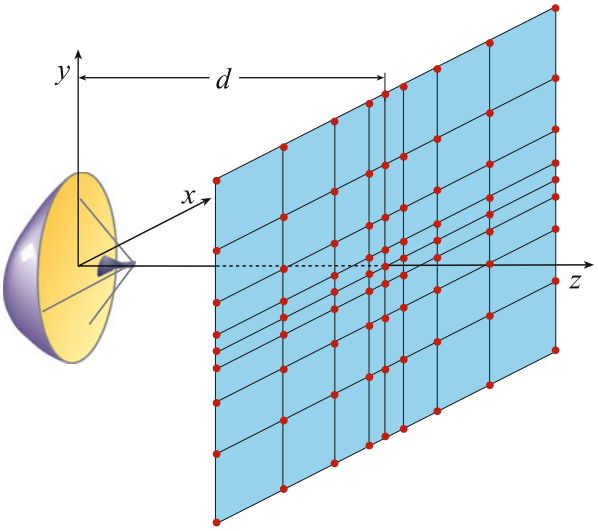


Fig. 9 Planar wide-mesh scanning

They are based on the spatial band limitation properties of radiated EM fields (Bucci and Franceschetti 1987, 1989), on their nonredundant sampling representations (Bucci et al. 1998c; Bucci and Gennarelli 2012), and on the optimal sampling interpolation (OSI) expansions of central type (Bucci et al. 1991a, b; Gennarelli et al. 1994). In particular, a significant reduction of the number of required NF data (and, as a consequence, of the corresponding measurement time) has been obtained for all the conventional scannings (see (Bucci et al. 1991a, 1998a, 2000) for the plane-polar, (D'Agostino et al. 2003) for the bipolar, (Bucci et al. 1998b; D'Agostino et al. 2002, 2012a) for the cylindrical, and (Bucci et al. 2001a; D'Agostino et al. 2011, 2013a) for the spherical scanning). In fact, the NF data needed by the corresponding traditional NF-FF transformation technique are accurately recovered by interpolating a minimum set of measurements via OSI expansions. A remarkable measurement time saving can be so obtained making these nonredundant transformations more and more appealing, since nowadays such a time is very much greater than the computational one. The mathematical justification for these results relies on the abovementioned band limitation properties and nonredundant sampling representations of EM fields. In fact, the EM fields radiated by antennas, enclosed in a convex domain bounded by a rotational surface Σ and observed on surface $\mathcal{M}$ with the same rotational symmetry, can be very well approximated by spatially band-limited functions when a proper phase factor is singled out from the field expression and proper parameterizations are used to describe $\mathcal{M}$ (Bucci et al. 1998c). Since the voltage acquired by a nondirective probe has the same effective spatial bandwidth of the AUT field, these representations can be, obviously, applied to the voltage too. The application of these nonredundant sampling representations has allowed also the development of an innovative and efficient planar NF-FF transformation using the planar wide-mesh scanning (PWMS) (Ferrara et al. 2007; D'Agostino et al. 2014a). Such a nonconventional plane-rectangular scanning technique is so named, since the sample grid is characterized by meshes wider and wider when going away from the center of the scanning region (see Fig. 9).

The use of the modulated scattering technique employing arrays of scattering probes, which allows a very fast electronic scanning, has been also proposed in (Bolomey et al. 1988) to reduce the time required for the acquisition of the NF data. However, apart from measurement precision issues, antenna testing facilities based on such a technique are not very flexible. Anyway, exploitation of the nonredundant sampling representations could allow to reduce the number of needed probes.

DOI 10.1007/978-981-4560-75-7_117-1

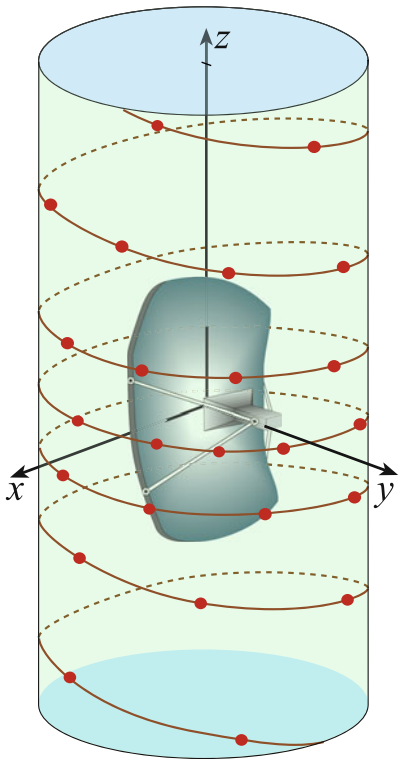


Fig. 10 Helicoidal scanning

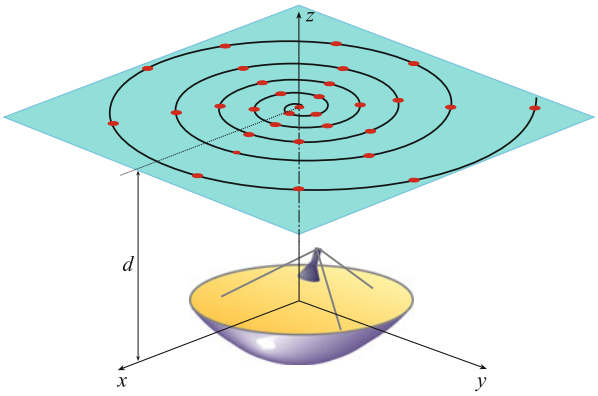


Fig. 11 Planar spiral scanning

A more viable way to reduce the time required for the NF data acquisition is the employment of innovative spiral scanning techniques. They have been implemented, as suggested in (Yaccarino et al. 1996), by means of continuous and synchronized movements of the positioning systems of the probe and AUT. Accurate, stable, and efficient NF-FF transformations with helicoidal (Bucci et al. 2001b; D'Agostino et al. 2008a, 2009a, b, 2012b) (see Fig. 10), planar (Bucci et al. 2002; D'Agostino et al. 2008b; Fig. 11), and spherical (Bucci et al. 2003; D'Agostino et al. 2009c, 2012c, 2013b, 2014b; Fig. 12) spiral scanning have been developed in the last years. They rely on nonredundant sampling representations and reconstruct the NF data needed by the classical NF-FF transformation corresponding to the adopted scanning surface, by interpolating, via appropriate OSI formulas, the nonredundant

DOI 10.1007/978-981-4560-75-7_117-1

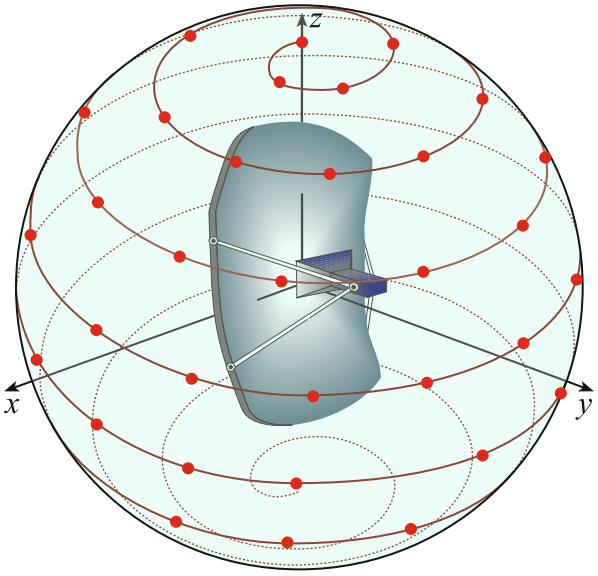


Fig. 12 Spherical spiral scanning

samples acquired by the measurement probe on the considered curve (helix or spiral). Other NF-FF transformation techniques with spiral scannings have been also proposed (Yaccarino et al. 1996; Costanzo and Di Massa 2004, 2007). However, since these approaches do not exploit the nonredundant representations of EM fields, they need a useless large amount of NF measurements.

The chapter is organized as follows. The classical NF-FF transformations with plane-rectangular, cylindrical, and spherical scannings, in their probe-uncompensated and probe-compensated versions, are summarized in section "Classical NF-FF Transformation Techniques." The nonredundant sampling representations of the electromagnetic fields are introduced in section "Nonredundant NF-FF Transformation Techniques," highlighting the role of the optimal parameterization and phase factor. In the same section, the application of these representations to the NF-FF transformations with conventional scannings, allowing a drastic measurement time saving, is also described. The theoretical foundations of the NF-FF transformations with spiral scanning for both quasispherical antennas (D'Agostino et al. 2006, 2009d) and those having two dimensions very different from the third one (D'Agostino et al. 2009e; Cicchetti et al. 2014) are presented in the subsequent section "NF-FF Transformation Techniques with Spiral Scannings" by providing an efficient sampling representation, which allows the reconstruction of the EM field (probe voltage) on a quite arbitrary rotational surface from a nonredundant number of its samples collected on a spiral wrapping it.

Classical NF-FF Transformation Techniques

NF-FF Transformation with Plane-Rectangular Scanning

The NF-FF transformation with plane-rectangular scanning (see Fig. 4) is without doubt the most simple and efficient one from the analytical and computational viewpoints. Its main drawback is that the antenna far field can be reconstructed only in a cone with an apex angle less than 180°. Therefore, it can be conveniently employed for highly directive antennas which radiate pencil beam patterns well within the solid angle specified by the edges of the AUT and those of the scanning area. In the plane-rectangular scanning, the probe is mounted on a $x - y$ positioner so that it can acquire the NF amplitude and phase data

DOI 10.1007/978-981-4560-75-7_117-1

on the wanted plane-rectangular grid. From these data, measured for two different orientations of the probe, one can compute the FF pattern of the AUT (Kerns 1970, 1981; Joy and Paris 1972; Paris et al. 1978; Joy et al. 1978).

In the first part of this subsection, the use of an ideal probe able to measure in its two orientations the tangential components E_x, E_y of the AUT electric field on the scanning plane is assumed and NF-FF transformation formulas derived. Then, such a hypothesis is removed and probe-corrected formulas given.

The starting point is the plane wave spectrum representation of EM fields (Clemmow 1966), summarized in the paragraph "Plane Wave Expansion" for readers' convenience. As shown in it, the tangential components E_x, E_y of the electric field radiated by the AUT can be represented on the scan plane at $z = d$ as a superposition of elementary plane waves, i.e.,

$$E_{x,y}(x,y,d) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left[\hat{E}_{x,y}(k_x,k_y)\mathrm{e}^{-\mathrm{j}\,k_z d}\right]\mathrm{e}^{-\mathrm{j}\left(k_x x + k_y y\right)}\mathrm{d}k_x\mathrm{d}k_y \tag{1}$$

By Fourier inverse transforming, it results

$$\hat{E}_{x,y}(k_x,k_y) = \frac{1}{4\pi^2}\mathrm{e}^{\mathrm{j}k_z d}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} E_{x,y}(x,y,d)\mathrm{e}^{\mathrm{j}\left(k_x x + k_y y\right)}\mathrm{d}x\mathrm{d}y \tag{2}$$

Accordingly, by assuming that the tangential components of the electric field are zero out of the measurement region on the scanning plane, it is possible to obtain the x and y components of the plane wave spectrum by a two-dimensional fast Fourier transform (FFT) algorithm. The other component of the spectrum can be determined by means of Eq. 75.

As shown in (Joy and Paris 1972), the sample spacings of the NF data are

$$\Delta x \le \lambda/2; \quad \Delta y \le \lambda/2 \tag{3}$$

In fact, if the scanning plane is located in a region of space where the EM field does not contain evanescent waves, then the plane wave spectrum is zero for $|k_x| > 2\pi/\lambda$ and $|k_y| > 2\pi/\lambda$. As a consequence of the two-dimensional Nyquist sampling theorem (Papoulis 1987), the EM field on the plane $z = d$ can be reconstructed from the knowledge of its samples at a rectangular lattice of points separated by grid spacings satisfying Eq. 3.

Once the plane wave spectrum has been determined, the FF components of the electric field in the spherical coordinate system (r, ϑ, φ) can be evaluated (see paragraph "Plane Wave Expansion") by using the relations

$$E_\vartheta(r,\vartheta,\varphi) = \mathrm{j}2\pi\beta\cos\vartheta\,\hat{E}_\vartheta(\beta\sin\vartheta\cos\varphi, \beta\sin\vartheta\sin\varphi)\frac{\mathrm{e}^{-\mathrm{j}\beta r}}{r} \tag{4}$$

$$E_\varphi(r,\vartheta,\varphi) = \mathrm{j}2\pi\beta\cos\vartheta\,\hat{E}_\varphi(\beta\sin\vartheta\cos\varphi, \beta\sin\vartheta\sin\varphi)\frac{\mathrm{e}^{-\mathrm{j}\beta r}}{r} \tag{5}$$

where $\hat{E}_\vartheta$, $\hat{E}_\varphi$ are related to $\hat{E}_x$ and $\hat{E}_y$ by

DOI 10.1007/978-981-4560-75-7_117-1

$$\hat{E}_{\varphi} = -\hat{E}_x \sin\varphi + \hat{E}_y \cos\varphi \tag{6}$$

$$\hat{E}_{\vartheta} = \left(\hat{E}_x \cos\varphi + \hat{E}_y \sin\varphi\right)/\cos\vartheta \tag{7}$$

As a matter of fact,

$$\hat{E}_{\vartheta} = \hat{E}_x \cos\vartheta\cos\varphi + \hat{E}_y \cos\vartheta \sin\varphi - \hat{E}_z \sin\vartheta$$

from which, by taking into account Eq. 75, Eq. 7 is easily obtained.

It is convenient to describe the employment of the FFT algorithm for computing the two-dimensional Fourier transforms in Eq. 2 from the knowledge of the acquired NF data. It can be useful to remember (Brigham 1974) that the discrete Fourier transform (DFT) is defined by

$$G\left(\frac{n}{NT}\right) = \sum_{i=0}^{N-1} g(iT)\mathrm{e}^{-\mathrm{j}2\pi ni/N} \quad n = 0, 1, \ldots, N-1 \tag{8}$$

whereas the inverse discrete Fourier transform (IDFT) is given by

$$g(iT) = \frac{1}{N}\sum_{n=0}^{N-1} G\left(\frac{n}{NT}\right)\mathrm{e}^{\,\mathrm{j}2\pi ni/N} \quad i = 0, 1, \ldots, N-1 \tag{9}$$

As well known, Eqs. 8 and 9 can be efficiently computed via the FFT algorithm. Moreover, they require both the functions to be periodic, namely,

$$G\left(\frac{n+pN}{NT}\right) = G\left(\frac{n}{NT}\right) \quad p = 0, \pm 1, \pm 2, \ldots \tag{10}$$

$$g[(i+pN)T] = g(iT) \quad p = 0, \pm 1, \pm 2, \ldots \tag{11}$$

The integration along x in Eq. 2 is now considered. By taking into account explicitly the truncation due to the finite size $2L_x \times 2L_y$ of the scanning plane and applying a straightforward approximation, it results

$$G(k_x) = \int_{-L_x}^{L_x} g(x)\mathrm{e}^{\mathrm{j}k_x x}\mathrm{d}x \approx \Delta x \sum_{n=0}^{N_x-1} g(x_n)\mathrm{e}^{\mathrm{j}k_x(n-N_x/2)\Delta x} \tag{12}$$

where $x_n = (n - N_x/2)\Delta x$, $N_x = 2L_x/\Delta x$ is the number of the measurement points along x (the overall number of measurement points is N_xN_y). When evaluating $G(k_x)$ at $k_{x_i} = i\Delta k_x = 2\pi i/(N_x\Delta x)$, it results

$$G(k_{x_i}) \approx \frac{2L_x}{N_x}\mathrm{e}^{-\mathrm{j}\pi i}\sum_{n=0}^{N_x-1} g(x_n)\mathrm{e}^{\mathrm{j}2\pi ni/N_x} \tag{13}$$

Accordingly, the integration over x in Eq. 2 can be efficiently performed via an inverse FFT algorithm, provided that the so obtained results are multiplied by the factor $2L_x\mathrm{e}^{-\mathrm{j}\pi i}$. The same considerations can be, obviously, applied to the integration over y.

DOI 10.1007/978-981-4560-75-7_117-1

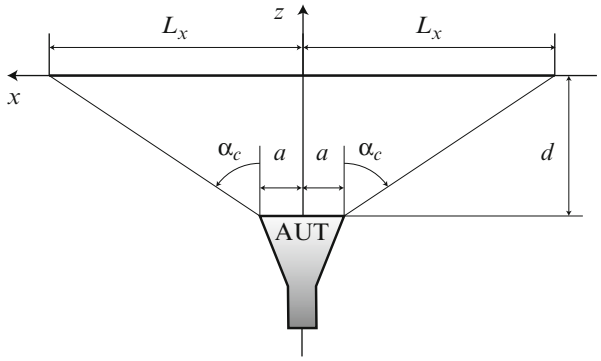


Fig. 13 Relevant to the evaluation of the validity angle

Summing up, for the evaluation of each of the Cartesian components $\hat{E}_x$ and $\hat{E}_y$ of the spectrum are needed N_y one-dimensional FFTs of size N_x and N_x FFTs of size N_y. This allows to obtain the FF components at the values of k_x and k_y given by $k_{x_i} = 2\pi i/(N_x \Delta x)$ and $k_{y_p} = 2\pi p/\left(N_y \Delta y\right)$. Note that the k_x and k_y vales such that $k_x^2 + k_y^2 > \beta^2$ must not be considered since the corresponding plane waves (called evanescent waves) do not contribute to the far field and represent NF reactive power storage.

To obtain the FF components with a greater resolution, it is sufficient "zero-filling" the NF data. Namely, to increase the number of output k_x values from N_x to N_x', the NF data must be increased with a proper number of zeros corresponding to fictitious $\left(N_x' - N_x\right)/2$ measurement points both before and after the effective ones. Similarly, to increase the number of output k_y values from N_y to N_y', $\left(N_y' - N_y\right)/2$ zeros must be added before and after the measured ones.

Since the measurement region is truncated in the plane-rectangular scanning, the reconstructed far field is affected by an inevitable truncation error, whose amount depends on the level of the neglected NF data external to the scanning area. In the following, the effect of the truncation due to the finite extension of the scanning plane along the x direction is analyzed. Quite analogous results hold also for that relevant to the y direction. When considering a scanning plane at distance d from the AUT, whose dimension along x is $2a$ (see Fig. 13), a convenient rule of thumb to predict the angular region of validity of the recovered FF pattern is given (Newell 1988) by

$$-\alpha_c \leqslant \vartheta \leqslant \alpha_c \tag{14}$$

where

$$\alpha_c = \tan^{-1}\left(\frac{L_x - a}{d}\right) \tag{15}$$

Such a validity angular region criterion was developed empirically from extensive NF measurements involving a large number of antenna and probe combinations (Yaghjian 1975) and derived using a theoretical analysis (Newell and Crawford 1974). Moreover, a ripple caused by the discontinuity of the near field at the edges of the scanning plane can appear even in the region of validity (Newell 1988).

It can be easily recognized that the NF tangential components E_x, E_y cannot be acquired when performing the measurement by means of a real, not ideal, probe. In fact, the probe sees the AUT center under different directions when moving on the scanning plane. Moreover, also at a fixed position, it sees each portion of the AUT under a different direction. As a consequence, the antenna far field cannot be

DOI 10.1007/978-981-4560-75-7_117-1

accurately recovered from the measured NF data by employing the previously described uncompensated NF-FF transformation.

The basic theory of probe-compensated NF measurements on a plane as proposed in (Paris et al. 1978; Joy et al. 1978) is based on the application of the Lorentz reciprocity theorem. The key relations in the here adopted reference system are

$$E_{\vartheta}(\vartheta,\varphi) = \left(I_H E'_{\varphi_V}(\vartheta,\ -\varphi) - I_V E'_{\varphi_H}(\vartheta,\ -\varphi)\right)/\Delta \tag{16}$$

$$E_{\varphi}(\vartheta,\varphi) = \left(I_H E'_{\vartheta_V}(\vartheta,\ -\varphi) - I_V E'_{\vartheta_H}(\vartheta,\ -\varphi)\right)/\Delta \tag{17}$$

where

$$\Delta = E'_{\vartheta H}(\vartheta,\ -\varphi)E'_{\varphi_V}(\vartheta,\ -\varphi) - E'_{\vartheta_V}(\vartheta,\ -\varphi)E'_{\varphi_H}(\vartheta,\ -\varphi) \tag{18}$$

and

$$I_{V,H} = A\cos\vartheta\ \ \mathrm{e}^{\,\mathrm{j}\beta d\cos\vartheta}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} V_{V,H}(x,y)\mathrm{e}^{\,\mathrm{j}\beta x\sin\vartheta\cos\varphi}\mathrm{e}^{\,\mathrm{j}\beta y\sin\vartheta\sin\varphi}\mathrm{d}x\ \ \mathrm{d}y \tag{19}$$

A being a proper constant. Namely, the antenna far field is related to (i) the two-dimensional Fourier transforms of the output voltages V_V and V_H of the probe for two independent sets of measurements (the probe is rotated by 90° in the second set); (ii) the FF components E'_{ϑ_V}, E'_{φ_V} and E'_{ϑ_H}, E'_{φ_H} radiated by the probe and the rotated probe when used as transmitting antennas.

In the following, a simple demonstration of the probe-compensated NF-FF transformation with plane-rectangular scanning is given. It is valid in the hypothesis that the mutual coupling effects be negligible and the AUT and probe be reciprocal. By expressing the electric field radiated by the AUT and impinging on the probe as superposition of elementary plane waves (see Eq. 80) and taking into account the relations between the AUT and the probe coordinate systems (see Fig. 14), it can be easily verified that the open circuit voltage at the probe terminals is given by

$$V(\underline{r}) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \underline{\hat{E}}(k_x,k_y)\cdot\underline{h}(k_x,\ -k_y)\mathrm{e}^{-\mathrm{j}\underline{k}\cdot\underline{r}}\mathrm{d}k_x\mathrm{d}k_y \tag{20}$$

where $\underline{h}$ is the receiving effective length of the probe and the symbol $(\cdot)$ denotes the inner product. By Fourier inverse transforming Eq. 20, it results

$$\underline{\hat{E}}(k_x,k_y)\cdot\underline{h}(k_x,\ -k_y) = \frac{1}{4\pi^2}\mathrm{e}^{\mathrm{j}k_z d}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} V(x,y)\mathrm{e}^{\mathrm{j}\left(k_x x+k_y y\right)}\mathrm{d}x\mathrm{d}y \tag{21}$$

For a reciprocal antenna the receiving and transmitting effective lengths coincide, and thus, by expressing the electric field $\underline{E}'$ radiated by the probe in the FF region as function of probe effective length (Franceschetti 1997)

DOI 10.1007/978-981-4560-75-7_117-1

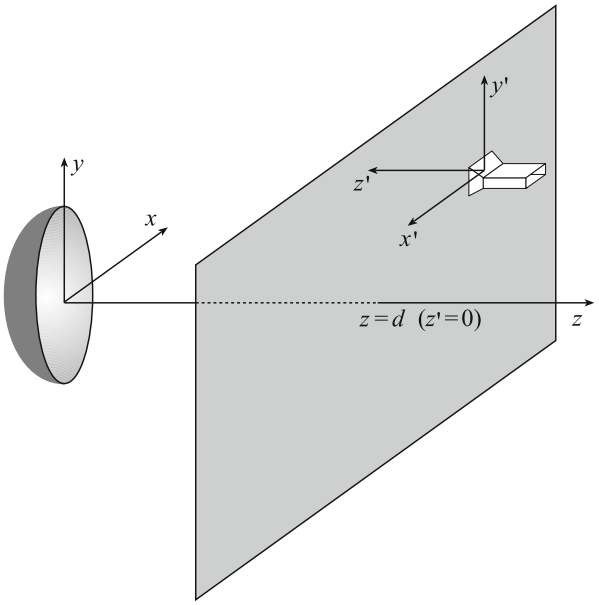


Fig. 14 Geometry relevant to the probe compensation

$$\underline{E}' = \mathrm{j}\frac{\zeta}{2\lambda r} I_0 \mathrm{e}^{-\mathrm{j}\beta r}\underline{h} \tag{22}$$

where ζ is the free-space impedance and I_0 is the antenna input current, it results

$$\hat{\underline{E}}(k_x, k_y)\cdot\underline{E}'(k_x, -k_y) = \frac{\mathrm{j}\zeta I_0 \mathrm{e}^{-\mathrm{j}\beta r}}{2\lambda r}\frac{\mathrm{e}^{\mathrm{j}k_z d}}{4\pi^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} V(x,y)\mathrm{e}^{\mathrm{j}\left(k_x x + k_y y\right)}\mathrm{d}x\mathrm{d}y \tag{23}$$

This last, by taking into account Eq. 81, becomes

$$\underline{E}(\vartheta,\varphi)\cdot\underline{E}'(\vartheta, -\varphi) = -\frac{2\pi\beta\zeta I_0 \mathrm{e}^{-\mathrm{j}2\beta r}}{2\lambda r^2}\frac{\mathrm{e}^{\mathrm{j}\beta d\cos\vartheta}}{4\pi^2}\cos\vartheta\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} V(x,y)\mathrm{e}^{\mathrm{j}\left(k_x x + k_y y\right)}\mathrm{d}x\mathrm{d}y \tag{24}$$

This last can be rewritten in a more convenient form by considering that the following relations between the AUT and the probe reference systems (see Fig. 14) hold:

$$\vartheta' = \vartheta; \quad \varphi' = -\varphi; \quad \hat{\vartheta}' = -\hat{\vartheta}; \quad \hat{\varphi}' = \hat{\varphi} \tag{25}$$

By properly taking into account these last, it is possible to rewrite Eq. 24 in terms of the spherical components of the field quantities, thus getting

$$\begin{aligned} &E_\vartheta(\vartheta,\varphi)E'_\vartheta(\vartheta, -\varphi) - E_\varphi(\vartheta,\varphi)E'_\varphi(\vartheta, -\varphi) \\ &= A\cos\vartheta\ \ \mathrm{e}^{\mathrm{j}\beta d\cos\vartheta}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} V(x,y)\ \ \mathrm{e}^{\mathrm{j}\beta x\sin\vartheta\cos\varphi}\mathrm{e}^{\mathrm{j}\beta y\sin\vartheta\sin\varphi}\mathrm{d}x\mathrm{d}y \end{aligned} \tag{26}$$

DOI 10.1007/978-981-4560-75-7_117-1

where

$$A = \frac{1}{4\pi^2}\frac{2\pi\beta\zeta I_0 \mathrm{e}^{-\mathrm{j}2\beta r}}{2\lambda r^2}$$

From Eq. 26, by assuming to perform two independent sets of measurements (the probe is rotated by 90° around its axis z' in the second set), the following linear system is obtained:

$$E_\vartheta(\vartheta,\varphi)E'_{\vartheta_V}(\vartheta,\ -\varphi) - E_\varphi(\vartheta,\varphi)E'_{\varphi_V}(\vartheta,\ -\varphi) = I_V \tag{27}$$

$$E_\vartheta(\vartheta,\varphi)E'_{\vartheta_H}(\vartheta,\ -\varphi) - E_\varphi(\vartheta,\varphi)E'_{\varphi_H}(\vartheta,\ -\varphi) = I_H \tag{28}$$

By solving such a system, Eqs. 16 and 17 are finally got.

NF-FF Transformation with Cylindrical Scanning

The NF-FF transformation with cylindrical scanning (see Fig. 7) allows one to reconstruct, from a single set of NF measurements, the antenna complete radiation pattern save for the zones surrounding the spherical polar angles and, accordingly, is particularly tailored for antennas that radiate mainly in an angular region centered on the horizontal plane. This, however, is obtained at the cost of a moderate increase in the analytical and computational complication with respect to that using the plane-rectangular scanning. In a cylindrical scanning facility, the AUT is mounted on a rotating table, whereas the probe moves along a line parallel to the rotation axis of the table. By properly matching these movements, the probe can acquire the NF amplitude and phase data on the wanted cylindrical grid. From these data, measured for two different orientations of the probe and accounting for the probe effects, the FF pattern can be evaluated (Joy et al. 1978; Leach and Paris 1973; Yaghjian 1977). As done in the plane-rectangular case, probe-uncorrected NF-FF transformation formulas, valid in the ideal probe assumption, are initially derived. Then, such an assumption is removed and probe-compensated formulas given.

As well known, in the cylindrical coordinate system (ρ, φ, z), the tangential components of the electric field radiated by the AUT can be represented (Leach and Paris 1973) on the scanning cylinder as a superposition of elementary cylindrical waves (see paragraph "Cylindrical Wave Expansion"), namely,

$$E_z(\varphi,z) = \sum_{v=-\infty}^{\infty}\int_{-\infty}^{\infty} b_v(\eta)\frac{\Lambda^2}{\beta}H_v^{(2)}(\Lambda d)\mathrm{e}^{\mathrm{j}v\varphi}\mathrm{e}^{-\mathrm{j}\eta z}\mathrm{d}\eta \tag{29}$$

$$E_\varphi(\varphi,z) = \sum_{v=-\infty}^{\infty}\int_{-\infty}^{\infty}\left[b_v(\eta)\frac{v\eta}{\beta d}H_v^{(2)}(\Lambda d) - a_v(\eta)\frac{\partial}{\partial\rho}H_v^{(2)}(\Lambda\rho)\Big|_{\rho=d}\right]\mathrm{e}^{\mathrm{j}v\varphi}\mathrm{e}^{-\mathrm{j}\eta z}\mathrm{d}\eta \tag{30}$$

where d is the cylinder radius, a_v and b_v are the modal expansion coefficients, $H_v^{(2)}(\cdot)$ is the Hankel function of second kind of order ν, and $\Lambda = (\beta^2 - \eta^2)^{1/2}$.

The modal coefficients are determined by Fourier inverse transforming Eqs. 29 and 30, thus obtaining

DOI 10.1007/978-981-4560-75-7_117-1

$$b_v(\eta)\frac{\Lambda^2}{\beta}H_v^{(2)}(\Lambda d)=\frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\pi}^{\pi}E_z(\varphi,z)\mathrm{e}^{-\mathrm{j}v\varphi}\mathrm{e}^{\mathrm{j}\eta z}\mathrm{d}\varphi\,\mathrm{d}z \tag{31}$$

$$b_v(\eta)\frac{v\eta}{\beta d}H_v^{(2)}(\Lambda d)-a_v(\eta)\frac{\partial}{\partial\rho}H_v^{(2)}(\Lambda\rho)\bigg|_{\rho=d}=\frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\pi}^{\pi}E_\varphi(\varphi,z)\mathrm{e}^{-\mathrm{j}v\varphi}\mathrm{e}^{\mathrm{j}\eta z}\mathrm{d}\varphi\mathrm{d}z \tag{32}$$

In the classical approach (Leach and Paris 1973), the FFT is employed to evaluate the modal coefficients in an efficient way, and the sample spacings of the NF data are

$$\Delta z\leqslant\lambda/2;\quad \Delta\varphi\leqslant\pi/(\beta a')=\lambda/(2a') \tag{33}$$

where a' is the radius of the smallest cylinder enclosing the AUT. In fact, if the scanning cylinder is located in a region of space where the EM field does not contain evanescent waves, then the cylindrical wave coefficients are zero for $|\eta|>\beta$ and $|v|>\beta a'$. As a consequence of the two-dimensional Nyquist sampling theorem (Papoulis 1987), the EM field on the cylinder at $\rho=d$ can be reconstructed from the knowledge of its samples at a regular lattice of points separated by grid spacings satisfying Eq. 33.

Once the modal coefficients are determined, the FF components of the electric field in the spherical coordinate system (r,ϑ,φ) can be evaluated (see paragraph "Cylindrical Wave Expansion") by

$$E_\vartheta(r,\vartheta,\varphi)=-\mathrm{j}2\beta\frac{\mathrm{e}^{-\mathrm{j}\beta r}}{r}\sin\vartheta\sum_{v=-\infty}^{\infty}\mathrm{j}^v b_v(\beta\cos\vartheta)\mathrm{e}^{\mathrm{j}v\varphi} \tag{34}$$

$$E_\varphi(r,\vartheta,\varphi)=-2\beta\frac{\mathrm{e}^{-\mathrm{j}\beta r}}{r}\sin\vartheta\sum_{v=-\infty}^{\infty}\mathrm{j}^v a_v(\beta\cos\vartheta)\mathrm{e}^{\mathrm{j}v\varphi} \tag{35}$$

It is worth noting that the summations in the above equations can be efficiently performed via the FFT.

The use of the FFT for the efficient evaluation of the two-dimensional Fourier transforms in Eqs. 31 and 32 is now briefly described. As regards the integration with respect to φ, by a straightforward approximation, it results

$$F_v=\int_0^{2\pi}f(\varphi)\mathrm{e}^{-\mathrm{j}v\varphi}\mathrm{d}\varphi\approx\frac{2\pi}{N}\sum_{m=0}^{N-1}f\left(\frac{2\pi m}{N}\right)\mathrm{e}^{-\mathrm{j}2\pi vm/N} \tag{36}$$

where $N=2M_1$, $M_1=\pi/\Delta\varphi$ is the highest order v of the angular harmonics. Accordingly, save for the factor $2\pi/N$, such an integration can be efficiently performed via a direct FFT algorithm. The integration over z can be tackled in the same way as that along x or y of the plane-rectangular case. Accordingly, by taking into account the truncation due to the finite height $2h$ of the measurement cylinder, it results

$$G(\eta_i)=G\left(\frac{2\pi i}{N_1\Delta z}\right)\approx\frac{2h}{N_1}\mathrm{e}^{-\mathrm{j}\pi i}\sum_{n=0}^{N_1-1}g(z_n)\mathrm{e}^{\mathrm{j}2\pi ni/N_1} \tag{37}$$

where $N_1=2h/\Delta z$ is the number of the measurement rings. Thus, the integration over z can be efficiently

DOI 10.1007/978-981-4560-75-7_117-1

performed via an inverse FFT algorithm, provided that the so obtained results are multiplied by the factor $2he^{-j\pi i}$.

By summing up, the two-dimensional Fourier transforms of the tangential components of the electric field can be efficiently computed through the following steps: (i) on each measurement ring, the NF data are directly transformed via FFT and then multiplied by the factor $2\pi/N$; (ii) for each index v, the so obtained data are inversely transformed via FFT and then multiplied by $2he^{-j\pi i}$.

Note that the evaluation of the Fourier transform integrals in Eqs. 31 and 32 via the FFT allows one to get the modal expansion coefficients a_v and b_v and, as a consequence, the FF components, at the values of ϑ given by

$$\vartheta_i = \cos^{-1}(\eta_i/\beta) = \cos^{-1}(i\lambda/(N_1\Delta z)) \tag{38}$$

To obtain the FF components with a greater resolution in ϑ, it is sufficient "zero-filling" the NF data. Namely, to get the FF components at the values $\vartheta_i = \cos^{-1}(i\lambda/(N_2\Delta z))$, the NF data must be increased with a proper number of zeros corresponding to fictitious $(N_2 - N_1)/2$ rings at both the ends of the scanning cylinder. In a similar way, to obtain the FF components at $2M_2$ values of φ, the positive and negative angular harmonics must be both augmented with $M_2 - M_1$ zeros before performing the summations in Eqs. 34 and 35 via the FFT.

Since the height of the scanning cylinder is obviously finite, an inevitable truncation error occurs, whose amount depends on the level of the neglected NF data. The angular region wherein the recovered FF pattern is accurate can be determined as in the plane-rectangular case. The truncation error arising when the scanning along φ does not cover a whole 2π range can be handled (Francis 2012) as that occurring in a truncated spherical scan (Hald et al. 1988).

It can be easily recognized that the NF tangential components E_z and E_φ cannot be acquired when performing the measurement by means of a real probe. In fact, it sees the AUT center under different directions when moving on the cylinder. Moreover, also at a fixed position, the probe sees each portion of the AUT under a different direction. As a consequence, the antenna far field cannot be accurately recovered from the measured NF data by employing the uncompensated NF-FF transformation.

The basic theory of probe-compensated NF measurements over a cylinder was developed by Leach and Paris (1973) and is based on an application of the Lorentz reciprocity theorem. They demonstrated rigorously that the modal coefficients a_v and b_v of the cylindrical wave expansion of the field radiated by the AUT are related to (i) the two-dimensional Fourier transform of the probe voltage for two independent sets of measurements (the probe is rotated 90° about its longitudinal axis in the second set); (ii) the coefficients of the cylindrical wave expansion of the field radiated by the probe and the rotated probe, when used as transmitting antennas. The key relations to obtain the modal expansion coefficients are

$$a_v(\eta) = \frac{\beta^2}{\Lambda^2\Delta_v(\eta)}\left[I_v(\eta)\sum_{m=-\infty}^{\infty} d'_m(-\eta)H^{(2)}_{v+m}(\Lambda d) - I'_v(\eta)\sum_{m=-\infty}^{\infty} d_m(-\eta)H^{(2)}_{v+m}(\Lambda d)\right] \tag{39}$$

$$b_v(\eta) = \frac{\beta^2}{\Lambda^2\Delta_v(\eta)}\left[I'_v(\eta)\sum_{m=-\infty}^{\infty} c_m(-\eta)H^{(2)}_{v+m}(\Lambda d) - I_v(\eta)\sum_{m=-\infty}^{\infty} c'_m(-\eta)H^{(2)}_{v+m}(\Lambda d)\right] \tag{40}$$

DOI 10.1007/978-981-4560-75-7_117-1

$$I_v(\eta) = \int_{-\infty}^{\infty}\int_{-\pi}^{\pi} V(\varphi, z)\ \mathrm{e}^{-\mathrm{j}v\varphi}\mathrm{e}^{\mathrm{j}\eta z}\mathrm{d}\varphi\mathrm{d}z; \quad I'_v(\eta) = \int_{-\infty}^{\infty}\int_{-\pi}^{\pi} V'(\varphi, z)\ \mathrm{e}^{-\mathrm{j}v\varphi}\mathrm{e}^{\mathrm{j}\eta z}\mathrm{d}\varphi\mathrm{d}z \tag{41}$$

$$\begin{aligned}\Delta_v(\eta) = &\sum_{m=-\infty}^{\infty} c_m(-\eta)H_{v+m}^{(2)}(\Lambda d)\sum_{m=-\infty}^{\infty} d'_m(-\eta)H_{v+m}^{(2)}(\Lambda d) + \\ &- \sum_{m=-\infty}^{\infty} c'_m(-\eta)H_{v+m}^{(2)}(\Lambda d)\sum_{m=-\infty}^{\infty} d_m(-\eta)H_{v+m}^{(2)}(\Lambda d)\end{aligned} \tag{42}$$

where V and V' are the voltages measured by the probe and the rotated probe at the point (d, φ, z). Quite analogous results have been obtained (Yaghjian 1977) by using the source scattering matrix formulation. In fact, as shown in (Appel-Hansen 1980), this last formulation leads, except for a normalization constant, to the same expressions.

The integrals I_v and I'_v are again efficiently evaluated via FFT and the NF data are spaced according to Eq. 33. The modal coefficients of the probe (c_m, d_m) and the rotated probe (c'_m, d'_m) can be evaluated from the measured amplitude and phase of the FF components radiated by them (Leach and Paris 1973). Once a_v and b_v have been determined, the FF components of the electric field in the spherical coordinate system (r, ϑ, φ) can be evaluated by means of Eqs. 34 and 35.

NF-FF Transformation with Spherical Scanning

The NF-FF transformation with spherical scanning (see Fig. 8) gives the full antenna pattern coverage, even though the data processing is considerably more complicated than that required by planar and cylindrical NF facilities. A tangible amount of work has been done in the past years for formulating and solving the problem of the FF reconstruction from the NF data acquired on a spherical scanning surface (Hald et al. 1988; Wacker 1975; Larsen 1980; Yaghjian and Wittmann 1985). In this framework, a comprehensive book (Hald et al. 1988), which deals with the theoretical as well as the practical aspects of the spherical NF scanning, was published by J.E. Hansen.

There are several ways to perform the scanning in a spherical NF facility (Hald et al. 1988). The most commonly adopted one is the roll-over-azimuth configuration (see Figs. 15 and 16). In such a configuration, the probe stays fixed, whereas the AUT is mounted on a roll positioner (φ axis), which is anchored by means of an L-shaped bracket to an azimuth rotating table (ϑ axis). In the elevation-over-azimuth configuration, the probe stays fixed and the AUT is mounted on an elevation positioner (φ axis), which is mounted on an azimuth rotating table (ϑ axis). On the contrary, in the azimuth-over-elevation setup, the AUT is anchored to an azimuth rotator (φ axis), mounted on an elevation positioner (ϑ axis). Another possibility is to mount the AUT on an azimuth rotating table for selecting the scanning meridian and to move the probe along an arc to perform the scanning along it.

Whatever the configuration has been chosen, the probe can acquire the NF amplitude and phase data on the wanted spherical grid. From these data, measured for two different orientations of the probe and accounting for the probe effects, the FF pattern can be evaluated (Hald et al. 1988). As done in the plane-rectangular and cylindrical scanning cases, probe-uncorrected NF-FF transformation formulas, valid in the ideal probe assumption, are initially derived. Then, such an assumption is removed and probe-compensated formulas given.

As well known, in the spherical coordinate system (r, ϑ, φ), the transverse electric field radiated by an AUT can be expressed on a sphere of radius $r = d$ containing it as a superposition of elementary spherical waves (see paragraph "Spherical Wave Expansion"), namely,

DOI 10.1007/978-981-4560-75-7_117-1

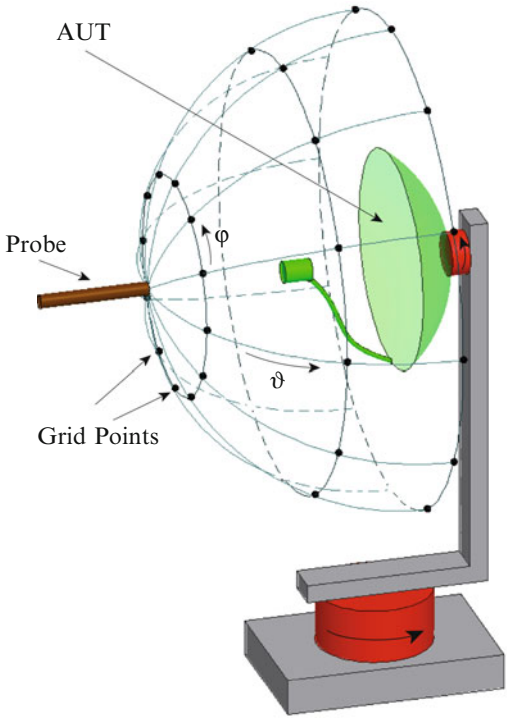


Fig. 15 Roll-over-azimuth NF spherical facility

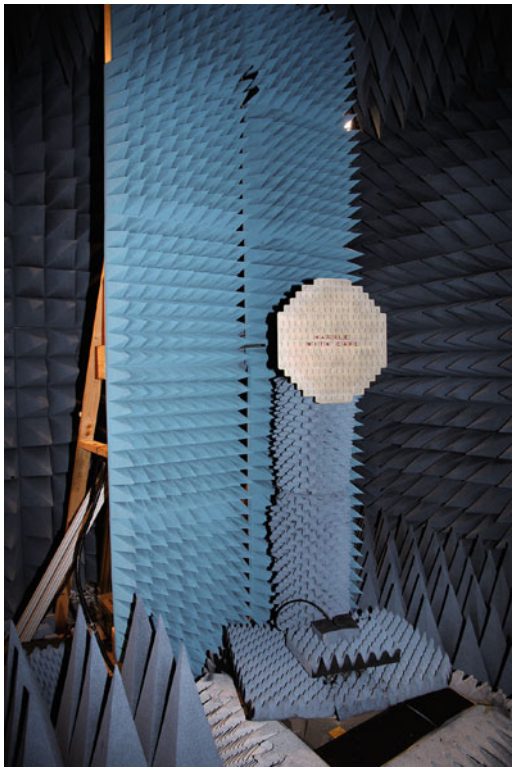


Fig. 16 Photo of the versatile NF system at the UNISA antenna characterization LAB: roll-over-azimuth spherical setup

DOI 10.1007/978-981-4560-75-7_117-1

$$\underline{E}_t(d,\vartheta,\varphi) = \beta \sum_{n=1}^{N_{\max}} \sum_{m=-n}^{n} \left[a_{1nm} g_{1n}(\beta d) \underline{F}_{1nm}(\vartheta,\varphi) + a_{2nm} g_{2n}(\beta d) \underline{F}_{2nm}(\vartheta,\varphi) \right] \tag{43}$$

where a_{1nm}, a_{2nm} are the spherical wave expansion coefficients,

$$\underline{F}_{1nm}(\vartheta,\varphi) = \underline{f}_{1nm}(\vartheta) \mathrm{e}^{\mathrm{j}m\varphi};\ \underline{F}_{2nm}(\vartheta,\varphi) = \underline{f}_{2nm}(\vartheta) \mathrm{e}^{\mathrm{j}m\varphi} \tag{44}$$

$$\underline{f}_{1nm}(\vartheta) = \left(\frac{-m}{|m|}\right)^m \frac{1}{\sqrt{2\pi n(n+1)}} \left[\frac{\mathrm{j}m}{\sin\vartheta} \overline{P}_n^{|m|}(\cos\vartheta)\hat{\vartheta} - \frac{\mathrm{d}}{\mathrm{d}\vartheta} \overline{P}_n^{|m|}(\cos\vartheta)\hat{\varphi} \right] \tag{45}$$

$$\underline{f}_{2nm}(\vartheta) = \left(\frac{-m}{|m|}\right)^m \frac{1}{\sqrt{2\pi n(n+1)}} \left[\frac{\mathrm{d}}{\mathrm{d}\vartheta} \overline{P}_n^{|m|}(\cos\vartheta)\hat{\vartheta} + \frac{\mathrm{j}m}{\sin\vartheta} \overline{P}_n^{|m|}(\cos\vartheta)\hat{\varphi} \right] \tag{46}$$

$$g_{1n}(\beta d) = h_n^{(2)}(\beta d);\quad g_{2n}(\beta d) = \frac{1}{\beta d} \frac{\mathrm{d}}{\mathrm{d}(\beta r)} \left(\beta r h_n^{(2)}(\beta r) \right) \Big|_{r=d} \tag{47}$$

$h_n^{(2)}(x)$ being the spherical Hankel function of second kind and order n.

In the classical approach (Hald et al. 1988), the choice of the highest spherical wave is usually determined according to the following rule of thumb:

$$N_{\max} = \mathrm{Int}(\beta a) + 10 \tag{48}$$

where a is the radius of the smallest sphere enclosing the AUT. In the approach (Bucci et al. 2001a), it is rigorously fixed by the spatial band limitation properties of the radiated EM fields (Bucci and Franceschetti 1987) and it results

$$N_{\max} = \mathrm{Int}(\chi' \beta a) + 1 \tag{49}$$

where χ' is an enlargement bandwidth factor slightly larger than unity for electrically large antennas (Bucci and Franceschetti 1987).

In Eqs. 45 and 46, $\overline{P}_n^{|m|}(x)$ (see Eq. 107) is the normalized associated Legendre function as defined by Belousov (1962).

The expansion coefficients a_{1nm}, a_{2nm} can be evaluated from the knowledge of the tangential electric field on the scanning sphere by taking into account that the spherical vector wave functions $\underline{F}_{1,2nm}(\vartheta,\varphi)$ are orthonormal. As a matter of fact, from Eq. 43 it results

$$\left\langle \underline{E}_t, \underline{F}_{1,2nm} \right\rangle = \int_0^{2\pi} \int_0^{\pi} \underline{E}_t(d,\vartheta,\varphi) \cdot \underline{F}^*_{1,2nm}(\vartheta,\varphi) \sin\vartheta \mathrm{d}\vartheta \mathrm{d}\varphi = \beta a_{1,2nm}\, g_{1,2n}(\beta d) \tag{50}$$

where the asterisk (*) indicates the complex conjugation, and therefore

$$a_{1,2nm} = \frac{1}{\beta g_{1,2n}(\beta d)} \int_0^{2\pi} \int_0^{\pi} \underline{E}_t(d,\vartheta,\varphi) \cdot \underline{f}^*_{1,2nm}(\vartheta) \mathrm{e}^{-\mathrm{j}m\varphi} \sin\vartheta \mathrm{d}\vartheta d\varphi \tag{51}$$

DOI 10.1007/978-981-4560-75-7_117-1

Once the spherical wave expansion coefficients have been determined, the transverse electric field radiated by the AUT in the FF region can be evaluated (see paragraph "Spherical Wave Expansion") by the FF spherical wave expansion:

$$\underline{E}_t(r,\vartheta,\varphi) = \frac{\mathrm{e}^{-\mathrm{j}\beta r}}{r}\sum_{n=1}^{N_{\max}}\sum_{m=-n}^{n}\left[\mathrm{j}^{n+1}a_{1nm}\,\underline{f}_{1nm}(\vartheta) + \mathrm{j}^{n}a_{2nm}\,\underline{f}_{2nm}(\vartheta)\right]\mathrm{e}^{\mathrm{j}m\varphi} \tag{52}$$

As will be shown in the following, the FFT is used to evaluate the modal coefficients in an efficient way, and the sample spacings of the NF data in the classical approach (Hald et al. 1988) are

$$\Delta\vartheta \leqslant 2\pi/(2N_{\max}+1); \quad \Delta\varphi = \Delta\vartheta \tag{53}$$

The number of NF data along φ can be reduced if the radius a' of the smallest cylinder enclosing the AUT and having its axis coincident with the z one is smaller than the radius a of the minimum sphere. In such a case, it results (Hald et al. 1988)

$$\Delta\varphi \leqslant 2\pi/(2M+1) \tag{54}$$

where $M = \mathrm{Int}\,(\beta a') + 10$.

In the approach (Bucci et al. 2001a), the sample spacings are rigorously fixed by the spatial band limitation properties of the radiated EM fields (Bucci and Franceschetti 1987) and, accordingly, $N_{\max}$ is now given by Eq. 49, whereas $\Delta\varphi$ is given by Eq. 54, where M decreases when moving from the sampling parallels near the equator to those near the poles according to the law

$$M = \mathrm{Int}(\chi^{*}\beta a \sin\vartheta) + 1 \tag{55}$$

In this last relation,

$$\chi^{*} = \chi^{*}(\vartheta) = 1 + (\chi' - 1)[\sin\vartheta]^{-2/3} \tag{56}$$

is the azimuthal enlargement bandwidth factor.

The efficient evaluation of Eq. 51 is now briefly described. By expanding the tangential electric field components on the measurement sphere in Fourier series with respect to φ, namely,

$$\underline{E}_t(d,\vartheta,\varphi) = \sum_{k=-M}^{M}\underline{G}_k(\vartheta)\mathrm{e}^{\mathrm{j}k\varphi} \tag{57}$$

the coefficients can be rewritten as follows:

$$a_{1,2nm} = \frac{2\pi}{\beta g_{1,2n}(\beta d)}\int_0^{\pi}\underline{G}_m(\vartheta)\cdot\underline{f}^{*}_{1,2nm}(\vartheta)\sin\vartheta\,\mathrm{d}\vartheta \tag{58}$$

The integration over ϑ in Eq. 58 can be efficiently carried out by expanding in Fourier series the components of $\underline{G}_m(\vartheta)$ and $\underline{f}_{1,2nm}(\vartheta)$ via FFT. Accordingly, it results

DOI 10.1007/978-981-4560-75-7_117-1

$$a_{1,2nm} = \frac{2\pi}{\beta g_{1,2n}(\beta d)} \sum_{\ell=-N_{\max}}^{N_{\max}} \sum_{i=-N_{\max}}^{N_{\max}} \underline{G}_{m\ell} \cdot \underline{f}^{*}_{1,2nmi} \int_0^{\pi} \mathrm{e}^{\mathrm{j}(\ell-i)\vartheta} \sin\vartheta \, \mathrm{d}\vartheta \tag{59}$$

The integral in Eq. 58 is so transformed in a double summation, of indexes ℓ and i, involving the Fourier series coefficients (in ϑ) of the components of $\underline{G}_m(\vartheta)$ and $\underline{f}_{1,2nm}(\vartheta)$, and the integrals

$$\int_0^{\pi} \mathrm{e}^{\mathrm{j}(\ell-i)\vartheta} \sin\vartheta \, \mathrm{d}\vartheta$$

Note that, in order to evaluate the Fourier series coefficients of the components of $\underline{G}_m(\vartheta)$ and $\underline{f}_{1,2nm}(\vartheta)$, it is necessary to extend the components of $\underline{G}_m(\vartheta)$ and $\underline{f}_{1,2nm}(\vartheta)$, from $[0, \pi]$ on the range $[0, 2\pi]$. This can be easily done by taking into account that (i) the components of $\underline{f}_{1,2nm}(\vartheta)$ are even about $\vartheta = \pi$ when m is odd, and vice versa; (ii) the components of $\underline{G}_m(\vartheta)$ have the same parity as those of $\underline{f}_{1,2nm}(\vartheta)$.

To take advantage of the numerical efficiency of the standard FFT algorithm, the number of NF parallels to be considered in the NF-FF transformation and the number of samples on them must be the first power of two greater or equal to $N_{\max}$ and $2M$, respectively.

Moreover, by inverting the summation order in Eq. 52, it can be rewritten in the form

$$\underline{E}_t(r,\vartheta,\varphi) = \frac{\mathrm{e}^{-\mathrm{j}\beta r}}{r} \sum_{m=-M}^{M} \sum_{\substack{n=|m| \\ (n\neq 0)}}^{N_{\max}} \left[\mathrm{j}^{n+1} a_{1nm} \underline{f}_{1nm}(\vartheta) + \mathrm{j}^{n} a_{2nm} \underline{f}_{2nm}(\vartheta)\right] \mathrm{e}^{\mathrm{j}m\varphi} \tag{60}$$

which allows an efficient evaluation of the antenna far field at the considered elevation angle ϑ by performing the summation via FFT.

It can be easily recognized that the NF tangential components E_ϑ and E_φ cannot be acquired when performing the measurement with a real probe. In fact, it sees each portion of the AUT under a different direction. As a consequence, the antenna far field cannot be accurately recovered from the measured NF data by using the uncompensated NF-FF transformation.

The probe-compensated NF-FF transformation with spherical scanning has been developed by using the source scattering matrix formulation. As shown in (Hald et al. 1988), when using a probe with a first-order azimuthal dependence FF pattern (e.g., an open-ended circular waveguide), the modal expansion coefficients can be determined from the knowledge of the voltages V and V' measured by the probe and rotated probe, respectively, and are given by

$$a_{1nm} = \frac{2n+1}{16\pi} \frac{\int_0^{\pi}\int_0^{2\pi} [I_1(\vartheta)V - \mathrm{j}\, I_2(\vartheta)V']\mathrm{e}^{-\mathrm{j}m\varphi} \sin\vartheta \;\; \mathrm{d}\varphi \, \mathrm{d}\vartheta}{\sum_{\nu=1}^{\nu_{\max}} \left[a'_{1\nu 1} A^{n}_{\nu 1}(\beta d) - a'_{2\nu 1} B^{n}_{\nu 1}(\beta d)\right]} \tag{61}$$

DOI 10.1007/978-981-4560-75-7_117-1

$$a_{2nm} = \frac{2n+1}{16\pi} \frac{\int_0^{\pi}\int_0^{2\pi} [I_2(\vartheta)V - \mathrm{j}\ I_1(\vartheta)V']\mathrm{e}^{-\mathrm{j}m\varphi} \sin\vartheta\ \ \mathrm{d}\varphi\ \mathrm{d}\vartheta}{\sum_{v=1}^{v_{\max}} \left[a'_{1v1}B^n_{v1}(\beta d) - a'_{2v1}A^n_{v1}(\beta d)\right]} \tag{62}$$

where $v_{\max}$ is the highest spherical wave of the probe, $a'_{1,2v1}$ are its expansion coefficients,

$$I_1(\vartheta) = \left[d^n_{1m}(\vartheta) - d^n_{-1m}(\vartheta)\right]; \quad I_2(\vartheta) = \left[d^n_{1m}(\vartheta) + d^n_{-1m}(\vartheta)\right] \tag{63}$$

and

$$d^n_{\mu m}(\vartheta) = \sqrt{\frac{(n+\mu)!(n-\mu)!}{(n+m)!(n-m)!}} \sum_{\sigma} \binom{n+m}{n-\mu-\sigma} \binom{n-m}{\sigma} (-1)^{n-\mu-\sigma} \cdot \left(\cos\tfrac{\vartheta}{2}\right)^{2\sigma+\mu+m} \left(\sin\tfrac{\vartheta}{2}\right)^{2n-2\sigma-\mu-m} \tag{64}$$

are the rotation coefficients (Hald et al. 1988; Edmonds 1974), the summation over σ involving all terms in which the binomial coefficients do not lead to negative arguments for the factorials. Moreover,

$$A^n_{v\mu}(\beta d) = \sqrt{\frac{(2n+1)\,(2v+1)}{n(n+1)v(v+1)}} \sqrt{\frac{(v+\mu)!(n-\mu)!}{(v-\mu)!(n+\mu)!}} (-1)^{\mu} \frac{1}{2} \mathrm{j}^{n-v} \cdot \sum_{p=|n-v|}^{n+v} \left\{\mathrm{j}^{-p}[n(n+1) + v(v+1) - p(p+1)]\ \tau\ (\mu, n, -\mu, v, p)\ h^{(2)}_p(\beta d)\right\} \tag{65}$$

$$B^n_{v\mu}(\beta d) = \sqrt{\frac{(2n+1)\,(2v+1)}{n(n+1)v(v+1)}} \sqrt{\frac{(v+\mu)!(n-\mu)!}{(v-\mu)!(n+\mu)!}} (-1)^{\mu} \frac{1}{2} \mathrm{j}^{n-v} \cdot \sum_{p=|n-v|}^{n+v} \left[\mathrm{j}^{-p}(2\mathrm{j}\mu\beta d)\ \tau\ (\mu, n, -\mu, v, p)\ h^{(2)}_p(\beta d)\right] \tag{66}$$

are the translation coefficients (Hald et al. 1988; Larsen 1980; Bruning and Lo 1971), wherein $\tau\ (\mu, n, -\mu, v, p)$ are the linearization coefficients defined by the expansion of the product of two unnormalized associated Legendre functions

$$P^{\mu}_n(x)P^{-\mu}_v(x) = \sum_{p=|n-v|}^{n+v} \tau(\mu, n, -\mu, v, p)P_p(x) \tag{67}$$

The probe expansion coefficients $a'_{1,2v1}$ in Eqs. 61, 62 can be evaluated from the knowledge of its tangential electric field $\underline{E}'_t$ on a scanning sphere when it is used as transmitting antenna

DOI 10.1007/978-981-4560-75-7_117-1

$$a'_{1,2v1} = \frac{1}{\beta g_{1,2v}(\beta d)} \int_0^{2\pi} \int_0^{\pi} \underline{E}'_t(d,\vartheta,\varphi) \cdot \underline{f}^*_{1,2v1}(\vartheta) \mathrm{e}^{-\mathrm{j}\varphi} \sin\vartheta \, \mathrm{d}\vartheta \, \mathrm{d}\varphi \tag{68}$$

Classical Wave Expansions of the Field Radiated by an Antenna

Some analytical details on the wave expansions commonly employed to represent the antenna field in the region external to it are collected in this paragraph. In any case, its reading is not strictly necessary for the comprehension of the Chapter and its aim is to enable the interested reader to achieve a more in-depth knowledge of the topic.

Plane Wave Expansion

As well known, in a linear, homogeneous, and isotropic medium, an EM field

$$\underline{E} = \underline{\hat{E}} \mathrm{e}^{-\mathrm{j}\underline{k}\cdot\underline{r}}; \; \underline{H} = \underline{\hat{H}} \mathrm{e}^{-\mathrm{j}\underline{k}\cdot\underline{r}} \tag{69}$$

represents a time-harmonic plane wave whose direction of propagation is specified by the vector $\underline{k} = k_x\hat{x} + k_y\hat{y} + k_z\hat{z}$.

Since the operator ∇ transforms into $-\mathrm{j}\underline{k}$ for such a field, Maxwell's equations in a region free of sources become

$$\underline{k} \times \underline{\hat{E}} = \omega\mu\underline{\hat{H}} \tag{70}$$

$$\underline{k} \times \underline{\hat{H}} = -\omega\varepsilon\underline{\hat{E}} \tag{71}$$

$$\underline{k} \cdot \underline{\hat{E}} = 0 \tag{72}$$

$$\underline{k} \cdot \underline{\hat{H}} = 0 \tag{73}$$

From Eqs. 70 and 71, it results that $\underline{\hat{E}}$ and $\underline{\hat{H}}$ cannot be chosen independently, for instance, $\underline{\hat{H}}$ is related to $\underline{\hat{E}}$ by

$$\underline{\hat{H}} = \frac{1}{\omega\mu}\underline{k} \times \underline{\hat{E}} \tag{74}$$

From Eq. 72, it results that, for any k_x, k_y, and k_z, only two components of $\underline{\hat{E}}(\underline{k})$ are independent. Let these be $\hat{E}_x$ and $\hat{E}_y$, then

$$\hat{E}_z = -\left(\hat{E}_x k_x + \hat{E}_y k_y\right)/k_z \tag{75}$$

From Eq. 70, it results that

$$\underline{k} \times \left(\underline{k} \times \underline{\hat{E}}\right) = \omega\mu\underline{k} \times \underline{\hat{H}} \tag{76}$$

Such a relation, taking into account Eqs. 71, 72 and the vector identity $\underline{a} \times (\underline{b} \times \underline{c}) = \underline{b}(\underline{a} \cdot \underline{c}) - \underline{c}(\underline{a} \cdot \underline{b})$, becomes

DOI 10.1007/978-981-4560-75-7_117-1

$$\underline{\hat{E}}\left[(\underline{k}\cdot\underline{k})-\omega^2\varepsilon\mu\right]=0 \tag{77}$$

In order that such an equation have a solution different from the trivial one $\underline{\hat{E}}=0$, it is necessary that

$$\underline{k}\cdot\underline{k}=k_x^2+k_y^2+k_z^2=\omega^2\varepsilon\mu=\beta^2 \tag{78}$$

Thus, at a fixed frequency, only two of the components of $\underline{k}$ can be independently specified. Let these be k_x and k_y, then the third component is related to them via the relation

$$k_z=\begin{cases}\left(\beta^2-k_x^2-k_y^2\right)^{1/2}, & \text{if } k_x^2+k_y^2\leqslant\beta^2\\ -\mathrm{j}\left(k_x^2+k_y^2-\beta^2\right)^{1/2}, & \text{if } k_x^2+k_y^2>\beta^2\end{cases} \tag{79}$$

The negative sign in the second of Eq. 79 is chosen in order to ensure that the plane wave Eq. 69 is bounded at infinity.

Since the medium is linear as well as the field equations, the general solution for $\underline{E}$ and $\underline{H}$ can be constructed as linear combination over all k_x and k_y:

$$\underline{E}(\underline{r})=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\underline{\hat{E}}(\underline{k})\mathrm{e}^{-\mathrm{j}\underline{k}\cdot\underline{r}}\,\mathrm{d}k_x\mathrm{d}k_y;\ \underline{H}(\underline{r})=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{\omega\mu}\underline{k}\times\underline{\hat{E}}(\underline{k})\mathrm{e}^{-\mathrm{j}\underline{k}\cdot\underline{r}}\,\mathrm{d}k_x\mathrm{d}k_y \tag{80}$$

Namely, any arbitrary monochromatic wave can be represented as a superposition of plane waves with different amplitudes and propagating in different directions. The amplitude function $\underline{\hat{E}}(\underline{k})$ is called the plane wave spectrum of the field. It is worthy to note that the dimensions of $\underline{\hat{E}}$ in Eqs. 69 and 80 are different although the same notation has been used.

In general, the evaluation of the double integral in Eq. 80 is difficult. However, it can be shown (Franceschetti 1997; Balanis 1997), by applying the stationary phase method, that in the far-field region such an integral can be represented by the asymptotic expansion

$$\underline{E}(r,\vartheta,\varphi)=\mathrm{j}2\pi k_z\underline{\hat{E}}(k_x,k_y)\frac{\mathrm{e}^{-\mathrm{j}\beta r}}{r} \tag{81}$$

where

$$k_x=\beta\sin\vartheta\cos\varphi;\quad k_y=\beta\sin\vartheta\sin\varphi;\quad k_z=\beta\cos\vartheta \tag{82}$$

Cylindrical Wave Expansion

In the cylindrical wave expansion (CWE), the field is represented as superposition of transverse electric (TE) and transverse magnetic (TM) cylindrical modes (Stratton 1941), which are elementary solutions of the homogeneous wave equation in the cylindrical coordinates (ρ, φ, z). In fact, an arbitrary field in a homogeneous source-free region can be always expressed as the sum of a TE field and a TM field (Harrington 1961). The expression for these modes can be easily derived by using the auxiliary vector potentials $\underline{A}$ and $\underline{F}$, which, in such a hypothesis, are solutions of the homogeneous vector Helmoltz equations:

DOI 10.1007/978-981-4560-75-7_117-1

$$\nabla^2 \underline{A} + \beta^2 \underline{A} = 0;\ \nabla^2 \underline{F} + \beta^2 \underline{F} = 0 \tag{83}$$

As well known, the fields $\underline{E}$ and $\underline{H}$ are related to them by

$$\underline{E} = -\mathrm{j}\omega \underline{A} + \frac{\nabla\nabla \cdot \underline{A}}{\mathrm{j}\omega\varepsilon\mu} - \frac{1}{\varepsilon}\nabla \times \underline{F};\quad \underline{H} = -\mathrm{j}\omega \underline{F} + \frac{\nabla\nabla \cdot \underline{F}}{\mathrm{j}\omega\varepsilon\mu} + \frac{1}{\mu}\nabla \times \underline{A} \tag{84}$$

From Eq. 84, it can be easily recognized that, to derive the field expressions that are TE to z, it is sufficient to assume (Balanis 1989)

$$\underline{A} = 0;\ \underline{F} = \hat{z}F(\rho, \varphi, z) \tag{85}$$

whereas

$$\underline{A} = \hat{z}A(\rho, \varphi, z);\ \underline{F} = 0 \tag{86}$$

for the TM to z modes (Balanis 1989). In both the cases, the vector Helmoltz equations reduce to the scalar ones. Their elementary solution is obtainable by using the separation of variables method (Balanis 1989) and, in the considered case of observation region external to smallest cylinder enclosing the AUT, is given by $H_v^{(2)}(\Lambda\rho)\mathrm{e}^{\mathrm{j}v\varphi}\mathrm{e}^{-\mathrm{j}\eta z}$, where v is an integer, η is a real number, $\Lambda = \left(\beta^2 - \eta^2\right)^{1/2}$, and $H_v^{(2)}(\Lambda\rho)$ is the Hankel function of second kind and order v.

By substituting such an elementary solution in Eq. 84, it results

$$\underline{E} = (-1/\varepsilon)\underline{M}_{v\eta} \tag{87}$$

for the TE case and

$$\underline{E} = (\beta/\mathrm{j}\omega\varepsilon\mu)\underline{N}_{v\eta} \tag{88}$$

for the TM one, wherein

$$\underline{M}_{v\eta}(\rho, \varphi, z) = \left(\frac{\mathrm{j}v}{\rho}H_v^{(2)}(\Lambda\rho)\hat{\rho} - \frac{\partial}{\partial\rho}H_v^{(2)}(\Lambda\rho)\hat{\varphi}\right)\mathrm{e}^{\mathrm{j}v\varphi}\mathrm{e}^{-\mathrm{j}\eta z} \tag{89}$$

$$\underline{N}_{v\eta}(\rho, \varphi, z) = \left(-\frac{\mathrm{j}\eta}{\beta}\frac{\partial}{\partial\rho}H_v^{(2)}(\Lambda\rho)\hat{\rho} + \frac{v\eta}{\beta\rho}H_v^{(2)}(\Lambda\rho)\hat{\varphi} + \frac{\Lambda^2}{\beta}H_v^{(2)}(\Lambda\rho)\hat{z}\right)\mathrm{e}^{\mathrm{j}v\varphi}\mathrm{e}^{-\mathrm{j}\eta z} \tag{90}$$

Accordingly, the electric field radiated by the AUT at a point external to the smallest cylinder enclosing it can be represented (Stratton 1941) as a linear combination of the elementary fields Eqs. 87 and 88 involving an integral over all η values, and a sum over all v:

$$\underline{E}(\rho, \varphi, z) = \sum_{v=-\infty}^{\infty}\int_{-\infty}^{\infty}\left[a_v(\eta)\underline{M}_{v\eta}(\rho, \varphi, z) + b_\eta(\eta)\underline{N}_{v\eta}(\rho, \varphi, z)\right]\mathrm{d}\eta \tag{91}$$

therefore

DOI 10.1007/978-981-4560-75-7_117-1

$$E_\rho(\rho,\varphi,z)=\sum_{v=-\infty}^{\infty}\int_{-\infty}^{\infty}\left[a_v(\eta)\frac{\mathrm{j}v}{\rho}H_v^{(2)}(\Lambda\rho)-\frac{\mathrm{j}\eta}{\beta}b_v(\eta)\frac{\partial}{\partial\rho}H_v^{(2)}(\Lambda\rho)\right]\mathrm{e}^{\mathrm{j}v\varphi}\mathrm{e}^{-\mathrm{j}\eta z}\mathrm{d}\eta \tag{92}$$

$$E_\varphi(\rho,\varphi,z)=\sum_{v=-\infty}^{\infty}\int_{-\infty}^{\infty}\left[b_v(\eta)\frac{v\eta}{\beta\rho}H_v^{(2)}(\Lambda\rho)-a_v(\eta)\frac{\partial}{\partial\rho}H_v^{(2)}(\Lambda\rho)\right]\mathrm{e}^{\mathrm{j}v\varphi}\mathrm{e}^{-\mathrm{j}\eta z}\mathrm{d}\eta \tag{93}$$

$$E_z(\rho,\varphi,z)=\sum_{v=-\infty}^{\infty}\int_{-\infty}^{\infty}b_v(\eta)\frac{\Lambda^2}{\beta}H_v^{(2)}(\Lambda\rho)\mathrm{e}^{\mathrm{j}v\varphi}\mathrm{e}^{-\mathrm{j}\eta z}\mathrm{d}\eta \tag{94}$$

The evaluations of the FF components of the electric field in the spherical coordinate system (r, ϑ, φ) from the above expressions involve the following steps. The first step is to replace in the CWE the Hankel function and its first derivative by their asymptotic expansions

$$H_v^{(2)}(x)\underset{x\to\infty}{\approx}\sqrt{\frac{2\mathrm{j}}{\pi x}}\mathrm{j}^v\mathrm{e}^{-\mathrm{j}x};\quad\frac{\mathrm{d}}{\mathrm{d}x}H_v^{(2)}(\alpha x)\underset{x\to\infty}{\approx}\sqrt{\frac{2\mathrm{j}\alpha}{\pi x}}\mathrm{j}^{v-1}\mathrm{e}^{-\mathrm{j}\alpha x} \tag{95}$$

The obtained expressions are then rewritten in terms of the spherical coordinates r, ϑ, φ. The next step is the asymptotic evaluation of the integrals over η by means of the method of the stationary phase (James 1980). At last, the FF components of the electric field in the cylindrical coordinate system are transformed in the spherical ones, thus obtaining Eqs. 34 and 35.

Spherical Wave Expansion

In the spherical wave expansion (SWE), the field is represented as superposition of transverse electric (to the radial direction) spherical modes (TE^r) and transverse magnetic ones (TM^r) (Stratton 1941), which are elementary solutions of the homogeneous wave equation in the spherical coordinates (r, ϑ, φ). In fact, an arbitrary field in a homogeneous source-free region can be expressed as the sum of a TE field and a TM field (Harrington 1961). The expression for these modes can be determined by using the auxiliary vector potentials $\underline{A}$ and $\underline{F}$. In particular, to derive the field expressions that are TE to r, it is sufficient to assume (Balanis 1989)

$$\underline{A}=0;\quad\underline{F}=\hat{r}F(r,\vartheta,\varphi) \tag{96}$$

whereas

$$\underline{A}=\hat{r}A(r,\vartheta,\varphi);\quad\underline{F}=0 \tag{97}$$

for the TM to r modes (Balanis 1989).

It must be stressed that, unlike the cylindrical scanning case, it is not convenient now to determine the auxiliary potentials by means of the vector Helmholtz equations (Eq. 83), valid in a source-free region when assuming that the vector potentials $\underline{A}$, $\underline{F}$ and the scalar ones Φ, Ψ satisfy the Lorentz conditions:

$$\nabla\cdot\underline{A}+\mathrm{j}\omega\varepsilon\mu\Phi=0;\quad\nabla\cdot\underline{F}+\mathrm{j}\omega\varepsilon\mu\Psi=0 \tag{98}$$

DOI 10.1007/978-981-4560-75-7_117-1

In fact, since $\nabla^2(\hat{r}G) \neq \hat{r}\nabla^2 G$, the vector Helmholtz equations (Eq. 83) do not reduce any longer to the corresponding scalar ones. Obviously, in such a case, also the expressions (Eq. 84), relating the fields to the auxiliary potentials, are no longer valid. Different relations for the fields and potentials that are not based on Lorentz conditions are now derived.

The TE^r case ($\underline{A} = 0, \underline{F} = \hat{r}F(r, \vartheta, \varphi)$) is first considered. It can be easily recognized that

$$\underline{E} = -\frac{1}{\varepsilon}\nabla \times \underline{F}; \quad \underline{H} = \frac{1}{\mathrm{j}\omega\varepsilon\mu}\nabla \times \nabla \times \underline{F} \tag{99}$$

By substituting the former of Eq. 99 in the second of Maxwell's equations, it results

$$\nabla \times (\underline{H} + \mathrm{j}\omega\underline{F}) = 0 \tag{100}$$

and, accordingly,

$$\underline{H} = -\mathrm{j}\omega\underline{F} - \nabla\Psi \tag{101}$$

By substituting Eq. 101 into the latter of Eq. 99, it is obtained

$$\nabla \times \nabla \times \underline{F} - \beta^2\underline{F} = -\mathrm{j}\omega\varepsilon\mu\nabla\Psi \tag{102}$$

By expanding Eq. 102 in spherical coordinates and imposing that $\Psi = (-1/\mathrm{j}\omega\varepsilon\mu)\ \partial F/\partial r$, it can be shown (Balanis 1989) that F can be determined by solving the scalar Helmholtz equation

$$\nabla^2\psi + \beta^2\psi = 0 \tag{103}$$

where

$$\psi(r, \vartheta, \varphi) = F(r, \vartheta, \varphi)/r \tag{104}$$

The elementary solution to Eq. 103 can be obtained by using the separation of variables method (Balanis 1989) and, in the considered case of an observation region external to the smallest sphere enclosing the AUT, it is given by

$$\psi(r, \vartheta, \varphi) = h_n^{(2)}(\beta r)\overline{P}_n^{|m|}(\cos\vartheta)\mathrm{e}^{\mathrm{j}m\varphi} \tag{105}$$

where n and m are integers, $\overline{P}_n^m(x)$ is the normalized associated Legendre function as defined by Belousov (1962), and $h_n^{(2)}(x)$ is the spherical Hankel function of second kind and order n, which is related to the ordinary Hankel function of second kind by

$$h_n^{(2)}(x) = \sqrt{\frac{\pi}{2x}}H_{n+1/2}^{(2)}(x) \tag{106}$$

The normalized associated Legendre function $\overline{P}_n^m(x)$ is related to the associated Legendre function $P_n^m(x)$ (solution of the associated Legendre differential equation (Balanis 1989)) by

DOI 10.1007/978-981-4560-75-7_117-1

$$\overline{P}_n^m(x) = \sqrt{\frac{2n+1}{2}\frac{(n-m)!}{(n+m)!}}P_n^m(x) \tag{107}$$

A similar procedure can be applied in the TM^r case ($\underline{A} = \hat{r}A(r,\vartheta,\varphi), \underline{F} = 0$), thus obtaining

$$\underline{H} = \frac{1}{\mu}\nabla \times \underline{A}; \quad \underline{E} = \frac{1}{\mathrm{j}\omega\varepsilon\mu}\nabla \times \nabla \times \underline{A} \tag{108}$$

and

$$\nabla \times \nabla \times \underline{A} - \beta^2 \underline{A} = -\mathrm{j}\omega\varepsilon\mu\nabla\Phi \tag{109}$$

By expanding Eq. 109 in spherical coordinates and imposing that $\Phi = (-1/\mathrm{j}\omega\varepsilon\mu)\ \partial A/\partial r$, it can be shown (Balanis 1989) that A can be determined by solving the scalar Helmholtz equation (Eq. 103), wherein now

$$\psi(r,\vartheta,\varphi) = A(r,\vartheta,\varphi)/r \tag{110}$$

The transverse electric field $\underline{E}_t$ related to an elementary spherical TE^r wave can be easily found from the former of Eq. 99, by expanding $\nabla \times \underline{F}$ in spherical coordinates and taking into account Eqs. 104 and 105, thus obtaining

$$\underline{E}_t = -\frac{1}{\varepsilon}g_{1n}(\beta r)\widetilde{\underline{F}}_{1nm}(\vartheta,\varphi) = -\frac{1}{\varepsilon}g_{1n}(\beta r)\widetilde{\underline{f}}_{1nm}(\vartheta)\mathrm{e}^{\mathrm{j}m\varphi} \tag{111}$$

where

$$g_{1n}(\beta r) = h_n^{(2)}(\beta r); \quad \widetilde{\underline{f}}_{1nm}(\vartheta) = \frac{\mathrm{j}m}{\sin\vartheta}\overline{P}_n^{|m|}(\cos\vartheta)\hat{\vartheta} - \frac{\mathrm{d}}{\mathrm{d}\vartheta}\overline{P}_n^{|m|}(\cos\vartheta)\hat{\varphi} \tag{112}$$

In a similar way, the transverse electric field related to an elementary spherical TM^r wave can be determined from the latter of Eq. 108, by expanding $\nabla \times \nabla \times \underline{A}$ in spherical coordinates and taking into account relations Eqs. 105 and 110, thus getting

$$\underline{E}_t = \frac{\beta}{\mathrm{j}\omega\varepsilon\mu}g_{2n}(\beta r)\widetilde{\underline{F}}_{2nm}(\vartheta,\varphi) = \frac{\beta}{\mathrm{j}\omega\varepsilon\mu}g_{2n}(\beta r)\widetilde{\underline{f}}_{2nm}(\vartheta)\mathrm{e}^{\mathrm{j}m\varphi} \tag{113}$$

where

$$g_{2n}(\beta r) = \frac{1}{\beta r}\frac{\mathrm{d}}{\mathrm{d}(\beta r)}\left(\beta r h_n^{(2)}(\beta r)\right); \quad \widetilde{\underline{f}}_{2nm}(\vartheta) = \frac{\mathrm{d}}{\mathrm{d}\vartheta}\overline{P}_n^{|m|}(\cos\vartheta)\hat{\vartheta} + \frac{\mathrm{j}m}{\sin\vartheta}\overline{P}_n^{|m|}(\cos\vartheta)\hat{\varphi} \tag{114}$$

By expressing the transverse electric field radiated by the AUT on a sphere of radius r containing it as a superposition of elementary TE^r and TM^r spherical waves, it results

$$\underline{E}_t = \sum_{n=1}^{N_{\max}}\sum_{m=-n}^{n}\left[\tilde{a}_{1nm}g_{1n}(\beta r)\widetilde{\underline{F}}_{1nm}(\vartheta,\varphi) + \tilde{a}_{2nm}g_{2n}(\beta r)\widetilde{\underline{F}}_{2nm}(\vartheta,\varphi)\right] \tag{115}$$

DOI 10.1007/978-981-4560-75-7_117-1

where $N_{\max}$ is the index of the highest spherical wave to be considered. It is worth noting that the summation on m extends from $-n$ to n, since $\overline{P}_n^{|m|}(\cos\vartheta)$ vanishes for $|m| > n$, whereas that on n starts from 1, since $\underline{\widetilde{F}}_{100}$ and $\underline{\widetilde{F}}_{200}$ are both null (Hald et al. 1988).

It can be shown (Hald et al. 1988) that

$$\left\langle \underline{\widetilde{F}}_{1nm}, \underline{\widetilde{F}}_{2\overline{nm}} \right\rangle = \int_0^{2\pi}\int_0^{\pi} \underline{\widetilde{F}}_{1nm}(\vartheta,\varphi)\cdot \underline{\widetilde{F}}^{*}_{2\overline{nm}}(\vartheta,\varphi)\sin\vartheta \mathrm{d}\vartheta\mathrm{d}\varphi = 0 \tag{116}$$

According to Eq. 116, the spherical wave functions $\underline{\widetilde{F}}_{1nm}$ and $\underline{\widetilde{F}}_{2\overline{nm}}$ are always orthogonal on the sphere even when $\overline{n} = n$ and $\overline{m} = m$. Moreover, it can be shown (Hald et al. 1988) that

$$\left\langle \underline{\widetilde{F}}_{1nm}, \underline{\widetilde{F}}_{1\overline{nm}} \right\rangle = \left\langle \underline{\widetilde{F}}_{2nm}, \underline{\widetilde{F}}_{2\overline{nm}} \right\rangle = 2\pi n(n+1)\delta_{m\overline{m}}\delta_{n\overline{n}} \tag{117}$$

where

$$\delta_{m\overline{m}} = \begin{cases} 1 & \text{for} \quad \overline{m} = m \\ 0 & \text{for} \quad \overline{m} \neq m \end{cases} \tag{118}$$

is the Kronecker delta.

In order to have orthonormal spherical wave functions and power-normalized spherical waves, it is convenient to introduce the new functions

$$\underline{F}_{1nm}(\vartheta,\varphi) = \left(\frac{-m}{|m|}\right)^m \frac{1}{\sqrt{2\pi n(n+1)}} \underline{\widetilde{F}}_{1nm}(\vartheta,\varphi) \tag{119}$$

$$\underline{F}_{2nm}(\vartheta,\varphi) = \left(\frac{-m}{|m|}\right)^m \frac{1}{\sqrt{2\pi n(n+1)}} \underline{\widetilde{F}}_{2nm}(\vartheta,\varphi) \tag{120}$$

The factor $(-m/|m|)^m$ ensures that the phase of the modes follows the phase of the spherical harmonics as defined by Edmonds (1974).

The transverse electric field radiated by the AUT on a sphere of radius r can be then rewritten in the form

$$\underline{E}_t = \beta \sum_{n=1}^{N_{\max}} \sum_{m=-n}^{n} \left[a_{1nm}\, g_{1n}(\beta r)\underline{F}_{1nm}(\vartheta,\varphi) + a_{2nm}\, g_{2n}(\beta r)\underline{F}_{2nm}(\vartheta,\varphi) \right] \tag{121}$$

The expression of the transverse electric field radiated by the AUT in its far-field region can be easily obtained from Eq. 121 by taking into account that

$$h_n^{(2)}(\beta r) \underset{\beta r\to\infty}{\approx} \mathrm{j}^{n+1}\frac{\mathrm{e}^{-\mathrm{j}\beta r}}{\beta r}; \quad \frac{1}{\beta r}\frac{\mathrm{d}}{\mathrm{d}(\beta r)}\left(\beta r h_n^{(2)}(\beta r)\right) \underset{\beta r\to\infty}{\approx} \mathrm{j}^{n}\frac{\mathrm{e}^{-\mathrm{j}\beta r}}{\beta r} \tag{122}$$

thus obtaining Eq. 52.

DOI 10.1007/978-981-4560-75-7_117-1

It is worth noting that, according to the Sommerfeld radiation conditions, in the far-field region the longitudinal electric field component E_r must vanish. Indeed, this is the case. As a matter of fact, each elementary TE^r spherical wave has no longitudinal component, whereas that relevant to a TM^r one decreases asymptotically as $1/r^2$ (Hald et al. 1988).

Nonredundant NF-FF Transformation Techniques

Nonredundant Sampling Representations of EM Fields

The main theoretical results concerning the nonredundant sampling representations of EM fields (Bucci et al. 1998c; Bucci and Gennarelli 2012) are summarized in this section. To this end, the field radiated by an arbitrary finite-size source enclosed in a convex domain $\mathcal{D}$, bounded by a surface Σ with rotational symmetry and observed on a regular surface $\mathcal{M}$ external to $\mathcal{D}$ and having the same symmetry, is considered. Due to its rotational symmetry, the surface $\mathcal{M}$ can be described by meridian curves and azimuthal circumferences, so that the field representation over a regular curve $\mathcal{C}$ described by a proper analytical parameterization $\underline{r} = \underline{r}(\xi)$ is considered in the following. Moreover, according to (Bucci et al. 1998c), it is convenient to define the reduced electric field

$$\underline{F}(\xi) = \underline{E}(\xi)\ e^{j\gamma(\xi)} \tag{123}$$

where $\gamma(\xi)$ is a proper analytical phase function to be determined. As shown in (Bucci and Franceschetti 1987; Bucci et al. 1998c), the "band limitation" error, occurring when the reduced field $\underline{F}(\xi)$ is approximated by a spatially band-limited function, becomes negligible as the bandwidth exceeds the critical value

$$W_\xi = \max_\xi [w(\xi)] = \max_\xi \left[\max_{\underline{r}'} \left| \frac{d\gamma(\xi)}{d\xi} - \beta \frac{\partial R(\xi, \underline{r}')}{\partial \xi} \right| \right] \tag{124}$$

where $\underline{r}'$ denotes the source point and $R = \left| \underline{r}(\xi) - \underline{r}' \right|$. As a matter of fact, for large sources and observation domains not too near them, such an error exhibits a step-like behavior, decreasing more than exponentially as the bandwidth exceeds W_ξ (Bucci and Franceschetti 1987; Bucci et al. 1998c). As a consequence, it can be effectively controlled by choosing the bandwidth of the approximating function equal to $\chi' W_\xi$, where χ' is an enlargement bandwidth factor (slightly greater than unity for electrically large antennas).

In order to obtain a nonredundant representation, namely, a representation requiring a minimum number of samples, first of all, the "local" bandwidth $w(\xi)$ must be minimized for each ξ. This is accomplished by choosing γ such that its derivative is given by

$$\frac{d\gamma}{d\xi} = \frac{\beta}{2} \left[\max_{\underline{r}'} \frac{\partial R}{\partial s} + \min_{\underline{r}'} \frac{\partial R}{\partial s} \right] \frac{ds}{d\xi} = \frac{\beta}{2} \left[\max_{\underline{r}'} \hat{R} \cdot \hat{t} + \min_{\underline{r}'} \hat{R} \cdot \hat{t} \right] \frac{ds}{d\xi} \tag{125}$$

namely, $d\gamma/d\xi$ is the average between the maximum and minimum value of $\beta \partial R/\partial \xi$, when $\underline{r}'$ varies in the source domain $\mathcal{D}$. In Eq. 125, s is the curvilinear abscissa on $\mathcal{C}$, $\hat{t}$ is the unit vector tangent to $\mathcal{C}$ at the observation point P, and $\hat{R}$ is the unit vector pointing from the source point to P. Accordingly, the optimal phase factor to be used is given by

DOI 10.1007/978-981-4560-75-7_117-1

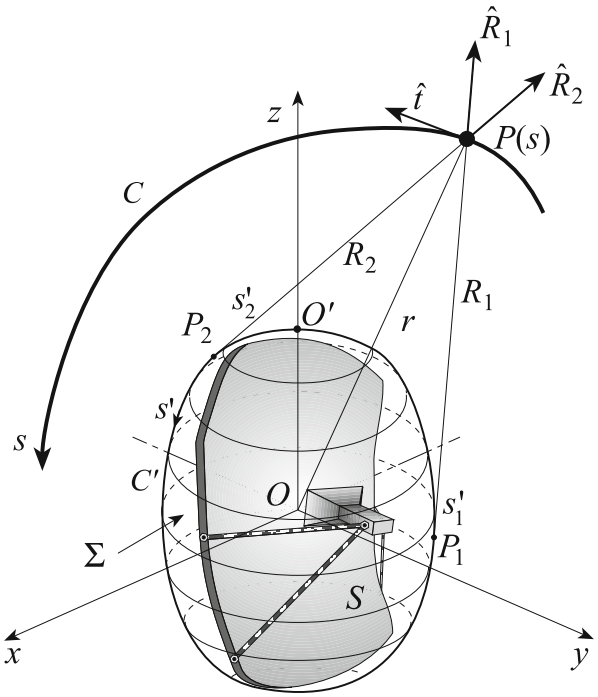


Fig. 17 Relevant to a meridian curve

$$\gamma(\xi) = \frac{\beta}{2} \int_0^{s(\xi)} \left[\max_{\underline{r}'} \frac{\partial R}{\partial s} + \min_{\underline{r}'} \frac{\partial R}{\partial s} \right] \mathrm{d}s = \frac{\beta}{2} \int_0^{s} \left[\max_{\underline{r}'} \hat{R} \cdot \hat{t} + \min_{\underline{r}'} \hat{R} \cdot \hat{t} \right] \mathrm{d}s \tag{126}$$

It can be easily verified that, by choosing $\mathrm{d}\gamma/\mathrm{d}\xi$ according to Eq. 125, it results

$$w(\xi) = \frac{\beta}{2} \left[\max_{\underline{r}'} \frac{\partial R}{\partial s} - \min_{\underline{r}'} \frac{\partial R}{\partial s} \right] \frac{\mathrm{d}s}{\mathrm{d}\xi} = \frac{\beta}{2} \left[\max_{\underline{r}'} \hat{R} \cdot \hat{t} - \min_{\underline{r}'} \hat{R} \cdot \hat{t} \right] \frac{\mathrm{d}s}{\mathrm{d}\xi} \tag{127}$$

With reference to the parameter ξ, for avoiding redundancy, it must be determined by requiring that the local bandwidth is constant. In fact, if $w(\xi)$ is variable with ξ, the sample spacing, which is dictated by the bandwidth W_ξ, becomes unnecessarily small in the zones wherein $w(\xi)$ is smaller than its maximum value, giving rise to redundancy in the sampling representation. Accordingly, by imposing that $w(\xi)$ is constant and equal to W_ξ, it results

$$\xi = \frac{\beta}{2W_\xi} \int_0^{s} \left[\max_{\underline{r}'} \frac{\partial R}{\partial s} - \min_{\underline{r}'} \frac{\partial R}{\partial s} \right] \mathrm{d}s = \frac{\beta}{2W_\xi} \int_0^{s} \left[\max_{\underline{r}'} \hat{R} \cdot \hat{t} - \min_{\underline{r}'} \hat{R} \cdot \hat{t} \right] \mathrm{d}s \tag{128}$$

When $\mathcal{C}$ is a meridian curve (Fig. 17) and $\hat{t}$ is external to the cone of vertex P, tangent to Σ, the extreme values of $\hat{R} \cdot \hat{t}$ occur at the two tangency points $P_{1,2}$ on $\mathcal{C}'$ (intersection curve between the meridian plane and Σ). By taking into account that in such a case (Bucci et al. 1998c)

DOI 10.1007/978-981-4560-75-7_117-1

$$\left.\frac{\partial R}{\partial s}\right|_{s'_{1,2}} = \frac{\mathrm{d}R_{1,2}}{\mathrm{d}s} \pm \frac{\mathrm{d}s'_{1,2}}{\mathrm{d}s} \tag{129}$$

where $s'_{1,2}$ are the arclength coordinates of $P_{1,2}$ and $R_{1,2}$ the distances from P to $P_{1,2}$ (Fig. 17) and choosing $W_\xi = \beta\ell'/2\pi$ (ℓ' being the length of C'), it results

$$\gamma = \frac{\beta}{2}\left[R_1 + R_2 + s'_1 - s'_2\right] \tag{130}$$

$$\xi = \frac{\pi}{\ell'}\left[R_1 - R_2 + s'_1 + s'_2\right] \tag{131}$$

Note that the angular-like parameter ξ covers a 2π range when P encircles the source once and the number of samples at Nyquist spacing ($\Delta\xi = \pi/W_\xi$) on a closed meridian curve $\mathcal{C}$ (also unbounded) is always finite and equal to $N_\xi = 2\pi/\Delta\xi = 2\ell'/\lambda$.

It can be shown (Bucci et al. 1998c) that in any meridian plane the curves $\gamma = \text{const}$ and $\xi = \text{const}$ are orthogonal.

When $\mathcal{C}$ is an azimuthal circumference of radius ρ, the extreme values of $\hat{R}\cdot\hat{t}$ are opposite and constant along it. It follows from Eqs. 126 and 128 that γ is constant and any parameter proportional to the arclength is optimal. As a consequence, it is convenient to use the azimuthal angle φ as parameter and to choose for γ the value relevant to any meridian curve passing through the observation point on $\mathcal{C}$. The corresponding bandwidth is (Bucci et al. 1998c)

$$W_\varphi = \frac{\beta}{2}\max_{z'}\left(R^+ - R^-\right) = \frac{\beta}{2}\max_{z'}\left(\sqrt{(z-z')^2 + (\rho+\rho'(z'))^2} - \sqrt{(z-z')^2 + (\rho-\rho'(z'))^2}\right) \tag{132}$$

where R^+ and R^- are the maximum and minimum distance, respectively, from each circumference describing the surface Σ to the observation point P and $\rho'(z')$ is the equation of Σ in cylindrical coordinates. As shown in (Bucci et al. 1998c), the maximum is attained on that zone of the surface Σ lying on the same side of the observation circumference with respect to the maximum transverse circle of Σ. When the radius of the observation circle goes to infinity, it results $R^+ = R^- + 2\rho'\sin\vartheta$ and, accordingly,

$$W_\varphi = \beta\rho'_{\max}\sin\vartheta \tag{133}$$

ϑ being the polar angle of the circle points and $\rho'_{\max}$ the maximum transverse radius of Σ.

It can be shown (Bucci et al. 1998c) that the number of samples N_s at Nyquist spacing on any closed observation surface (also unbounded) surrounding the source is

$$N_s \cong (\text{area of } \Sigma)/(\lambda/2)^2 \tag{134}$$

Any finite-size source can be always considered as enclosed in a spherical surface. Accordingly, by choosing Σ coincident with a sphere of radius a, for any meridian curve, it results

$$R_1 = R_2 = \sqrt{r^2 - a^2} \tag{135}$$

$$s'_1 = (\vartheta - \alpha)a; \quad s'_2 = (\vartheta + \alpha)a; \quad \alpha = \cos^{-1}(a/r) \tag{136}$$

DOI 10.1007/978-981-4560-75-7_117-1

Accordingly, since $\ell' = 2\pi a$, it results $W_\xi = \beta a$ and

$$\gamma = \beta\sqrt{r^2 - a^2} - \beta a \cos^{-1}(a/r); \quad \xi = \vartheta \tag{137}$$

When considering an azimuthal circumference, by putting $z' = a \cos \vartheta'$ and $\rho' = a \sin \vartheta'$, it can be shown (Bucci et al. 1998c) that the maximum in Eq. 132 is attained at $\vartheta' = \cos^{-1}(a \cos \vartheta/r)$ and, accordingly,

$$W_\varphi = \beta a \sin \vartheta \tag{138}$$

Eq. 134 highlights the role that the source modeling plays in minimizing the overall number of samples: the surface Σ must fit very well the antenna geometry, moreover it must be analytically regular and such that the corresponding phase factor and parameterization are simple to be determined. To this end, ellipsoidal geometries (oblate and prolate) are considered in the following.

An effective modeling for elongated antennas is obtained by choosing the surface Σ coincident with a rotational prolate ellipsoid, having major and minor semiaxes equal to a and b (Fig. 18), whereas a rotational oblate ellipsoid is an appropriate modeling for quasiplanar sources. In both the cases, it can be shown (Bucci et al. 1998c) that in any meridian plane the curves $\gamma = \text{const}$ and $\xi = \text{const}$ are respectively ellipses and hyperbolas confocal to $\mathcal{C}'$. This implies that ξ and γ are functions of the elliptic coordinates $u = (r_1 - r_2)/2f$ and $v = (r_1 + r_2)/2a$, respectively, where $r_{1,2}$ are the distances from the observation point P on the meridian plane to the foci of $\mathcal{C}'$ and $2f$ is the focal distance. By taking into account that the length of the ellipse $\mathcal{C}'$ is $\ell' = 4aE(\pi/2|\varepsilon^2)$, the bandwidth W_ξ becomes

$$W_\xi = (4a/\lambda)E\left(\pi/2|\varepsilon^2\right) \tag{139}$$

where $E(\cdot|\cdot)$ denotes the elliptic integral of second kind and $\varepsilon = f/a$ is the eccentricity of $\mathcal{C}'$.

The evaluation of γ at P can be simplified by considering the intersection point between the confocal ellipse through P and the symmetry axis, instead of P itself, since the same value of γ corresponds to all points lying on a confocal ellipse. By straightforward but lengthy computations, from Eq. 130, it results (Bucci et al. 1998c)

$$\gamma = \beta a \left[v\sqrt{\frac{v^2 - 1}{v^2 - \varepsilon^2}} - E\left(\cos^{-1}\sqrt{\frac{1 - \varepsilon^2}{v^2 - \varepsilon^2}}|\varepsilon^2 \right) \right] \tag{140}$$

With reference to ξ, since the same value corresponds to all points on the confocal hyperbola branch passing through P, it can be convenient to consider, instead of P itself, the intersection point P_0 between the considered hyperbola branch and $\mathcal{C}'$. Accordingly from Eq. 131, it results $\xi = 2\pi s_0'/\ell'$, s_0' being the curvilinear abscissa of P_0. By straightforward computations, the following expression is finally obtained (Bucci et al. 1998c):

$$\xi = \frac{\pi}{2} \begin{cases} E(\sin^{-1}u|\varepsilon^2)/E(\pi/2|\varepsilon^2) + 1 & \text{prolate ellipsoid} \\ E(\sin^{-1}u|\varepsilon^2)/E(\pi/2|\varepsilon^2) & \text{oblate ellipsoid} \end{cases} \tag{141}$$

Such a relation is valid when the angle ϑ corresponding to P belongs to the range $[0, \pi/2]$. The case in which ϑ belongs to $[\pi/2, \pi]$ can be easily handled by determining the value ξ' corresponding to the point specified by the angle $\pi - \vartheta$ and then putting $\xi = \pi - \xi'$.

DOI 10.1007/978-981-4560-75-7_117-1

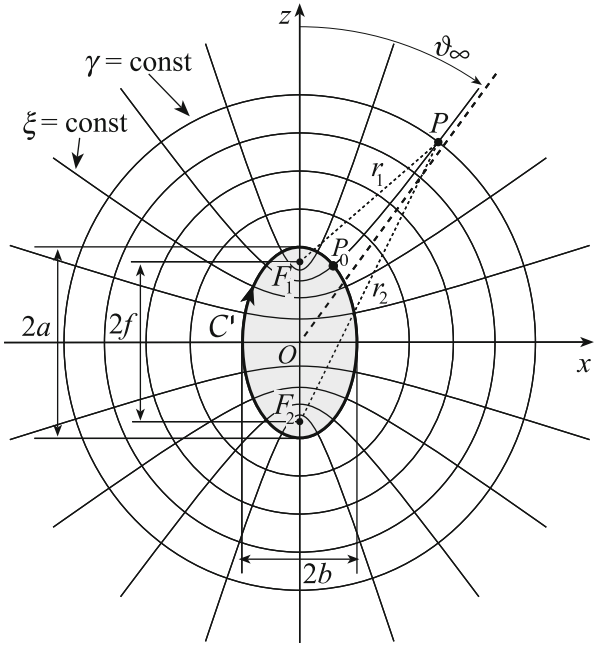


Fig. 18 Ellipsoidal source modeling: prolate case

Since all transverse circles belonging to the hyperboloid of rotation fixed by a value of ξ are characterized by the same value of W_φ (Bucci et al. 1998c), the bandwidth value corresponding to an azimuthal circumference can be evaluated by moving such a circle to infinity along the hyperbola $\xi = \text{const}$. Accordingly, by taking Eq. 133 into account, it results

$$W_\varphi(\xi) = \begin{cases} \beta b \sin \vartheta_\infty(\xi) & \text{prolate ellipsoid} \\ \beta a \sin \vartheta_\infty(\xi) & \text{oblate ellipsoid} \end{cases} \tag{142}$$

where

$$\vartheta_\infty = \begin{cases} \sin^{-1} u + \pi/2 & \text{prolate ellipsoid} \\ \sin^{-1} u & \text{oblate ellipsoid} \end{cases} \tag{143}$$

is the polar angle of the asymptote to the hyperbola through P (see Fig. 18).

It can be easily shown (Bucci et al. 1998c) that the spherical modeling can be obtained from the prolate or oblate ellipsoidal one by considering an ellipsoid with eccentricity $\varepsilon = 0$.

Another effective source modeling for long antennas is obtained by choosing the surface Σ coincident with a rounded cylinder (Fig. 19), namely, a cylinder of height h' ended in two half-spheres of radius a' (Bucci et al. 1998b; D'Agostino et al. 2011). Such a modeling is quite general and contains the spherical one as particular case. When considering quasiplanar antennas, a convenient alternative to the oblate ellipsoidal modeling is the two-bowls one (Fig. 20), wherein Σ is a surface formed by two circular bowls with the same aperture diameter, but with bending radii c and c' of the upper and lower arcs eventually different to fit better the actual AUT geometry (Bucci et al. 1998a; D'Agostino et al. 2011). It can be easily recognized that the surface Σ coincides with a sphere if $c = c' = a$, it becomes a half-sphere if $c = 0$ and $c' = a$, and it reduces to a circular dish for $c = c' = 0$.

DOI 10.1007/978-981-4560-75-7_117-1

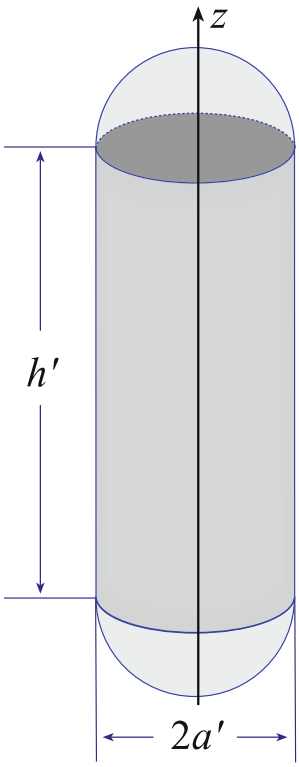


Fig. 19 Rounded cylinder modeling

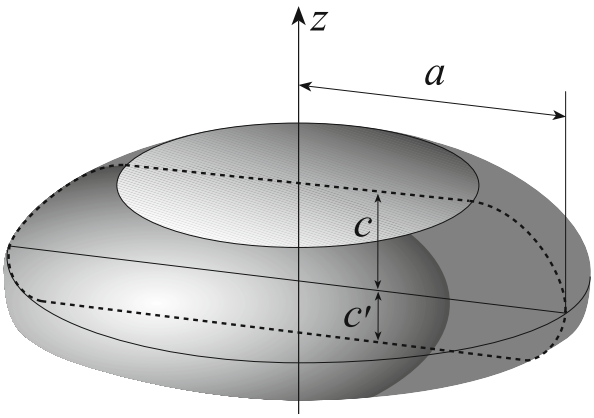


Fig. 20 Two-bowls modeling

The real part of the electric field and reduced electric field y-component radiated by an antenna along a line is reported in Fig. 21 to show the effect of multiplying the field by the factor $e^{j\gamma(\xi)}$. The considered antenna is a uniform planar circular array with radius equal to 20λ, lying in the plane $z = 0$, and has been modeled by an oblate ellipsoid with $2a = 40\lambda$ and $2b = 5\lambda$. The array elements, radially and azimuthally spaced of 0.8λ, are elementary Huygens sources linearly polarized along the y axis. The considered straight line is the x-directed line at $y = 0$, lying in a plane at distance $d = 12\lambda$ from the AUT center. As can be seen, the spatial variations of the reduced field component are slower than the electric field ones, since the local bandwidth has been minimized. Quite analogous results are obtained for the imaginary part.

DOI 10.1007/978-981-4560-75-7_117-1

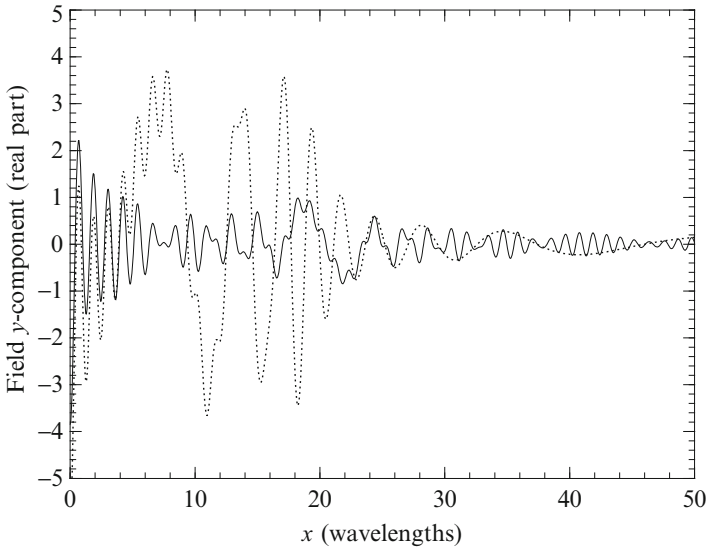


Fig. 21 Real part of the NF *y*-component on the given line. *Solid line*: electric field. *Dashed line*: reduced field

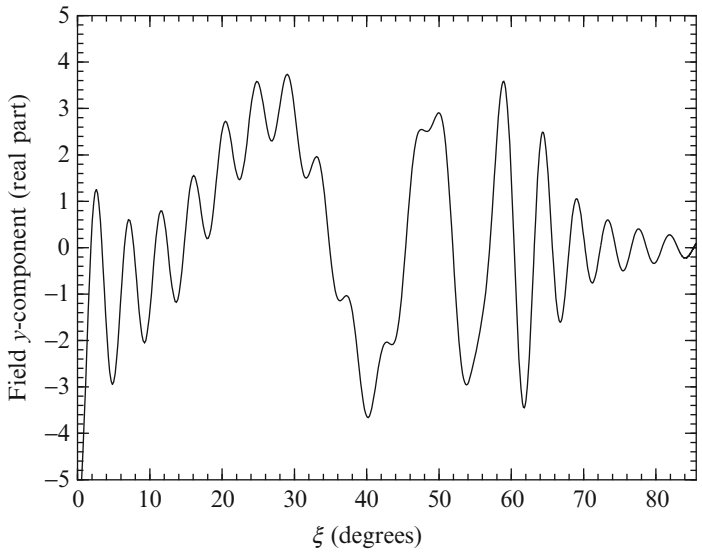


Fig. 22 Real part of the reduced electric field *y*-component on the given line as function of the optimal parameter ξ

In order to highlight the role of the optimal parameter for a convenient representation of the reduced field on the observation curve, the real part of the reduced electric field *y*-component along the same straight line is plotted as function of ξ in Fig. 22.

As can be seen, when using the optimal parameter, the regions wherein the reduced field exhibits fast changes are represented in an enlarged scale, whereas those characterized by slow variations are shown in a reduced scale. As a consequence, a uniform sampling in ξ gives rise to a denser sample distribution where the local bandwidth w is greater and to a sparser one in the zones characterized by smaller values of w (see Fig. 23).

DOI 10.1007/978-981-4560-75-7_117-1

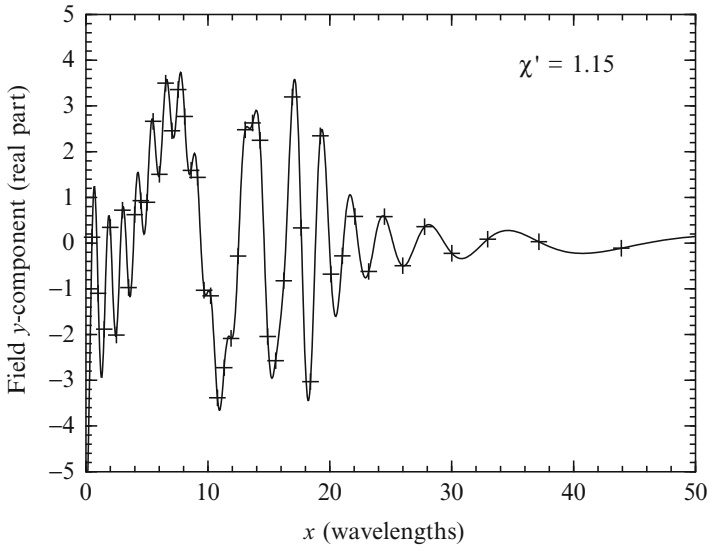


Fig. 23 Samples distribution on the given line

Application of the Nonredundant Sampling Representations to the NF-FF Transformations

Let an antenna be considered as enclosed in a convex domain bounded by a rotational surface Σ, and the surface scanned by a nondirective probe be obtained by rotating a meridian curve always external to the cone of vertex at the observation point P and tangent to Σ. Since the voltage V measured by such a kind of probe has practically the same effective spatial bandwidth of the AUT field (Bucci et al. 1998d), the above described nonredundant sampling representations of EM fields can be applied to it and, accordingly, it is convenient to introduce the reduced voltage $\tilde{V}(\xi) = V(\xi)e^{j\gamma(\xi)}$.

Therefore, the use of cardinal series (CS) expansions (Whittaker 1915) springs out naturally as an appropriate tool to conveniently represent it. However, the use of the CS representation, while completely satisfactory from the accuracy point of view, has the drawback that all samples (or, at least, all the relevant ones) must be considered in evaluating the voltage at each output point, otherwise a relatively large truncation error is introduced, due to the slow decay of the sampling functions. As a consequence, its use would lead to unacceptably large computational times. Such a slow decay leads to a further, subtler difficulty, if it is taken into account that, in any practical instance, the measured data are inaccurate. Provided that their values are well above the noise level, the measured results are usually affected by an (approximately) constant relative, not absolute, error, so that the absolute error corresponding to the highest voltage values can be relatively large. This error is spread out by the sampling functions without a severe attenuation, and this gives rise to a remarkable relative error in the zones where the voltage level is low (Bucci et al. 1991b). These difficulties have been overcome by resorting to an OSI expansion (Bucci et al. 1991a, b), in which only relatively few samples in the neighborhood of the output point are used in the reconstruction. These algorithms minimize the truncation error for a given number of retained samples and are more stable than the CS expansions with respect to random errors affecting the data.

In the light of the above discussion, the voltage at P $(\xi(\vartheta), \varphi)$ on the meridian curve fixed by φ can be efficiently evaluated via the OSI expansion (Bucci et al. 1998c; Bucci and Gennarelli 2012):

DOI 10.1007/978-981-4560-75-7_117-1

$$V(\xi(\vartheta),\varphi)=\mathrm{e}^{-\mathrm{j}\gamma(\xi)}\tilde{V}(\xi,\varphi)=\mathrm{e}^{-\mathrm{j}\gamma(\xi)}\sum_{m=m_0-q+1}^{m_0+q}\tilde{V}(\xi_m,\varphi)\Omega_M\left(\xi-\xi_m,\overline{\xi}\right)D_{M''}(\xi-\xi_m) \tag{144}$$

where $m_0=\mathrm{Int}(\xi/\Delta\xi)$ is the index of sample nearest (on the left) to P, $2q$ is the number of the retained intermediate samples $\tilde{V}(\xi_m,\varphi)$, namely, the reduced voltages at the intersection points between the sampling azimuthal circumferences and the meridian curve, $\overline{\xi}=q\Delta\xi$, $M=M''-M'$, and

$$\xi_m=m\Delta\xi=2\pi m/(2M''+1);\quad M''=\mathrm{Int}(\chi M')+1;\quad M'=\mathrm{Int}(\chi' W_\xi)+1 \tag{145}$$

χ being an oversampling factor required to control the truncation error (Bucci et al. 1991a, 1998c), and $\mathrm{Int}(x)$ denoting the integer part of x. Moreover,

$$D_{M''}(\xi)=\frac{\sin\left[(2M''+1)\xi/2\right]}{(2M''+1)\sin(\xi/2)};\quad \Omega_M\left(\xi,\overline{\xi}\right)=\frac{T_M\left[2\cos^2(\xi/2)/\cos^2\left(\overline{\xi}/2\right)-1\right]}{T_M\left[2/\cos^2\left(\overline{\xi}/2\right)-1\right]} \tag{146}$$

are the Dirichlet and Tschebyscheff sampling functions, respectively, $T_M(\cdot)$ being the Tschebyscheff polynomial of degree M.

It is worth noting that the weight function $\Omega_M(\cdot,\cdot)$ has been obtained in (Bucci et al. 1991a, b) by paralleling the properties of the so-called sampling window function introduced by Knab (1983), which (in the square norm) represents the practically optimal weight function for the case of an indefinite observation domain, where the kernel of the CS expansion is the $\sin(x)/x$ function instead of the Dirichlet one.

The intermediate samples $\tilde{V}(\xi_m,\varphi)$ can be determined by means of the OSI formula:

$$\tilde{V}(\xi_m,\varphi)=\sum_{n=n_0-p+1}^{n_0+p}\tilde{V}\left(\xi_m,\varphi_{n,m}\right)\Omega_{N_m}\left(\varphi-\varphi_{n,m},\overline{\varphi}\right)D_{N''_m}\left(\varphi-\varphi_{n,m}\right) \tag{147}$$

where $n_0=\mathrm{Int}(\varphi/\Delta\varphi_m)$, $\tilde{V}(\xi_m,\varphi_{n,m})$ are the reduced samples on the azimuthal circumference fixed by ξ_m, $2p$ is the retained samples number, and

$$\varphi_{n,m}=n\Delta\varphi_m=2\pi n/\left(2N''_m+1\right);\quad N''_m=\mathrm{Int}\left(\chi N'_m\right)+1;\quad N'_m=\mathrm{Int}\left[\chi^* W_\varphi(\xi_m)\right]+1 \tag{148}$$

$$\overline{\varphi}=p\Delta\varphi_m;\quad N_m=N''_m-N'_m;\quad \chi^*=1+(\chi'-1)\left[\sin\vartheta(\xi_m)\right]^{-2/3} \tag{149}$$

The variation of the azimuthal enlargement bandwidth factor χ^* with ξ is required to ensure a band limitation error constant with respect to ξ (Bucci and Franceschetti 1987).

By properly matching Eqs. 144 and 147, the two-dimensional OSI expansion is obtained:

$$\begin{aligned}V(\xi(\vartheta),\varphi)=\mathrm{e}^{-\mathrm{j}\gamma(\xi)}\sum_{m=m_0-q+1}^{m_0+q}&\left\{\Omega_M\left(\xi-\xi_m,\overline{\xi}\right)D_{M''}(\xi-\xi_m)\right.\\&\left.\cdot\sum_{n=n_0-p+1}^{n_0+p}\tilde{V}\left(\xi_m,\varphi_{n,m}\right)\Omega_{N_m}\left(\varphi-\varphi_{n,m},\overline{\varphi}\right)D_{N''_m}\left(\varphi-\varphi_{n,m}\right)\right\}\end{aligned} \tag{150}$$

DOI 10.1007/978-981-4560-75-7_117-1

which makes possible to reconstruct efficiently and accurately the NF data needed to carry out the traditional NF-FF transformation corresponding to the considered scanning geometry from a nonredundant, i.e., minimum, number of NF measurements.

Efficient and accurate NF-FF transformation techniques with plane-polar (Bucci et al. 1998a, 2000), bipolar (D'Agostino et al. 2003), cylindrical (Bucci et al. 1998b; D'Agostino et al. 2002), and spherical (Bucci et al. 2001a; D'Agostino et al. 2011, 2013a) scannings have been so developed. In all the cases, a remarkable reduction of the number of the required NF data and of the related measurement time has been achieved.

The effectiveness of the OSI algorithms depends on the choice of the following parameters: the enlargement bandwidth factor, the oversampling factor, and the retained samples number. The enlargement bandwidth factor χ' allows to control the aliasing error and values slightly greater than unity are enough to ensure small errors in the case of electrically large antennas (Bucci and Franceschetti 1987, 1989). A numerical procedure to properly select the χ' value is described in (Gennarelli et al. 1994). As regards the choice of the parameters controlling the truncation error, namely, the oversampling factor χ, and the numbers p, q of the retained nearest samples, it can be numerically made as in (Bucci et al. 1998c). Note that such an error can be decreased on increasing p, q, and/or χ. An increase of p, q implies a growth of the interpolation time, whereas an increase of χ reflects in a growth of the required NF data and, as a consequence, of the measurement time. In practice, the choice of the above parameters is done in such a way that the error related to the interpolation (aliasing plus truncation) must be smaller than the measurement one (background noise plus measurement uncertainties). It must stressed that, due to the filtering properties of the interpolation functions, it is possible also to cut away the spatial harmonics relevant to the noise sources outside the AUT spatial bandwidth.

NF-FF Transformation Techniques with Spiral Scannings

The scanning of a nondirective probe along a proper spiral wrapping an arbitrary rotational surface $\mathcal{M}$, obtained by rotating a meridian curve always external to the cone having the vertex at P and tangent to the rotational surface Σ enclosing the AUT, is considered in the following. The Cartesian coordinates of a generic point on such a spiral are

$$\begin{cases} x = r(\theta)\sin\theta\cos\phi \\ y = r(\theta)\sin\theta\sin\phi \\ z = r(\theta)\cos\theta \end{cases} \tag{151}$$

where $r(\theta)$ is specified by the meridian curve generating the surface $\mathcal{M}$, ϕ is the angular parameter describing the spiral, and the angle θ is a monotonic increasing function of ϕ. It can be easily verified that $r(\theta) = d/\cos\theta$ in the case of a planar spiral lying on a plane at distance d, $r(\theta) = d/\sin\theta$ for a helix wrapping a cylinder of radius d, whereas $r(\theta) = d$ when a spherical spiral is considered. It is worth noting that the angle θ, unlike the zenithal angle ϑ, can assume negative values. As a matter of fact, when the spiral describes a complete round on the surface $\mathcal{M}$, moving from the south pole to the north pole and then returning to the south one, θ varies in the range $[-\pi, \pi]$. Moreover, ϕ is always continuous, whereas, according to Eq. 151, the azimuthal angle φ displays a discontinuity jump of π when the spiral crosses the poles.

In order to obtain a nonredundant sampling representation of the probe voltage on the surface $\mathcal{M}$ from its samples collected along the spiral, it is necessary (D'Agostino et al. 2009e; Cicchetti et al. 2014):

DOI 10.1007/978-981-4560-75-7_117-1

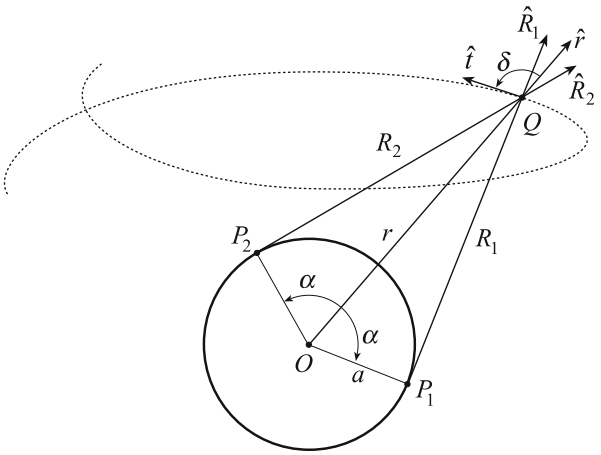


Fig. 24 Geometry of the problem in the plane $\hat{t}$, $\hat{r}$

(a) to choose the spiral in such a way that its pitch, specified by two consecutive intersections (at ϕ and $\phi + 2\pi$) with a meridian curve, be equal to the sample spacing needed for the interpolation along this curve;
(b) to develop a nonredundant sampling representation along the spiral.

According to condition (a), the optimal parameter ξ relevant to the meridian curve and the angular one ϕ describing the spiral are related by

$$\xi = k\phi \tag{152}$$

where k must be such that the interval $\Delta\xi = 2\pi k$, corresponding to the spiral pitch, is equal to the required sample spacing $\Delta\xi = 2\pi/(2\,M'' + 1)$ (see Eq. 145). Accordingly, $k = 1/(2\,M'' + 1)$. The scanning spiral can be so viewed as the projection on the surface $\mathcal{M}$, via the curves at $\xi = \text{const}$, of the corresponding spiral wrapping with the same pitch the modeling surface Σ.

The development of a nonredundant sampling representation of the voltage along the spiral is a more complex task, which has been heuristically solved in (D'Agostino et al. 2009e) by paralleling the rigorous procedure (D'Agostino et al. 2006) valid when the antenna is modeled by a spherical surface. Therefore, the main results of the unified theory of spiral scannings for antennas enclosed in a spherical surface (D'Agostino et al. 2006) are reviewed in the following.

In such a case, the optimal parameter to describe a meridian curve is the angle θ (which coincides in the range $[0, \pi]$ with the zenithal one ϑ) and the related bandwidth is βa. Moreover, the optimal phase factor ψ and parameter η to obtain a nonredundant representation along the spiral can be obtained from Eqs. 126 and 128 by observing that the extreme values of $\hat{R} \cdot \hat{t}$ occur (D'Agostino et al. 2006) at the two tangency points $P_{1,2}$ (Fig. 24) between the sphere modeling the AUT and the straight lines passing through the point Q on the spiral and belonging to the plane specified by the unit vectors $\hat{t}$ (tangent to the spiral at Q) and $\hat{r}$ (pointing from the origin to Q). Denoting by $\hat{R}_{1,2}$ the related unit vectors and by δ the angle between $\hat{r}$ and $\hat{t}$ (see Fig. 24), it results (D'Agostino et al. 2006)

$$\left(\hat{R}_1 + \hat{R}_2\right)/2 = \hat{r}\sin\alpha = \hat{r}\sqrt{1 - a^2/r^2} \tag{153}$$

$$\left(\hat{R}_1 - \hat{R}_2\right)/2 = \hat{n}\cos\alpha \tag{154}$$

where $\hat{n}$ is the unit vector parallel to the plane $\hat{r}$, $\hat{t}$ and orthogonal to $\hat{r}$. Accordingly

$$\left(\hat{R}_1 - \hat{R}_2\right) \cdot \hat{t}/2 = (\hat{n} \cdot \hat{t})\,(a/r) = (a/r)\sin\delta \tag{155}$$

By substituting Eq. 153 in Eq. 126 and taking into account that $\mathrm{d}r = \hat{r} \cdot \hat{t}\,\mathrm{d}s$, it results

$$\psi = \beta \int_0^r \sqrt{1 - a^2/r^2}\,\mathrm{d}r = \beta\sqrt{r^2 - a^2} - \beta a \cos^{-1}(a/r) \tag{156}$$

Namely, the phase function ψ relevant to the sampling representation along the spiral coincides with that γ related to the representation on a meridian curve.

By differentiating Eq. 151, it can be easily verified that

$$\mathrm{d}s = \sqrt{r^2 \sin^2\theta + k^2 r^2 + k^2 \dot{r}^2}\,\mathrm{d}\phi \tag{157}$$

where $\dot{r} = \mathrm{d}r/\mathrm{d}\theta$. Moreover,

$$\hat{r} \cdot \hat{t} = \frac{\mathrm{d}r}{\mathrm{d}s} = \frac{\mathrm{d}r}{\mathrm{d}\phi}\frac{\mathrm{d}\phi}{\mathrm{d}s} = \left[\frac{\mathrm{d}r}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}\phi}\right]\frac{\mathrm{d}\phi}{\mathrm{d}s} = k\dot{r}\frac{\mathrm{d}\phi}{\mathrm{d}s} = \frac{k\dot{r}}{\sqrt{r^2 \sin^2\theta + k^2 r^2 + k^2 \dot{r}^2}} \tag{158}$$

and, accordingly,

$$\sin\delta = \sqrt{1 - (\hat{r} \cdot \hat{t})^2} = \sqrt{\frac{r^2 \sin^2\theta + k^2 r^2}{r^2 \sin^2\theta + k^2 r^2 + k^2 \dot{r}^2}} \tag{159}$$

By substituting Eqs. 155 and 157 in Eq. 128 and taking into account Eq. 159, it results

$$\eta = \frac{\beta a}{W_\eta} \int_0^\phi \sqrt{k^2 + \sin^2 k\phi}\;\mathrm{d}\phi \tag{160}$$

that is, the optimal parameter η for describing the spiral is β/W_η times the arclength of the projecting point that lies on the spiral wrapping the sphere Σ. Since such a spiral is a closed curve, it is convenient to choose the bandwidth W_η in such a way that the angular-like parameter η covers a 2π range when the entire projecting spiral is described. Therefore,

$$W_\eta = \frac{\beta a}{\pi} \int_0^{(2M''+1)\pi} \sqrt{k^2 + \sin^2 k\phi}\;\mathrm{d}\phi \tag{161}$$

namely, the bandwidth W_η is β/π times the length of the spiral which wraps from pole to pole the sphere Σ.

The more general case wherein the AUT is no longer modeled as enclosed in a sphere is now considered. The parameterization η for describing the scanning spiral, the related phase factor ψ, and bandwidth W_η can be heuristically obtained (D'Agostino et al. 2009e) by generalizing the corresponding ones for the spherical modeling case (see Eqs. 160, 156, and 161). In particular, η is β/W_η times the

DOI 10.1007/978-981-4560-75-7_117-1

arclength of the projecting point that lies on the spiral wrapping the surface Σ (the projection is obtained by the curves at $\xi = \text{const}$ that, in such a case, take the role of the radial lines of the spherical modeling), ψ coincides with the phase function γ for a meridian curve, and the bandwidth W_η is β/π times the length of the spiral wrapping Σ from pole to pole. In other words, the spiral, ψ, and η coincide with those relevant to the spherical modeling when the surface Σ approaches a sphere.

In the light of the above results, the reduced voltage at any point Q of the spiral can be reconstructed by the following OSI expansion (D'Agostino et al. 2006, 2009e; Cicchetti et al. 2014):

$$\tilde{V}(\eta) = \sum_{n=n_0-p+1}^{n_0+p} \tilde{V}(\eta_n)\Omega_N(\eta - \eta_n, \overline{\eta})D_{N''}(\eta - \eta_n) \tag{162}$$

where $2p$ is the number of retained samples $\tilde{V}(\eta_n)$, $n_0 = \text{Int}(\eta/\Delta\eta)$ is the index of the sample nearest (on the left) to the point Q, $\overline{\eta} = p\Delta\eta$, $N = N'' - N'$, and

$$\eta_n = n\Delta\eta = 2\pi n/(2N'' + 1); \quad N'' = \text{Int}(\chi N') + 1; \quad N' = \text{Int}\left(\chi' W_\eta\right) + 1 \tag{163}$$

It must be stressed that small variations of η correspond to very large changes of ϕ in the neighborhood of the poles ($\vartheta = 0$ and $\vartheta = \pi$), so that, when interpolating the voltage in these zones, the enlargement bandwidth factor χ' must be properly increased to avoid a significant growth of the band limitation error.

The expansion (Eq. 162) can be used to determine the "intermediate samples," namely, the reduced voltage values at the intersection points between the meridian curve passing through the observation point P and the spiral. Once they have been determined, the voltage at the point P can be recovered by means of the following OSI expansion:

$$V(\xi(\vartheta), \varphi) = \mathrm{e}^{-\mathrm{j}\gamma(\xi)} \sum_{m=m_0-q+1}^{m_0+q} \tilde{V}(\xi_m)\Omega_M\left(\xi - \xi_m, \overline{\xi}\right)D_{M''}(\xi - \xi_m) \tag{164}$$

where $2q$ is the number of retained intermediate samples $\tilde{V}(\xi_m)$, $m_0 = \text{Int}\,[(\xi - \xi_0)/\Delta\xi]$ is the index of the sample nearest (on the left) to the point P, $M = M'' - M'$, $\overline{\xi} = q\Delta\xi$, and

$$\xi_m = \xi_m(\varphi) = k\varphi + m\Delta\xi = \xi_0 + m\Delta\xi \tag{165}$$

Summary

The aim of this chapter is to provide a complete description of the NF antenna measurement techniques to students, junior engineers, researchers, and physicists that wish to tackle such a topic. To this end, a wide introduction resumes the state of the art by supplying also a rich and updated bibliography. Then, the classical NF-FF transformations with plane-rectangular, cylindrical, and spherical scannings, in their probe-uncompensated and probe-compensated versions, are summarized by pointing out their advantages and drawbacks. Moreover, some analytical details on the wave expansions commonly employed to represent the antenna field in the region external to it are reported in order to enable the interested reader to achieve a more in-depth knowledge of the topic. The nonredundant sampling representations of the electromagnetic fields are then introduced, highlighting how their application to the NF-FF transformations with conventional scannings allows a drastic measurement time reduction with respect to the classical ones, without any loss in accuracy of the FF reconstruction. At last, the NF-FF transformations

DOI 10.1007/978-981-4560-75-7_117-1

using innovative spiral scannings, which allow a further measurement time saving by exploiting continuous and synchronized movements of the positioning systems, are described. They employ efficient sampling representations to accurately reconstruct the probe voltage on a quite arbitrary rotational surface from a nonredundant number of its samples collected on a proper spiral wrapping it.

Cross-References

- ▶ Anechoic Chamber Design
- ▶ Antenna Measurement Setups-Introduction
- ▶ EMI/EMC Chamber Design, Measurement and Instrument
- ▶ Mm-wave Sub-mm-wave Antenna Measurement
- ▶ Radiation Efficiency Measurements of Small Antennas

References

Alvarez Y, Las-Heras F, Pino M (2008) Probe-distortion correction for the sources reconstruction method. Antennas Propag Mag IEEE 50(6):117–124

Appel-Hansen J (1980) On cylindrical near-field scanning techniques. Antennas Propag IEEE Trans 28(2):231–234

Appel-Hansen J, Dyson J, Gillespie E, Hickman T (1982) Antenna measurements, chap. 8. In: Rudge A, Milne K, Olver A, Knight P (eds) The handbook of antenna design, Electromagnetic waves. Peter Peregrinus, London, pp 584–694

Balanis C (1989) Advanced engineering electromagnetics. Wiley, New York

Balanis C (1997) Antenna theory – analysis and design, 2nd edn. Wiley, New York

Belousov S (1962) Tables of normalized associated legendre polynomials. Pergamon Press, Oxford

Bennett J, Anderson A, McInnes P, Whitaker A (1976) Microwave holographic metrology of large reflector antennas. Antennas Propag IEEE Trans 24(3):295–303

Bolomey J, Cown B, Fine G, Jofre L, Mostafavi M, Picard D, Estrada J, Friederich P, Cain F (1988) Rapid near-field antenna testing via arrays of modulated scattering probes. Antennas Propag IEEE Trans 36(6):804–814

Brigham E (1974) The fast fourier transform. Prentice Hall, Englewood Cliffs

Bruning J, Lo Y (1971) Multiple scattering of EM waves by spheres part I – multipole expansion and ray-optical solutions. Antennas Propag IEEE Trans 19(3):378–390

Bucci O, Franceschetti G (1987) On the spatial bandwidth of scattered fields. Antennas Propag IEEE Trans 35(12):1445–1455

Bucci O, Franceschetti G (1989) On the degrees of freedom of scattered fields. Antennas Propag IEEE Trans 37(7):918–926

Bucci O, Gennarelli C (2012) Application of nonredundant sampling representations of electromagnetic fields to NF-FF transformation techniques. Int J Antennas Propag 2012, 319856, 14 pages

Bucci O, D'Elia G, Leone G, Pierri R (1990) Far-field pattern determination from the near-field amplitude on two surfaces. Antennas Propag IEEE Trans 38(11):1772–1779

Bucci O, Gennarelli C, Savarese C (1991a) Fast and accurate near-field – far-field transformation by sampling interpolation of plane-polar measurements. Antennas Propag IEEE Trans 39(1):48–55

Bucci O, Gennarelli C, Savarese C (1991b) Optimal interpolation of radiated fields over a sphere. Antennas Propag IEEE Trans 39(11):1633–1643

Bucci O, Gennarelli C, Riccio G, Savarese C (1998a) Near-field – far-field transformation from nonredundant plane-polar data: effective modellings of the source. Microwaves Antennas Propag IEE Proc 145(1):33–38

Bucci O, Gennarelli C, Riccio G, Savarese C (1998b) NF-FF transformation with cylindrical scanning: an effective technique for elongated antennas. Microwaves Antennas Propag IEE Proc 145(5):369–374

Bucci O, Gennarelli C, Savarese C (1998c) Representation of electromagnetic fields over arbitrary surfaces by a finite and nonredundant number of samples. Antennas Propag IEEE Trans 46(3):351–359

Bucci O, D'Elia G, Migliore M (1998d) Advanced field interpolation from plane-polar samples: experimental verification. Antennas Propag IEEE Trans 46(2):204–210

Bucci O, D'Elia G, Migliore M (1999) An effective near-field far-field transformation technique from truncated and inaccurate amplitude-only data. Antennas Propag IEEE Trans 47(9):1377–1385

Bucci O, D'Agostino F, Gennarelli C, Riccio G, Savarese C (2000) NF-FF transformation with plane-polar scanning: ellipsoidal modelling of the antenna. Automatika 41(3−4):159–164

Bucci O, D'Agostino F, Gennarelli C, Riccio G, Savarese C (2001a) Data reduction in the NF-FF transformation technique with spherical scanning. J Electromagn Waves Appl 15(6):755–775

Bucci O, Gennarelli C, Riccio G, Savarese C (2001b) Nonredundant NF-FF transformation with helicoidal scanning. J Electromagn Waves Appl 15(11):1507–1519

Bucci O, D'Agostino F, Gennarelli C, Riccio G, Savarese C (2002) Probe compensated far-field reconstruction by near-field planar spiral scanning. Microwaves Antennas Propag IEE Proc 149(2):119–123

Bucci O, D'Agostino F, Gennarelli C, Riccio G, Savarese C (2003) Near-field – far-field transformation with spherical spiral scanning. Antennas Wirel Propag Lett IEEE 2(1):263–266

Cicchetti R, D'Agostino F, Ferrara F, Gennarelli C, Guerriero R, Migliozzi M (2014) Near-field to far-field transformation techniques with spiral scannings: a comprehensive review. Int J Antennas Propag 2014, 143084, 13 pages

Clemmow P (1966) The plane wave spectrum representation of electromagnetic fields. Pergamon, London

Corona P, Ferrara G, Gennarelli C (1989) Measurement distance requirements for both symmetrical and antisymmetrical aperture antennas. Antennas Propag IEEE Trans 37(8):990–995

Costanzo S, Di Massa G (2002) An integrated probe for phaseless near-field measurements. Measurement 31(2):123–129

Costanzo S, Di Massa G (2004) Far-field reconstruction from phaseless near-field data on a cylindrical helix. J Electromagn Waves Appl 18(8):1057–1071

Costanzo S, Di Massa G (2007) Near-field to far-field transformation with planar spiral scanning. Prog Electromagn Res 73:49–59

Costanzo S, Di Massa G, Migliore D (2005) A novel hybrid approach for far-field characterization from near-field amplitude-only measurements on arbitrary scanning surfaces. Antennas Propag IEEE Trans 53(6):1866–1874

D'Agostino F, Ferrara F, Gennarelli C, Riccio G, Savarese C (2002) NF-FF transformation with cylindrical scanning from a minimum number of data. Microwave Opt Technol Lett 35(4):264–270

D'Agostino F, Gennarelli C, Riccio G, Savarese C (2003) Data reduction in the NF-FF transformation with bi-polar scanning. Microwave Opt Technol Lett 36(1):32–36

D'Agostino F, Gennarelli C, Riccio G, Savarese C (2006) Theoretical foundations of nearfield – far-field transformations with spiral scannings. Prog Electromagn Res 61:193–214

D'Agostino F, Ferrara F, Gennarelli C, Guerriero R, Migliozzi M (2008a) Near-field – far-field transformation technique with helicoidal scanning for elongated antennas. Prog Electromagn Res B 4:249–261

D'Agostino F, Ferrara F, Gennarelli C, Guerriero R, Migliozzi M (2008b) An effective NF-FF transformation technique with planar spiral scanning tailored for quasi-planar antennas. Antennas Propag IEEE Trans 56(9):2981–2987

D'Agostino F, Ferrara F, Gennarelli C, Guerriero R, Migliozzi M (2009a) Laboratory tests assessing the effectiveness of the NF-FF transformation with helicoidal scanning for electrically long antennas. Prog Electromagn Res 98:375–388

D'Agostino F, Ferrara F, Fordham J, Gennarelli C, Guerriero R, Migliozzi M, Riccio G, Rizzo C (2009b) An effective near-field – far-field transformation technique for elongated antennas using a fast helicoidal scan [measurements corner]. Antennas Propag Mag IEEE 51(4):134–141

D'Agostino F, Ferrara F, Gennarelli C, Guerriero R, Migliozzi M, Riccio G (2009c) A nonredundant near-field to far-field transformation with spherical spiral scanning for nonspherical antennas. Open Electr Electron Eng J 3:1–8

D'Agostino F, Ferrara F, Gennarelli C, Guerriero R, Migliozzi M, Riccio G (2009d) Near field – far-field transformation techniques with spiral scannings. CUES, Salerno

D'Agostino F, Ferrara F, Gennarelli C, Guerriero R, Migliozzi M (2009e) The unified theory of near-field-far-field transformations with spiral scannings for nonspherical antennas. Prog Electromagn Res B 14:449–477

D'Agostino F, Ferrara F, Gennarelli C, Guerriero R, Migliozzi M (2011) Effective antenna modellings for NF-FF transformations with spherical scanning using the minimum number of data. Int J Antennas Propag 2011, 936781, 11 pages

D'Agostino F, Ferrara F, Gennarelli C, Gennarelli G, Guerriero R, Migliozzi M (2012a) On the direct non-redundant near-field-to-far-field transformation in a cylindrical scanning geometry. Antennas Propag Mag IEEE 54(1):130–138

D'Agostino F, Ferrara F, Gennarelli C, Guerriero R, Migliozzi M (2012b) An innovative direct NF-FF transformation technique with helicoidal scanning. Int J Antennas Propag 2012, 912948, 9 pages

D'Agostino F, Ferrara F, Gennarelli C, Guerriero R, Migliozzi M (2012c) Far-field reconstruction from a minimum number of spherical spiral data using effective antenna modelings. Prog Electromagn Res B 37:43–58

D'Agostino F, Ferrara F, Gennarelli C, Guerriero R, Migliozzi M (2013a) Non-redundant spherical NF-FF transformations using ellipsoidal antenna modeling: experimental assessments [measurements corner]. Antennas Propag Mag IEEE 55(4):166–175

D'Agostino F, Ferrara F, Gennarelli C, Guerriero R, Migliozzi M (2013b) Far-field reconstruction from near-field data acquired via a fast spherical spiral scan: experimental evidences. Prog Electromagn Res 140:719–732

D'Agostino F, De Colibus I, Ferrara F, Gennarelli C, Guerriero R, Migliozzi M (2014a) Far-field pattern reconstruction from near-field data collected via a nonconventional plane-rectangular scanning: experimental testing. Int J Antennas Propag 2014, 763687, 9 pages

D'Agostino F, Ferrara F, Gennarelli C, Guerriero R, Migliozzi M (2014b) Efficient reconstruction of the pattern radiated by a long antenna from data acquired via a spherical-spiral-scanning near-field facility [measurements corner]. Antennas Propag Mag IEEE 56(2):146–153

Edmonds A (1974) Angular momentum in quantum mechanics. Princeton University Press, Princeton

Ferrara F, Gennarelli C, Guerriero R, Riccio G, Savarese C (2007) An efficient near-field to far-field transformation using the planar wide-mesh scanning. J Electromagn Waves Appl 21(3):341–357

Franceschetti G (1997) Electromagnetics. Plenum Press, New York

Francis M (2012) IEEE recommended practice for near-field antenna measurements. IEEE Std 17202012:1−102

Francis M, Wittmann R (2008) Near-field scanning measurements: theory and practice, chap. 19. In: Balanis C (ed) Modern antenna handbook. Wiley, Hoboken, pp 929–976

Gatti M, Rahmat-Samii Y (1988) FFT applications to plane-polar near-field antenna measurements. Antennas Propag IEEE Trans 36(6):781–791

Gennarelli C, Riccio G, Speranza V, Savarese C (1994) Fast and accurate interpolation of radiated fields over a cylinder. Prog Electromagn Res 8:349–375

Gennarelli C, Riccio G, D'Agostino F, Ferrara F (2004) Near-field – far-field transformation techniques, vol 1. CUES, Salerno

Gennarelli C, Capozzoli A, Foged L, Fordham J, van Rensburg D (2012) Recent advances in near-field to far-field transformation techniques. Int J Antennas Propag 2012:3

Gillespie E (1988) Special issue near-field scanning techniques. Antennas Propag IEEE Trans 36(6):725–901

Gregson S, McCormick J, Parini C (2007) Principles of planar near-field antenna measurements, Electromagnetics and radar series. Institution of Engineering and Technology (IET), London

Hald J, Hansen J, Jensen F, Larsen F (1988) Spherical near-field antenna measurements. IEE electromagnetic waves series, vol 26. Peter Peregrinus, London

Hansen R (1984) Measurement distance effects on low sidelobe patterns. Antennas Propag IEEE Trans 32(6):591–594

Hansen T (2011) Spherical near-field scanning with higher-order probes. Antennas Propag IEEE Trans 59(11):4049–4059

Harrington R (1961) Time-harmonic electromagnetic fields. McGraw-Hill, New York

Hollis J, Lyon T, Clayton L, Hansen C (1972) Microwave antenna measurements. Scientific, Atlanta

Isernia T, Leone G, Pierri R (1996) Radiation pattern evaluation from near-field intensities on planes. Antennas Propag IEEE Trans 44(5):701–710

James G (1980) Geometrical theory of diffraction for electromagnetic waves. Peter Peregrinus, London

Joy E, Paris D (1972) Spatial sampling and filtering in near-field measurements. Antennas Propag IEEE Trans 20(3):253–261

Joy E, Leach WM, Rodrigue G, Paris D (1978) Applications of probe-compensated nearfield measurements. Antennas Propag IEEE Trans 26(3):379–389

Kerns D (1970) Correction of near-field antenna measurements made with an arbitrary but known measuring antenna. Electron Lett 6(11):346–347

Kerns D (1981) Plane-wave scattering-matrix theory of antennas and antenna-antenna interactions, no. 162 in NBS monograph. U.S. Government Printing Office, Washington, DC

Knab J (1983) The sampling window. Inf Theory IEEE Trans 29(1):157–159

Larsen F (1980) Probe-corrected spherical near-field antenna measurements, PhD dissertation, Technical University of Denmark, Rep. LD36

Las-Heras F, Sarkar T (2002) Radial field retrieval in spherical scanning for current reconstruction and NF-FF transformation. Antennas Propag IEEE Trans 50(6):866–874

Las-Heras F, Pino M, Loredo S, Alvarez Y, Sarkar T (2006) Evaluating near-field radiation patterns of commercial antennas. Antennas Propag IEEE Trans 54(8):2198–2207

Leach WM, Paris D (1973) Probe compensated near-field measurements on a cylinder. Antennas Propag IEEE Trans 21(4):435–445

Newell A (1988) Error analysis techniques for planar near-field measurements. Antennas Propag IEEE Trans 36(6):754–768

Newell A, Crawford M (1974) Planar near-field measurements on high performance array antennas. Nat Bur Stand (US) NBSIR 74:74–380

Papoulis A (1987) Signal analysis. McGraw-Hill, New York

Paris D, Leach WM, Joy E (1978) Basic theory of probe-compensated near-field measurements. Antennas Propag IEEE Trans 26(3):373–379

Petre P, Sarkar T (1992) Planar near-field to far-field transformation using an equivalent magnetic current approach. Antennas Propag IEEE Trans 40(11):1348–1356

Pierri R, D'Elia G, Soldovieri F (1999) A two probes scanning phaseless near-field far-field transformation technique. Antennas Propag IEEE Trans 47(5):792–802

Qureshi M, Schmidt C, Eibert T (2013) Adaptive sampling in spherical and cylindrical near-field antenna measurements. Antennas Propag Mag IEEE 55(1):243–249

Rahmat-Samii Y, Galindo-Israel V, Mittra R (1980) A plane-polar approach for far-field construction from near-field measurements. Antennas Propag IEEE Trans 28(2):216–230

Sarkar T, Taaghol A (1999) Near-field to near/far-field transformation for arbitrary near-field geometry utilizing an equivalent electric current and MoM. Antennas Propag IEEE Trans 47(3):566–573

Silver S (ed) (1984) Microwave antenna theory and design, IEE electromagnetic waves. Peregrinus, London

Stratton J (1941) Electromagnetic theory. McGraw-Hill, New York

Taaghol A, Sarkar T (1996) Near-field to near/far-field transformation for arbitrary near-field geometry, utilizing an equivalent magnetic current. Electromagn Compat IEEE Trans 38(3):536–542

Wacker P (1975) Non-planar near field measurements: spherical scanning. NBSIR 75-809, Boulder

Whittaker E (1915) On the functions which are represented by the expansions of the interpolation theory. Proc R Soc Edinb 35:181–194

Williams L, Rahmat-Samii Y, Yaccarino R (1994) The bi-polar planar near-field measurement technique, part I: implementation and measurement comparisons. Antennas Propag IEEE Trans 42(2):184–195

Yaccarino R, Rahmat-Samii Y (1999) Phaseless bi-polar planar near-field measurements and diagnostics of array antennas. Antennas Propag IEEE Trans 47(3):574–583

Yaccarino R, Rahmat-Samii Y, Williams L (1994) The bi-polar planar near-field measurement technique, part II: near-field to far-field transformation and holographic imaging methods. Antennas Propag IEEE Trans 42(2):196–204

Yaccarino R, Williams L, Rahmat-Samii Y (1996) Linear spiral sampling for the bipolar planar near-field antenna measurement technique. Antennas Propag IEEE Trans 44(7):1049–1051

Yaghjian A (1975) Upper-bound errors in far-field antenna parameters determined from planar near-field measurements. Part 1: analysis, National Bureau Standards(US), Technical Note 667, 76

Yaghjian A (1977) Near-field antenna measurements on a cylindrical surface: a source scattering matrix formulation, NBS Tech. note 696, U.S. Government Printing Office, Washington, DC

Yaghjian A (1986) An overview of near-field antenna measurements. Antennas Propag IEEE Trans 34(1):30–45

Yaghjian A, Wittmann R (1985) The receiving antenna as a linear differential operator: application to spherical near-field scanning. Antennas Propag IEEE Trans 33(11):1175–1185

Yaghjian A, Woodworth M (1996) Sampling in plane-polar coordinates. Antennas Propag IEEE Trans 44(5):696–700

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_119-1

Reconfigurable Antennas for Wireless Communications

Y. Jay Guo* and Pei-Yuan Qin
University of Technology, Sydney, Australia

Abstract

With the capability to dynamically change their radiation characteristics, reconfigurable antennas will become indispensable parts for the next-generation wireless communications and sensing systems where the RF front-ends are required to be cognitive in nature. Compared to traditional antennas whose characteristics are fixed, reconfigurable antennas pose new challenges to antenna researchers and designers, such as tuning the operating frequency of an antenna whilst maintaining its radiation pattern. In the last two decades, substantial progress has been made in the development of reconfigurable antennas from both academia and industry. This chapter provides an overview of the state-of-the-art of reconfigurable antennas by elaborating the basic concepts of different reconfigurable antennas and practical techniques to realize them. It is focused on both reconfigurable antenna elements and arrays, and outlines some directions for future research.

Keywords

Reconfigurable antennas; Frequency reconfigurable antennas; Polarisation reconfigurable antennas; Pattern reconfigurable antennas; Leaky wave antennas; Phase shifters; Arrays

Introduction

Owing to the rapid proliferation of wireless communications systems, the limited electromagnetic spectrum has become more and more congested. To address this challenge, future wireless communications systems need to be cognitive and reconfigurable. They will have the intelligence and capacity to adopt the most appropriate communications strategy based on the feedback from channel sensing activities and signal quality assessment, including the operating frequency, the main beam direction and modulation schemes. In order to achieve these flexibilities, reconfigurable antennas (RAs) with adaptive radiation characteristics are required instead of traditional antennas whose structures are tailored to fit particular applications. RAs can also be used to avoid interference, save energy, enhance security, and mitigate signal quality deterioration caused by multipath fading.

Typical parameters of an antenna that can be reconfigured include frequency, polarization, radiation pattern, or a combination of the above. Depending on the reconfiguration mechanisms, RAs can also be classified into three main groups, namely, those using electrical devices, such as PIN diodes, varactor diodes and radio frequency micro electromechanical system (RF-MEMS), and those using mechanical changes and using material changes. Although this topic is a relatively new one, there are several books (Bernhard 2005; Bernhard and Volakis 2007) dedicated to it and a large number of papers reported

*Email: jay.guo@uts.edu.au

DOI 10.1007/978-981-4560-75-7_119-1

(Christodoulou et al. 2012). This chapter presents recent progress of RAs using electrical methods to achieve reconfiguration. Regarding those RAs using mechanical changes, readers can refer to (Tawk et al. 2011; Bernhard et al. 2001) for frequency RAs, (Barrera and Huff 2014) for polarization RAs and (Rodrigo et al. 2012; Sievenpiper et al. 2002) for pattern RAs.

The remainder of the chapter is organised as follows. Sections "Frequency Reconfigurable Antennas (RAs)," "Polarization Reconfigurable Antenna (RA)," "Pattern Reconfigurable Antenna (RA)," and "Compound Reconfigurable Antennas (RAs)" review the designs of frequency RAs, polarization RAs, pattern RAs and combined frequency and polarization RAs, respectively. Section "Reconfigurable Phase Shifter and Its Application in Beamforming Antennas" describes a novel phase shifter based on reconfigurable defected microstrip structure and its application in a beam-scanning antenna arrays. Section "Reconfigurable Leaky Wave Antennas" presents some designs of reconfigurable leaky wave antennas. Section "Summary" summarizes this chapter and gives some directions for future research.

Frequency Reconfigurable Antennas (RAs)

Frequency RAs can change the antenna operating frequency while keeping the polarization and radiation pattern stable across the entire frequency tuning range. Currently, it is a common practice for a single radio device to handle several services over a wide range of frequencies. For example, smart mobile phones or laptops may be required to support different standards, such as wireless local area network (WLAN), Worldwide Interoperability for Microwave Access (WiMAX), Bluetooth, Global Positioning System (GPS), 3G and 4G. To this end, the antennas need to cover multiple frequency bands. From the antenna perspective, multiband, wideband and frequency RAs are three potential candidates to be employed in systems requiring multiple operating bands. However, if only a portion of these operating bands is required at any given time, then frequency RAs would be the most appropriate choice. Compared to multiband and wideband antennas, one of the merits of frequency RAs is that the antenna can provide noise rejection in the bands that are not in use, so that the filter requirements of the front-end circuits can be greatly reduced. Furthermore, compared to multiband or wideband antennas, frequency RAs can be made much more compact.

Frequency RAs can be classified according to the basic antenna elements employed including microstrip antennas, dipole antennas, planar inverted F antennas (PIFA) and slot antennas. Alternatively, frequency RAs can also be classified into two types, namely, frequency continuous tuning and discrete tuning. The following introduction will be presented based on this classification.

Continuous Tuning

Varactor diodes are usually employed to achieve continuous frequency tuning. Initial work in this regard employed a method to load the radiating edges of patch antennas with varactor diodes to realize frequency agility (Bhartia and Bahl 1982; Waterhouse and Shuley 1994). By changing the bias voltages of the varactor diodes, the effective electrical size of the patch can be varied, which can produce a 1.1–1.2 frequency tuning ratio. Hum et al. proposed a differentially-fed frequency-agile microstrip patch antenna (Hum and Xiong 2010), as shown in Fig. 1. By loading three pairs of varactor diodes on the microstrip patches, the antenna can achieve a 2.0 frequency tuning ratio. Microstrip slot antennas also serve as good candidates for frequency RA designs. By using varactor diodes to change the length of the slot, the frequency of the antenna can be tuned. Generally speaking, slot frequency RAs can accomplish a wider tuning ratio, such as 3.52 (Li et al. 2010a), but they have the drawbacks of lower gain and efficiency.

Owing to the advantages of high gain, wide impedance bandwidth and ease of integration with microstrip-based monolithic-microwave integrated-circuits (MMIC), frequency reconfigurable printed

DOI 10.1007/978-981-4560-75-7_119-1

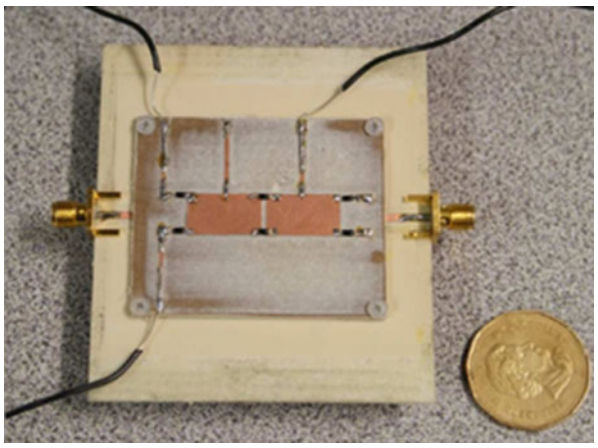


Fig. 1 Differentially-fed frequency-agile microstrip patch antenna

Fig. 2 Photo of the frequency reconfigurable UHF antenna (with a 60 cm long metallic ruler)

quasi-Yagi dipole antennas have attracted significant attention (Deal et al. 2000). Compared to a broadband printed dipole antenna with fixed frequency response, a frequency reconfigurable design offers frequency selective functionality which can reduce the adverse effects of co-site interference and jamming. The mechanism of frequency reconfiguration is to change the electrical length of the dipole arms by using varactor diodes or switches. In (Cai et al. 2012), a frequency reconfigurable high-gain quasi-Yagi dipole antenna operating over the 478–741 MHz UHF TV band was proposed for cognitive radio applications. The antenna consists of two main parts: a printed Yagi antenna with varactor-loaded dipole elements and a metallic corner reflector combined with a metallic hollow box for mechanical support, as shown in Fig. 2. The antenna metallization is etched on both sides of a Rogers RO3035 printed circuit board (PCB) with dimensions of 745 mm $\times$ 360 mm. The thickness of the substrate is 1.524 mm and the dielectric constant is 3.55. The top metallization consists of a driven dipole element, four parasitic directors, a broadband microstrip-to-coplanar stripline (CPS) transition and resistive biasing lines. The bottom side is a truncated microstrip ground, which serves as the reflector element for the antenna. Two varactor diodes are placed in each arm of the driven dipole and in each director. As the capacitance of the varactor diodes in the driven dipole is varied, the effective length of the dipole is changed, thereby altering the antenna resonant frequency. In the meantime, the varactor diodes in the directors are tuned to maximise the gain and to realize an acceptable impedance match for each operating frequency. A simulated and measured frequency tuning range from 478 to 741 MHz is achieved, which is shown

DOI 10.1007/978-981-4560-75-7_119-1

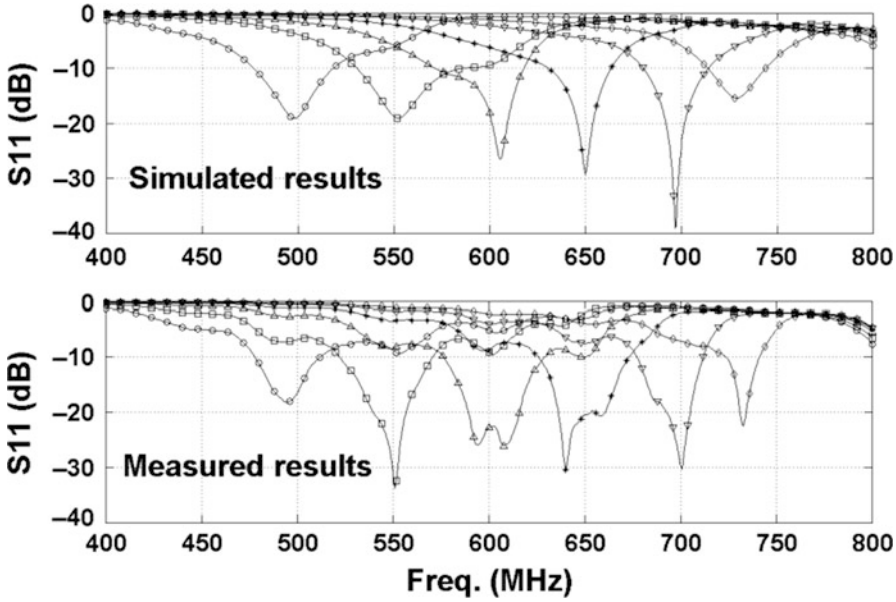


Fig. 3 Simulated and measured reflection coefficients in the six tuning sub-bands

in Fig. 3. In order to bias the diodes embedded in the driver and directors that are located on the ungrounded substrate, a low-cost and effective resistive biasing circuitry is employed, in which thin (1 mm width) but long metallic strip lines are broken down in short sections. The gaps between the metallic sections are bridged with high value surface mount resistors. The resistive nature of the biasing circuitry chokes high frequency currents effectively.

Discrete Tuning

PIN diode and MEMS switches are usually employed to realize discrete frequency tuning. A typical example is the compact frequency-agile microstrip patch antenna presented in (Genovesi et al. 2014). The antenna consists of a central patch connected to four different peripheral elements using PIN diodes. Each one of the four peripheral elements operates at a certain frequency independent from the others that can be designed according to the specifications. The patch antenna can achieve an overall of $2^4 = 16$ different states within a wide frequency range from 0.8 to 3.0 GHz.

In addition, a frequency reconfigurable folded dipole quasi-Yagi antenna was presented in (Qin et al. 2010b). The antenna is printed on two sides of a RO4003 substrate (0.813 mm thick and dielectric constant 3.55). The configuration of the antenna is shown in Fig. 4. The top side of the substrate consists of a microstrip feed, a broad-band microstrip-to-CPS balun, a folded dipole driver element fed by the CPS and a dipole parasitic director element. The bottom side is a truncated microstrip ground serving as the reflector element for the antenna. The combination of the parasitic director and reflector element directs the radiation of the antenna towards the end-fire direction. The folded dipole driver element is printed with six 0.5 mm gaps. Six PIN diodes are mounted across the gaps using electrically conductive silver epoxy. The length of the folded dipole element can be changed by switching between the different states of the diodes. When diodes 1 and 2 are on and all the other diodes are off, the length of the folded dipole is L_7. In this case, the length of the folded dipole is short and the proposed antenna resonates at a high operating frequency (denoted as State I). Changing the polarity of the dc voltage turns diodes 1 and 2 off and all the other diodes on. In this case, the length of the folded dipole is increased to $L_7 + 2\,W_{10} + 2\,L_{10}$ and the antenna resonates at a lower frequency (denoted as State II). Figure 5 shows the simulated and measured reflection coefficients versus frequency for State I and State II. From State II to State I, the

DOI 10.1007/978-981-4560-75-7_119-1

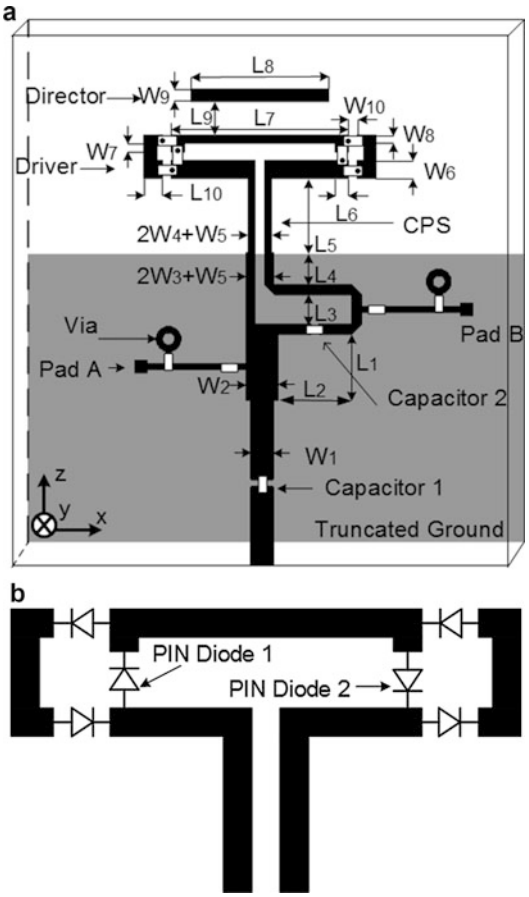


Fig. 4 (**a**) Configuration of the reconfigurable quasi-Yagi folded dipole Antenna. (**b**) Orientation of the PIN diodes in the folded dipole

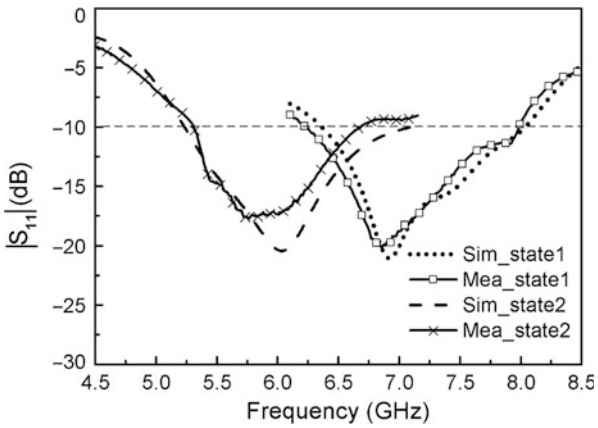


Fig. 5 Simulated and measured input reflection coefficients for the different states of Antenna

resonant frequency shifts from 5.95 to 7.2 GHz, corresponding to a frequency ratio of 1.21. The dc bias voltages are applied to the diodes through the metallization of the balun, CPS and folded dipole. It can be noted that, by using a folded dipole, a closed dc circuit is formed, thereby eliminating the need for extra biasing lines to be attached to the folded dipole. This, in turn, helps maintain the radiation pattern across the frequency tuning range.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_119-1

Polarization Reconfigurable Antenna (RA)

Polarization reconfiguration can take place in the form of changing angles of linear polarizations, or switching between left-hand circular polarization (LHCP) and right-hand circular polarization (RHCP), as well as between linear and circular polarizations. Such antennas can provide polarization diversity to mitigate signal fading in multipath propagation environments. Moreover, polarization diversity can be employed to increase the capacity of a multiple-input-multiple-output (MIMO) system (Qin et al. 2010a, b). The main challenge in achieving polarization agility is that it must be accomplished without significant changes in the antenna input impedance characteristics. As a result, it is challenging to design a polarization RA that can switch between linear and circular polarizations because it is difficult to simultaneously realize a good impedance match for these two polarizations. The reason is that circular polarization (CP) is typically generated by two degenerate orthogonal linear modes, and its input impedance is significantly different from that of the one resonant mode used to generate linear polarization (LP). Furthermore, it is even more difficult to design a dual-band or a multi-band polarization RA as the interdependence between the frequency response and the polarization characteristic is much stronger than that of a single-band antenna. In other words, it is more challenging to change the polarization states for the dual/multiple bands simultaneously while keeping the frequency response stable. In this section, some novel techniques are presented that can realize polarization switching between linear and circular senses and dual-band polarization RAs.

Single Band Polarization RA

Several interesting polarization RA designs have been proposed to switch between circular and linear polarizations. In (Sung et al. 2004), four pin-diodes were used on a corner-truncated square patch to produce LP and CP radiation with a small impedance bandwidth (2.5 %). A perturbed square-ring slot antenna using four pin diodes was designed that allows operation in both CP and LP modes (Dorsey and Zaghloul 2009). Unfortunately, the authors did not introduce the physical biasing and control circuits. In (Chen and Row 2008), a ring-slot-coupled microstrip circular patch antenna, as shown in Fig. 6, was proposed that can switch between linear and circular polarizations with an overlapped impedance bandwidth of 2.2 %. It was fabricated on two single FR4 substrates separated by a piece of foam, thus it may be difficult to integrate the antenna into a compact wireless device due to its large volume.

Another example is the design of a microstrip U-slot patch antenna that can switch between linear and circular polarizations (Qin et al. 2010c). The configuration of the proposed antenna is shown in Fig. 7. A U-slot is inserted into a rectangular patch which is printed on a 3.175-mm-thick RT/duroid 5880 substrate (dielectric constant 2.2). It should be noted that in the antenna prototype, another thinner slot is cut on the top of the U-slot, which separates the patch into two parts to ensure dc isolation. Three 30 pF capacitors are placed across this thinner slot to maintain RF continuity. The outer part of the patch is dc grounded by a shorting pin through an inductor used as an RF choke. Both the dc bias voltage and the RF signal are simultaneously fed through the coaxial probe by using a bias-tee. Beam lead PIN diodes are used as switching elements in the U-slot. The length of the U-slot arm can be changed by changing between the different states of the diodes. When the left diode is on and the right diode is off, the RF current can flow across the left arm of the U-slot. In this case, the left arm of the U-slot is shorter than the right arm. The asymmetrical U-slot can excite two orthogonal modes in the patch. Adjusting the location of the PIN diodes can make the two modes have the same magnitude and a phase difference of 90° at a given frequency, thus enabling the antenna to generate CP radiation with an acceptable axial ratio.

DOI 10.1007/978-981-4560-75-7_119-1

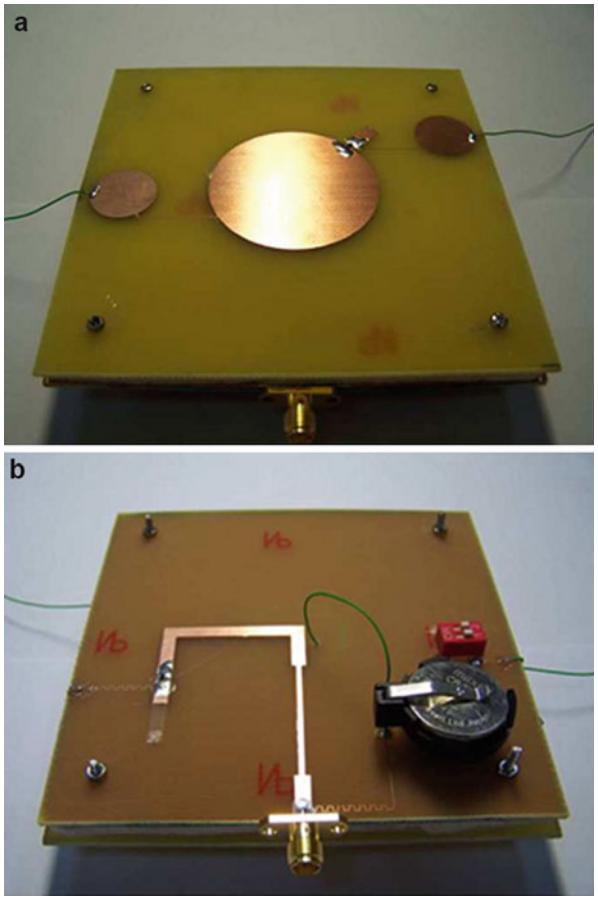


Fig. 6 Circular patch polarization reconfigurable antenna: (**a**) Top layer; (**b**) Bottom layer

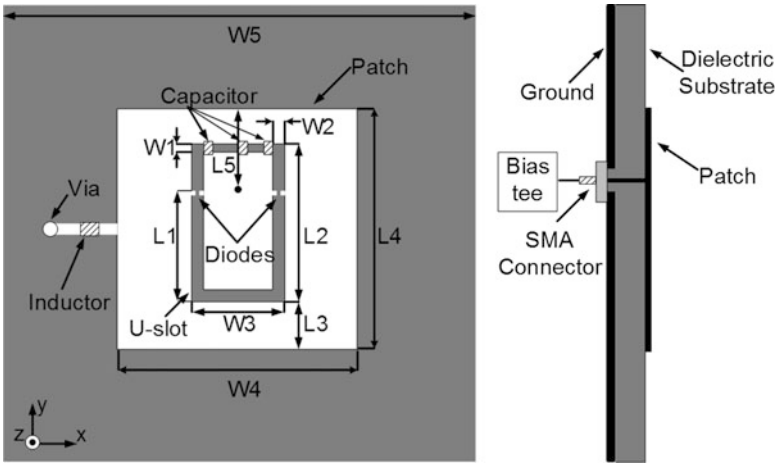


Fig. 7 Schematics of the reconfigurable U-slot antenna

The antenna radiates LHCP when the left arm of the U-slot is longer than the right arm. RHCP can be achieved if the right arm is longer than the left arm. The U-slot becomes symmetrical when both of the diodes are on or off, which enables the antenna to radiate linear polarization. In this case, the electric field polarization is parallel to the y-axis in Fig. 7.

DOI 10.1007/978-981-4560-75-7_119-1

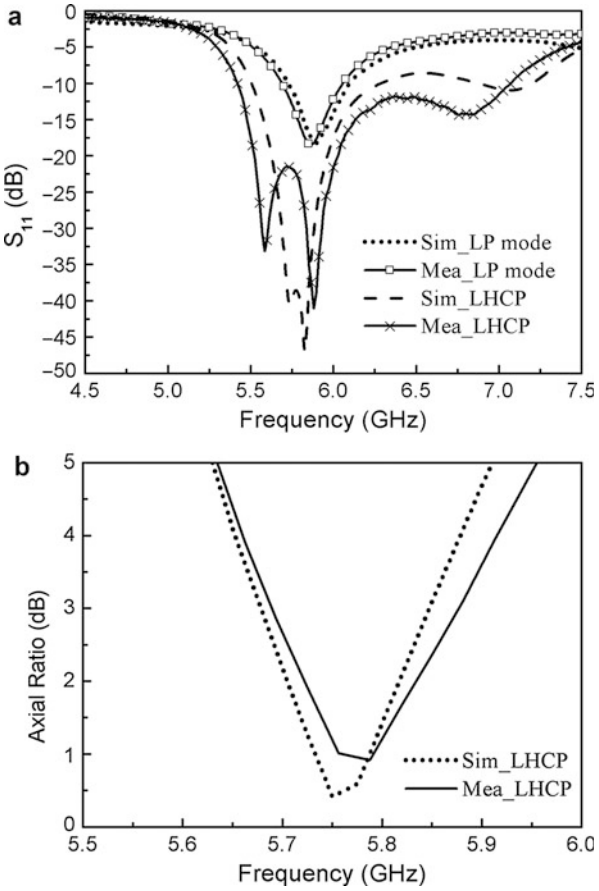


Fig. 8 Simulated and measured (**a**) input reflection coefficients for the LP and CP modes (**b**) axial ratio for the CP mode

Simulated and measured input reflection coefficients versus frequency for the CP and LP modes are shown in Fig. 8a. Simulated and measured axial ratios at boresight for the CP mode are given in Fig. 8b. From the experimental results, it is observed that the impedance bandwidths for the LP and CP modes are 6.1 % and 13.5 %, respectively, with almost the same center frequency of 5.9 GHz, which can cover the entire 5.725–5.85 GHz wireless local area network (WLAN) band. The measured 3 dB axial ratio bandwidth at boresight extends from 5.7 to 5.86 GHz (2.8 %).

Dual-band Polarization RA

Recently, many single-band polarization reconfigurable antennas have been developed for WLAN systems. Hsu et al. designed a polarization reconfigurable microstrip antenna using piezoelectric transducers (PETs) (Hsu and Chang 2007). Either RHCP or LHCP can be obtained by changing the bias voltage of two PETs at 5.8 GHz. An E-shaped patch antenna was studied in (Khidre et al. 2013), which is capable of radiating either RHCP or LHCP targeting the 2.4 GHz band. Li et al. described a microstrip square slot antenna with a coplanar waveguide-to-slot line transition (Li et al. 2010b). It can switch between horizontal and vertical polarizations at 2.4 GHz. For the above reported designs, the antennas can reconfigure their polarizations at only one frequency band of the WLAN standards. Since dual-band operation is commonplace for today's WLAN devices, a polarization reconfigurable antenna operating at both 2.4 and 5.8 GHz bands is highly desirable. To meet this requirement, Qin et al. proposed a single-aperture-fed polarization reconfigurable antenna enabling the switch between horizontal, vertical, and 45° linear polarizations in both 2.4 and 5.8 GHz bands (Qin et al. 2013a).

DOI 10.1007/978-981-4560-75-7_119-1

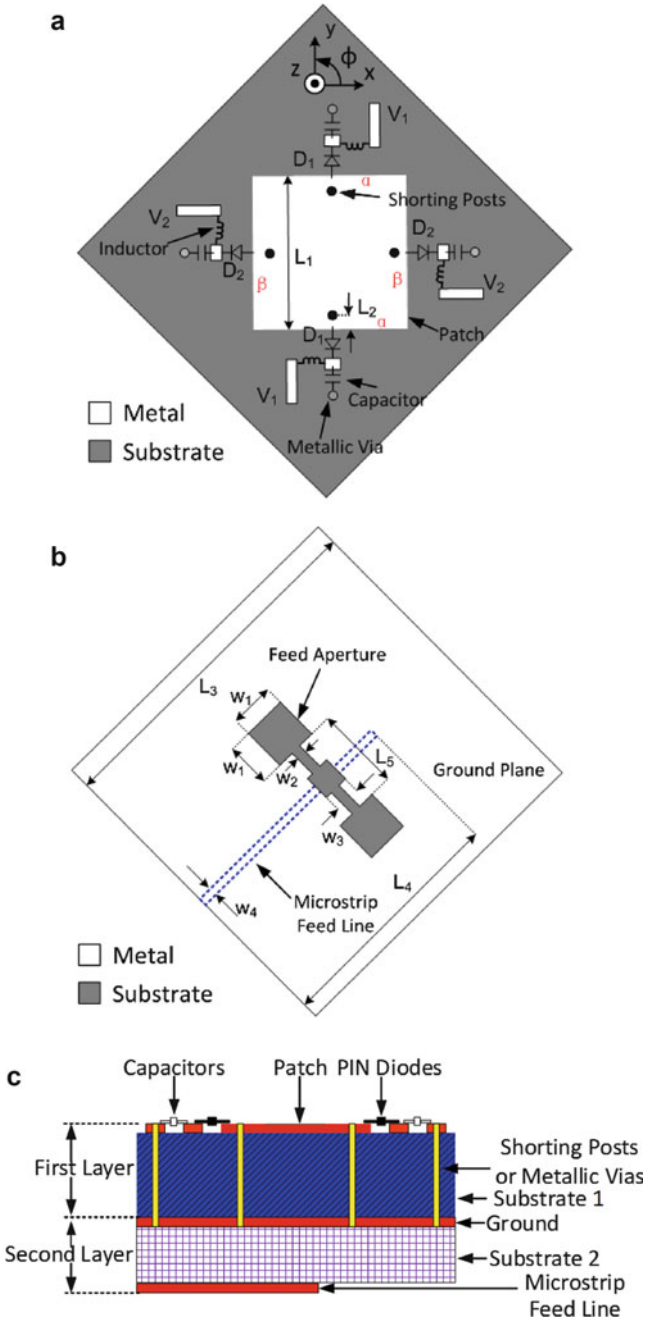


Fig. 9 Schematics of the proposed antenna: (**a**) First layer; (**b**) Second layer; (**c**) Side view

The configuration of the proposed antenna is shown in Fig. 9. The antenna consists of two substrate layers as shown in Fig. 9a, b, respectively. A side view of the antenna structure is given in Fig. 9c. The first layer is a 4.75-mm-thick RT/Duriod 5880 substrate (dielectric constant 2.2). A square patch inserted with four shorting posts is printed on the top side of this layer. There is no metallization on the bottom side of this layer. The TM_{10} and TM_{30} modes of the microstrip patch are selected to make the antenna operate in the 2.4 and 5.8 GHz bands, respectively. Generally speaking, the resonant frequency of TM_{30} mode is three times as high as that of TM_{10} mode. In order for the frequency ratio of the two modes to meet the WLAN dual-band (2.4 and 5.8 GHz) requirement, shorting posts are inserted at the nulls of the electric field of the TM_{30} mode. The shorting posts have a significant effect on the resonant frequency of TM_{10}

DOI 10.1007/978-981-4560-75-7_119-1

mode whilst having little effect on the TM_{30} mode. In this way, the frequency gap between the two modes is reduced.

The second layer is a 1.524-mm-thick RO4003C substrate (dielectric constant 3.55) with metallization on both sides. On the top side, a modified H-shaped feed aperture consisting of an ordinary H-shaped slot and a smaller centrally located square slot are etched in the ground plane (see Fig. 9b). The patch is coupled to the microstrip feed line through this aperture. The dimensions of the modified H-shaped aperture are optimized to achieve good impedance match for each polarization state of the dual bands. A 50-Ω-microstrip feed line (blue dashed line in Fig. 9b) is printed on the bottom side of this layer.

As the microstrip feed line is located along the diagonal line of the patch, both the *x* - and *y*- polarized modes with the same resonant frequency are excited for both TM_{10} and TM_{30} modes. By connecting the middle of the patch edges α (Fig. 9a) to the ground through metallic vias, the resonant frequency of the *y*-polarized mode is changed without affecting the *x*-polarized mode. This is because edges α are the radiating edges for the *y*-polarized mode. Similarly, by shorting the middle of the patch edges β to the ground, the resonant frequency of the *x*-polarized mode is changed without affecting the *y*-polarized mode. Therefore, a single mode (*x*- or *y*- polarized) can be selected by shifting the resonant frequency of another mode away from targeted bands. For the proposed antenna, PIN diodes are placed between the vias and the middle of each patch edge so that the connection between the patch edges and ground can be controlled. According to the orientation of the PIN diodes in Fig. 9a, the edges α of the patch are shorted to the ground when PIN diodes D1 are turned on and D2 are turned off. In this case, the frequency of *y*-polarized mode is shifted to above the *x*-polarized one. The *x*-polarized one is chosen as the antenna operating mode (State I). When PIN diodes D1 are turned off and D2 are turned on, the *y*-polarized mode is chosen as the antenna operating mode (State II). When all the diodes are turned off, the antenna radiates 45° linear polarization (State III) that is parallel to the feed line. When all the diodes are switched on, it is found that the input reflection coefficient of the antenna is too high to be acceptable for WLAN applications.

Figures 10, 11 and 12 show the simulated and measured input reflection coefficients versus frequency for States I, II and III, respectively. It can be observed that for States I and II two bands are found at 2.4 and 5.8 GHz. As discussed in last paragraph, the mechanism of the polarization reconfiguration is to separate the *x*- and *y*-polarized modes by moving the resonant frequency of one mode away from the targeted bands. As a result, in Figs. 10 and 11, it can be seen that the undesired modes are around 2.7 GHz. In addition, for the three polarization states, there is a parasitic resonance around 3.8 GHz, which is caused by the feed aperture. In practice, filters would be employed in a wireless communications system to reject the interference signals from the unwanted frequency bands.

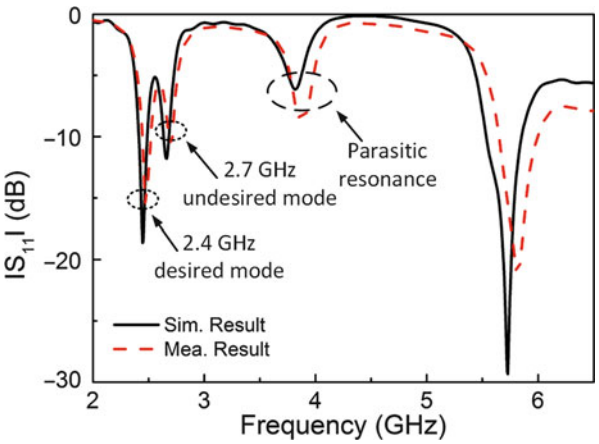


Fig. 10 Simulated and measured input reflection coefficients for State I (*x*-oriented polarization)

DOI 10.1007/978-981-4560-75-7_119-1

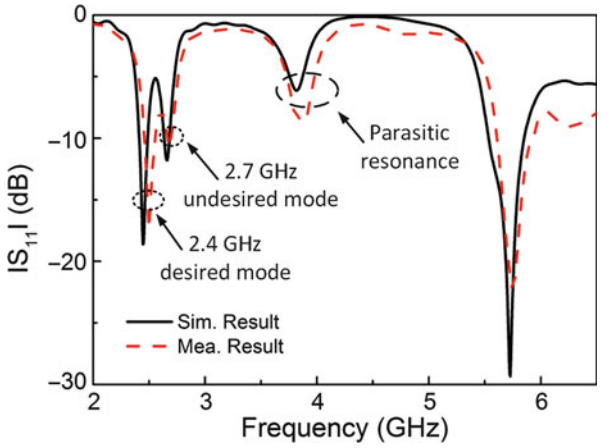


Fig. 11 Simulated and measured input reflection coefficients for State II (*y*-oriented polarization)

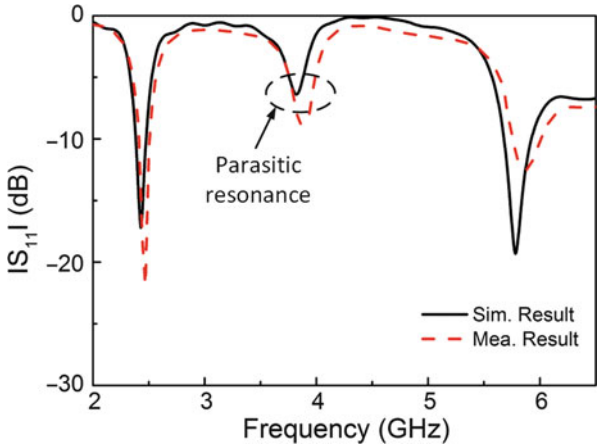


Fig. 12 Simulated and measured input reflection coefficients for State III (45^0-oriented linear polarization)

Pattern Reconfigurable Antenna (RA)

Pattern RAs have the capability to change the main-beam shape or provide main-beam scanning. They have the potential to reduce interference by altering the null positions, to save energy by directing the signal toward intended users and to provide large coverage by steering the main beam. In addition, pattern diversity provided by pattern reconfigurable antennas can be exploited by the multiple-input-multiple-output (MIMO) systems to increase the system capacity and/or link quality.

The frequency characteristic should be maintained nearly unchanged for the different radiation patterns of the antenna. Since the currents on the antenna structure directly determine the antenna radiation pattern, the reconfigurability of the radiation pattern is usually realized by manipulating the current distribution. However, as the current distribution also has a strong impact on the antenna frequency response, it is very challenging to deliver pattern reconfigurability without significant changes in the operating frequency. Several methods have been employed to overcome this challenge. One of them is to use specific antenna structures, such as reflector antennas or parasitically coupled antennas, so that the input feed port can be almost independent from the reconfigured part of the structure, allowing frequency characteristics to remain almost stable. Another measure is to compensate for the changes in the antenna input impedance by using some additional structures or matching circuits.

DOI 10.1007/978-981-4560-75-7_119-1

Change of Main-Beam Shape

An example is a wideband circular patch antenna that can switch between boresight and conical radiation patterns by using dual L-probe feeds (Yang and Luk 2006). TM_{11} mode for boresight radiation and TM_{01} for conical radiation are excited by the two feeds, respectively. Four shorting posts are added to the patch to reduce mainly the resonant frequency of the conical pattern mode, thereby enlarging the overlapped operating frequency of the two modes. In order to reconfigure the radiation pattern electronically, an integrated matching network consisting of switches is required. In (Wu and Ma 2008; Li et al. 2011), wideband bow-tie pattern reconfigurable antennas with tunable coplanar waveguide (CPW)-to-Slotline transition feed were presented. PIN diodes are used to reconfigure the feed mode between CPW mode, left slotline (LS) mode, and right slotline (RS) mode. The CPW feed mode can generate an almost omni-directional pattern while the LS mode and the RS mode can excite two end-fire patterns whose main beams are directed to exactly opposite directions.

Another example is a pattern reconfigurable U-slot antenna that is shown in Fig. 13 (Qin et al. 2012). A U-slot is inserted into a square patch of dimensions $L1 \times L1$ on a 3.175-mm-thick RT/Duroid 5880 substrate (dielectric constant 2.2). Each side of the patch is connected with two shorting posts via PIN diodes. The orientation of the diodes is also shown in Fig. 13. As all PIN diodes are mounted across the ground and the center patch, only a bias tee attached to the SMA connector is needed to control the PIN diodes. When the bias voltage is supplied from the coaxial probe, opposite bias conditions are applied to diodes in group A and B due to their reversed orientation. When the dc voltage is zero, all diodes are turned off. In this case, the antenna operates in the normal patch mode and radiates a boresight pattern (State I). When the dc voltage is negative, diodes in group B are on, and the other diodes are off. In this case, the antenna has four shorting posts connected and can be regarded as a monopolar patch antenna, which radiates a conical pattern with the maximum power level in the *z*-*y* plane (State II). Changing the

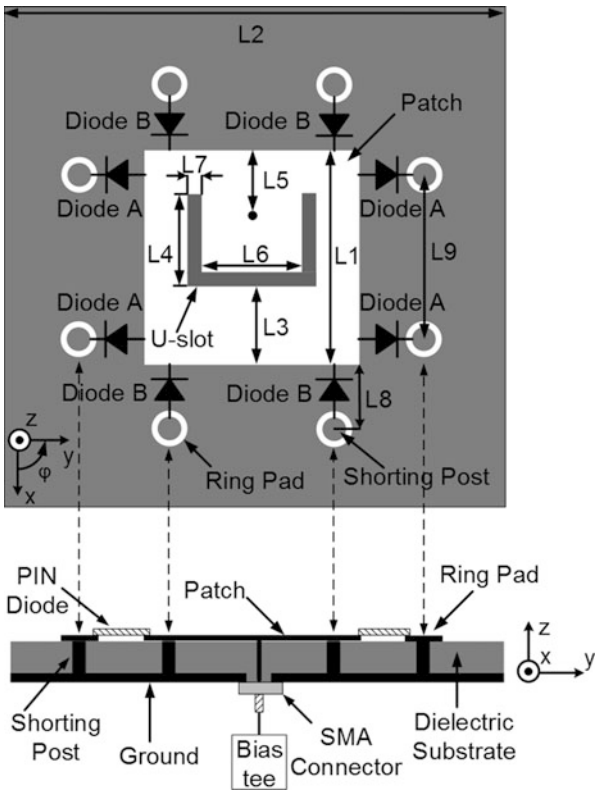


Fig. 13 Schematics of the pattern reconfigurable U-slot antenna

DOI 10.1007/978-981-4560-75-7_119-1

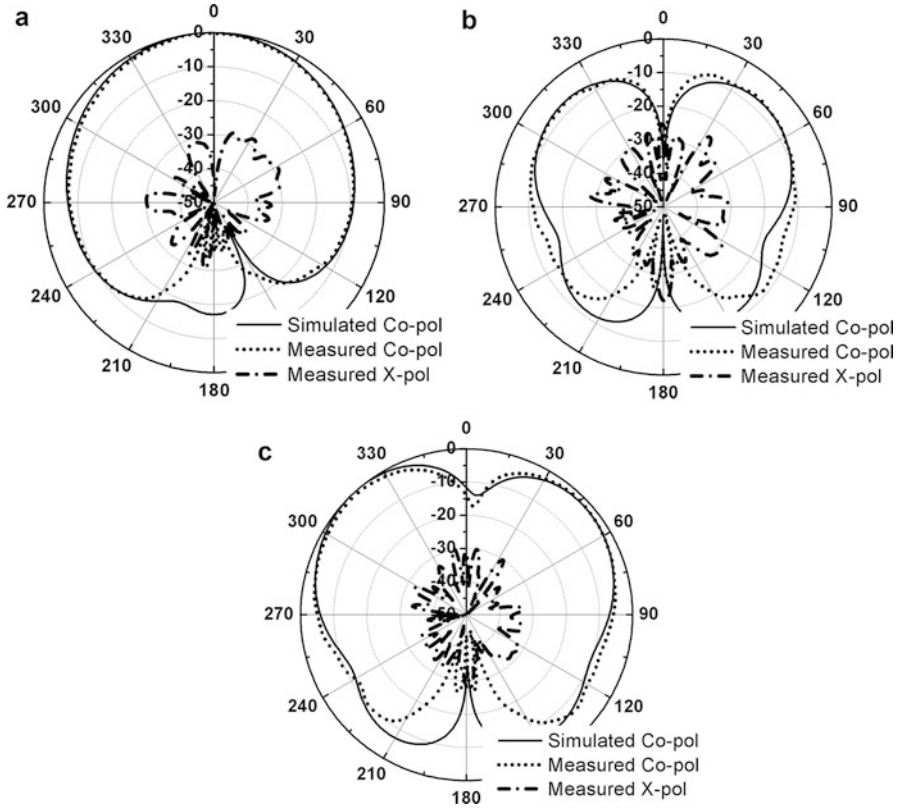


Fig. 14 Measured and simulated *z-x* plane normalized radiation patterns of the proposed antenna at 5.3 GHz (**a**) State I (**b**) State II (**c**) State III

polarity of the dc voltage from negative to positive, diodes in group A are on, and all the other diodes are off. In this case, a similar conical pattern can be observed with the maximum power level in the *z-x* plane (State III). Two shorting posts are used at each side of the patch instead of one in order to realize good impedance match for all operating states.

Simulated and measured normalized radiation patterns at 5.3GHz for the *z-x* and *z-y* planes are displayed in Figs. 14 and 15, respectively. For State I, boresight radiation patterns with a maximum cross-polarization level of −20 dB are shown in Figs. 14a and 15a. For State II, a symmetrical conical pattern with the maximum power level in the *z-y* plane directed at (elevation angle) 44° is plotted in Figs. 14b and 15b. For State III, an asymmetrical conical pattern with the maximum power level in the *z-x* plane directed at (elevation angle) 45° is drawn in Figs. 14c and 15c. It can be seen from Fig. 14c that the pattern is asymmetrical and there is 1 dB difference between the left and right maximum power level of the conical pattern. This is due to the position of the probe feed. Simulation results show that if the probe feed is located at the center of the patch, the difference between the left and right maximum power level in Fig. 14c will become smaller. However, in that case the overlapped impedance bandwidth of the two modes will be reduced. This can be viewed as a compromise for the antenna to provide good overlapped impedance bandwidth and radiation patterns.

Main Beam Scanning

Various antennas have been proposed to steer the main beam to predefined directions by using electronic switches, such as PIN diodes, to activate one or several elements out of a few radiators. Lai et al. proposed a four-element L-shaped antenna array that can achieve beam steering over 360^0 in the azimuth plane with

DOI 10.1007/978-981-4560-75-7_119-1

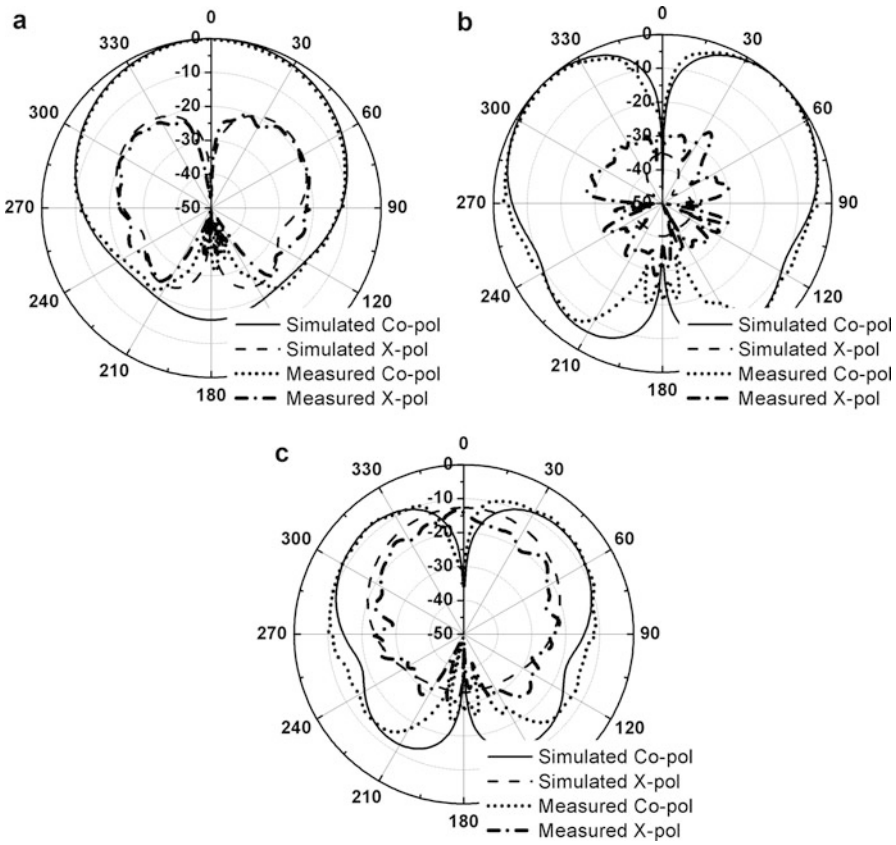


Fig. 15 Measured and simulated *z-y* plane normalized radiation patterns of the proposed antenna at 5.3 GHz (**a**) State I (**b**) State II (**c**) State III

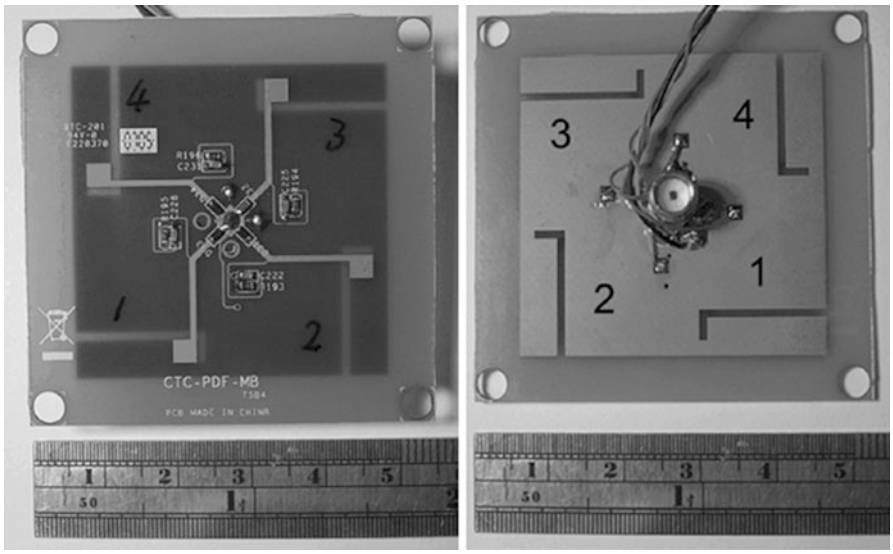


Fig. 16 L-shaped slot pattern reconfigurable antenna

a gain around −0.5–2.1 dBi (Lai et al. 2008). The configuration of the antenna is shown in Fig. 16. It is composed of four L-shaped quarter-wavelength slot antenna elements which are arranged towards 0^0, 90^0, 180^0, and 270^0, respectively. The maximum radiation direction is toward near the direction of the open

DOI 10.1007/978-981-4560-75-7_119-1

end of the slot. As a result, by using PIN diodes to enable one or more L-shaped slot antennas, several switchable patterns can be achieved. Also, spiral structure was employed to change the main beam direction by altering the length of the spiral. In (Jung et al. 2006; Huff et al. 2003; Nair and Ammann 2007), rectangular single-arm spiral antennas were designed to change the main beam over five directions, three directions and four directions, respectively. The gains of the antennas in (Jung et al. 2006; Huff et al.2003; Nair and Ammann 2007) are between 3–6 dBi, 4 dBi, and 1.1–4.6 dBi, respectively. In addition, work has gone into developing beam-steering antennas based on Yagi-Uda type array. Usually, such antennas have one driven element and several parasitic elements integrated with switches. By controlling the switches, the directive and reflective roles of the parasitic elements can be changed, thereby changing the main beam direction. The driven element can be either a microstrip dipole (Zhang et al. 2004), a microstrip patch (Yang et al. 2007; Donelli et al. 2007), or a wire antenna (Lim and Ling 2007). While there have been substantial advances in the design of beam-steering pattern reconfigurable antennas, it is found that most of the reported antennas suffer from low realized gain, which may significantly limit their applications.

In (Qin et al. 2013b), a high gain beam-switching pattern reconfigurable quasi-Yagi dipole antenna was proposed. It is capable of directing the E-plane main beam direction towards either 20^0, -20^0, or 0^0 with a realized gain between 7.5 and 10 dBi. The antenna was etched on a 1.27-mm-thick Rogers 6010 substrate (dielectric constant 10.9), as shown in Fig. 17. As shown in Fig. 17a, the top side of the substrate consists of a microstrip feed, an impedance transformer, a broad-band microstrip-to-coplanar stripline (CPS) balun, a dipole driver fed by the CPS, and two tilted rows of directive strips. The bottom side is a truncated ground, serving as the reflector. Usually, a design for maximizing the gain of a Yagi-Uda antenna requires a non-uniform director length and spacing. In this work, the directive strips are of the same length and the spacing between two adjacent strips is identical in order to reduce the design complexity.

The microstrip-to-CPS balun is used to introduce a phase difference for the currents on the two arms of the dipole. Details of the balun are shown in Fig. 17b. It can be seen that the balun is a symmetrical structure with respect to the line AA′. Therefore, only the balun above the line AA′ is described. The right-hand part of the balun is split into three sections with lengths of L_1, L_2, L_3, respectively, and the left-hand part is split into two sections with lengths of L_1 and L_4, respectively. The gap between each section is connected by PIN diodes. The width of the gap for the PIN diodes is W_{PIN}. In addition, the second section of the right-hand part of the balun with the length of L_2 is split into two smaller sections by a capacitor with a gap W_c. This capacitor is used for biasing purpose.

Conventionally, the directors of a quasi-Yagi dipole antenna are placed horizontally (parallel to the dipole driver) to direct the maximum beam towards the end-fire direction ($\varphi = 0^0$). However, it is found that when the beam is steered away from the end-fire direction, the original horizontally located directors can reduce the beam scanning range and the antenna gain. This is because the directors are not parallel to the titled beam direction any more. In order to increase the beam steering range and maintain the antenna gain, two tilted rows of metal strips are placed in front of the dipole driver with 6 strips on each row, which is shown in Fig 17a. To a certain extent, the tilted angle of the directive strips determines the maximum steering range. In this work, it is made to be close to the desired antenna beam tilted angle. Furthermore, a small gap is etched on each strip to split it into two short parts with PIN diodes inserted into the gaps. When the PIN diodes are switched on, the two short strips are connected to perform as a director. When the PIN diodes are switched off, the two short strips are disconnected and they do not serve as directors, thereby having little effect on the far-field radiation pattern. In this way, the PIN diodes can be used to choose the proper directors for a certain main beam direction. To be specific, when the PIN diodes *L* are switched on and the diodes *R* are switched off, only the left-hand row of strips perform as directors and they facilitate the beam tilt towards the left-hand side with respect to the end-fire direction. Similarly, when the PIN diodes *R* are switched on and diodes *L* are switched off, the beam tilt towards the right-hand

DOI 10.1007/978-981-4560-75-7_119-1

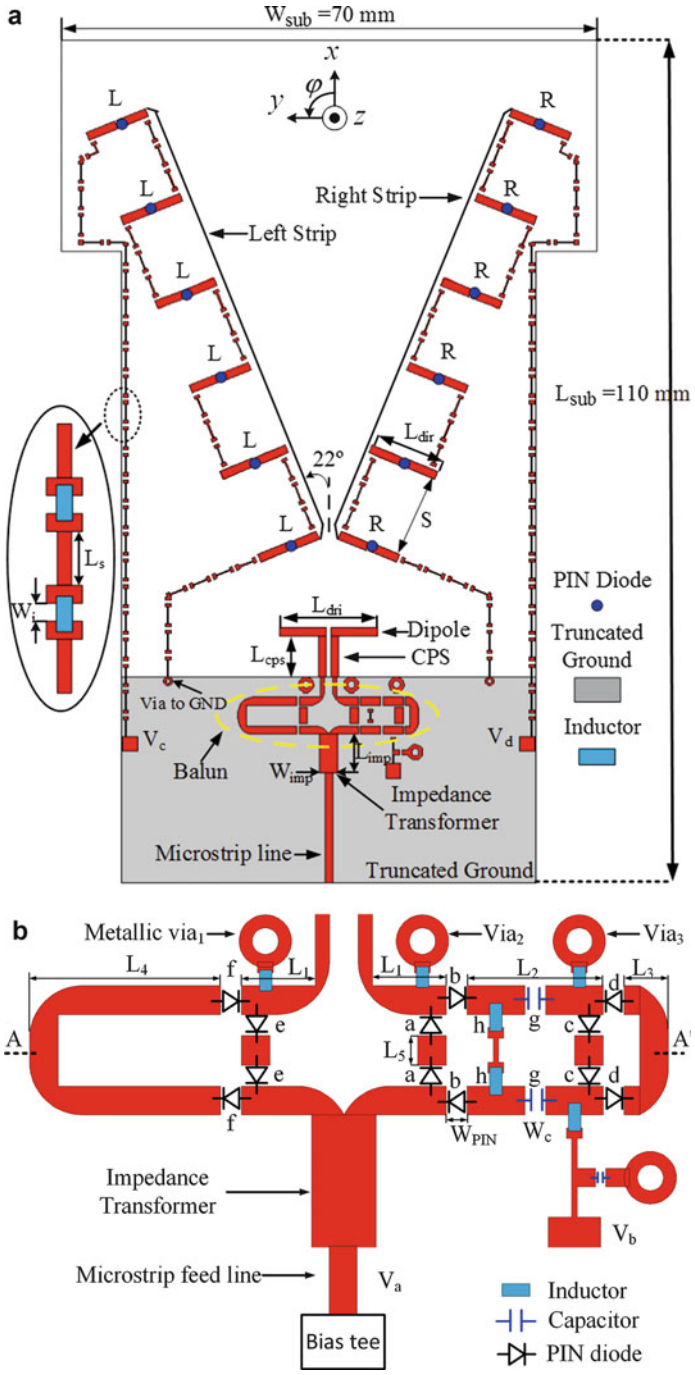


Fig. 17 Schematics of the proposed antenna: (**a**) Whole structure; (**b**) Balun of the antenna

side is enhanced. When both sides of the diodes are switched on, two rows of directors maintain the beam towards the end-fire direction.

By switching between the different states of the PIN diodes on the balun, the lengths of the current path on the right-hand and left-hand parts of the balun can be changed, thereby altering the phase difference of the currents on the dipole arms. According to the PIN diodes orientation shown in Fig. 17b, the antenna can operate in three states. For State I, diodes *e*, *b*, *c* and the diodes on the two rows of directive strips (diodes *L* and *R*) are switched on, and all the others are switched off. In this case, the maximum beam directs at $\varphi = 0^\circ$ (end-fire direction). For State II, diodes *e*, *b*, *d* and the diode group *L* are on, and all the

DOI 10.1007/978-981-4560-75-7_119-1

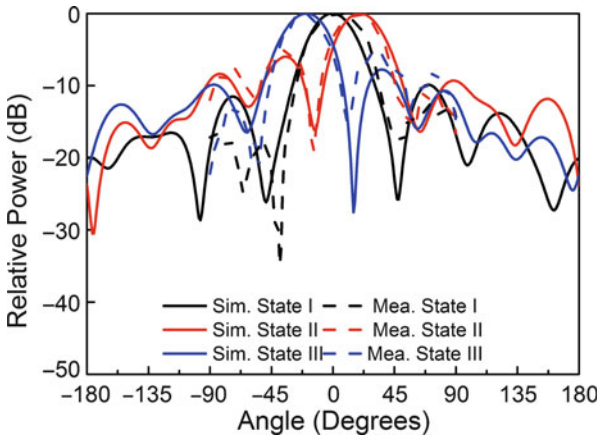


Fig. 18 Simulated and measured E-plane normalized radiation patterns at 5.2 GHz

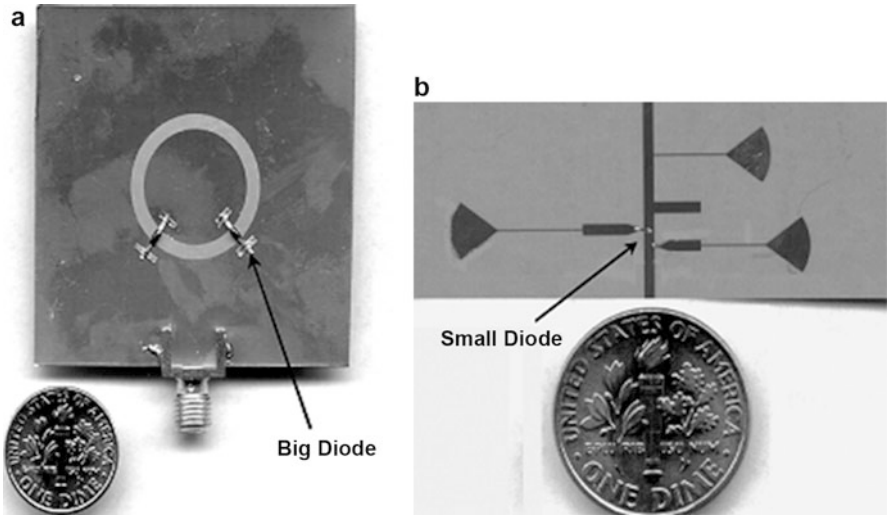


Fig. 19 (**a**) Front side of the annular slot antenna; (**b**) Back side, the impedance matching network

others are off. In this case, the maximum beam in E plane (*x*-*y* plane) radiate towards $\varphi = 20^\circ$ direction. For State III, diodes *f*, *a*, and diode group *R* are on, and all the others are off. For this state, the maximum beam direction can be directed towards $\varphi = -20^\circ$. Figure 18 shows the simulated and measured far-field radiation patterns at 5.2 GHz. As seen from the figure, the measured main beam for State I is pointed at end fire direction, while it is steered to 19^0 and -20^0 for State II and State III, respectively.

Compound Reconfigurable Antennas (RAs)

Compound RAs have the ability to independently change the operating frequency, polarization, and radiation pattern of an antenna, which is the ultimate goal of reconfigurable antenna design. Since a compound RA can deliver more flexibility and diversity than a single characteristic reconfigurable antenna, it can bring significant benefits to wireless communication systems. Recently, Nikolaou et al. have achieved combined frequency and radiation pattern reconfigurability (Nikolaou et al. 2006). Figure 19 shows the antenna configuration. By loading PIN diodes across the slot at specific locations

DOI 10.1007/978-981-4560-75-7_119-1

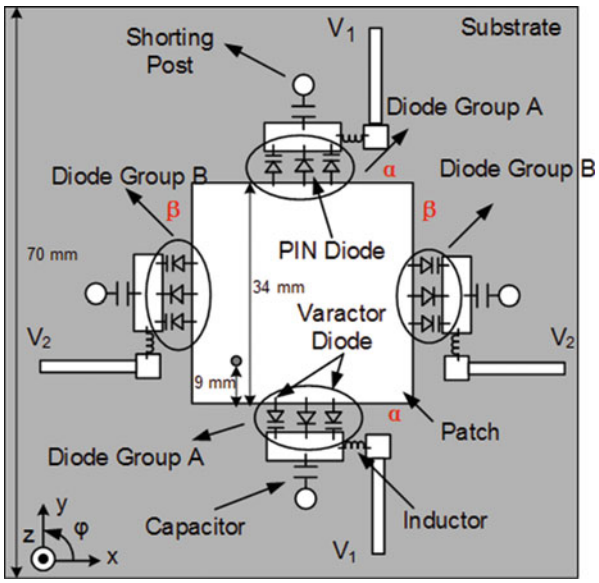


Fig. 20 Schematics of the proposed reconfigurable antenna

(Fig. 19a), the direction of the null of the radiation pattern can be changed. In addition, by reconfiguring the matching network shown in Fig. 19b, the operating frequency of the antenna can also be tuned.

In addition, Qin et al. proposed a single-fed combined frequency and polarization reconfigurable microstrip patch antenna with a maximum 1.67 frequency tuning ratio (Qin et al. 2011). The antenna can radiate one of the three linear polarizations (horizontal, vertical, and 45^0 linear polarizations) with a wide independent frequency tuning range for each polarization. The layout of the antenna is shown in Fig. 20. The lengths of the square substrate and the patch are 70 mm and 34 mm, respectively. The feed point is located 9 mm from the bottom of the patch edge along the diagonal line. The centre of each edge of the patch is connected to a shorting post via a PIN diode. Two varactor diodes (with a 0.1–1.0 pF junction capacitance tuning range for a corresponding voltage from 20 to 2 V) are located beside the PIN diode at each edge.

A square patch antenna fed along the diagonal line without any connection to the ground radiates *x*- and *y*- oriented modes, which have the same resonant frequency. By adding shorting posts in the middle of the patch edges α (Fig. 20), the resonant frequency of the *y*-oriented mode is increased without affecting the *x*-oriented mode. This is because edges α are the radiating edges for the *y*-oriented mode. By attaching shorting posts which can be represented by inductors, the resonant frequency of the *y*-oriented mode is increased. But the *x*-oriented mode will not be affected since its radiating edges are β. Similarly, by adding shorting posts in the middle of the patch edges β, the resonant frequency of the *x*-oriented mode is raised without affecting the *y*-oriented mode. Therefore, a single mode (*x*- or *y*- oriented) can be selected by shifting the resonant frequency of the undesired mode away from that of the desired one. For each mode, two varactor diodes are mounted along the corresponding radiating edges to change the resonant frequency.

Another recent design with full polarization diversity and frequency agility was presented in (Ho and Rebeiz 2014). MEMS switches are integrated into a feed network to provide four states of polarization control: vertical, horizontal, left-hand circular polarization, and right-hand circular polarization. Hyper-abrupt silicon junction tuning diodes are used to tune the antenna center frequency for each polarization. The tuning ranges are from 0.9 to 1.55 GHz and 1.1–1.5 GHz for linear and circular polarizations, respectively.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_119-1

Reconfigurable Phase Shifter and Its Application in Beamforming Antennas

Phase Shifter Design Based on Reconfigurable Defected Microstrip Structure (RDMS)

Phased array antennas can achieve analogue beamforming with high gains and are extensively utilized in satellite communications, radar systems, and other military applications (Parker and Zimmermann 2002). Usually, a large number of phase shifters are employed in a phased array antenna, and thus their cost, size, and integration method are of great concerns (Hansen 1998). The most popular phase shifters used in phased array antennas are diode ones (White 1974; Davis 1975) and ferrite phase shifters (Whicker 1973). Diode phase shifters have fast switching time, a low weight, a low cost but a high insertion loss, whereas ferrite phase shifters are relatively bulky and heavy, and require significant switching power compared with diode ones.

Recently, phase shifters based on defected ground structure (DGS) are developed (Patil et al. 2012; Han et al. 2005; Shafai et al. 2004). For most of the designs, multi-layer structures and micro electromechanical systems (MEMS) are used. Therefore, they suffer from a high cost and high fabrication complexity, and they are difficult to be integrated into microstrip systems, which substantially limits their applications. As a dual structure of DGS, the defected microstrip structure (DMS) also has the potential to provide a phase shift. In (Ye et al. 2012), C-shaped slots are etched on microstrip lines in the feed network for phase alignment. However, each slot only has a fixed phase delay of 3^0 at 12.5 GHz. In order to have controllable phase shift, Can et al. proposed compact phase shifters using reconfigurable defected microstrip structure (RDMS) for phased array antennas (Ding et al. 2014; Ding et al. 2015).

Figure 21 shows the structure of the RDMS unit based on a 50-Ω microstrip line for the phase shifter design. As depicted in the figure, a rectangular slot with a size of $W_{slot} \times L_{slot}$ is etched on the microstrip line to introduce a defect. PIN diodes are inserted into the edges of the slot area, which enables structural reconfiguration and results in a RDMS unit. Capacitors and metal stubs used for capacitor mounting are placed in the middle of the slot to achieve RF continuity as well as to provide DC isolation. The RDMS unit has two working states, the "All-on" and "All-off" states, when the diodes are turned "on" and "off," respectively. In the All-on state, both the diodes and capacitors allow the currents to go through them with minor losses. Numerical simulations find that the RDMS unit behaves like a uniform microstrip line in this state. In the All-off state, the currents across the diodes are blocked, resulting in a different current distribution as shown in Fig. 21b. It is noted that in the All-off state, the current path is longer than that in the All-on state, which produces a phase shift. The phase shift value is in proportional to the difference of the electrical length of the current path in the two states, which is affected by W_{slot} and L_{slot}. An RDMS

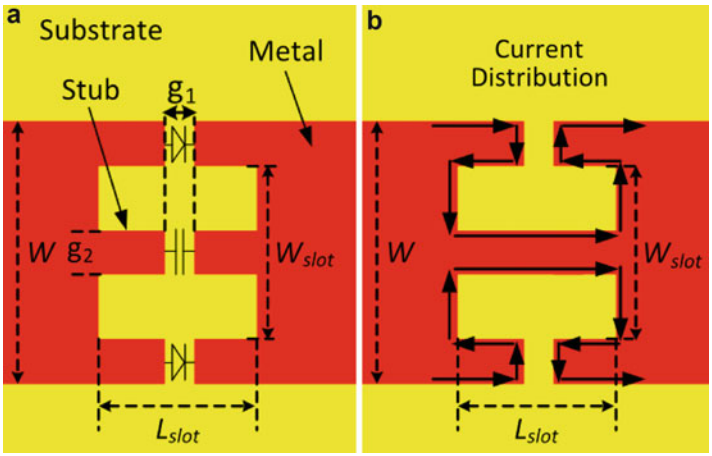


Fig. 21 (**a**) Structure of the RDMS unit. (**b**) All-off state current distribution

DOI 10.1007/978-981-4560-75-7_119-1

unit with $\{W_{slot}, L_{slot}\} = \{2\text{ mm}, 3.3\text{ mm}\}$ was fabricated as an example. Figure 22a, b, and c shows the picture of the fabricated RDMS unit, the measured insertion losses, and phase shifts, respectively. According to the measured results, the fabricated RDMS unit produces a phase shift of 20^0 with an insertion loss of 1.1 dB at 5.2 GHz.

Larger phase shift values can be achieved by cascading several RDMS units described above. A 2-RDMS phase shifter cascading two RDMS units is shown in Fig. 23a, which can realize a 40^0 phase shift. Furthermore, a phase shifter cascading four RDMS units is shown in Fig. 23b. By using two biasing voltages, phase shifts of 44° and 88° can be obtained at 5.2 GHz.

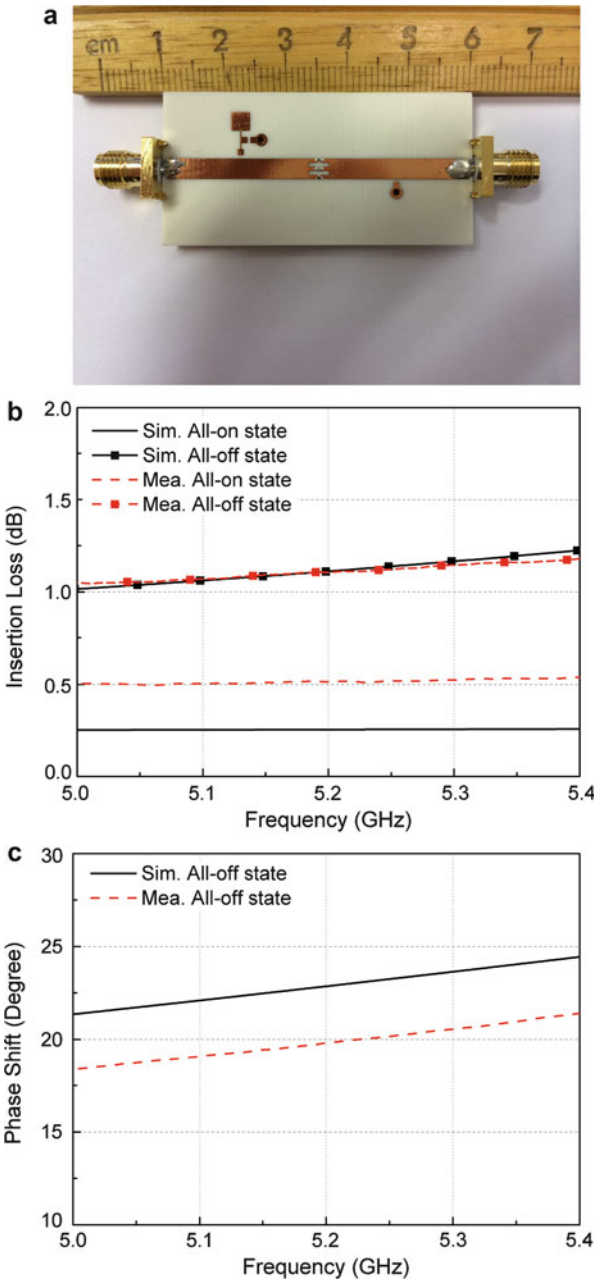


Fig. 22 (**a**) Photo of the fabricated 1-bit phase shifter. The simulated and measured results of the (**b**) insertion losses and (**c**) phase shift

DOI 10.1007/978-981-4560-75-7_119-1

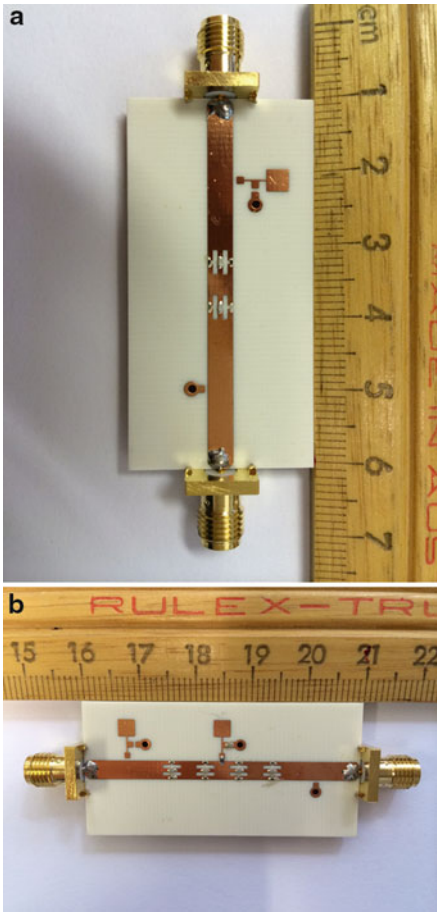


Fig. 23 (**a**) 2-RDMS phase shifter; (**b**) 4-RDMS phase shifter

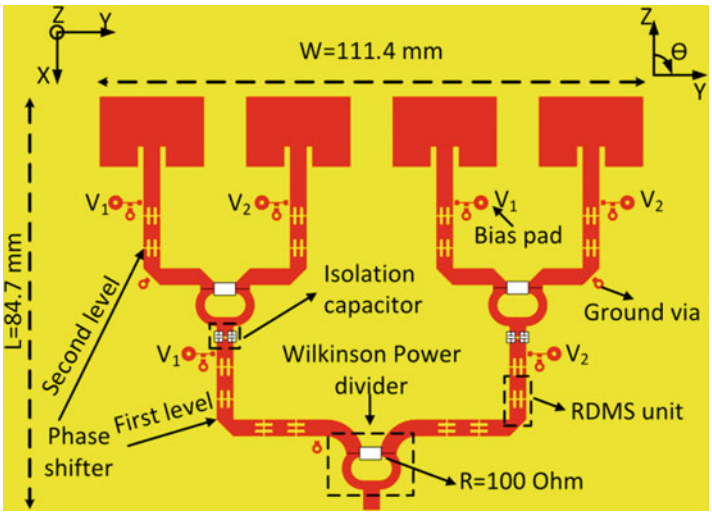


Fig. 24 The 4-element phased array prototype

Application in a 4-Element Phased Array

The above described phase shifters have been employed in a 4-element phased array feed network as shown in Fig. 24. The array is designed to be able to switch its main radiation beam to 0^0, -15^0 and 15^0 in

DOI 10.1007/978-981-4560-75-7_119-1

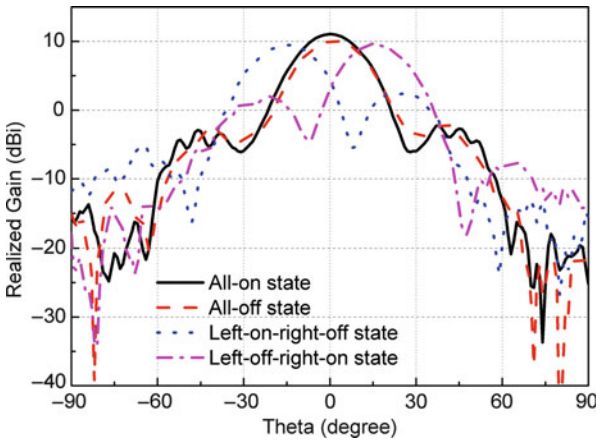


Fig. 25 The measured far-field pattern of the phased array

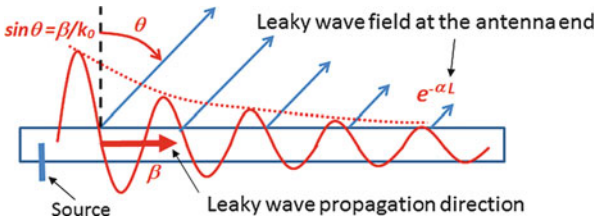


Fig. 26 Scheme of a leaky wave antenna

the H plane. In this design, 3 Wilkinson power dividers are employed to split the power while 8 phase shifters are integrated in the feed network to realize 50° progressive phase differences between the array elements at 5.2 GHz. Each phase shifter is composed of two RDMS units with $W_{slot} = 2\ mm$ and $L_{slot} = 3.6\ mm$. For the first-level power divider, there are 2 phase shifters at each of the 2 branches. For the second-level, there is 1 phase shifter at each of the 4 branches. Two bias DC voltages labelled by *V1* and *V2* are employed to control the phase shifters. There are 4 different working states of the phased array. When voltages *V1* and *V2* are "+,+," "–,–," "+,–," and "–,+," the phased array can work in the "All-on," "All-off," "left-on-right-off," and "Left-off-right-on" states. In the All-on and All-off states, the array elements are in phase so that the beam is not tilted. In the left-on-right-off and Left-off-right-on states, phase advances and delays of 50^0 between the array elements are obtained, respectively. Therefore, beam tilts of -15^0 and 15^0 are realized. The measured far-field pattern of the array is given in Fig. 25.

Reconfigurable Leaky Wave Antennas

Introduction

Leaky wave antennas can be regarded as transmission lines that gradually leak the energy out into free space. The direction in which the energy is radiated from a leaky wave antenna, θ, is determined by the phase constant β, of the leaky wave along the transmission line (Goldstone and Oliner 1959), as shown in Fig. 26. The beamwidth of a leaky wave antenna is determined by the length of the leaky aperture, which must be illuminated by proper control over the leakage rate α (Goldstone and Oliner 1959).

The concept of leaky-wave antennas has been around for decades (Horn et al. 1980). Compared with reflector and lens antennas, it has the advantage that no protruding feed is needed so the antenna can have

DOI 10.1007/978-981-4560-75-7_119-1

a very low profile. Compared with array antennas, it does not need a feed network which can become very lossy for large arrays. The phase constant of a leaky wave antenna is typically a function of the operating frequency, and this fact can be used to steer the beam direction by sweeping the operating frequency.

In the last decade, the fastest growth of the wireless communications industry and a multitude of wireless communications system standards have been witnessed. Wireless communication systems normally operate in a given frequency band defined by the spectrum regulators. Therefore, the inherent frequency scanning nature of leaky wave antennas has very limited use. To this end, there is a great need to develop pattern reconfigurable (beam-steering) leaky wave antennas which can be used to exploit the complexity of the propagation channel. Furthermore, leaky wave antennas have relatively narrow bandwidth, typically of only a few percent. However, given that wireless communications systems do not need a simultaneous wide bandwidth, a frequency reconfigurable (tunable) leaky wave antenna would serve as a good candidate for many applications.

The frequency-dependent nature of leaky wave antennas has limited their applications in modern communication systems, which generally require fixed frequency operation for effective channelization. In the past, significant efforts have been directed toward developing frequency-independent leaky wave antennas. Horn et al. (Horn et al. 1980) used PIN diodes as switches and electrically changed the radiation angle by controlling the guided wavelength. In their approach, however, only two discrete radiation angles were present because diodes have only two states, namely biased and unbiased. Maheri et al. (Maheri et al. 1988) reported a magnetically scannable leaky-wave antenna built on a ferrite slab structure, in which the radiation angle is scanned by tuning the DC magnetic field. In (Huang et al. 2000), Huang et al. applied PIN diodes as switches to control the period of the structure, and the reconfigurability was limited to two discrete radiation angles.

In (Sievenpiper 2005), Sievenpiper employed a high impedance surface which called a textured surface to develop a leaky wave antenna which can scan from backward to forward directions from -50^0 to 50^0. In the design, varactor diodes incorporated into the structure allow electronic control of the reflection phase and the surface wave properties. This tunable textured surface is then used as an electronically steerable leaky wave antenna by coupling energy into a leaky wave band using a flared notch antenna. The textured surface consists of a periodic lattice of small mushroom-shaped protrusions made of square metal plates, connected to a common ground plane by vertical metal pins. These pins are alternatively connected to and protruded through the ground-plane to allow the controlling voltage to be applied. The square plates are all connected to each other with varactor diodes. By exploiting multiple degrees of freedom which the surface geometry provided, the author achieved independent control of the magnitude and phase of the surface wave radiation, so the antenna can be programmed to have a large effective aperture over the entire scan range.

Left-handed (LH) materials characterised by simultaneously negative permittivity and permeability were introduced theoretically by Veselago (Veselago 1968) and investigated experimentally by Shelby et al. (Shelby et al. 2001). The propagation constant of a left-handed material is negative, representing a phase advance, whereas that of a right handed material is positive, representing a phase lag. Combining structures with LH and RH contributions results in a composite RH-LH (CRLH) structure, which can be used to realize both forward and backward scanning. This concept was first introduced using a transmission line structure in (Liu et al. 2002). In (Lim et al. 2005), a metamaterial-based electronically controlled transmission line structure incorporating varactor diodes was proposed as a leaky-wave antenna with tunable radiation angle and beamwidth functionalities. This structure is, in essence, a CRLH microstrip structure incorporating varactor diodes for fixed-frequency voltage-controlled operation. Angle scanning at a fixed frequency is achieved by modulating the capacitances of the structure by adjusting a uniform bias voltage applied to the varactors. Beamwidth tuning is obtained by making the structure non-uniform by the application of a non-uniform bias voltage distribution of the varactors. A 30-cell leaky wave

antenna structure, incorporating both series and shunt varactors for optimal impedance matching and maximal tuning range, was designed. This prototype exhibits continuous scanning capability from $+50^0$ to -49^0 by tuning the bias voltages from 0 to 21 V at 3.33 GHz. A maximum gain of 18 dBi at broadside is achieved, but the gain variation with scan angle is large. In addition, it provides half-power beamwidth variation of up to 200 % with comparison to the case of uniform biasing.

Employing tunable partially reflective surfaces and high impedance surfaces in Fabry-Pérot structures proved to be another way of achieving frequency reconfigurability and pattern reconfigurability in leaky wave antennas effectively. The authors of this chapter have conducted a systematic study in this regard in recent years which will be presented in the following sections.

Frequency Reconfigurable Fabry Pérot Antenna

Fabry-Pérot leaky wave antennas have the advantages of low profile, simple construction and high directivity. They are created by placing a partially reflective surface (PRS) around half a wavelength above a ground plane containing a low directivity source antenna (Trentini 1956; Feresidis and Vardaxoglou 2001; Feresidis et al. 2005; Wang et al. 2006; Weily et al. 2008). The PRS is usually a periodic array of dipoles, patches or slots on a dielectric substrate. If the ground plane is replaced with a high impedance surface (HIS), the profile of the antenna can be significantly reduced (Feresidis et al. 2005; Wang et al. 2006). A significant drawback of the Fabry-Pérot leaky wave antenna is the narrow operating bandwidth, due to the high Q factor of the Fabry Pérot cavity. However, the utility can be improved by making the operating frequency reconfigurable, through the use of a tunable HIS on the lower surface of the cavity where each cell of the HIS is tuned through the voltage applied to a pair of varactor diodes (Hum et al. 2005; Hum et al. 2007).

The geometry of the Fabry-Pérot leaky wave antenna, comprising a PRS, tunable HIS and stacked patch feed antenna, is shown in Fig. 27 (Weily et al. 2008). Its operating frequency was designed to be tuned from 5.2 to 5.775 GHz to enable use in WLAN applications. Referring to Fig. 28, the cavity height Lr can be expressed in terms of the operating wavelength λ_0, and the reflection phase of the PRS φ_1 and tunable HIS φ_2 as follows (Wang et al. 2006).

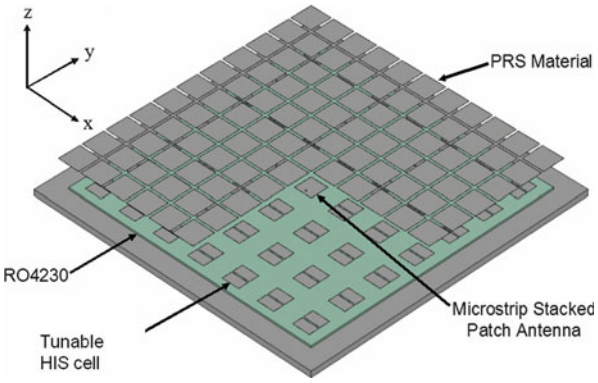


Fig. 27 Geometry of the frequency reconfigurable FP LWA

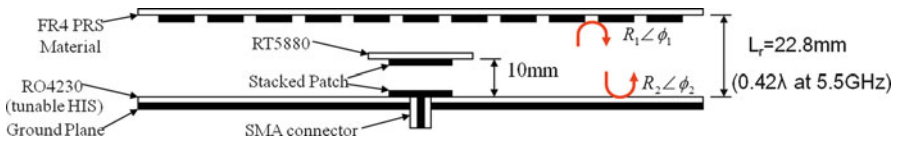


Fig. 28 Schematic of the frequency reconfigurable FP LWA showing reflection coefficients of the PRS and HIS

DOI 10.1007/978-981-4560-75-7_119-1

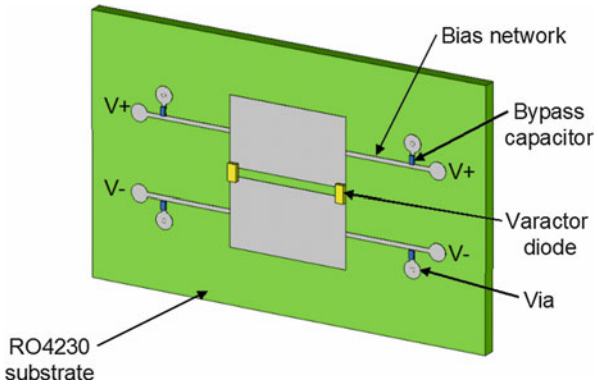


Fig. 29 Geometry of the tunable HIS unit cell

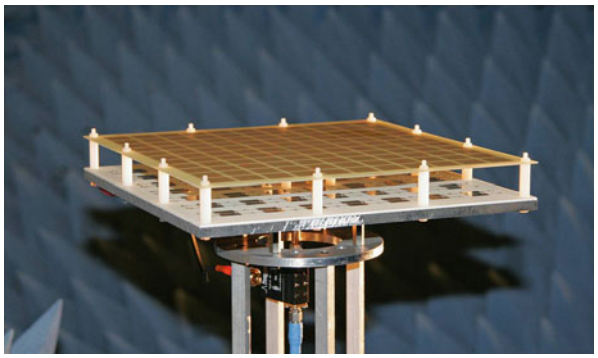


Fig. 30 Photograph of the frequency reconfigurable FP LWA prototype

$$L_r = \left(\frac{\varphi_1 + \varphi_2}{\pi}\right)\frac{\lambda_0}{4} + \frac{\lambda_0}{2} \tag{1}$$

Since φ_2 is a function of the varactor tuning voltage, it is possible to reconfigure the operating frequency of the antenna. Increasing the tuning voltage gives an increase in the antenna resonant frequency, while lowering the voltage reduces the resonant frequency. In this implementation the PRS is made from 18 mm $\times$ 18 mm square metallic patches placed on a 0.8 mm thick FR4 substrate ($\varepsilon_r = 4.4$, $tan\delta = 0.018$), with a periodicity of 20 mm in both x- and y-directions. A total of 48 tunable HIS cells are used on a 1.524 mm thick Rogers RO4230 substrate ($\varepsilon_r = 3.0$, $tan\delta = 0.0023$). Patch dimensions are 14 mm by 17 mm, with a 1 mm air gap between its two halves for placement of the varactor diode pair. Both substrates have lateral dimensions of 240 mm by 240 mm. A stacked probe-fed patch was chosen for the source because simulations showed it gave better matching as the HIS surface was tuned, compared to a single patch.

Figure 29 depicts the geometry of the tunable HIS cell used in the antenna, and includes the bias network which consists of 0.5 mm wide high impedance lines, 2.2 pF bypass capacitors and vias. The tunable HIS cell is based on a design from (Hum et al. 2005), used in a reconfigurable reflectarray (Hum et al. 2007). However, it has been modified to include a distributed element bias network to allow simultaneous biasing of a complete row of tunable HIS cells. A photograph of the fabricated prototype antenna is shown in Fig. 30. Various nylon spacers have been used to position the PRS above the tunable HIS, and the coupled patch above the driven patch. A low-loss bias tee, which may be observed behind the antenna, is used to provide the bias voltage to the tunable HIS cells, and thus reconfigure the antenna's operating frequency. The varactor diodes used in the prototype were also "matched" by the manufacturer

DOI 10.1007/978-981-4560-75-7_119-1

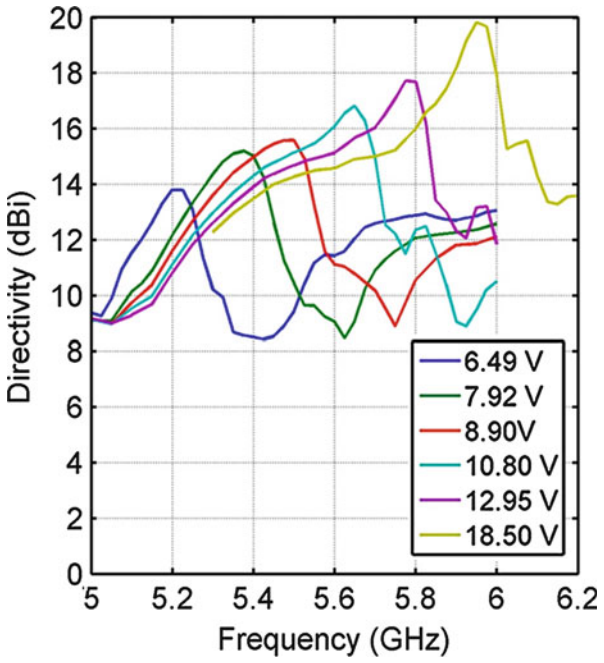


Fig. 31 Measured directivity of the frequency reconfigurable FP LWA for six different bias voltages

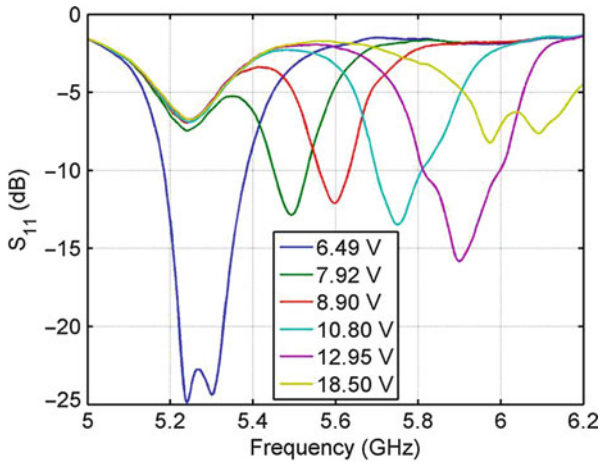


Fig. 32 Measured reflection coefficient of the frequency reconfigurable FP LWA for six different bias voltages

to ensure variations between the devices were as small as possible. The measured directivity versus frequency is plotted in Fig. 31 for six different bias voltages, which correspond to six different junction capacitances of the varactor diode. It is clear from Fig. 31 that the measured operating frequency tunes from 5.2 to 5.95 GHz. The measured reflection coefficient of the antenna is shown in Fig. 32 and tracks the directivity peaks of Fig. 31 for the corresponding six bias voltages.

Pattern Reconfigurable Leaky Wave Antennas

As mentioned earlier, one of the disadvantages of leaky wave antennas is that the radiation beam typically scans with the operating frequency. This is often problematic as, for many applications, it is required to have a fixed frequency band when the antenna beam is scanned. With one dimensional leaky wave antennas, this can be done by electronically reconfiguring the leaky-line boundary condition. Using active

DOI 10.1007/978-981-4560-75-7_119-1

devices such as varactor diodes or MEMS, the leaky-mode complex propagation constant can be altered, thus producing the desired control of the scanning beam. Using the Fabry-Pérot structure, a number of techniques have been developed to scan the beam of a leaky-wave antenna at a fixed frequency.

Half-space scanning 1D Fabry-Pérot leaky wave antenna

As the first step, a one-dimensional half-space scanning leaky wave antenna is designed, as shown in Fig.33a. This structure is inspired by the passive 1D Fabry-Pérot leaky wave antenna presented in (García-Vigueras et al. 2011), where a Fabry-Pérot cavity made of a top partially reflective surface (PRS) and a bottom high impedance surface (HIS) enables control of the leakage rate (α) and the phase constant (β) of the leaky mode through design of the physical length of the resonant patches. In this case, the structure provides electronic scanning of the main beam in the forward quadrant, at a fixed operating frequency (Guzmán-Quirós et al. 2012a), by replacing the passive HIS of (García-Vigueras et al. 2011) by a bottom

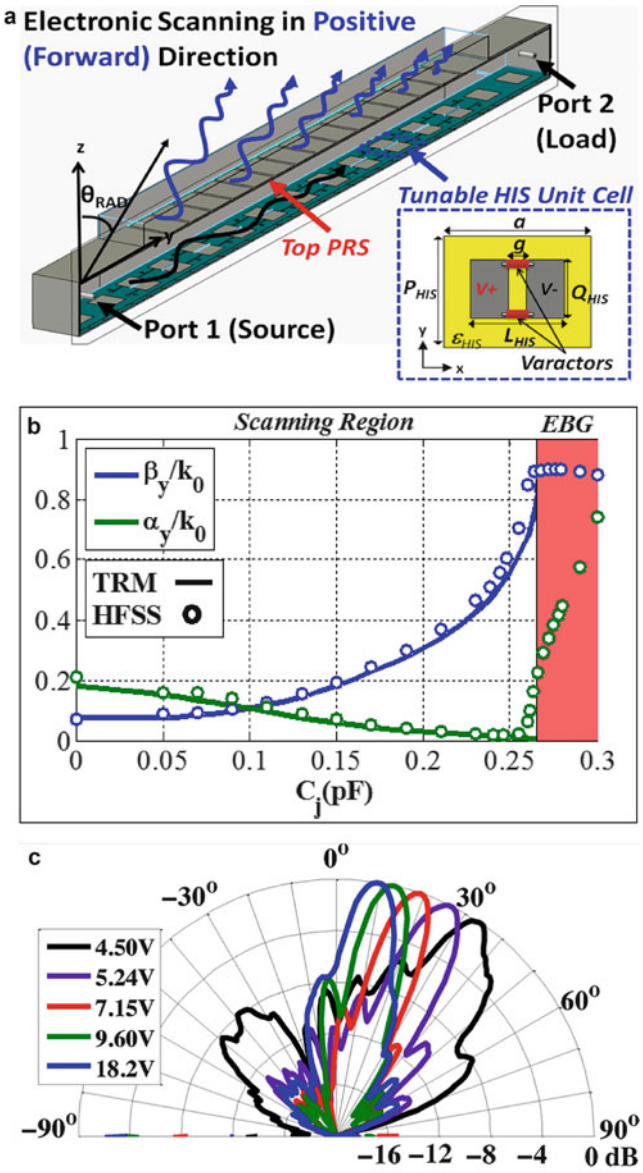


Fig. 33 (**a**) Scheme of a half-space scanning 1D FP LWA (**b**) Dispersion curves versus Cj and (**c**) Measured radiation patterns (H-plane) for different V_R at 5.5 GHz

DOI 10.1007/978-981-4560-75-7_119-1

tunable high impedance surface (HIS) loaded with varactor diodes (Guzmán-Quirós et al. 2012a). Now, the β of the leaky wave antenna can be controlled electronically as a function of the tunable junction capacitance (C_j) introduced by the varactor diodes which load the HIS patches (Fig. 33a) (Weily et al. 2008; Hum et al. 2005; Hum et al. 2007). This C_j is tuned by the DC bias voltage (V_R) applied to the diodes.

The control of the LM dispersion with C_j is demonstrated in Fig. 33b, showing the normalized phase (β/k_0) and leakage (α/k_0) rates as a function of C_j for a fixed operating frequency of 5.5 GHz, where k_0 is the free-space wave number in air. It is shown how the LM phase constant increases with C_j in the scanning region defined in the range $C_j = [0\text{ pF},0.27\text{ pF}]$. As β/k_0 is related with the main beam radiation angle θ_{RAD} by $sin(\theta_{RAD}) \approx \beta/k_0$ (Goldstone and Oliner 1959), a half-space electronic scanning of the fan beam direction with V_R is obtained in the positive quadrant from $\theta_{RAD} = +9°$ to $\theta_{RAD} = +34°$, as experimentally confirmed by the measured radiation patterns shown in Fig. 33c.

Full-space scanning 1D Fabry-Pérot leaky wave antennas

The structure of a symmetrically-fed 1D Fabry-Pérot leaky wave antenna is depicted in Fig. 34a. This antenna is conceived to extend the half-space scanning range obtained from the previous 1D Fabry-Pérot leaky wave antenna design, to a full-space scanning range. This is achieved by taking advantage of the electromagnetic band-gap (EBG) property that is presented in the Fabry-Pérot cavity for low values of the tuning voltage V_R (Guzmán-Quirós et al. 2012b). As shown in Fig. 34b, an EBG zone extends beyond the scanning region for $C_j = [0.27\text{ pF}, 0.3\text{ pF}]$. LM propagation along the FP waveguide is prevented in this zone. A central coaxial probe feed divides the original antenna in to two symmetric leaky lines: LWA_{LEFT} and $\text{LWA}_{\text{RIGHT}}$. Each one is separately biased by V_L or V_R. Therefore, two leaky waves which are oppositely launched by the central coaxial feed yield two radiation patterns whose main beam pointing angles can be independently controlled based on the dispersion properties of the 1D Fabry-Pérot leaky wave antenna (Guzmán-Quirós et al. 2012a).

Using the scanning and EBG properties of the HIS-loaded Fabry-Pérot cavity, the leaky wave antenna can be electronically tuned to propagate/radiate energy or to serve as a reflector, depending on the electronic tuning of the HIS. Hence, the symmetrically-fed 1D Fabry-Pérot leaky wave antenna can be electronically tuned to operate in three different regimes:

(1) Backward scanning regime ($\theta_{RAD} < 0°$): LWA_{LEFT} must be tuned inside the scanning region ($C_{jL} < 0.27$ pF) while $\text{LWA}_{\text{RIGHT}}$ is fixed at EBG region ($C_{jR} = 0.3$ pF). In this way the input signal is guided to the left side of the LWA (no energy travels to the right side), providing scan at negative angles.
(2) Forward scanning regime ($\theta_{RAD} > 0°$): This is the symmetrical case. LWA_{LEFT} is operated in the EBG region ($C_{jL} = 0.3$ pF) and $\text{LWA}_{\text{RIGHT}}$ is tuned in the scanning region to provide beam steering at positive angles (with $C_{jR} < 0.27$ pF).
(3) Broadside radiation ($\theta_{RAD} = 0°$): Radiation at boresight can be obtained by launching two oppositely-directed leaky waves. To this end, LWA_{LEFT} and $\text{LWA}_{\text{RIGHT}}$ are tuned at the same operating point inside the scanning region at or below the splitting condition $\beta = \alpha$ (Lovat et al. 2006) (which is obtained for $C_{jR} = C_{jL} = 0.1$ pF in this case), so that the fields on the two sides of the antenna merge to form a single beam radiating at $\theta_{RAD} = 0°$.

DOI 10.1007/978-981-4560-75-7_119-1

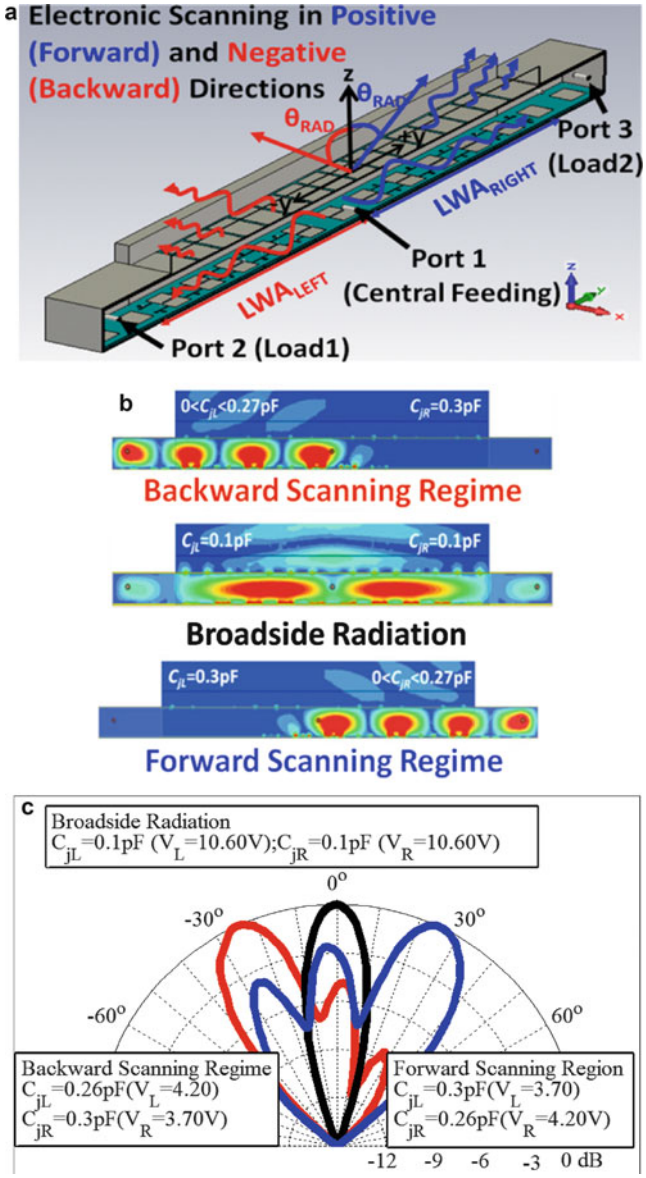


Fig. 34 (**a**) Scheme of full-space scanning 1D FP LWA (**b**) Electric field distribution along the 1D FP LWA structure and (**c**) Measured normalized radiation patterns (H-plane) at different operating regions (5.5 GHz)

For a better physical insight, the electric field distribution inside the Fabry-Pérot cavity is depicted in Fig. 34b for the three operating configurations, showing how the input energy can be routed to the right or left directions depending on the tuning voltage of the varactor diodes. Finally, measured radiation patterns of a fabricated prototype are shown in Fig. 34c, confirming full-space electronic scanning from −25° to +25° at 5.5 GHz.

Full-Space scanning 2D Fabry-Pérot leaky wave antennas

The EBG routing and scanning concepts described earlier can be extended from 1D to 2D Fabry-Pérot leaky wave antennas, so that 2D electronic scanning (azimuth and elevation) is obtained without using phased-arrays (Debogovic and Perruisseau-Carrier 2014). As a proof of the concept, a 2D Fabry-Pérot antenna was designed using a 2D PRS at the top and 2D tunable HIS at the bottom. The varactors have been arranged into four independently biased azimuthal sectors (SA to SD, as shown in the scheme of

DOI 10.1007/978-981-4560-75-7_119-1

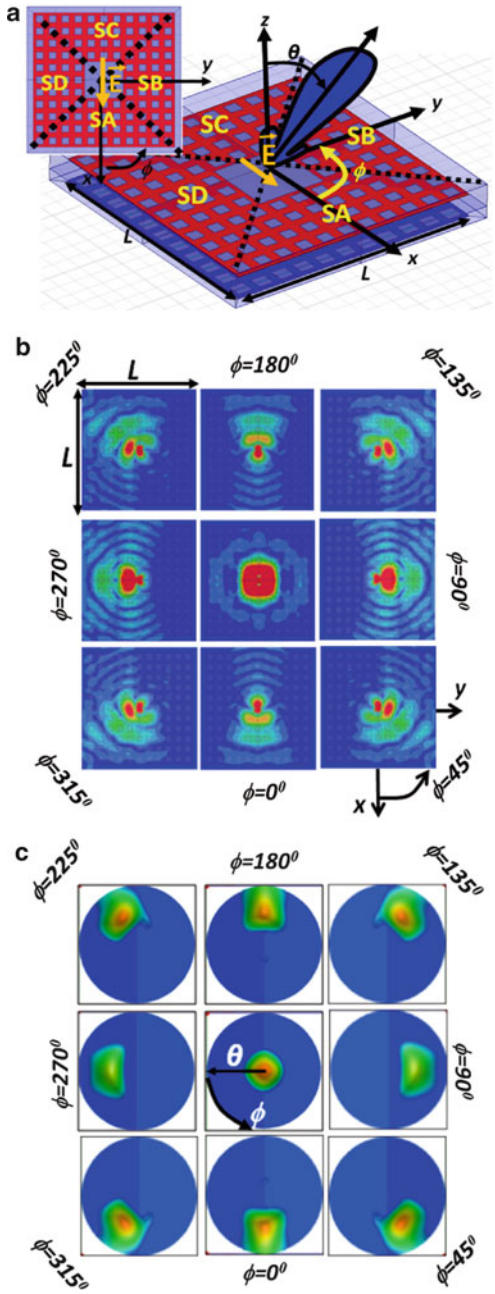


Fig. 35 Extension of the EBG routing mechanism for full-space electronic scanning in 2D FP LWAs: (**a**) HFSS 3D model of the tunable 2D FP ($A = 5.4\lambda_0 \times 5.4\lambda_0$), (**b**) Near fields inside the 2D FP cavity for several configurations of the sectors and (**c**) Radiation pattern in U-V coordinates (θ range represented [0°,45°])

Fig. 35a). In this case, a cylindrical LM is excited at the centre of the 2D Fabry-Pérot cavity by a horizontal dipole oriented along the *x* axis (Fig. 35a). This cylindrical LM is thus TM-polarized in sectors SA and SC (*x*-axis), TE-polarized in sectors SB and SD (*y*-axis), and a hybrid TE/TM LM for other directions (Ip and Jackson 1990). As in the 1D case, the LM will be routed towards a given sector by tuning it into the scanning region while the rest of the sectors are tuned to the EBG region.

Therefore, pencil beams which are scanned in the elevation plane (θ_{RAD}) can be created for discrete azimuthal (ϕ_{RAD}) angles defined by the four sectors. As a result, four radiation regions can be defined according to the direction of propagation of the wave:

DOI 10.1007/978-981-4560-75-7_119-1

(1) Radiation by TM LM propagating along the $\pm x$ axis. SA or SC is active (tuned at the scanning region) while all other sectors are inactive (tuned at EBG region).
(2) Radiation by TE LM propagating along the $\pm y$ axis. SB or SD is active and all other sectors are inactive.
(3) Radiation by hybrid TE/TM LM propagating along the oblique directions when two adjacent sectors are active (configurations SA&SB, SB&SC, SC&SD or SD&SA) and the other two are tuned to the EBG region.
(4) Broadside radiation. Optimal directive pencil beam radiating at boresight can be obtained when all the sectors are active and tuned at or below the splitting condition (Lovat et al. 2006).

Four different configurations of the sectors are illustrated in Fig. 35b, c), showing the simulated near fields inside the 2D Fabry-Pérot cavity and their respective radiation patterns. From the guided fields depicted in Fig. 35b, it is observed how the antenna aperture illumination can be modified by changing the activation of the azimuthal sectors. These illuminations lead to the synthesis of radiated pencil beams which can be steered to discrete azimuthal angles $\phi_{RAD} = [0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°]$. This is demonstrated by the position of the pencil beam in the radiation patterns obtained in Fig. 35c for each sector configuration. Moreover, for any constant azimuthal angle ϕ_{RAD}, tuning of the elevation scanning angle in the range $\theta_{RAD} = [5°, 25°]$ can be obtained, by properly tuning the active sector inside the scanning range. Finally, the broadside case is also obtained when the four sectors are tuned to (or below) the splitting condition, resulting in a cylindrical LM covering all sectors (Fig. 35b) and a pencil beam pointing at boresight (Fig. 35c).

From these results, it is demonstrated that continuous elevation scanning and discrete azimuthal sectorization can be electronically obtained. This requires accurate control of the dispersion properties of TE, TM and hybrid TE/TM cylindrical LMs.

Reconfigurable 1D SIW LWA

Reconfigurable LWAs can also be realized in substrate integrated waveguide (SIW), taking advantage of the planar, low cost, low loss, easy integration with planar circuits, and simple excitation properties of this technology (Xu and Wu 2005). Several reconfigurable LWAs have been proposed in the last decade using metamaterial CRLH microstrip leaky lines (Liu et al. 2002; Lim et al. 2005) in order to obtain electronic beam steering (Liu et al. 2002) or even beamshaping (Lim et al. 2005). In this section, two structures are presented to operate at ISM frequency band (5.5 GHz). The first one is a reconfigurable 1D antenna in SIW technology (1D SIW LWA), which is based on the non-reconfigurable static design presented in (Martínez-Ros et al. 2012a). This design is modified to provide full reconfiguration of its radiation pattern (radiation angle and beamwidth) using two control lines. It must be highlighted that it is the first time that it is demonstrated simultaneous electronic reconfiguration over the scanning angle and the beamwidth, without using complicated metamaterial unit-cells (Lim et al. 2005). The second structure is a radial array of SIW LWAs as originally proposed in (Martínez-Ros et al. 2012b). The introduction of electronic reconfiguration in the 2D LWA provides electronic elevation scanning and azimuth sectorization (as in the 2D FP LWA studied in the previous section), as well as electronic shaping of the pencil beam.

A schematic of the proposed full-reconfigurable 1D SIW LWA is presented in Fig. 36. The antenna consists of a double-layer structure with a ground plane in between. The top layer holds the SIW, which is loaded with a tunable PRS and a tunable HIS in substrate integrated technology. Both PRS and HIS are SIW versions of the Fabry-Pérot sheets. They are made of arrays of metallic posts periodically arranged at a certain distance (called PRS posts and HIS posts in Fig. 36). The PEC plane needed for the HIS is made of a dense arrangement of via-holes at a certain distance D from the HIS posts. Each post is connected by a through via to a varactor diode loaded in a microstrip stub, which is etched to the second layer (the control

DOI 10.1007/978-981-4560-75-7_119-1

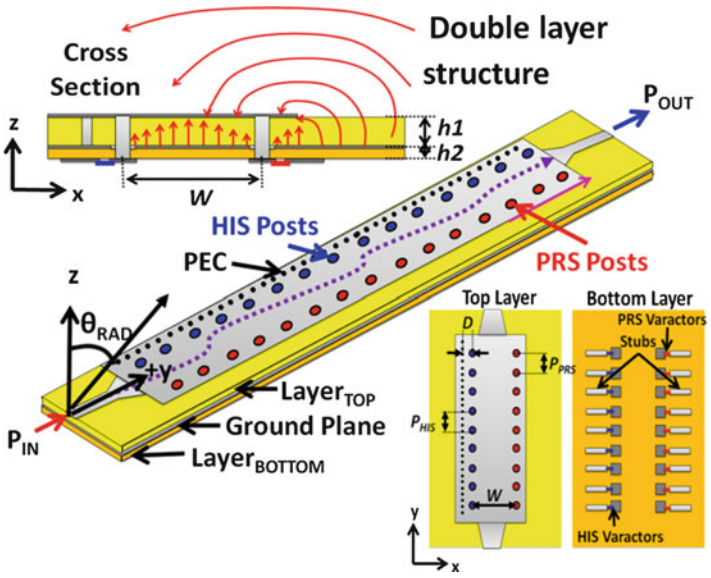


Fig. 36 Schematic of full-reconfigurable 1D SIW LWA: 3D model, cross section and front and rear views of the structure

layer). Teflon substrates with $\varepsilon_r = 2.2$ and thicknesses $h_1 = 3.17$ mm and $h_2 = 0.127$ mm are used for the radiation and control layers, respectively.

The PRS and HIS are separated by a distance W, comprising a resonant cavity that is analogous to the FP air-filled cavities of the antennas presented in the previous section, but in a substrate integrated technology. It is interesting to note that the radiation mechanism of this 1D SIW LWA is quite similar to the previous 1D FP LWA; according to the defined coordinate system, the FP antennas resonate between the PRS-HIS cavity, so the resonance is produced along the z-axis, whereas in the 1D SIW LWA case, this resonance arises along the x-axis. Moreover, in this case the PRS is also a tunable structure, so that it allows electronic control of the leakage rate (α) of the LM which is related to the half-power beamwidth and directivity (Goldstone and Oliner 1959). On the other hand, the tunable HIS controls the scanning angle by altering the LM phase constant (β), as described for the previous FP antennas. Both SIW PRS and HIS are independently biased, to enable independent and simultaneous electronic control of the radiation angle and beamwidth of the scanned fan beam.

A 1D SIW LWA has been simulated at a fixed frequency of 5.5 GHz in order to demonstrate the concept. The antenna length is $L_A = 3\lambda_0$, and it is terminated with a short-circuit load in order to estimate the variations in the leakage rate from the reflected lobe. The tuning parameters which reconfigure the antenna are the varactors' junction capacitances introduced at the HIS and the PRS circuits, C_{jHIS} and C_{jPRS}, respectively. Simulated radiation patterns at 5.5 GHz for six different configurations are presented in Fig. 37. The values of the scanning angle (θ_{RAD}), the normalized attenuation constants (α/k_0), and associated -3 dB beamwidth ($\Delta\theta$), are listed in Table 1 for each case. It is demonstrated that the scanning angle and the beamwidth can be simultaneously controlled by properly adjusting the two independently tunable parameters, namely C_{jHIS} and C_{jPRS}. Particularly, cases 1 and 2 show a beam scanned at $\theta_{RAD} \approx 15°$ with variable beamwidths $\Delta\theta = 11°$ and $\Delta\theta = 36°$, as it can be seen in the radiation patterns of Fig. 37a. The synthesis of narrow and broad beams is repeated for other scanning angles, namely $\theta_{RAD} \approx 24°$ and $\theta_{RAD} \approx 34°$, for cases 3–4 and 5–6, respectively. The normalized leakage rate is estimated for each case, by measuring the level between the main lobe and the mirrored reflected lobe (Goldstone and Oliner 1959). As expected, narrow beams correspond to low leakage rates, while broader beams arise

DOI 10.1007/978-981-4560-75-7_119-1

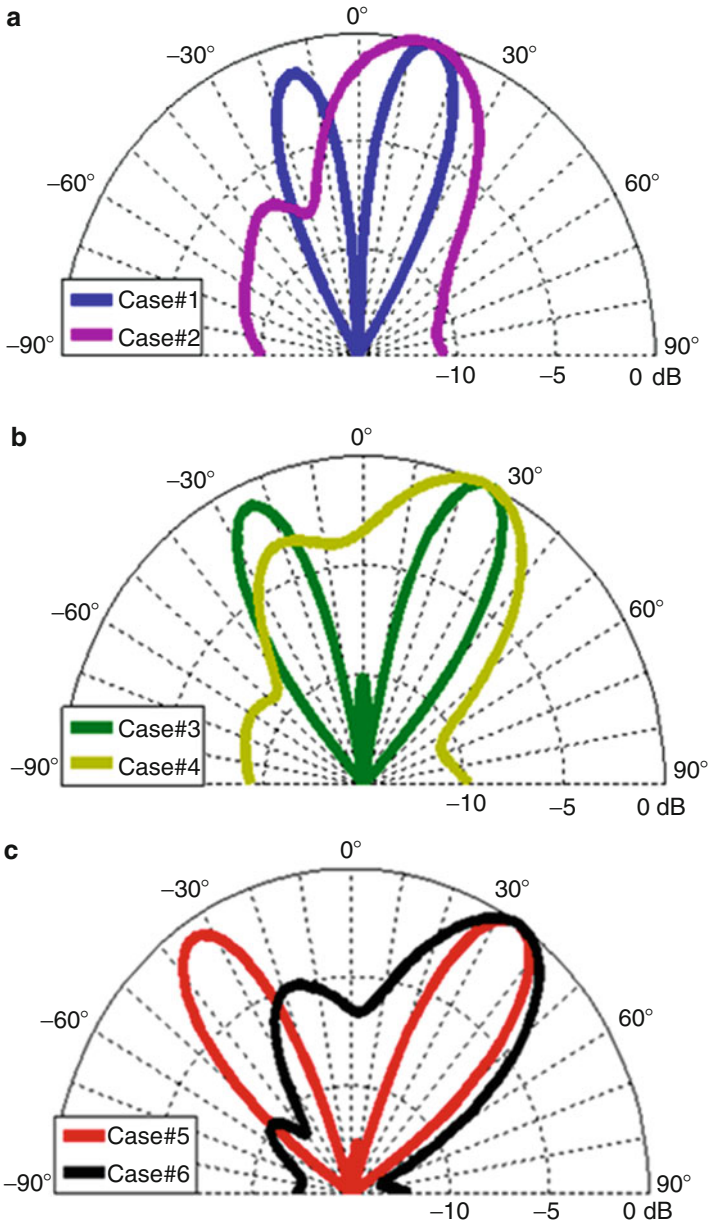


Fig. 37 Simulated radiation patterns (normalized directivity in dB) for different Cj_{HIS} & Cj_{PRS} configurations (5.5 GHz): (**a**) Cases 1 and 2, (**b**) Cases 3 and 4, and (**c**) Cases 5 and 6

Table 1 Results for different combinations of C_{jHIS} and C_{jPRS}

Case#	C_{jHIS}(pF)	C_{jPRS}(pF)	θ_{RAD}(deg)	α/k_0	$\Delta\theta$(deg)
1	0.45	0.2	15°	0.0036	11°
2	0.2	0.6	13°	0.0360	36°
3	0.5	0.2	24°	0.0017	13°
4	0.23	0.6	24°	0.0120	39°
5	0.6	0.2	34°	0.0009	16°
6	0.25	0.6	35°	0.0070	30°

DOI 10.1007/978-981-4560-75-7_119-1

from increased leakage rate. A one-order-of-magnitude variation in α/k_0 is obtained for all scanning angles. To conclude, these preliminary simulations verify that the proposed tunable 1D SIW LWA allows flexible electronic tuning of the scanning angle and the directivity at a fixed frequency of 5.5 GHz. Future work will provide a more detailed study of the dispersion curves of the antenna with C_{jPRS} and C_{jHIS}, and the optimal design and fabrication of a full prototype.

Summary

In the last decade, substantial achievements on reconfigurable antennas (RAs) have been made and several inherent challenges have been successfully overcome, including the design of dual-band polarization RAs and high-gain beam-steering pattern RAs. Furthermore, the authors of this chapter have employed novel reconfiguration techniques to achieve low-cost, low-loss phase shifters and reconfigurable leaky wave antennas. The reconfigurable defected microstrip structure (RDMS) based phase shifter has demonstrated its capability to realize a compact phased antenna array for beamforming applications. The reconfigurable leaky wave antennas are capable of solving two disadvantages of traditional leaky wave antennas, including the narrow impedance bandwidth and frequency dependent beam scanning. This, in turn, makes the leaky wave antenna more promising for today's wireless communications systems.

Future work in this field includes the design of a RA that can deliver full reconfiguration of the three characteristics of the antenna. Such a RA can bring considerable benefits to many wireless communication systems. However, the strong linkage between an antenna's frequency response and its radiation characteristics makes it extremely challenging to independently control of operating frequency, polarization and radiation pattern. Therefore, new and effective techniques are needed to break the linkage in order to achieve full reconfiguration of an antenna. Another significant direction for future work is the reconfiguration of the entire radio frequency (RF) front-end as well as the signal processing in the physical layer. It is well known that antennas are only one part of the entire transmitter and receiver sub-systems. If reconfigurable antennas are incorporated in wireless communications systems, the other parts of the system should also be capable of reconfiguring their characteristics in order to utilize the diversity derived from antenna agility. As research on these two topics progresses, RAs may play a major role to enhance the performance of the next-generation communications systems significantly.

Cross-References

► Reconfigurable Passive and Active Antennas/Tunable Antennas

References

Barrera JD, Huff GH (2014) A fluidic loading mechanism in a polarization reconfigurable antenna with a comparison to solid state approaches. IEEE Trans Antennas Propag 62:4008–4014

Bernhard JT (2005) Reconfigurable antennas. Wiley, New York

Bernhard JT, Volakis JL (2007) Antenna engineering handbook, 4th edn. McGraw-Hill, New York

Bernhard JT, Kiely E, Washington G (2001) A smart mechanically-actuated two-layer electromagnetically coupled microstrip antenna with variable frequency, bandwidth, and antenna gain. IEEE Trans Antennas Propag 49:597–601

Bhartia P, Bahl IJ (1982) Frequency agile microstrip antennas. Microw J :67–70
Cai Y, Guo YJ, Bird TS (2012) A frequency reconfigurable printed yagi-Uda dipole antenna for cognitive radio applications. IEEE Trans Antennas Propag 60:2905–2912
Chen RH, Row JS (2008) Single-fed microstrip patch antenna with switchable polarization. IEEE Trans Antennas Propag 56:922–926
Christodoulou CG, Tawk Y, Lane SA, Erwin SR (2012) Reconfigurable antennas for wireless and space applications. Proc IEEE 100:2250–2261
Davis ME (1975) Integrated diode phase-shifter elements for an X-band phased-array antenna. IEEE Trans Microw Theory Tech 23:1080–1084
Deal WR, Kaneda N, Sor J, Qian Y, Itoh T (2000) A new quasi-Yagi antenna for planar active antenna arrays. IEEE Trans Microw Theory Tech 48:910–918
Debogovic T, Perruisseau-Carrier J (2014) Array-fed partially reflective surface antenna with independent scanning and beamwidth dynamic control. IEEE Trans Antennas Propag 62:446–449
Ding C, Guo YJ, Qin PY, Bird TS, Yang Y (2014) A defected microstrip structure (DMS) based phase shifter and its application to beamforming antennas. IEEE Trans Antennas Propag 62:641–651
Ding C, Guo YJ, Qin PY, and Yang Y (2015) A compact phase shifter employing reconfigurable defected microstrip structure (RDMS) for phased array antennas. IEEE Trans. Antennas Propag
Donelli M, Azaro R, Fimognari L, Massa A (2007) A planar electronically reconfigurable Wi-Fi band antenna based on a parasitic microstrip structure. IEEE Antennas Wirel Propag Lett 6:623–626
Dorsey WM, Zaghloul AI (2009) Perturbed square-ring slot antenna with reconfigurable polarization. IEEE Antennas Wirel Propag Lett 8:603–606
Feresidis AP, Vardaxoglou JC (2001) High gain planar antenna using optimized partially reflective surfaces. IEE Proc Microw Antennas Propag 148:345–350
Feresidis AP, Goussetis G, Wang S, Vardaxoglou JC (2005) Artificial magnetic conductor surfaces and their application to low-profile high-gain planar antennas. IEEE Trans Antennas Propag 53:209–215
García-Vigueras M, Gómez-Tornero JL, Goussetis G, Weily AR, Guo YJ (2011) 1D-leaky wave antenna employing parallel-plate waveguide loaded with PRS and HIS. IEEE Trans Antennas Propag 59:3687–3694
Genovesi S, Candia AD, Monorchio A (2014) Compact and low profile frequency agile antenna for multistandard wireless communication systems. IEEE Trans Antennas Propag 62:1019–1026
Goldstone L, Oliner AA (1959) Leaky-wave antennas I: rectangular waveguides. IRE Trans Antennas Propag 7:307–319
Guzmán-Quirós R, Gómez-Tornero JL, Weily AR, Guo YJ (2012a) Electronically steerable 1D Fabry-Perot leaky-wave antenna employing tunable high impedance surface. IEEE Trans Antennas Propag 60:5046–5055
Guzmán-Quirós R, Gómez-Tornero JL, Weily AR, Guo YJ (2012b) Electronic full-space scanning with 1D Fabry-Pérot LWA using electromagnetic band gap. IEEE Antennas Wirel Propag Lett 11:1426–1429
Han SM, Kim CS, Ahn D, Itoh T (2005) Phase shifter with high phase shifts using defected ground structures. Electron Lett 41:196–197
Hansen RC (1998) Phased array antennas. Wiley, New York
Ho KM, Rebeiz GM (2014) A 0.9–1.5 GHz microstrip antenna with full polarization diversity and frequency agility. IEEE Trans. Antennas Propag 62:2398–2406
Horn RE, Jacobs H, Freibergs E, Klohn KL (1980) Electronic modulated beam steerable silicon waveguide array antenna. IEEE Trans Microw Theory Tech 28:647–653
Hsu SH, Chang K (2007) A novel reconfigurable microstrip antenna with switchable circular polarization. IEEE Antennas Wirel Propag Lett 6:160–162

Huang L, Chiao J, Lisio P (2000) An electronically switchable leaky wave antenna. IEEE Trans Antennas Propag 48:1769–1772

Huff GH, Feng J, Zhang S, Bernhard JT (2003) A novel radiation pattern and frequency reconfigurable single turn square spiral microstrip antenna. IEEE Microw Wirel Compon Lett 13:57–59

Hum SV, Xiong HY (2010) Analysis and design of a differentially-fed frequency agile microstrip patch antenna. IEEE Trans Antennas Propag 58:3122–3130

Hum SV, Okoniewski M, Davies RJ (2005) Realizing an electronically tunable reflectarray using varactor diode-tuned elements. IEEE Microw Wirel Compon Lett 15:422–424

Hum SV, Okoniewski M, Davies RJ (2007) Modeling and design of electronically tunable reflectarrays. IEEE Trans Antennas Propag 55:2200–2210

Ip A, Jackson DR (1990) Radiation from cylindrical leaky waves. IEEE Trans Antennas Propag 38:482–488

Jung CW, Lee M, Li GP, Flaviis FD (2006) Reconfigurable scan-beam single-arm spiral antenna integrated with RF-MEMS switches. IEEE Trans Antennas Propag 54:455–463

Khidre A, Lee KF, Yang F, Elsherbeni AZ (2013) Circular polarization reconfigurable wideband E-shaped patch antenna for wireless applications. IEEE Trans Antennas Propag 61:960–964

Lai MI, Wu TY, Wang JC, Wang CH, Jeng S (2008) Compact switched-beam antenna employing a four-element slot antenna array for digital home applications. IEEE Trans Antennas Propag 56:2929–2936

Li H, Xiong J, Yu Y, He S (2010a) A simple compact reconfigurable slot antenna with a very wide tuning range. IEEE Trans Antennas Propag 58:3725–3728

Li Y, Zhang Z, Chen W, Feng Z (2010b) Polarization reconfigurable slot antenna with a novel compact CPW-to-Slotline transition for WLAN application. IEEE Antennas Wirel Propag Lett 9:252–255

Li Y, Zhang Z, Zheng J, Feng Z, Iskander MF (2011) Experimental analysis of a wideband pattern diversity antenna with compact reconfigurable CPW-to-slotline transition feed. IEEE Trans Antennas Propag 59:4222–4228

Lim S, Ling H (2007) Design of electrically small pattern reconfigurable Yagi antenna. Electron Lett 43:1326–1327

Lim S, Caloz C, Itoh T (2005) Metamaterial-based electronically controlled transmission-line structure as a novel leaky-wave antenna with tunable radiation angle and beamwidth. IEEE Trans Microw Theory Tech 53:161–173

Liu L, Caloz C, Itoh T (2002) Dominant mode leaky-wave antenna with backfire-to-endfire scanning capability. Electron Lett 38:1414–1416

Lovat G, Burghignoli P, Jackson DR (2006) Fundamental properties and optimization of broadside radiation from uniform leaky-wave antennas. IEEE Trans Antennas Propag 54:1442–1452

Maheri H, Tsutsumi M, Kumagi N (1988) Experimental studies of magnetically scannable leaky-wave antennas having a corrugated ferrite slab/dielectric layer structure. IEEE Trans Antennas Propag 36:911–917

Martínez-Ros AJ, Gómez-Tornero JL, Goussetis G (2012a) Planar leaky-wave antenna with flexible control of the complex propagation constant. IEEE Trans Antennas Propag 60:1625–1630

Martínez-Ros AJ, Gómez-Tornero JL, and Goussetis G (2012b) Broadside radiation from radial arrays of substrate integrated leaky-wave antennas. In: Proceedings of the 6th European conference on antennas and propagation (EUCAP), pp 252–254

Nair SVS, Ammann MJ (2007) Reconfigurable antenna with elevation and azimuth beam switching. IEEE Antennas Wirel Propag Lett 9:367–370

Nikolaou S et al (2006) Pattern and frequency reconfigurable annular slot antenna using PIN diodes. IEEE Trans Antennas Propag 54:439–448

Parker D, Zimmermann DC (2002) Phased arrays – part I: theory and architecture. IEEE Trans Microw Theory Tech 50:688–698

Patil P, Khot UP, Bhujade S (2012) DGS based microstrip phase shifters. In: International conference on sensing technology, pp 723–728

Qin PY, Guo YJ, Liang CH (2010a) Effect of antenna polarization diversity on MIMO system capacity. IEEE Antennas Wirel Propag Lett 9:1092–1095

Qin PY, Weily AR, Guo YJ, Bird TS, Liang CH (2010b) Frequency reconfigurable quasi-Yagi folded dipole antenna. IEEE Trans Antennas Propag 58:2742–2747

Qin PY, Weily AR, Guo YJ, Liang CH (2010c) Polarization reconfigurable U-slot patch antenna. IEEE Trans Antennas Propag 58:3383–3388

Qin PY, Guo YJ, Cai Y, Dutkiewicz E, Liang CH (2011) A reconfigurable antenna with frequency and polarization agility. IEEE Antennas Wirel Propag Lett 10:1373–1376

Qin PY, Guo YJ, Weily AR, Liang CH (2012) A pattern reconfigurable U-slot antenna and its applications in MIMO systems. IEEE Trans Antennas Propag 60:516–528

Qin PY, Guo YJ, Ding C (2013a) A dual-band polarization reconfigurable antenna for WLAN systems. IEEE Trans Antennas Propag 61:5706–5713

Qin PY, Guo YJ, Ding C (2013b) A beaming steering pattern reconfigurable antenna. IEEE Trans Antennas Propag 61:4891–4899

Rodrigo D, Jofre L, Cetiner BA (2012) Circular beam-steering reconfigurable antenna with liquid metal parasitic. IEEE Trans Antennas Propag 60:1796–1802

Shafai C, Sharma SK, Shafai L, Chrusch DD (2004) Microstrip phase shifter using ground-plane reconfiguration. IEEE Trans Microw Theory Tech 52:144–153

Shelby RA, Smith DR, Shultz S (2001) Experimental verification of a negative index of refraction. Science 292:77–79

Sievenpiper DF (2005) Forward and backward leaky wave radiation with large effective aperture from an electronically tunable textured surface. IEEE Trans Microw Theory Tech 53:236–247

Sievenpiper D, Schaffner J, Lee JJ, Livingston S (2002) A steerable leaky-wave antenna using a tunable impedance ground plane. IEEE Antennas Wirel Propag Lett 1:179–182

Sung YJ, Jang TU, Kim YS (2004) A reconfigurable microstrip antenna for switchable polarization. IEEE Microw Wirel Compon Lett 14:534–536

Tawk Y, Costantine J, Avery K, Christodoulou CG (2011) Implementation of a cognitive radio front-end using rotatable controlled reconfigurable antennas. IEEE Trans Antennas Propag 59:1773–1778

Trentini GV (1956) Partially reflecting sheet arrays. IEEE Trans Antennas Propag 4:666–671

Veselago VG (1968) (Russian text 1967)) The electrodynamics of substances with simultaneously negative values of ε and μ. Sov Phys Usp 10:509–514

Wang S, Feresidis AP, Goussetis G, Vardaxoglou JC (2006) High-gain subwavelength resonant cavity antenna based on metamaterial ground planes. IEE Proc Microw Antennas Propag 153:1–6

Waterhouse R, Shuley N (1994) Full characterisation of varactor-loaded, probe-fed, rectangular, microstrip patch antennas. IEE Proc Microw Antennas Propag 141:367–373

Weily AR, Bird TS, Guo YJ (2008) A reconfigurable high gain partially reflecting surface antenna. IEEE Trans Antennas Propag 56:3382–3389

Whicker LR (1973) Review of ferrite phase shifter technology. IEEE MTT-S Int. Microwave Symp. Dig. 95–97

White JF (1974) Diode phase shifters for array antennas. IEEE Trans Microw Theory Tech 22:658–674

Wu SJ, Ma TG (2008) A wideband slotted bow-tie antenna with reconfigurable CPW-to-slotline transition for pattern diversity. IEEE Trans Antennas Propag 56:327–334

Xu F, Wu K (2005) Guided-wave and leakage characteristics of substrate integrated waveguide. IEEE Trans Microw Theory Tech 53:66–73

Yang SLS, Luk KM (2006) Design a wide-band L-probe patch antenna for pattern reconfigurable or diversity applications. IEEE Trans Antennas Propag 54:433–438

Yang XS, Wang BZ, Wu W, Xiao S (2007) Yagi patch antenna with dual-band and pattern reconfigurable characteristics. IEEE Antennas Wirel Propag Lett 6:168–171

Ye S, Wang XL, Wang WZ, Jin RH, Geng JP, Bird TS, Guo YJ (2012) High gain planar antenna arrays for mobile satellite communications. IEEE Antennas Propag Mag 54:256–268

Zhang S, Huff GH, Feng J, Bernhard JT (2004) A pattern reconfigurable microstrip parasitic array. IEEE Trans Antennas Propag 52:2773–2776

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_120-1

Physics and Mathematics of Radio Wave Propagation in Cellular Wireless Communications

Magdalena Salazar-Palma[a]*, Tapan K. Sarkar[b], Mohammad N. Abdallah[b], Walid Dyab[b], M. V. S. N. Prasad[c] and Sio Weng Ting[d]
[a]Department of Signal Theory & Communications, Universidad Carlos III de Madrid, Leganés, Madrid, Spain
[b]Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, USA
[c]National Physical Laboratory, New Delhi, India
[d]Department of Electrical Engineering and Computer Science, University of Macau, Macau, China

Abstract

The objective of this chapter is to illustrate that an electromagnetic macro model can accurately predict the dominant component of the propagation path loss for a cellular wireless communication. The reason a macro model can provide accurate results that agree with experiments is because the trees, buildings, and other man-made obstacles contribute second-order effects to the propagation path loss as the dominant component that affects propagation is the free-space propagation of the signal and the effect of the earth over which the signal is propagating. It is demonstrated using both measurements and an analytical theoretical model that the propagation path loss inside a cellular communication cell is first 30 dB per decade of distance, and later on, usually outside the cell, it is 40 dB per decade of the electrical distance between the transmitter and the receiver irrespective of their heights from the ground. This implies that the electric field decays first at a rate of $\rho^{-1.5}$ inside the cell and later on, usually outside the cell, as ρ^{-2}, where ρ stands for the distance between the transmitter and the receiver. This appears to be independent of the frequency of operation in the band of interest and the parameters of the ground. It is also illustrated that the so-called slow fading is due to the interference between the direct wave and the ground wave as introduced by Sommerfeld over 100 years ago. All these statements can be derived from the approximate integration of the Sommerfeld integrals using a modified path for the steepest descent method and also using a purely numerical methodology. Finally, an optical analog is described based on the image theory developed by Van der Pol to illustrate the mechanism of radio wave propagation in a cellular wireless communication system.

Keywords

Cellular wireless communication; Propagation over earth; Sommerfeld integrals; Schelkunoff integrals; Propagation models; Okumura's experiments; Propagation over water

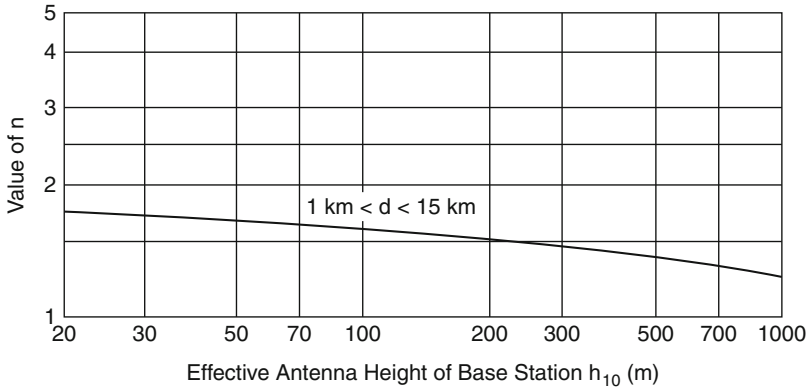
Introduction

In D. Gabor (1953), states:

> The wireless communication systems are due to the generation, reception and transmission of electro-magnetic signals. Therefore all wireless systems are subject to the general laws of radiation. Communication theory has up to now been

*Email: salazar@tsc.uc3m.es
*Email: m.salazar-palma@ieee.org

DOI 10.1007/978-981-4560-75-7_120-1



Distance dependence of median field strength in urban area ($E_m \propto d^{-n}$).

Fig. 1 The propagation path loss predicted by Okumura et al. and this is Fig. 12 of their paper

developed mainly along mathematical lines, taking for granted the physical significance of the quantities which are fundamental in its formalism. But communication is the transmission of physical effects from one system to another. Hence communication theory should be considered as a branch of physics.

This illustrates that to visualize the physics of the propagation mechanism in cellular wireless communication systems, it is necessary to apply the basic principles of electromagnetic theory as a statistical model is incapable of capturing the fundamentals of propagation with a field decay of $\rho^{-1.5}$ with distance ρ inside the cell. To illustrate this point, various experimental data are presented first. Then an analytical model (Abdallah et al. 2014; Sarkar et al. 2012a, b) is derived from the Sommerfeld integrals to reproduce the experimental data of Okumura et al. (1968) as this is one of the few papers which clearly state the nature of the transmitting and receiving antennas used and in which environment they were operating, thus making it possible for an interested reader to completely duplicate the experiment. Comparison is also made with other experimental data to illustrate the accuracy of the prediction carried out using an electromagnetic macro model based on available computational electromagnetic codes. Thus, it is shown that these electromagnetic analysis codes can provide path loss models which are as good as carrying out real-time path loss measurements using a van and other accompanying equipment which are both very expensive and time consuming. Finally, an optical analog is provided to visualize the actual mode of propagation for the nature of the wave propagation in a cellular environment (Abdallah et al. 2014; Sarkar et al. 2012a, b; De et al. 2010).

Description and Analysis of Measured Data

Figure 12 of Okumura et al.'s classic paper (Okumura et al. 1968), which is presented here as Fig. 1, illustrates quite clearly that for most practical heights of the base station antenna, the decay of the electric field strength is at a rate of $\rho^{-1.5}$ where ρ stands for the distance from the transmitting antenna. As described by Okumura et al. in Fig. 1, it is seen that the decay of the electric field strength inside the cell is at a rate of $\rho^{-1.5}$ or equivalently the power decays as ρ^{-3}, where ρ represents the distance from the transmitter. This translates into a propagation path loss of 30 dB per decade of distance traveled. This conclusion appears to be independent of the frequency of operation.

In Fig. 2, the data from Ericsson for an in-building path loss model is presented. This was shown in chapter 3 of *Mobile Antenna Systems Handbook* (Fujimoto 2008). Figure 2 shows that for the distance between 1 m and 10 m, the path loss is typically 30 dB at 900 MHz, the same value stated earlier by

DOI 10.1007/978-981-4560-75-7_120-1

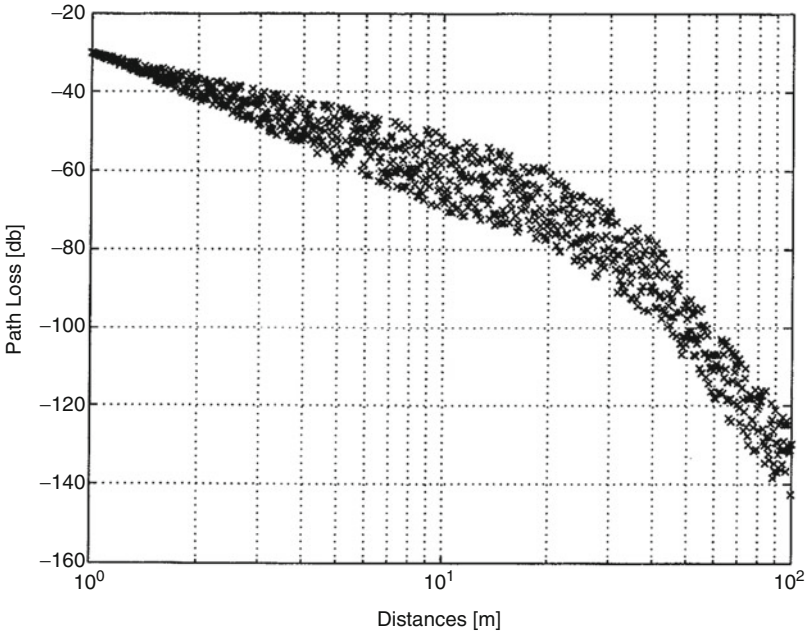


Fig. 2 Prediction from Ericsson in-building path loss model (Reproduced by permission from Simon R. Saunders, *Advances in Mobile Propagation Prediction Methods*, Chapter 3 of Fujimoto (2008))

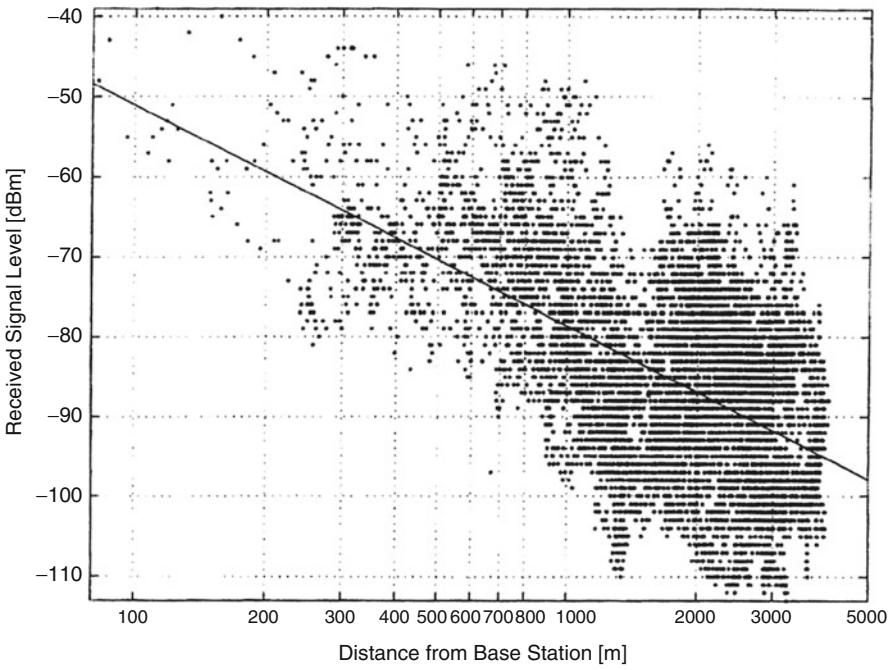


Fig. 3 Empirical model of macro cell propagation at 900 MHz, the dots are measurements taken in suburban area, whereas the *solid line* represents a best fit empirical model (Reproduced by permission from Simon R. Saunders, *Advances in Mobile Propagation Prediction Methods*, Chapter 3 Fujimoto (2008))

Okumura et al. even for different frequencies of operation. In Fig. 3, an empirical model for macro cell propagation at 900 MHz is presented (Fujimoto 2008). The dots are measurements taken in a suburban area, whereas the solid line represents a best fit empirical model (this solid line was not drawn by the

DOI 10.1007/978-981-4560-75-7_120-1

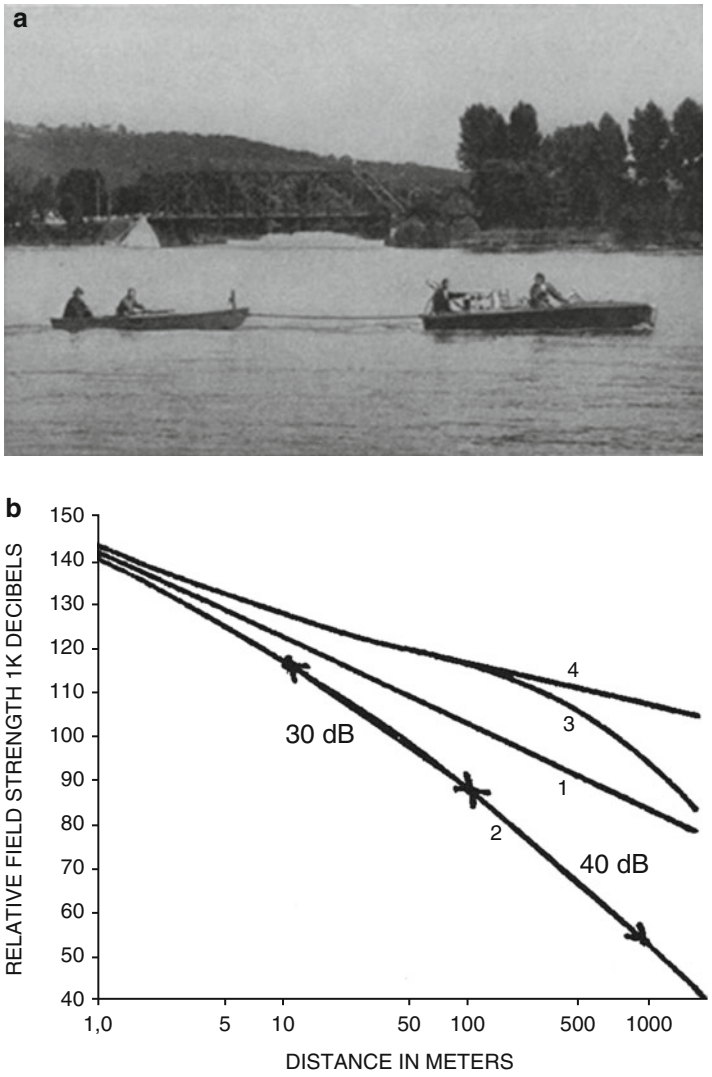


Fig. 4 Plot of the variation of the field strength as a function of radial distance over the lake. (**a**) Experimental setup and (**b**) variation of the path loss as a function of distance

authors of this article but by the authors of that chapter). If one goes from 100 m to 1000 m, it is seen that the path loss is again 30 dB for the best fit line at 900 MHz as presented in Fujimoto (2008) even though the parameters of the experiment are different from that of Fig. 3.

In 1937, C. R. Burrows from Bell Labs (Burrows 1937) carried out the measurements over Seneca Lake, New York, USA, to find the nature of electromagnetic wave propagation over water at 150 MHz. The transmitting and receiving antennas were placed on two different boats as seen in Fig. 4a. The measured propagation path loss is illustrated in Fig. 4b. Again, it is seen that as one moves from 10 m to 100 m, the path loss is 30 dB and forms 100 m to 1000 m, away from the transmitter, the path loss is 40 dB as mentioned before but this time at 150 MHz.

Next, results are presented displaying actual measurements carried out at different frequencies for three base stations located in dense urban, urban, and suburban environments. The special feature of the data taken in an urban environment was that the houses were not uniformly spaced. First, the results are presented for data taken at the carrier signals of 1800-MHz GSM base station transmitters located in the dense urban/urban environments of New Delhi area belonging to *Idea Cellular Network* which have been monitored with Nokia GSM receiver (model 6150) generally used as a drive-in tool for planning cellular

DOI 10.1007/978-981-4560-75-7_120-1

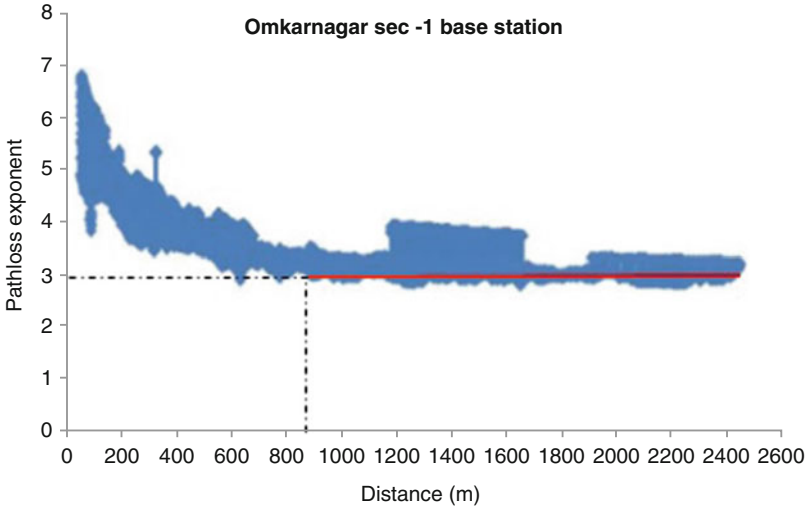


Fig. 5 Variation of path loss exponent with distance for OM-1 base station (1800 MHz). Base station height 24 m

networks along with a GPS receiver to know the latitude and longitude of the mobile. The sensitivity of the receiver is –102 dBm. The transmitting powers of all the base stations are +43 dBm. The receiving equipment was installed in a vehicle along with the data acquisition system. The vehicle moved on a normal road at a permissible speed in the traffic and the downlink signal strength was monitored by the receiver. The observed signal levels were converted into path loss values based on the received signal level, the transmitter power, and the transmitting and receiving antenna gains for further analysis. The gain of the transmitting antenna was 18 dBi. The estimated measurement rms (root mean square) error is around 1.5 dB. The height of the mobile receiving antenna was 1.5 m from the ground.

For the case of the 900-MHz results, the experiment is carried out with the help of *Aircom International Limited*, a UK company based in India. The transmitting power for all the base stations used in this study was 43.8 dBm, and the transmitting antenna gain was 2 dBi for all the base stations. The gain of the receiving antenna was 0 dBi and the height from the ground was 1.5 m. The receiver is a standard Nokia equipment used in drive-in tools used for field trials. For all the measurements, the position of the mobile is determined from the GPS (global positioning system) receiver, and this information with the coordinates of the base station was utilized to deduce the distance traveled by the mobile from the base station. The signal strength information recorded in dBm was converted into path loss values utilizing the gains of the antenna. The data was recorded with 512 samples in one second on a laptop computer, and the number of samples collected for each site varied from 1×10^5 to 2×10^5. Measured rms error was around 1.5 dB. The data was averaged over a conventional range of 40 $\mathscr{W}^o$ where $\mathscr{W}^o$ represents the wavelength.

The first base station operating at 1800 MHz called OM-1 was located in a dense urban environment in the Indian capital city of Delhi, and the height of the transmitting base station antenna was at 24 m from the ground. The variation of the path loss is shown in Fig. 5. It is seen that inside the cell, the path loss has settled down approximately to a factor of 3.

Next, the data at 900 MHz for the base station UA located in a medium urban environment of the Indian capital city of Delhi is shown. The base station antenna is located at a height of 24 m. From Fig. 6, it appears that the path loss exponent factor in the cell has settled down to a value of 3 at around $\frac{4H_{TX}H_{RX}}{\mathscr{W}^o} = \frac{4\times24\times1.5}{1/3} = 432$ m as predicted by the theory (H_{TR} represents the height of the transmitting antenna over the ground, H_{RX} represents height of the receiving antenna over the ground, and $\mathscr{W}^o$ is the wavelength of operation) (De et al. 2010).

DOI 10.1007/978-981-4560-75-7_120-1

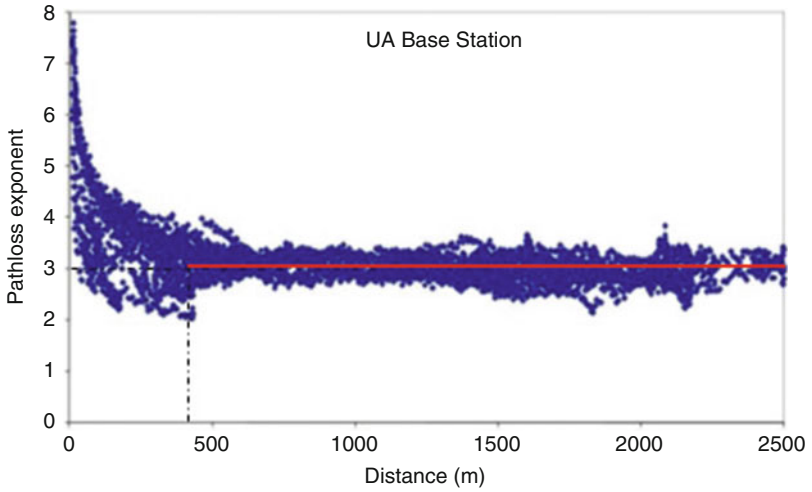


Fig. 6 Variation of path loss exponent with distance for UA base station (900 MHz). Base station height 24 m

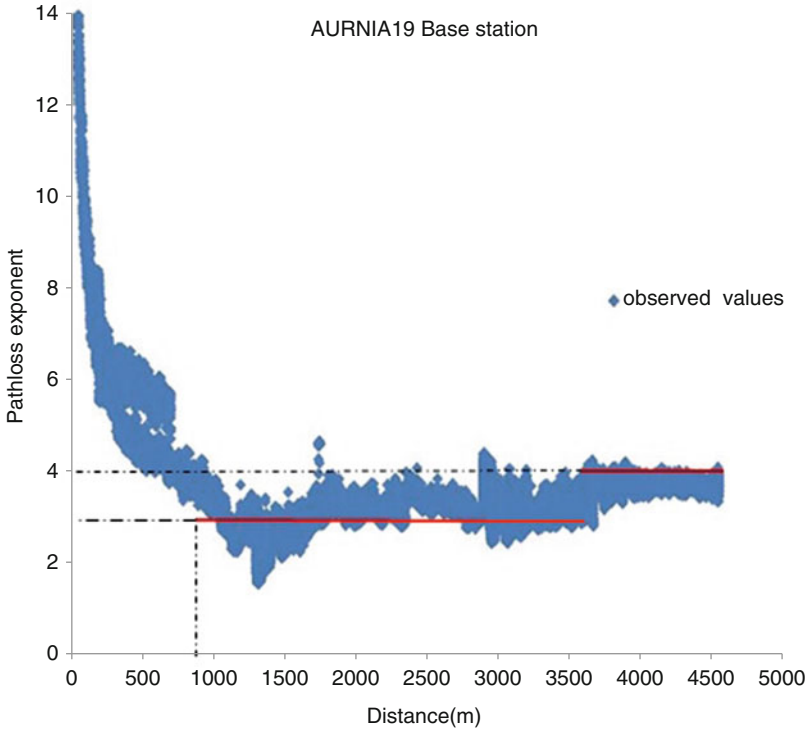


Fig. 7 Variation of path loss exponent with distance for AURNIA 19 base station (900 MHz). Base station height 50 m. Near the transmitter, the path loss exponent has a value of 3 and changing to 4 as one moves further away from the base station

Figure 7 shows the variation of the path loss exponent from the base station AURNIA 19. This station is located in an industrial area in the city of Aurangabad, in the Maharashtra state of western India. The transmitter height was 50 m from the ground and was operating at 900 MHz. The path loss exponent is expected to start at 3 at approximately a distance of $\frac{4H_{TX}H_{RX}}{\mathcal{W}^{\circ}} = \frac{4\times50\times1.5}{1/3} = 900\,\text{m}$ (De et al. 2010) as the operating frequency was 900 MHz. This region starts where the slow fading region ends as illustrated in

DOI 10.1007/978-981-4560-75-7_120-1

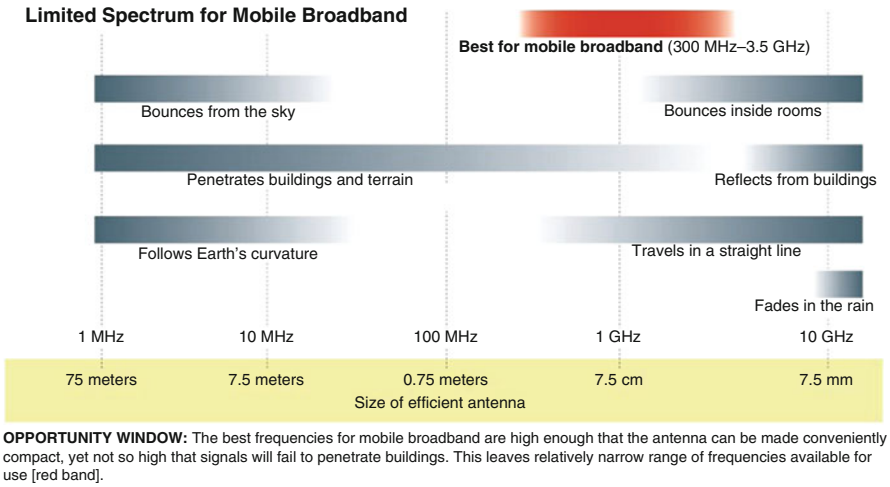


Fig. 8 Suitable bands for mobile communications in the electromagnetic spectrum (Lazarus 2010)

De et al. (2010). After some distance, the path loss exponent gradually increases to 4 in the far-field region, as predicted by theory (De et al. 2010).

All three measurements show that in an intermediate region from the base station antenna, the propagation path loss exponent is 3. Outside that region and far away from the transmitter, the path loss exponent is 4. This is independent of the frequency of operation and also of the ground parameters. This implies that the electric field varies as $\rho^{-1.5}$ inside the cell and in the far-field region, usually outside the cell as ρ^{-2}. Another observation to be made is that at 1000 m from the transmitter, the path loss due to the decay of the free-space fields propagating over earth is 10 $\log_{10}$ ρ^{-3} which is equal to −90 dB. This is quite a large value. Also, the large variation in the experimental data reveals that due to buildings, trees, and so on, the curve is not smooth but has the variability in the path loss exponent as illustrated in the measured data. This variation due to environments is approximately of the order of 30–40 dB. Hence, it is clear that the path loss due to the free-space propagation which is of ~90 dB is of prime importance and the effects of the buildings and the nature of the terrain are secondary as they are ~30–40 dB. Also, the path loss exponent seems to be insensitive to the nature of the ground, be it an urban and suburban grounds or propagation over water. It is also insensitive to the height of the transmitting antenna.

From the above data, it is now asserted that a physics-based electromagnetic macro model can provide accurate predictions for the path loss exponent in a cellular network using electromagnetic simulation tools that depend only on some physical parameters of the macro model and primarily those related to the antennas and the electrical parameters of the ground. In a macro model, one needs to include only the electrical parameters of the environment without including the clutter-effect factors such as buildings, trees, and so on.

Furthermore as illustrated in Lazarus (2010) and displayed in Fig. 8, the reason that the 0.8–2.1-GHz cellular band was chosen for mobile broadband is that the reflection from buildings is negligible, and yet the signals can penetrate buildings and do not significantly bounce inside the rooms. Furthermore in Fig. 9, the one-way attenuation through common building materials illustrates that at the cellular band, signals can penetrate buildings very easily McMillan (2006). Hence, the clutter effects generated by buildings or trees are considered to be second-order effects, as the primary factor influencing propagation is the effect of the imperfectly conducting ground which however is seldom accounted for in any propagation model. Information in Figs. 8 and 9 makes it clear that the use of an electromagnetic macro model for the analysis of the propagation path loss in mobile communication using physical parameters like height, gain, and tilt of an antenna over an imperfect ground can accurately duplicate

DOI 10.1007/978-981-4560-75-7_120-1

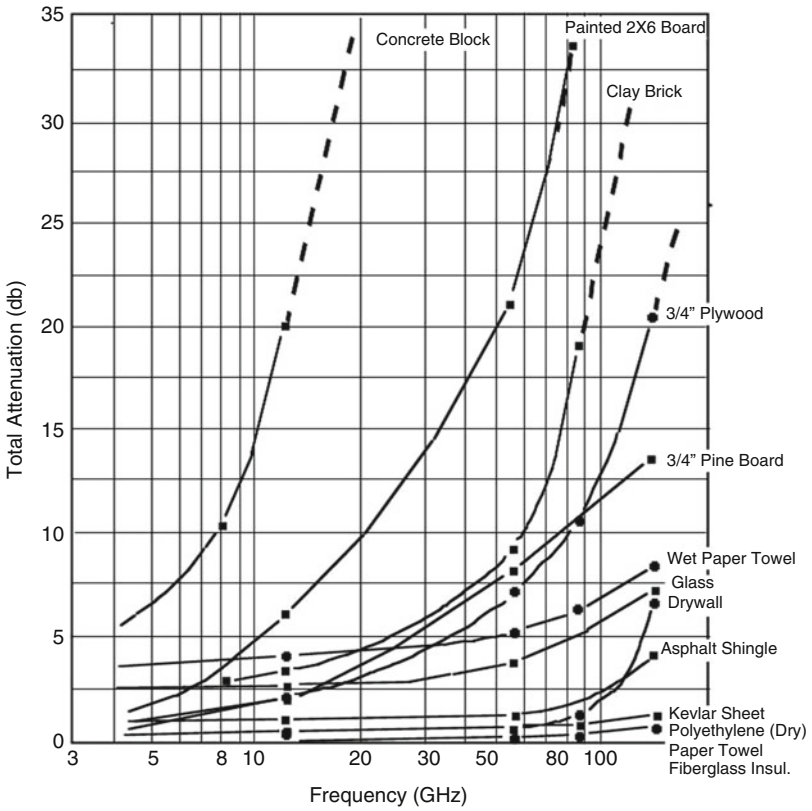


Fig. 9 One-way attenuation through common building materials as a function of frequency (McMillan 2006)

experimental data taken in a real environment. The nonplanar nature of the real ground is deemed not to be a show stopper for the propagation analysis using a macro model as illustrated in Abdallah et al. (2014), Sarkar et al. (2012a, b), and De et al. (2010). It is primarily because of the preponderance of physical evidence that wireless signals penetrate through buildings and do not get reflected significantly from them that an electromagnetic macro model has been proposed based on fundamental physics replacing the statistical models which do not address the underlying electromagnetics, namely, the effect of the propagation over earth!

Typically, in the current state of the art, the drive test data are fit in a least-squares fashion to a statistical model or a ray-tracing scenario is used. The fact of the matter is that current propagation modeling tools cannot capture the fundamental physics, namely, inside a cell, the path loss is typically 30 dB per decade (Abdallah et al. 2014; Sarkar et al. 2012a, b) or equivalently, the electric field strength varies as $\rho^{-1.5}$ with the distance from the antenna ρ, and outside the cell, the propagation path loss is approximately 40 dB per decade or equivalently, the electric field strength varies as ρ^{-2}. Thus, statistical-based models do not capture the basic physics as they are generally devoid of basic physics. Furthermore, it does not make sense to have the path loss exponent tied to a value of a reference distance in the current state-of-the-art ad hoc models which is quite meaningless from a scientific point of view as the propagation effects should be dependent on the physical parameters of the system. In addition, the value of the reference distance should be such that it is located in the far field of the transmitting antenna. But seldom is this ad hoc rule of thumb respected!

Next, the electromagnetic macro model that is going to predict the path loss exponents is presented.

DOI 10.1007/978-981-4560-75-7_120-1

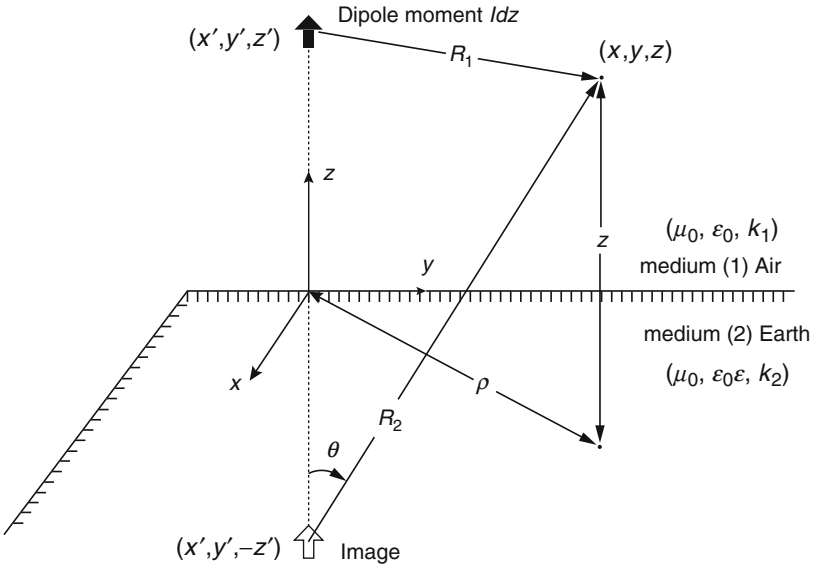


Fig. 10 A vertical dipole over a horizontal imperfect ground plane

Theoretical Analysis of Propagation over Earth

To simulate the electromagnetic propagation path loss in a cellular environment accurately using the physical parameters related to the environment, a method of moment code based on Sommerfeld Green's function to treat the imperfectly conducting planar ground has been used (Sommerfeld 1909; Sarkar 1977; Djordjevic et al. 2002; Dyab et al. 2013; Sarkar et al. 2014). However, an approximate analytical formulation based on integrating the semi-infinite integrals by a modified method of steepest descent is also going to be used in addition to an accurate numerical evaluation which also reaches the same conclusion.

In the original Sommerfeld formulation, one considers an elementary dipole of moment *Idz* oriented along the *z*-direction and located at (x', y', z'). The dipole is situated over an imperfect ground plane characterized by a complex relative dielectric constant ε as seen in Fig. 10. The complex relative dielectric constant is given by $\varepsilon = \varepsilon_r - \frac{j\sigma}{\omega \varepsilon_0}$ where ε_r represents the relative permittivity of the medium, ε_0 is the permittivity of vacuum, σ is the conductivity of the medium, ω stands for the angular frequency, and j is the imaginary unit, i.e., $j = \sqrt{-1}$. It is possible to formulate a solution to the problem of radiation from the dipole operating in the presence of the imperfect ground in terms of a single Hertzian vector potential $\hat{u}_z \Pi_z$ of the electric type, where $\hat{u}_z$ stands for the unit vector in the positive *z*-direction. A time variation of exp ($j\,\omega\,t$) is assumed throughout the analysis, where t is the time variable. The Hertzian vector potential $\hat{u}_z \Pi_z$ in this case satisfies the wave equation

$$\left(\nabla^2 + k_1^2\right)\Pi_{1z} = \frac{-I\,dz}{j\omega\,\varepsilon_0}\,\delta(x - x')\delta(y - y')\delta(z - z') \tag{1}$$

$$\left(\nabla^2 + k_2^2\right)\Pi_{2z} = 0 \tag{2}$$

where

$$k_1^2 = \omega^2 \mu_0 \varepsilon_0 \tag{3}$$

$$k_2^2 = \omega^2 \mu_0 \varepsilon_0 \varepsilon \tag{4}$$

and δ represents the delta function. The primed and unprimed coordinates are for the source and field points, respectively. The subscript 1 denotes the upper half space which is air, and the subscript 2 denotes the lower half space which is the imperfectly conducting earth characterized by a complex relative dielectric constant ε. The electric and the magnetic field vectors are derived from the Hertzian vector potential using

$$\vec{E}_i = \vec{\nabla}\left(\vec{\nabla} \bullet \vec{\Pi}_i\right) + k_i^2 \overrightarrow{\Pi_i} \tag{5}$$

and

$$\vec{H}_i = j\omega\varepsilon_0\varepsilon_i\left(\vec{\nabla} \times \overrightarrow{\Pi_i}\right) \tag{6}$$

respectively, with $i = 1, 2$.

In medium 1, $\varepsilon_1 = 1$, and for medium 2, $\varepsilon_2 = \varepsilon$. So that the propagation constants in mediums 1 and 2, called k_1 and k_2, respectively, are related by $\frac{k_2}{k_1} = \sqrt{\varepsilon}$. At the interface $z = 0$, the tangential electric and magnetic field components must be continuous, conditions which in terms of the Hertzian vector potential components can be written as

$$\frac{\partial \Pi_{1z}}{\partial y} = \varepsilon \frac{\partial \Pi_{2z}}{\partial y} \tag{7a}$$

$$\frac{\partial \Pi_{1z}}{\partial x} = \varepsilon \frac{\partial \Pi_{2z}}{\partial x} \tag{7b}$$

$$\frac{\partial}{\partial y}\left(\frac{\partial \Pi_{1z}}{\partial z}\right) = \frac{\partial}{\partial y}\left(\frac{\partial \Pi_{2z}}{\partial z}\right) \tag{7c}$$

$$\frac{\partial}{\partial x}\left(\frac{\partial \Pi_{1z}}{\partial z}\right) = \frac{\partial}{\partial x}\left(\frac{\partial \Pi_{2z}}{\partial z}\right) \tag{7d}$$

Since all the boundary conditions must hold at $z = 0$ for all x and y, the x and y dependence of the fields on either side of the interface must be the same. Therefore,

$$\Pi_{1z} = \varepsilon \Pi_{2z} \tag{8a}$$

$$\frac{\partial \Pi_{1z}}{\partial z} = \frac{\partial \Pi_{2z}}{\partial z} \tag{8b}$$

The complete solutions for the Hertz vector potentials satisfying the wave Eqs. 1 and 2 and the boundary conditions (8) have been derived by many researchers over the last century. An incomplete partial list (Sommerfeld 1909; Sarkar 1977; Djordjevic et al. 2002; Dyab et al. 2013; Sarkar et al. 2014) that will be

DOI 10.1007/978-981-4560-75-7_120-1

important to our discussions is provided starting with Sommerfeld. The solutions of the Hertz potentials are

$$\Pi_{1z} = P\left[\left[\frac{\exp(-jk_1R_1)}{R_1} + \int_0^\infty \frac{J_0(\lambda\rho)}{\sqrt{\lambda^2 - {k_1}^2}} \frac{\varepsilon\sqrt{\lambda^2 - {k_1}^2} - \sqrt{\lambda^2 - {k_2}^2}}{\varepsilon\sqrt{\lambda^2 - {k_1}^2} + \sqrt{\lambda^2 - {k_2}^2}} \exp\left(-\sqrt{\lambda^2 - {k_1}^2}(z+z')\right)\lambda\, d\lambda\right]\right] \tag{9}$$

and

$$\Pi_{2z} = 2P\int_0^\infty \frac{J_0(\lambda\rho)\exp\left(\sqrt{\lambda^2 - {k_2}^2}z - \sqrt{\lambda^2 - {k_1}^2}z'\right)}{\varepsilon\sqrt{\lambda^2 - {k_1}^2} + \sqrt{\lambda^2 - {k_2}^2}}\lambda\, d\lambda \tag{10}$$

for Real $\left[\sqrt{\lambda^2 - k_{1,2}^2}\right] > 0$. Here,

$$P = \frac{I\, dz}{j\omega 4\pi\varepsilon_0} \tag{11}$$

$$\rho = \sqrt{(x-x')^2 + (y-y')^2} \tag{12}$$

$$R_1 = \sqrt{\rho^2 + (z-z')^2} \tag{13}$$

and λ is the variable of integration. For Π_{1z}, the first term inside the brackets can be interpreted as the particular solution or the direct line-of-sight (LOS) contribution from the dipole antenna source, i.e., a spherical wave or a direct wave originating from the source and reaching to the observation point, and the second term can be interpreted as the complementary solution or a reflection term (reflection from the imperfect ground plane). This second term in the potential, Π_{1z}, is responsible for the fields of the *ground wave*, as per IEEE Standard Definitions of Terms for Radio Wave Propagation (Standards 1998). Observe in Eq. 9 that the second term of this potential is the strongest one near the surface of the earth and exponentially decays as one goes away from the interface.

Similarly, the solution for Π_{2z} can be interpreted as a partial transmission of the wave from medium 1 into medium 2. With these thoughts in mind, the potential Π_{1z} can be split up into two terms or equivalently the potentials responsible for the direct and the ground wave as

$$\Pi_{1z} = \Pi_{1z}^{\text{direct}} + \Pi_{1z}^{\text{reflected}} = P(g_0 + g_s) \tag{14}$$

where

$$\Pi_{1z}^{\text{direct}} = P\exp(-jk_1R_1)/R_1 = Pg_0 \tag{15}$$

$$\Pi_{1z}^{\text{reflected}} = P\int_0^\infty \left(\frac{\varepsilon\sqrt{\lambda^2 - {k_1}^2} - \sqrt{\lambda^2 - {k_2}^2}}{\varepsilon\sqrt{\lambda^2 - {k_1}^2} + \sqrt{\lambda^2 - {k_2}^2}}\right)\frac{J_0(\lambda\rho)\exp\left[-\sqrt{\lambda^2 - {k_1}^2}(z+z')\right]}{\sqrt{\lambda^2 - {k_1}^2}}\lambda\, d\lambda = Pg_s \tag{16}$$

DOI 10.1007/978-981-4560-75-7_120-1

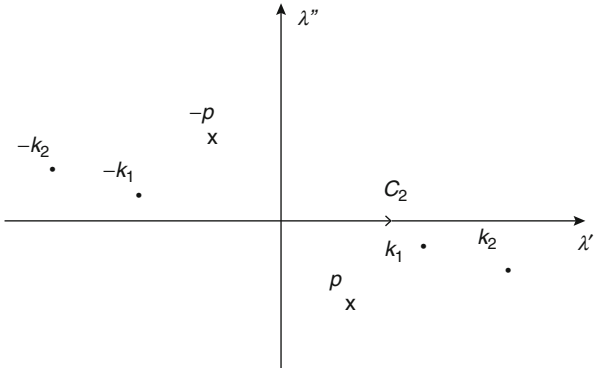


Fig. 11 The contour of integration along the real axis from 0 to ∞ in the complex λ-plane

The path of integration for the semi-infinite integral is labeled C_2 along the positive real axis and is depicted in Fig. 11 along with the singularities of the multivalued function, two branch points at k_1 and k_2 and a pole p arising from the ratio of two functions placed inside brackets in Eq. 16. A physical explanation to the two components of the Hertz potential Π_{1z} can now be given. The first one Π_{1z}^{direct} can be explained as a spherical wave originating from the source dipole. This term is easy to deal with. The difficult problem lies in the evaluation of $\Pi_{1z}^{\text{reflected}}$. Therefore, $\Pi_{1z}^{\text{reflected}}$ is interpreted as a superposition of plane waves resulting from the reflection of the various plane waves into which a spherical wave from the image point can be expanded. This arises from the identity

$$\frac{\exp(-jk_1R_2)}{R_2} = \int_0^\infty \frac{J_0(\lambda\rho)\exp\left[-\sqrt{\lambda^2 - {k_1}^2}(z+z')\right]}{\sqrt{\lambda^2 - {k_1}^2}}\lambda\, d\lambda \tag{17}$$

and

$$R_2 = \sqrt{\rho^2 + (z+z')^2} \tag{18}$$

The term under the integral sign in Eq. 17 can be recognized as a multiple plane-wave decomposition of the spherical wave source. Upon reflection of the plane waves from the dipole source as expressed in $\Pi_{1z}^{\text{reflected}}$, the amplitude of each wave must be multiplied by the reflection coefficient $\Gamma(\lambda)$. The complex reflection coefficient $\Gamma(\lambda)$ takes into account the phase change as the wave travels from the source (x', y', z') to the boundary and then to the point of observation (x, y, z). The reflection coefficient $\Gamma(\lambda)$ is then defined as the term inside the brackets in Eq. 16 as

$$\Gamma(\lambda) = \frac{\varepsilon\sqrt{\lambda^2 - {k_1}^2} - \sqrt{\lambda^2 - {k_2}^2}}{\varepsilon\sqrt{\lambda^2 - {k_1}^2} + \sqrt{\lambda^2 - {k_2}^2}} \tag{19}$$

where the semi-infinite integral over λ in $\Pi_{1z}^{\text{reflected}}$ takes into account all the possible plane waves. As $\varepsilon \to \infty$, i.e., a perfect conductor for the earth, then g_s of Eq. 16 reduces to Eq. 17 and represents a simple spherical wave originating at the image point. The reflection coefficient takes into account the effects of the ground plane in all the waves' decomposition of the spherical wave and sums it up as a ray originating from the image of the source dipole but multiplied by a specular reflection coefficient $\Gamma(\theta)$, where $\Gamma(\theta) = \frac{\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}$ and θ is illustrated in Fig. 10.

DOI 10.1007/978-981-4560-75-7_120-1

It is now important to point out that there are two forms of $\Pi_{1z}^{\text{reflected}}$ that may be used interchangeably as the two expressions are mathematically identical in nature (but have different asymptotic properties as will be illustrated).

The first one is defined as

$$\Pi_{1z}^{\text{reflected}} = P\left[\frac{\exp(-jk_1R_2)}{R_2} - 2\int_0^\infty \frac{\sqrt{\lambda^2-k_2{}^2}}{\sqrt{\lambda^2-k_1{}^2}} \frac{J_0(\lambda\rho)\exp\left[-\sqrt{\lambda^2-k_1{}^2}(z+z')\right]}{\varepsilon\sqrt{\lambda^2-k_1{}^2}+\sqrt{\lambda^2-k_2{}^2}}\lambda\, d\lambda\right] \underline{\Delta}\, P[g_1 - g_{sV}] \tag{20}$$

where g_1 represents the spherical wave originating from the image of the source and g_{sV} represents the correction factor to accurately characterize the effects of the ground.

Equivalently, one can rewrite the same expression as

$$\Pi_{1z}^{\text{reflected}} = P\left[-\frac{\exp(-jk_1R_2)}{R_2} + 2\varepsilon\int_0^\infty \frac{J_0(\lambda\rho)\exp\left[-\sqrt{\lambda^2-k_1{}^2}(z+z')\right]}{\varepsilon\sqrt{\lambda^2-k_1{}^2}+\sqrt{\lambda^2-k_2{}^2}}\lambda\, d\lambda\right] \underline{\Delta}\, P[-g_1 + G_{sV}] \tag{21}$$

Now, the image from the source has a negative sign along with the correction factor. This expansion is useful when both the transmitter and the receiver are close to the ground, since the reflection coefficient $\Gamma(\theta)$, is -1 for grazing angle of incidence where $\theta \approx \pi/2$. Then the direct term g_0 cancels the image term g_1 leaving only the correction factor G_{sv}.

In Sommerfeld's 1909 paper, he changed the limits of the integration from 0 to ∞ in Eq. 20 to an integral from $-\infty$ to $+\infty$ by transforming the Bessel function of the first kind and zeroth order to a combination of Hankel functions of the first and second kinds and zeroth order through the use of the following identity

$$J_0(x) = \frac{1}{2}\left[\mathrm{H}_0{}^{(1)}(x) + \mathrm{H}_0{}^{(2)}(x)\right] \tag{22}$$

and also utilizing

$$\mathrm{H}_0{}^{(1)}\left(xe^{j\pi}\right) = -\mathrm{H}_0{}^{(2)}(x) \tag{23}$$

where $H_0{}^{(1)}$ and $H_0{}^{(2)}$ are the Hankel functions of zeroth order and of first and second kinds, respectively. Thus, Sommerfeld closed the path integral C_2 of Fig. 11 by using the contour as shown in Fig. 12 where he also plotted the locations of the branch points k_0 and k_1, and the pole p for the term $\Pi_{1z}^{\text{reflected}}$. Also shown in Fig. 12 are the two appropriate branch cuts and a path encircling the pole.

The crux of the problem lies in the characterization of the various branch points and singularities associated with Eq. 19. The first point to observe is that the second term of the Hertz potential denoted by a complex integral and particularly $\Gamma(\lambda)$ has four branch points located at $\pm k_1$ and $\pm k_2$. Associated with these branch points are four branch cuts and this gives rise to four Riemann sheets. On the four Riemann sheets, the following conditions are satisfied:

Sheet 1: Real $(\sqrt{\lambda^2-k_1{}^2}) > 0$ and Real $(\sqrt{\lambda^2-k_2{}^2}) > 0$
Sheet 2: Real $(\sqrt{\lambda^2-k_1{}^2}) < 0$ and Real $(\sqrt{\lambda^2-k_2{}^2}) > 0$

DOI 10.1007/978-981-4560-75-7_120-1

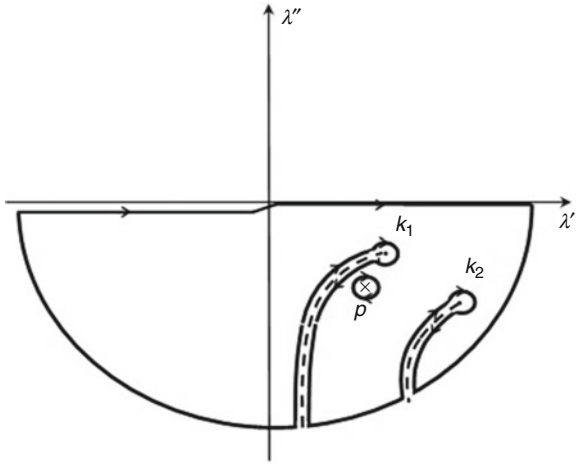


Fig. 12 Actual location of the pole in the lower complex λ-plane

Sheet 3: Real $(\sqrt{\lambda^2 - {k_1}^2}) > 0$ and Real $(\sqrt{\lambda^2 - {k_2}^2}) < 0$
Sheet 4: Real $(\sqrt{\lambda^2 - {k_1}^2}) < 0$ and Real $(\sqrt{\lambda^2 - {k_2}^2}) < 0$

Sheet 1 is the proper Riemann sheet. Now, the function described in Eq. 19 has two zeros corresponding to the zeros of the numerator, usually called Brewster zeros, and two poles corresponding to the zeros of the denominator, usually called the surface wave poles. The zeros and poles occur exactly at the same location $\lambda_S = \pm \frac{k_1 k_2}{\sqrt{k_1^2 + k_2^2}}$. However, on some Riemann sheets, they appear as poles and on other Riemann sheets as zeros. So, the four poles and zeros are distributed on the four Riemann sheets. The most confusing stuff is that on the proper Riemann sheet whether it will be a pole or zero depends on the value of the dielectric constant ε. For example, examining the denominator of $\Gamma(\lambda)$, one can observe that for real values of ε on the proper Riemann sheet, there is a zero at λ_s. Whereas for a complex value of the dielectric constant ε, that zero becomes a pole on the proper Riemann sheet. Therefore, unless one specifies the values for the dielectric constant and chooses the proper Riemann sheet, it is difficult to know whether the pole will occur or not on the proper Riemann sheet in the presentation of Eq. 19.

The zero of the reflection coefficient in Eq. 19 illustrates the Brewster's phenomenon (i.e., the wave goes into the second medium for a particular angle of incidence without reflecting any energy), and a pole for the reflection coefficient illustrates the presence of a surface wave (i.e., a wave propagating close to the interface). Also, in general, it is difficult to distinguish between a Zenneck wave and a surface wave as both decay exponentially as one moves away from the planar interface and the wave propagates with a low loss along the radial direction (Sarkar et al. 2014). In addition, it is well known that the Brewster's angle which illustrates that a wave will penetrate into the second medium without reflection is independent of frequency, whereas the surface wave phenomenon is highly dependent on frequency. As the frequency increases, the fields of the wave are more confined to the planar boundary. In addition, a surface wave does not radiate, whereas a Zenneck wave does. These points have been illustrated in Sarkar et al. (2014).

In the original Sommerfeld formulation (Sommerfeld 1909), there was no error in the sign, but the presentation by Sommerfeld was not complete. First, of all Sommerfeld demonstrated in his expression, there was a pole, and he wrote the partial solution – the contribution from the pole and associated with that some physical properties. However, as has been pointed out in Abdallah et al. (2014), Sarkar et al. (2012a, b, 2014), and Dyab et al. (2013) and illustrated in Fig. 12 that when $\varepsilon_r > 0$, the pole should lie within the circle defined by the locus whose center is at $\lambda = 0$ and of radius k_1. In addition, when the integration along

DOI 10.1007/978-981-4560-75-7_120-1

the branch cut is evaluated using the saddle point method, part of this branch cut integral actually cancels the contribution from the pole, and so in the final solution, the effect of the pole is not seen! This has been explicitly demonstrated in Collin (2004), and so the conjecture that there was an error in sign in Sommerfeld's formulation as initially intimated by Norton is a myth! The fact that the pole does not contribute to the total solution makes sense as in some cases the poles may not exist on the proper Riemann sheet, as can be demonstrated particularly for real values of the dielectric constant. Also, when the hyperbolic branch cuts in the Sommerfeld's solution are replaced by vertical branch cuts, the poles migrate into a different Riemann sheet and are not relevant (Ishimaru 1991; Baños 1966)!

The integral in Eq. 20 can be simplified by making the following substitutions (Sarkar et al. 2012a, b, 2014):

$$\lambda = k_1 \sin\beta \tag{24}$$

$$\rho = R_2 \sin\theta \tag{25}$$

$$z + z' = R_2 \cos\theta \tag{26}$$

The interpretation of the angle θ is shown in Fig. 10. Hence, the application of Eqs. 23, 24, 25, and 26 to 19 yields

$$g_{sV} \approx \int_{\Gamma_1} \left[\frac{2k_1 \sin\beta}{\pi R_2 \sin\theta}\right]^{1/2} \frac{\sqrt{\varepsilon - \sin^2\beta}}{\varepsilon\cos\beta + \sqrt{\varepsilon - \sin^2\beta}} \cdot \exp[j\{\{-\pi/4 - k_1 R_2 \cos(\beta - \theta)\}\}]d\beta \tag{27}$$

where Γ_1 is a path in the complex β plane as shown in Fig. 13. The application of Eqs. 24 to 26 to 29 yields

$$g_{sV} \approx \int_{\Gamma_1} e^{-j\pi/4} \sqrt{\frac{2k_1 \sin\beta}{\pi R_2 \sin\theta}} \frac{\sqrt{\varepsilon - \sin^2\beta}}{\varepsilon\cos\beta + \sqrt{\varepsilon - \sin^2\beta}} \exp[-jk_1 R_2 \cos(\beta - \theta)]d\beta \tag{28}$$

where Γ_1 is a path in the complex β plane as shown in Fig. 13. The path of steepest descent never crosses any of the poles.

In order to solve for the total field near the interface, a modified saddle point method as explained in Sarkar et al. (2012a, b) is applied to take into account the effect of the pole β_P near the saddle point. In the expression of both g_{sV} and G_{sV} in Eqs. 20 and 21, there is a pole β_P, resulting in

$$\frac{1}{\left[\varepsilon\cos\beta + \sqrt{\varepsilon - \sin^2\beta}\right]} = \frac{1}{\varepsilon^2 - 1} \frac{\sqrt{\varepsilon - \sin^2\beta} - \varepsilon\cos\beta}{\sin(\beta + \beta_P)\sin(\beta - \beta_P)} \tag{29}$$

where $\varepsilon\cos\beta_P + \sqrt{\varepsilon - \sin^2\beta_P} = 0$ with $\sin\beta_P = \pm\sqrt{\frac{\varepsilon}{\varepsilon+1}}$ and $\cos\beta_P = -\sqrt{\frac{1}{\varepsilon+1}}$. One obtains for G_{sV}, applying a modified saddle point integration (Sarkar et al. 2012a, b) to obtain

DOI 10.1007/978-981-4560-75-7_120-1

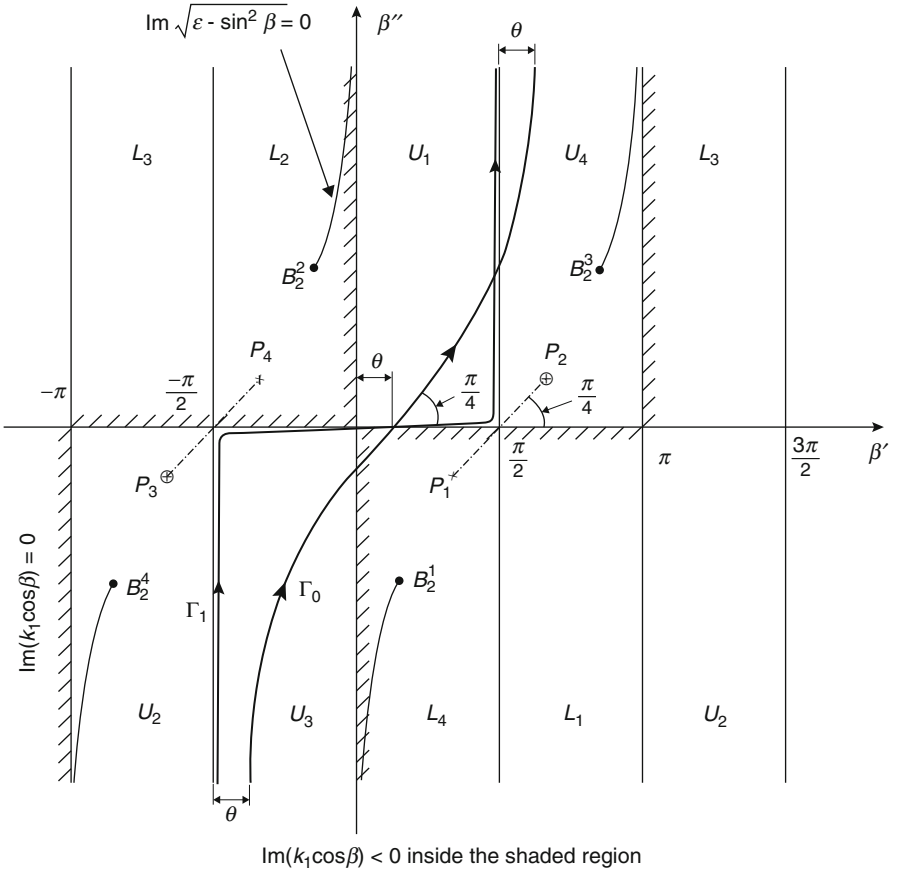


Fig. 13 The complex β plane showing possible branch points, branch cuts, poles, and the path of steepest descent for an imperfect ground plane with the material parameters $\varepsilon = \varepsilon'(1-j)^2$ and $|\varepsilon| >> 1$

$$G_{sV} = \varepsilon \exp\left(-j\tfrac{\pi}{4}\right) \int_{\Gamma_1} \left(\frac{2k_1 \sin\beta}{\pi R_2 \sin\theta}\right)^{1/2} \frac{\exp[-jk_1R_2\cos(\beta-\theta)]\cos\beta}{\varepsilon\cos\beta + \sqrt{\varepsilon - \sin^2\beta}}\, d\beta$$
$$\approx \varepsilon\sqrt{\frac{4\pi k_1 j}{R_2}}\, \frac{\cos\theta}{\cos\theta - \dfrac{1}{\sqrt{\varepsilon+1}}}\, \frac{\sqrt{\varepsilon - \sin^2\theta} - \varepsilon\cos\theta}{\varepsilon^2 - 1}\, \frac{\exp\left[-jk_1R_2 - W^2\right] erfc(jW)}{\sqrt{1 + \dfrac{\cos\theta}{\sqrt{\varepsilon+1}} + \dfrac{\sqrt{\varepsilon}\sin\theta}{\sqrt{\varepsilon+1}}}} \tag{30}$$

where

$$W^2 = -jk_1R_2\, 2\sin^2\left(\frac{\theta - \beta_P}{2}\right) = -jk_1R_2\left[1 + \frac{\cos\theta}{\sqrt{\varepsilon+1}} - \frac{\sqrt{\varepsilon}\sin\theta}{\sqrt{\varepsilon+1}}\right] \tag{31}$$

Here, W was termed by Sommerfeld (1909) the numerical distance. If $|\varepsilon| >> 1$, $\theta \approx \pi/2$, and W is very small, then one obtains $\exp\left[-W^2\right] erfc(jW) \approx 1$. Under this assumption, a simplified expression for G_{sV} can be obtained and is given by Sarkar et al. (2012a)

$$G_{sV} \approx -\sqrt{\frac{2\pi k_1 j}{R_2}} \exp[-jk_1R_2]\, \frac{(z+z')}{R_2}\, \frac{\varepsilon}{\sqrt{\varepsilon^2 - 1}} \approx -\sqrt{2\pi k_1 j}\, \frac{(z+z')\exp[-jk_1R_2]}{R_2^{1.5}}. \tag{32}$$

DOI 10.1007/978-981-4560-75-7_120-1

Equation 32 thus illustrates that when $\theta \approx \pi/2$, the dominant term of the potential $\Pi_{1z} \propto {R_2}^{-1.5}$ and therefore the leading term for the fields will also be varying approximately as $\rho^{-1.5}$, if $(z + z')$ is small compared to ρ in Eq. 32. It is interesting to observe that Eq. 32 is not a function of the ground parameters nor of the frequency of operation. So under these conditions, the path loss exponent factor should be 3 near the ground, as the field varies as $\rho^{-1.5}$. Also, the reflection coefficient method is not applicable, under those circumstances. This is the situation corresponding to cellular communication systems. These results are in accordance with the case of an imperfectly conducting ground as mentioned by Stratton (1941) since the reflection coefficient is approximately +1 for a perfectly conducting ground when the fields are observed far from the ground and it transforms to -1 when the fields are observed near the ground, i.e., for $\theta \approx \pi/2$. This particular variation of the field near an imperfect ground will be verified by an accurate numerical analysis and experimental data in the next sections.

The previous results are valid for small values of W. However, as W becomes large, then

$$\exp\left[-W^2\right] erfc(jW) \approx \frac{-j}{W\sqrt{\pi}}\left[1 + \frac{1}{2\,W^2}\right] \tag{33}$$

for

$$|W| \to \infty \tag{33}$$

and

$$|\arg\, W| < \frac{3\pi}{4} \tag{33}$$

and for $|\varepsilon| >> 1$, $W^2 \approx \frac{-j\,k_1 R_2}{2\,\varepsilon}$. Under this condition,

$$G_{sV} \approx 2\sqrt{\varepsilon}\exp[-jk_1R_2]\,\frac{(z+z')}{R_2^2}\left[1 - \frac{\varepsilon}{j\,k_1R_2}\right] \tag{34}$$

Thus, the total Hertz potential in medium 1, which is valid near the interface, for $|\varepsilon| >> 1$ and $\theta \approx \pi/2$ becomes

$$\Pi_{1z} \approx \begin{cases} P\left[\dfrac{\exp(-jk_1R_1)}{R_1} - \dfrac{\exp(-jk_1R_2)}{R_2} - \sqrt{j2\pi k_1}(z+z')\dfrac{\exp(-jk_1R_2)}{{R_2}^{1.5}}\right], & W < 1 \\ P\left[\dfrac{\exp(-jk_1R_1)}{R_1} - \dfrac{\exp(-jk_1R_2)}{R_2} + 2\sqrt{\varepsilon}(z+z')\dfrac{\exp(-jk_1R_2)}{{R_2}^2}\left[1 - \dfrac{\varepsilon}{jk_1R_2}\right]\right], & W > 1 \end{cases}$$

Next, one can observe that as one moves further away from the source (i.e., for $W > 1$), the dominant term of the field near the interface decays asymptotically as $1/R^2$ as the first two terms cancel each other. A wave with such a decay can be recognized as a Norton surface wave. It is important to note that the third term for $W > 1$, i.e., the Norton surface wave term, shows up only in the far-field region. Note also that it provides the so-called height gain for the transmitting and receiving antennas and that, consequently, this height gain only applies to the intermediate and in the far-field regions. Also, in the far field, the ground parameters do influence the result. Finally, it can also be seen that there is a higher-order term which decays as $1/R^3$, which also shows the height gain and the influence of the ground parameters.

As mentioned before, in the intermediate region, i.e., the region before the far-field region, the fields decay as approximately $\rho^{-1.5}$ (when the height of the antenna is small compared to the distance from the antenna). Also, for $W < 1$, the above expression is independent of the ground parameters and of the

DOI 10.1007/978-981-4560-75-7_120-1

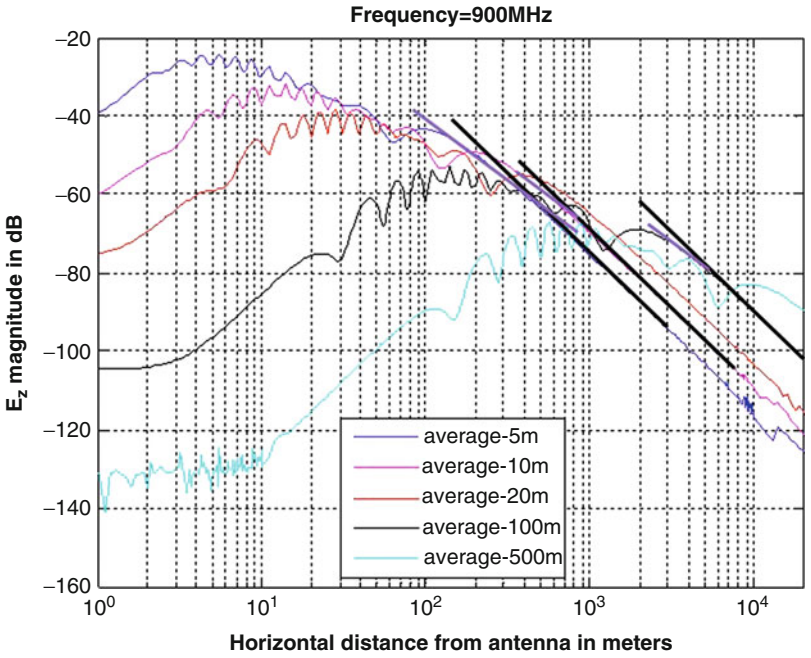


Fig. 14 Magnitude of the z-component of the electric field in dB radiated from a half-wave dipole antenna over an average imperfect ground $\varepsilon_r = 15$ and $\sigma = 0.005$ mhos/m, different curves belong to different heights above the imperfect ground, and the frequency of operation is 900 MHz

frequency of operation. This is the reason why for all the measurements carried out for a cellular wireless system the propagation path loss is 30 dB per decade as illustrated in section "Description and Analysis of Measured Data," irrespective of whether the system is located in an urban, suburban, or rural area, and this holds even when the transmit/receive system is located over water.

Next, the exact Green's function in Eq. 14 is calculated accurately using a computational electromagnetic code evaluating the semi-infinite integral in Eq. 9 in an essentially exact way. This is carried out by using the commercially available code AWAS (Djordjevic et al. 2002). As an example, consider a vertically oriented half-wave dipole located at different heights over an average ground with parameters for the permittivity $\varepsilon_r = 15$ and $\sigma = 0.005$ mhos/m, where the ground parameters were taken from De et al. (2010). The frequency of operation was 900 MHz. Figure 14 plots the field strength as a function of the radial distance for different heights of the antenna above the ground. In this figure, it is seen that near the transmitting antenna, there is the interference between the direct space wave and the field from the image produced by the imperfect ground, providing variation of the total field strength. This is often labeled as fading. This interference pattern stops at an approximate distance of $\frac{4H_{TR}H_{RX}}{\mathcal{W}^o}$, and a monotonic decay of the fields occurs with a slope of 30 dB per decade and continues approximately to a distance of $\frac{8H_{TR}H_{RX}}{\mathcal{W}^o}$ from where the far field of the antenna starts and the slope becomes roughly 40 dB per decade (De et al. 2010). In Fig. 14, the slope of 30 dB per decade is marked by a thick purple straight line, and the slope of 40 dB per decade is marked by a thick black straight line. Observations seem to verify this prediction carried out by the electromagnetic macro model (De et al. 2010). There is a height gain in the far field of the antenna, but in the near field which is of importance in cellular communication, there is actually a height loss. Hence, it is proposed that a better solution will then be to deploy the transmitting antenna closer to the ground. In that case, the region of the variation in the field strength would be quite small, and the field strength will decay monotonically inside the remainder of the cell minimizing fading. Since there will not be any interference pattern and thus no slow fading, then it is possible to reduce the transmitting power at least by a factor of 10 (say) providing a better safe and cheaper system as in most

DOI 10.1007/978-981-4560-75-7_120-1

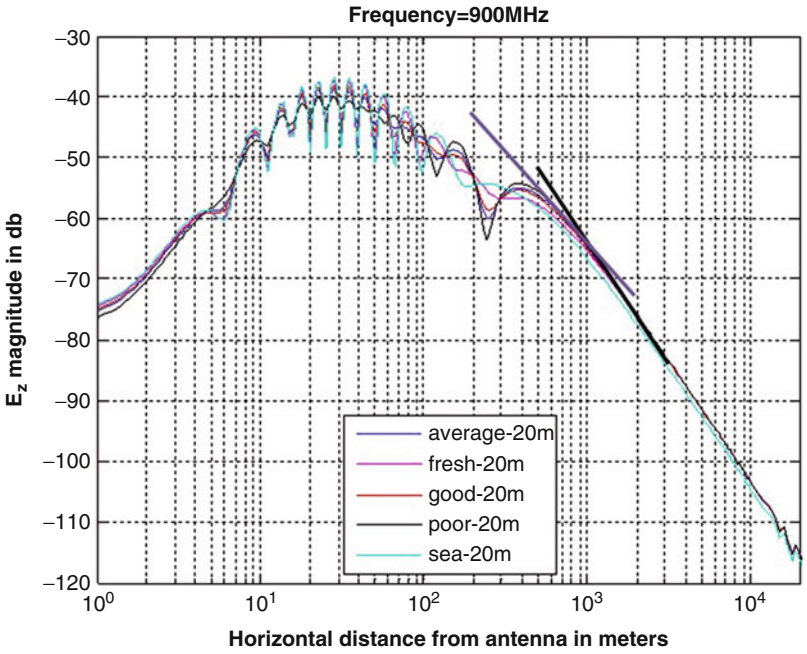


Fig. 15 Magnitude of the z-component of the electric field in dB radiated from a half-wave dipole antenna over different imperfect grounds, height of the antenna is 20 m, different curves belong to different imperfect grounds, and the frequency of operation is 900 MHz

cases the tower costs more than the antenna system. However, one could deploy more base stations as the transmitting power for each antenna is significantly reduced. An interesting scenario of this can be seen perhaps in some South American cities where Wi-Fi is delivered to individual houses by deploying base station antennas on every lamp post. Such a discussion is quite relevant as there is a second channel from the mobile to the base station in which there is no height gain as the mobile is near the ground and the mobile is transmitting a fraction of the power of the base station. Moreover, the antenna on the mobile can be oriented in any direction, and hence, one should question the validity of the state-of-the-art rules of thumb developed for deploying base station antennas as they do not consider the environment of the second mobile channel which is quite important and which does not satisfy any of the rules of thumb associated with the deployment of base stations!

Figure 15 shows another simulation result for a vertical half-wave dipole located 20 m above different types of imperfect grounds. The frequency of operation was 900 MHz. The ground parameters chosen were for the poor ground with a dielectric permittivity and a conductivity of $\varepsilon_r = 4$ and $\sigma = 0.001$ mhos/m, for the average ground, $\varepsilon_r = 15$ and $\sigma = 0.005$ mhos/m, for the good ground $\varepsilon_r = 25$ and $\sigma = 0.02$ mhos/m, for the sea water $\varepsilon_r = 81$ and $\sigma = 5$ mhos/m, and for the fresh water $\varepsilon_r = 81$ and $\sigma = 0.01$ mhos/m. The ground parameters were taken from De et al. (2010). By observing the plots in Fig. 15, it is seen that within the cell, the electrical properties of the ground have very little effects as for a fixed height of the transmitting antenna radiating over an imperfect ground, the ground parameters do not change the nature of the distant fields, whereas near the antenna, the shape of the interference pattern can be slightly different. In Fig. 15, the slope of 30 dB per decade is marked by a thick purple straight line, and the slope of 40 dB per decade is marked by a thick black straight line. This implies that whether we consider propagation in urban, suburban, industrial, rural, or over water, the results for the fields should not differ too much from each other, and furthermore, an electromagnetic macro model can accurately make such predictions!

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_120-1

Experimental Validation of the Theoretical Analysis of Sommerfeld Using Okumura et al.'s Measurement Data

The analysis presented is now used for the analysis of the propagation data measured by Okumura et al. (1968) in their classic propagation measurements in the city of Tokyo. Okumura et al. placed a transmitting antenna at different heights. The specific height that will be chosen for the simulation for comparison with experimental data is 140 m. The signal is received by another vertically polarized antenna located on top of a van 3 m above the ground. The receiving antenna had a gain of 1.5 dB. The transmitting antenna was a 5-element Yagi having a gain of approximately 11 dB and radiating 150 W of power. The van was then driven in the city of Tokyo from 1 km to 100 km from the transmitting antenna. Here, the measurements are carried out at 453 MHz. Since the 5-element Yagi was an antenna composed of wires, in the current simulations, an optimized 5-element Yagi antenna array which had a gain of 11 dB was used. In this analysis, the Yagi antenna was synthesized and was used in the computations. First, the integral equation using the Green's function was used to solve for the current distribution on the transmitting antenna, and then these currents were used to compute the radiating fields (Sarkar 1977). The radiating fields are calculated using the new Green's function (Schelkunoff formulation) (Dyab et al. 2013) instead of using the classical Sommerfeld Green's function in a code that has been already developed for the analysis of radiation over imperfect ground planes based on Sommerfeld integrals. An example for such a code is AWAS (Djordjevic et al. 2002). The parameters for the urban ground were relative permittivity of $\varepsilon_r = 4$ and $\sigma = 2 \times 10^{-4}$ mhos/m (De et al. 2010). However, since one did not know how Okumura et al. matched their antennas and how it was exactly fed, and with not knowing how much power was being radiated by their antenna, first, a simple 1 V excitation was applied to the Yagi antenna for the simulation. Then all of the computed filed points were scaled up simultaneously by adding a constant value of decibels (98 dB) to the computed results, so that the two different plots (the simulated and the measurement) exactly matched at one point at 7 km as shown in Fig. 16a. Then the various plots are overlaid without any further massaging of the data and all the plots of the theoretical predictions by both the new Schelkunoff formulation (Sarkar et al. 2014) and the classical Sommerfeld formulation (Djordjevic et al. 2002), all scaled up by the value of 98 dB, and plotted along with Okumura et al.'s experimental data. The plots show remarkable similarity. It is interesting to note that the simulation using the Schelkunoff Green's function provided a more stable qualitative plot of the field from 1 km to 100 km from the base station antenna. The results were computed for this analysis by using the commercially available computer code called *analysis of wire antennas and scatterers* (AWAS) (Djordjevic et al. 2002) using the Sommerfeld formulation and the new Schelkunoff formulation (Sarkar et al. 2014). In summary, the three plots in Fig. 16 compare between the experimental (Okumura et al. 1968) and theoretical predictions (Schelkunoff and Sommerfeld formulations) computed through a macro model for predicting propagation path loss in an urban environment at (a) 453 MHz, (b) 922 MHz, and (c) 1920 MHz. In AWAS, the results for the fields became somewhat unstable when the horizontal distance from the transmitting antenna becomes quite large, say greater than 10 km. This is due to the Sommerfeld integral tail problem which is totally eliminated in the new formulation (Dyab et al. 2013). It is also important to point out that for both the theoretical and experimental data, the slope for the path loss exponent between 1 km and 10 km is about 30 dB per decade, which was expected from the theoretical analysis using the saddle point method in Sarkar et al. (2012a, b). The slope between 10 km and 100 km is 40 dB per decade, as predicted by the Norton ground wave for the far field. This illustrates that an accurate electromagnetic macro modeling of the environment is sufficient to predict the path loss as evidenced by the comparison between theory and experiment. To compare with the experimental data, a power adjustment was made to the theoretical graphs to compensate for the different input power levels. In Fig. 16b, 130 dB is added to the theoretically computed data, and in Fig. 16c, 125 dB is added to the theoretically computed data to

DOI 10.1007/978-981-4560-75-7_120-1

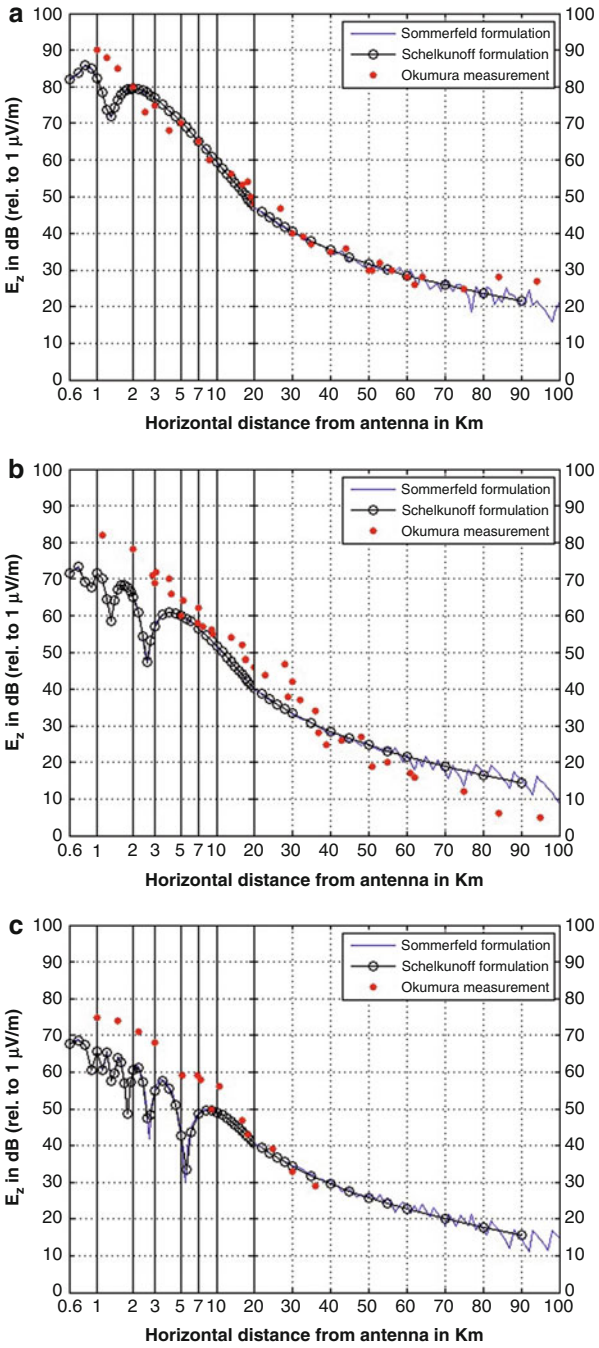


Fig. 16 Comparison between the experimental (Okumura et al. 1968) and theoretical predictions (Schelkunoff and Sommerfeld formulations) computed through a macro model for predicting propagation path loss in an urban environment at (**a**) 453 MHz, (**b**) 922 MHz, and (**c**) 1920 MHz

compare with the experimental data. It is important to note that those numbers include the adjustment of the reference field from 1 V/m to 1 μV/m. For the last two frequencies, the plots in Fig. 16b and c are not in perfect match with the experiment because in the simulations, at these two frequencies, a simple half-wave dipole antenna was used as the transmitter, whereas Okumura et al. (1968) used a highly directive antenna like a parabolic reflector whose dimensions were not reported in their paper.

DOI 10.1007/978-981-4560-75-7_120-1

Fig. 17 Elongated image of the sun over a rippled lake (http://fineartamerica.com/featured/alexandria-bay-sunset-steve-ohlsen.html. Accessed 25 Mar 2015)

Fig. 18 Waves on a wet ground due to the partial reflectivity of the surface. The figure gives an impression on the physical propagation mechanism to be expected in cellular environments at frequencies where a wet earth represents a complex impedance surface

DOI 10.1007/978-981-4560-75-7_120-1

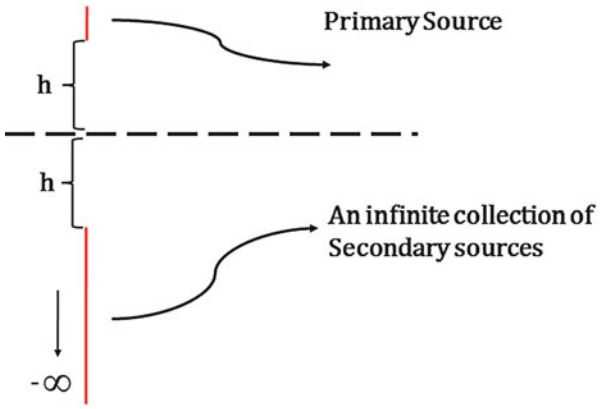


Fig. 19 Equivalent image sources that generate the desired radiation fields at the interface

Next, the nature of the mechanism of radio wave propagation over a single boundary layer generated by two dissimilar mediums on each side is illustrated.

Visualization of the Propagation Mechanism

How does the signal actually propagate over a single boundary generated by two different mediums can be visualized by inspecting the path of each of the waves represented by the integrands in Eq. 9, for example, especially those waves represented by the main part of the contour going from 0 to k_1. Those waves when they hit the ground, which is imperfectly conducting, they suffer from different reflection coefficients which depend on their different angles of incidence. On reflection, the rays diverge forming a semi-infinite image. The rays will never converge to form a perfect image of the same size as the source except in case that the second medium is perfectly conducting and perfectly smooth. This fact was predicted in 1935 by Van der Pol (1935) which explains the elongated image of the sun or moon on a clear day on a wavy lake.

Van der Pol (1935) points out that the results of Sommerfeld (1926) and Weyl (1919) were, as a rule, not very transparent on account of the fact that the approximations or developments used were more of a mathematical nature. This leads Van der Pol to recast the same problem without any approximation whatsoever which leads to a solution in the form of a simple space integral which allows a direct physical interpretation. The integration related to the source of the fields from a transmitting dipole located over an imperfect ground extends over the part of space occupied by the second medium below the geometrical image of the source. This scenario is seen in Fig. 19, which is similar to Fig. 10 in many respects. It is shown that the field in the first medium where the dipole is located, apart from the direct radiation from an elementary dipole, can be described as due to a secondary wave originating in the integration space extending from its image from $-h$ to $-\infty$. In other words, the image of a source over an imperfect ground consists of a line source starting from its image and continuing to $-\infty$ as shown in Fig. 19. The amplitude of the wave from the image is determined by the amplitude of the primary wave. It can be considered to spread from the geometrical image of the point source with the propagation constant and absorption of the second medium. The higher the conductivity of this second medium, the more the primary wave is concentrated near the image of the point source until, for infinite conductivity, it is wholly concentrated at the image itself.

This result that the fields produced by a dipole over an imperfect ground plane are due to the direct contribution from the source plus the effect of the ground represented by an equivalent image which consists of a line source that extends from $-h$ to $-\infty$ has been used by other researchers. In fact, the same

DOI 10.1007/978-981-4560-75-7_120-1

physical picture is available in Booker and Clemmow (Booker and Clemmow 1950). Sommerfeld also writes a similar expression (Sommerfeld 1909, Eq. 10d on p. 250). The detailed derivation is available in Van der Pol's work (1935), and one obtains for the Hertz potential due to a normalized point source located at a height z_a, while the observation point is located at z_b, the following expression:

$$\Pi_{1z} \approx \frac{\exp(-jk_1R_1)}{R_1} + \frac{\exp(-jk_1R_2)}{R_2} - \frac{2jk_1^2}{k_2}\exp[-a(z_a+z_b)] \int_{\zeta=z_a+z_b}^{\zeta=z_a+z_b+\varepsilon\times\infty} \frac{\exp\left(-jk_1\sqrt{\rho^2+\zeta^2}\right)}{\sqrt{\rho^2+\zeta^2}}\exp(a\zeta)\,d\zeta \tag{35}$$

where

$$a = -\frac{jk_1^2}{k_2}$$

and

$$\zeta = z_a + z_b + \varepsilon z$$

As stated by Van der Pol, the third term in Eq. 35 can be interpreted as a wave spreading from the geometrical image of the point source while all points of the second medium below the level of the geometrical image, apparently, send secondary waves to the observer. When $\varepsilon \gg 1$, the distance from an arbitrary point in the second medium to the observer, as given in the exponent of the third term, contains z multiplied by ε, i.e., to the observer, the vertical part of the distance below the image of the point source is multiplied by ε. The wave originating from the image will therefore be observed as elongated vertically, like the vertically very elongated image of the sun over a wind-rippled lake, as shown in Fig. 17 (Van der Pol 1935; Sarkar et al. 2014).

Next, consider a rainy night, when the ground is wet representing a partially reflective surface, and then try to observe how the city lights are stretched along the ground by waves propagating over the surface. Compare such a view as shown in Fig. 18 to the situation as described by Van der Pol in Eq. 35. Imagine that the existence of the green traffic light is the information to be transmitted by the source, and now notice how a wet ground helps to transfer the information to a point where there is no line-of-sight path. In fact, in cellular systems, the ground is always an imperfect reflecting surface, and the way light propagates in Fig. 19 is what we should expect as a propagation mechanism in a cellular system. In other words, the transmitting antennas (usually down tilted to the ground) excite a radiating field in the cell, which in the intermediate region behaves as a radiating field from a line source and in the far-field region behaves as a Norton surface wave. This wave then represents one of the main dominant means by which the base station antenna communicates with the mobile device. This mechanism of propagation is shown in Figs. 17, 18, and 19. Hence, the stretching of the lights in Figs. 17 and 18 essentially represents the image of the primary source.

Figures 17, 18, and 19 provide a physical insight into how waves propagate in mobile communications systems and thus illustrates why smart antennas and beamforming are not very successful in cellular communications till now despite all of the research efforts done in those fields. If the real scenario of propagation is something similar to what is seen in Figs. 17, 18, and 19, then one should change their outlook at the implementation of multiple antennas and adaptive arrays in cellular systems. The important point is that the effect of the ground plays a dominant role. This equivalent line source formed from the images generates a field in the air which decays as 30 dB per decade. The decay of the fields as 30 dB per

decade with distance is generated in the intermediate region approximately where the slow fading region ends which is approximately at a distance of about $\frac{4H_{TX}H_{RX}}{\mathscr{W}^{o}}$ (De et al. 2010) from the base station antenna. Here, H_{TX} and H_{RX} represent the height of the transmitting and the receiving antennas over the ground, and $\mathscr{W}^{o}$ λ is the wavelength of operation.

A Note on the Conventional Propagation Models

Typically, in the current state of the art, the drive test data are fit in a least-squares fashion to a statistical model or an empirical-based ray-tracing scenario is often used. The fact of the matter is that current propagation modeling tools cannot capture the fundamental physics, namely, inside a cell, the path loss is typically 30 dB per decade (Abdallah et al. 2014) or equivalently the electric field strength varies as $\rho^{-1.5}$ with the distance ρ from the antenna inside the cell, and outside the cell, the propagation path loss is approximately 40 dB per decade or equivalently the electric field strength varies as ρ^{-2} as a consequence of the addition of the direct wave from the source and the infinite image that has been discussed. Thus, statistical-based models do not capture the basic physics.

Conventionally, in wireless communications textbooks, such as Goldsmith (2005), this problem of propagation modeling is tackled first by explaining the two-ray model over a flat perfectly conducting earth. In the two-ray model – the incident and the reflected rays – the reflection coefficient is taken to be -1 and θ (the angle of incidence) is always $\frac{\pi}{2} = 90^{\circ}$ (i.e., perfect reflecting earth is considered). However, the two-ray model illustrates the intermediate path loss to be 20 dB per decade since the field decays ρ^{-1} for free-space propagation. This is in contrast to the path loss of 30 dB per decade as has been presented in this chapter. After explaining the two-ray model, empirical models are usually presented, such as the well-known Okumura-Hata model (Fujimoto 2008). Although empirical models (Ji et al. 2001, 2002) have been extensively applied with good results, they suffer from some disadvantages. The main disadvantage is that empirical models provide no physical insight into the mechanism by which propagation occurs. In addition, these models are limited to the specific environments and parameters used in the measurements.

Next, a brief description of the Hata model (an example for the empirical models) is provided and that of the two-ray model. Then one can make a comparison between the proposed macro model described in this chapter and these two empirical models.

Hata model is an empirical formulation of the graphical path loss data provided by Okumura's model. The formula for the median path loss in urban areas is given by

$$L(\text{urban})(\text{dB}) \;=\; 69.55 + 26.16\log f_c - 13.82\log h_{te} - a(h_{re}) + (44.9 - 6.55\log h_{te})\log d \qquad (36)$$

where f_c is the frequency in MHz and varies from 150 to 1500 MHz, h_{te} and h_{re} are the effective heights of the base station and the mobile antennas (in meters), respectively, d is the distance from the base station to the mobile antenna in kilometers, and $a(h_{re})$ is the correction factor for the effective antenna height of the mobile which is a function of the size of the area of coverage. For small- to medium-sized cities, the mobile antenna correction factor is given by

$$a(h_{re}) = (1.1\log f_c - 0.7)h_{re} - (1.56\log f_c - 0.8)\ \text{dB} \qquad (37)$$

For a large city, it is given by

DOI 10.1007/978-981-4560-75-7_120-1

$$a(h_{re}) = 8.29(\log 1.54h_{re})^2 - 1.1 \text{ dB for } f_c \leq 300\,\text{MHz} \tag{38}$$

$$a(h_{re}) = 3.2(\log 11.75h_{re})^2 - 4.97\,\text{dB for } f_c \geq 300\,\text{MHz} \tag{39}$$

To obtain the path loss in a suburban area, the standard Hata formula is modified as

$$L(\text{dB}) = L(\text{urban}) - 2\left[\log(f_c/28)\right]^2 - 5.4 \tag{40}$$

The path loss in open rural areas is expressed through

$$L(\text{dB}) = L(\text{urban}) - 4.78(\log f_c)^2 - 18.33\log f_c - 35.94 \tag{41}$$

If one is interested in calculating the received power instead of the path loss, then one can use the following formula to calculate the received power in dBm knowing the path loss L(dB), the transmitted power P_t (dBm), and the gain G_t (dB) through

$$P_r(d) = P_t + G_t - L \tag{42}$$

This model is quite suitable for large cell mobile systems, but not for personal communications systems, which cover a circular area of approximately a few kilometers in radius (Goldsmith 2005; Ji et al. 2001, 2002; Sarkar et al. 2003).

The two-ray model is based on ray-tracing techniques. In the two-ray model, one only considers the direct ray and the ray reflected off the plane of the earth. In the model, the reflection coefficient is considered to be –1 which is that of a perfect conductor. The formula for the two-ray model is given by

$$P_r(d) = P_t G_t \left[\frac{\lambda}{4\pi}\right]^2 \left| \frac{1}{\sqrt{(h_t - h_r)^2 + d^2}} - \frac{e^{-j2\pi\left(\sqrt{(h_t+h_r)^2+d^2} - \sqrt{(h_t-h_r)^2+d^2}\right)}}{\sqrt{(h_t + h_r)^2 + d^2}} \right|^2 \tag{43}$$

where P_t is the transmitted power, G_t is the transmitting antenna gain, λ is the wavelength, h_t is the transmitting antenna height in meters, h_r is the receiver antenna height in meters, and d is the horizontal distance from the transmitting antenna in meters.

Figure 20 plots the results for the electromagnetic macro model generated by AWAS for a vertical half-wave dipole over an urban ground at a height of 20 m above the ground assuming that the receiving antenna is located at 1.5 m above the ground. The frequency is 900 MHz. In Fig. 20, in addition to the results generated by the electromagnetic macro model, the results for the two-ray model and that from the Hata model for the same transmitting and receiving antenna locations are also presented. One can clearly see that neither the two-ray model nor the Hata model predicts a decay of 30 dB per decade. Furthermore, the nulls in the two-ray model are much sharper than the AWAS results, and this is due to the assumption of a perfect ground for the earth used in the two-ray model. In fact, if one assumes a vertical half-wave dipole over a perfect ground using the electromagnetic analysis code AWAS, one will obtain a slope of 20 dB per decade in the path loss of the far field as the two-ray model suggests and not the usual 40 dB per decade which physics predicts! Again, this is because in the two-ray model, one assumes that the reflection coefficient is –1, while the truth is that the reflection coefficient is +1 for a vertical dipole radiating on top of a perfect ground and it is –1 for a horizontal dipole operating over a perfect ground. In

DOI 10.1007/978-981-4560-75-7_120-1

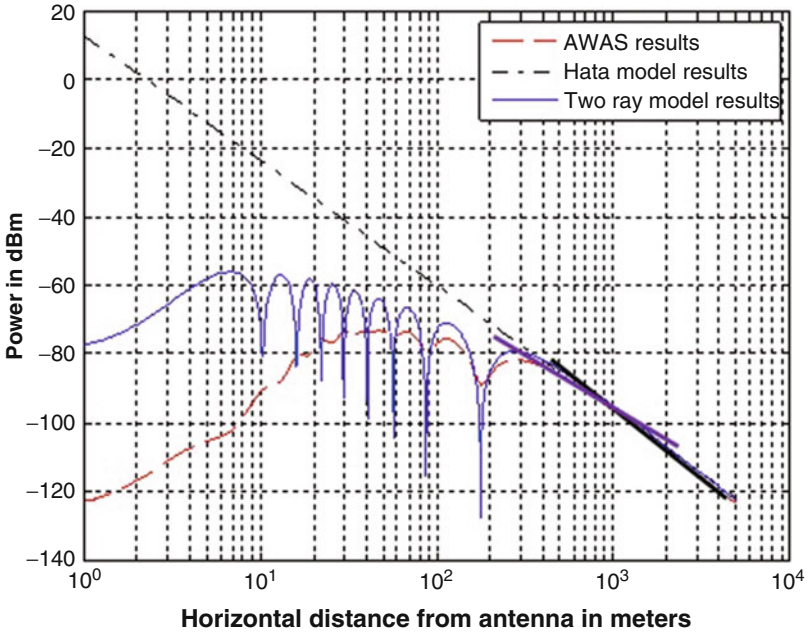


Fig. 20 The *dashed red curve* is the received power of a vertical half-wave dipole at 20 m above an urban ground using AWAS. The *dashed black curve* is the results from the Hata model for the same setup as in the AWAS simulation. The *blue curve* is the results for the two-ray model for the same setup as in the AWAS simulation

Fig. 20, the purple thick line represents the slope of 30 dB per decade and the black thick line represents the slope of 40 dB per decade.

So, to find more satisfactory models, researchers usually follow one of the two paths. Either they choose more sophisticated physical models which encounter other propagation mechanisms such as diffraction, scattering, and ray tracing (Fujimoto 2008; Ji et al. 2001, 2002; Sarkar et al. 2003) or they delve into statistical modeling (Goldsmith 2005). It is quite interesting to read the following conclusion from Fujimoto (2008):

> Although the plane earth model has a path loss exponent close to that observed in actual measurements (i.e., 4), the simple physical situation it describes is rarely applicable in practice. The mobile is always almost operated (at least in macro cells) in situations where it does not have a line-of-sight path to either the base station or to the ground reflection point, so the two-ray situation on which the plane earth model relies is hardly ever applicable. To find a more satisfactory physical propagation model, we examine diffraction as a potential mechanism.

Instead of examining diffraction as a potential mechanism or going to statistical modeling, the work presented in this chapter provides a rigorous mathematical solution based on the exact Sommerfeld formulation of the two-ray model but with imperfectly reflecting earth taken into consideration. This approach directly implies that the physical model of propagation in the cellular environment described above is the radiation field associated with a line source. Namely, the power decreases with the distance from the transmitter by 30 dB per decade for most of the practical area within a typical cell, as the antennas are located quite high from the ground. Then the fading and shadowing due to buildings and large obstacles appear as variations around the 30 dB per decade slope line, as seen in the measured data in the plots of Figs. 5–7. Almost none of the physical models take fields produced by a line source into consideration such as the dielectric canyon model, flat edge model, and sophisticated ray-tracing models (Fujimoto 2008; Ji et al. 2001, 2002).

In summary, the ray theory can never predict a decay of the fields as $\rho^{-1.5}$. The analysis of Van der Pol (1935) presented here provides a physical picture of how the wave, which is not a surface wave,

DOI 10.1007/978-981-4560-75-7_120-1

propagates over an imperfect ground from the base station antenna to the mobile device. In short, besides the direct ray from the source, there are fields from a line source generated by the image of the original source over the imperfectly conducting ground which can be easily visualized in Figs. 17 and 18.

Conclusion

The objective of this chapter has been to illustrate that an electromagnetic macro model can accurately predict the propagation path loss in a cellular wireless environment as the buildings, trees, and so on are the secondary source of the propagation path loss, the primary being the mechanism of electromagnetic wave propagation over an imperfect ground. Documentary evidence using both theory and experimental data from multiple sources has been provided to illustrate these subtle points.

In addition, it is shown that experimental data collected from various sources illustrate that the macro model is quite accurate and provide a basic physical understanding of the propagation mechanism. Some of the salient features are as follows: The nature of the ground like urban, rural, suburban, or water has essentially the same effect in a cellular environment. The larger height of the base station antenna is not a good choice, and if it is brought closer to the ground, it may provide a better and cheaper solution including less usage of power.

Finally, an optical analog is provided for the propagation mechanism to illustrate that the image of an antenna radiating over an imperfect ground can be characterized by an image which is essentially an infinite line source which extends from the same height at which the antenna is placed over the ground, but now it is located below the ground and extending to negative infinity. This line source is responsible for predicting the proper propagation path loss. Such a mechanism cannot be explained either through a statistical model, through a multiple ray model, or through diffraction as none of these modes of analysis deal with the imperfect ground representing earth which is the main source influencing the radio wave propagation in a cellular wireless environment.

References

Abdallah MN, Dyab W, Sarkar TK, Prasad MVSN, Misra CS, Lampérez AG, Salazar-Palma M, Ting SW (2014) Further validation of an electromagnetic macro model for analysis of propagation path loss in cellular networks using measured driving-test data. IEEE Antennas Propagat Mag 56:108–129

Baños A (1966) Dipole radiation in the presence of a conducting half-space. Pergamon Press, Oxford

Booker HG, Clemmow PC (1950) A relation between the Sommerfeld theory of radio propagation over a flat earth and the theory of diffraction at a straight edge. Proc IEE 97:18–27

Burrows CR (1937) The surface wave in radio propagation over plane earth. Proc IRE 25:219–229

Collin RE (2004) Hertzian dipole radiating over a lossy earth or sea: some early and late 20th-century controversies. IEEE Antennas Propagat Mag 46:64–79

De A, Sarkar TK, Salazar-Palma M (2010) Characterization of the far field environment of antennas located over a ground plane and implications for cellular communication systems. IEEE Antennas Propagat Mag 52:19–40

Djordjevic AR, Bazdar MB, Sarkar TK, Harrington RF (2002) AWAS version 2.0: analysis of wire antennas and scatterers, software and user's manual. Artech House, Norwood

Dyab WM, Sarkar TK, Salazar-Palma M (2013) A physics-based Green's function for analysis of vertical electric dipole radiation over an imperfect ground plane. IEEE Trans Antennas Propagat 61:4148–4157

Fujimoto K (2008) Mobile antenna systems handbook, 3rd edn. Artech House, Norwood

Gabor D (1953) Communication theory and physics. IRE Trans Informat Theory 1:48–59
Goldsmith A (2005) Wireless communications. Cambridge University Press, Cambridge
IEEE Standard Definitions of Terms for Radio Wave Propagation (1998) http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=705931&userType=inst&tag=1. Accessed Mar 2015
Ishimaru A (1991) Electromagnetic wave propagation, radiation, and scattering. Englewood Cliffs, Prentice Hall. Chapter 15 and Appendix to Chapter 15
Ji Z, Li BH, Wang HX, Chen HY, Sarkar TK (2001) Efficient ray-tracing methods for propagation prediction for indoor wireless. IEEE Antennas Propagat Mag 43:41–49
Ji Z, Sarkar TK, Li BH (2002) Methods for optimizing the location of base stations for indoor wireless communications. IEEE Trans Antennas Propagat 50:1481–1483
Lazarus M (2010) The great spectrum famine. IEEE Spectr 47:26–31
McMillan RW (2006) Terahertz Imaging, Millimeter-Wave Radar. Advances in Sensing with Security Applications, Springer Netherlands, 2:243–268
Okumura T, Ohmori E, Kawano T, Fukuda K (1968) Field strength and its variability in VHF and UHF land mobile service. Rev Elect Commun Lab 16:825–873
Sarkar TK (1977) Analysis of arbitrarily oriented thin wire antennas over a plane imperfect ground. AEÜ 31:449–457
Sarkar TK, Ji Z, Kim K, Medouri A, Salazar Palma M (2003) A survey of various propagation models for wireless communication. IEEE Antennas Propagat Mag 45:51–82
Sarkar TK, Dyab W, Abdallah MN, Salazar-Palma M, Prasad MVSN, Barbin S, Ting SW (2012a) Physics of propagation in a cellular wireless communication environment. Radio SciBull 343:5–21. http://www.ursi.org/files/RSBissues/RSB_343_2012_12.pdf. Accessed 1 Mar 2015
Sarkar TK, Dyab W, Abdallah MN, Salazar-Palma M, Prasad MVSN, Ting SW, Barbin S (2012b) Electromagnetic macro modeling of propagation in mobile wireless communication: theory and experiment. IEEE Antennas Propagat Mag 54:17–43
Sarkar TK, Dyab WM, Abdallah MN, Salazar-Palma M, Prasad MVSN, Ting SW (2014) Application of the Schelkunoff formulation to the Sommerfeld problem of a vertical electric dipole radiating over an imperfect ground. IEEE Trans Antennas Propagat 62:4162–4170
Sommerfeld AN (1909) Propagation of waves in wireless telegraphy. Ann Phys 28:665–736
Sommerfeld AN (1926) Propagation of waves in wireless telegraphy. Ann Phys 81:1135–1153
Stratton JA (1941) Electromagnetic theory. McGraw-Hill Book, New York
Van der Pol B (1935) Theory of the reflection of light from a point source by a finitely conducting flat mirror with application to radiotelegraphy. Physics 2:843–853
Weyl H (1919) Propagation of electromagnetic waves over a plane conductor. Ann Phys 60:481–500
www.prometheus-us.com/asi/sensors2005/papers/mcmillan.pdf. Accessed 21 Mar 2015

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_121-1

Antennas in MRI Systems

Sven Junge*
Bruker BioSpin MRI GmbH, Ettlingen, Germany

Abstract

Today the Nuclear Magnetic Resonance (NMR) and Magnetic Resonance Imaging are very well-established methods for noninvasive investigations of live objects, substances and materials. The radio frequency (RF) coil is the first component where the Magnetic Resonance (MR) signal is stimulated and received and therefore is one of the most important components of a Magnetic Resonance Imaging (MRI) system. The design of properly developed RF coils is the key to achieve the best clinical, preclinical or experimental result for MRI scientists or clinicians. In this chapter a detailed overview on RF-coil concepts for MRI is presented. This article contains some results of my personal work and the study of many articles, reference textbooks, and other people's work. The intention of this chapter is to give the reader a rough summary of the state-of-the-art knowledge and physical background of MRI RF coils engineering in an easy-to-understand format. This chapter contains three main sections. The first section is a simple introduction to MRI and will provide a basic understanding of MR physics behind the detection of RF signals for MRI for students who do not have any knowledge of MRI or NMR. The RF engineer must be familiar with these basic principles in order to design and build successfully a MRI RF coil. The next section is related to the basic types of RF coils. It is divided in multiple subsections covering volume coils and their basic design principles, local RF coils including their arrangement for so-called array RF coils, and last but not least cryogenically cooled RF coils. The last section discusses further directions and challenges in the field of MRI RF coil engineering and gives a brief description of active shaping of the RF field within a given volume and its challenges. While this section gives only a rough overview about the topics of MRI RF coil engineering, the interested reader will find the most interesting textbooks and articles in the reference section.

Keywords

MRI; RF coils; Birdcage; Array

Introduction

Magnetic Resonance Imaging – MRI

Today the Nuclear Magnetic Resonance (NMR) is a very well-established method for noninvasive investigation of substances and materials and plays a major role in the medical and preclinical community. NMR is based on nuclear magnetic resonance of the nucleus inside an external magnetic field $B_0 = |\boldsymbol{B}_0|$. For this discovery Bloch and Purcell (Bloch 1946; Purcell et al. 1946) received the Nobel Prize in 1952. With the introduction of an additional temporary magnet field gradient Lauterbur, Mansfield, and Damadian (Lauterbur 1973) generated the first spatial images in 1973 based on Fourier transformation of the nuclear magnetic resonance signals. From that time (Mansfield 1996; Damadian 1981) until today

*Email: sven.Junge@bruker-biospin.de

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_121-1

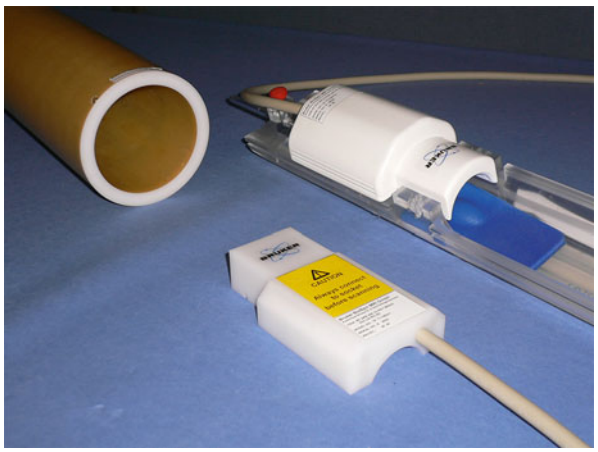


Fig. 1 A Set of RF coils for preclinical imaging of small objects consists of a volume coil for RF transmission and a two local RF coils for MRI signal detection

the development of this noninvasive method advanced quickly for medical diagnostics, material research, and preclinical imaging.

Today typically four types of MRI systems are distinguished by their magnetic field strength in Tesla (T):

1. Low-field systems: 1–0.5 T
2. Midfield systems: 0–1.5 T
3. High-field systems: 0–7.0 T
4. Ultrahigh-field systems: 0 T–21 T

Low-field systems with a proton-resonance frequency of 8.5–17 MHz are commonly equipped with recessive or permanent magnets and are generally used as open MRI systems, e.g. for interventional surgery. Today the low signal strength and the low raw image quality can be compensated with extensive postprocessing on high-performance computer systems or with the introduction of cryogenically cooled RF antennas. Mid-field systems with a proton-resonance frequency of 42–63 MHz provide a better image quality and have been commonly used as standard MRI systems in the clinical environment for a long time now. High-field MRI systems with a proton-resonance frequency of 125–300 MHz offer a higher signal-to-noise ratio (SNR) and potentially higher contrast-to-noise ratios, spectral resolution, and can provide a much better image quality for certain applications. With the introduction of 3 T and 7 T clinical MRI systems at the beginning of this century, new doors were opened for medical diagnostics. Ultrahigh-field systems are commonly used only for research and provide in specific circumstances a significantly increased SNR. MRI systems up to 21 T (950 MHz) are available and commonly operate in a frequency range of 200–600 MHz in the area of preclinical imaging. But the effort with respect to the RF engineering increased significantly to achieve high SNR and a homogenous image contrast.

The RF coil is the first component where the MR signal is stimulated and received and therefore one of the most important components of a MRI system. The design of properly developed RF coils is the key to achieve the best clinical, preclinical or experimental result for MRI scientists or clinicians.

Before going into details of RF coils for MRI, a basic understanding of the signal reception and excitation for NMR and MRI is necessary. The RF coils are not stand-alone devices and part of a complex imaging system. Full-length textbooks about the design of RF coils for MRI are already published

DOI 10.1007/978-981-4560-75-7_121-1

(Vaughan and Griffiths 2012; Mispelter et al. 2006; Jin 1999) and in this section a brief introduction into the design and theory of MRI coils is given (Fig. 1).

Requirements for Antennas in NMR and MRI Systems

Basic Introduction to the Principle of NMR and MRI

To understand the design goals of RF coils for MRI, it is first necessary to explain the basic principles of NMR and MRI. A correct explanation of a so-called spin ensemble inside an electromagnetic field requires a quantum mechanical prediction. Nevertheless the semiclassical prediction is sufficient for the description of the effects for NMR and MRI that are important for RF coil design. A more detailed explanation can be found, e.g., in (Lawrence 1996; Haacke et al. 1996). The fundamental characteristics of nuclei used for NMR and MRI are characterized by a nonzero nuclear spin **I** that is linked by

$$\boldsymbol{m} = \gamma \cdot \boldsymbol{I}$$

to the angular momentum $\boldsymbol{m}$. The so-called gyromagnetic ratio γ is specific for each nucleus. Some values $v_0 = \frac{2\pi}{\gamma}$ that are important for NMR and MRI are listed in Table 1.

Based on the so-called Zeeman effect (Zeeman 1882) the energy levels of $\boldsymbol{I}$ split into 2I + 1 discrete level inside an external magnetic field $\mathbf{B_0}$. By convention the $\mathbf{B_0}$-field points in the z–direction $\mathbf{B_0}$ = (0,0, B_0). The nuclei listed in Table 1 can adopt one of the two spin states 1/2 and -1/2. These spin states correspond to the two possible orientations of the magnetic momentum $\boldsymbol{m}$ parallel or anti-parallel to $\mathbf{B_0}$ with an angular orientation of about $\Omega_0 = 54.7°$ and can be evaluated by the magnetic momentum μ and its corresponding z-component (Fig. 2).

The energy difference *dE* between these two adjacent energy levels is

$$dE = \gamma h \cdot B_o \tag{1}$$

where *h* is the so-called Planck's constant. The nuclei can undergo a transition between these adjacent energy levels by absorbing or emitting of a photon, e.g., sending or receiving an electromagnetic wave. For example nuclei in the upper energy level emit an electromagnetic wave and end up in the lower energy state. The energy of this photon is E = h* *v* and matches exactly the energy difference between these two levels

$$dE = h\nu_0 = \hbar\omega_0 \tag{2}$$

where v_0 is the frequency of this photon and ω_0 is the angular frequency. Combining Eqs. 1 and 2 the so-called Larmor frequency

Table 1 List of most commonly used nuclei for MRI and NMR. The most significant nucleus for a large range of MRI studies is the proton ^{1}H because of the natural concentration of water in live objects under investigations, e.g., humans or animals

	^{1}H	^{31}P	^{13}C	^{19}F	^{3}He
$v_0 = \frac{2\pi}{\gamma}\left[\frac{MHz}{T}\right]$	42,57	17,23	10,71	40,05	32,43
Total spin number $\boldsymbol{I}$	1/2	1/2	1/2	1/2	1/2

DOI 10.1007/978-981-4560-75-7_121-1

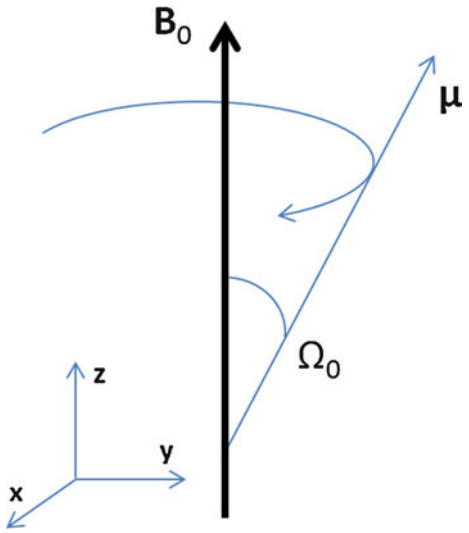


Fig. 2 Precessing of the magnetic momentum μ inside an external magnetic field $\boldsymbol{B}_0$

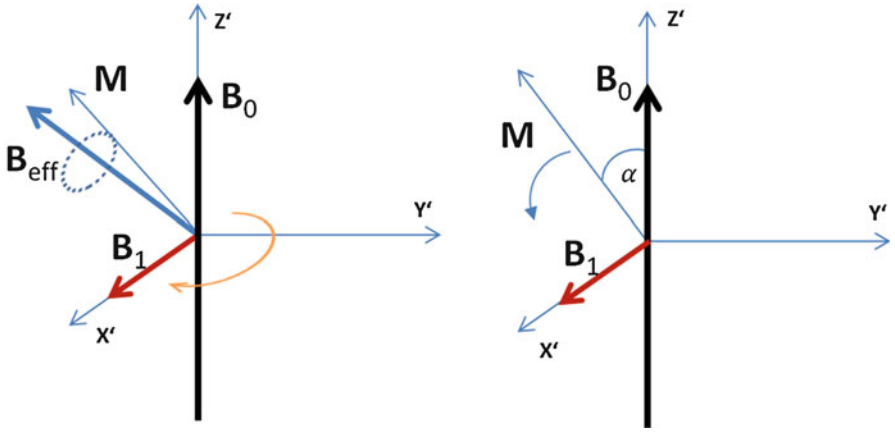


Fig. 3 *Left*: Net magnetization vector M in the rotating frame rotating with $\omega_0 = \gamma \cdot B_0$ and aligned with $\mathbf{B}_0$; *Right*: Flipping the magnetization for a required angle α by applying a transversal magnetic field $\mathbf{B}_1$

$$\omega_0 = 2 \cdot \pi \cdot \nu_0 = \gamma \cdot B_0 \tag{3}$$

is obtained which ranges for protons (^{1}H) commonly from about 42 MHz up to 900 MHz (1 T – 21 T field strength of the external magnetic field) at MRI. This is the frequency-range for operation of NMR and MRI RF coils.

In the object under investigation a bulk sample is considered which consists of many nuclear spins. The energy levels of the nuclear spins are governed by the Boltzmann statistics. The probability of the number of spins of the different energy levels is assigned by the so-called Boltzmann factor $e^{-\frac{dE}{kT}}$. At the equilibrium the lower energy level is more populated than the upper one (anti-parallel) and the small difference of the energy levels is responsible for the relatively intrinsic low sensitivity of magnetic resonance that can be increased by applying a high as possible external magnetic field (see Eq. 1). This is the major cause for the ongoing trend to increasingly higher magnet-field strength for MRI. The sum over all magnetic spins in the volume under investigations (V) is called magnetization

DOI 10.1007/978-981-4560-75-7_121-1

$$\boldsymbol{M} = \left(M_x, M_y, M_z\right) = \frac{1}{V}\sum \mu_i. \tag{4}$$

When a sample is placed in a uniform magnetic field $\boldsymbol{B}_0 = (0, 0, \mathrm{B}_0)$ aligned in the z-direction, it develops a net magnetization in the same direction. In addition a force on the magnetization is applied by $M \times B_0$ and results in a movement of the magnetization that can be described by the differential equation

$$\frac{d}{dt}\boldsymbol{M}(t) = \gamma \boldsymbol{M}(t) \times \boldsymbol{B}_0 \tag{5}$$

The movement of the magnetization is called precessing and moves with the Larmor-frequency ω_0 on a torque aligned with the external magnetic field $\boldsymbol{B}_0$.

A rotating coordinate-frame (laboratory frame) aligned with the external magnetic field $\boldsymbol{B}_0$ is introduced. By convention the magnetization vector is fixed. The magnetization in the rotating laboratory-system with the basis vector $\boldsymbol{e_x}$, $\boldsymbol{e_y}$, $\boldsymbol{e_z}$ rotates with the Larmor frequency $\omega_0 = \gamma \cdot B_0$ in the same direction as the spin-system. Without limiting the generality a second **B**-field (Fig. 3 left) is introduced that results in an effective magnetic **B**-field $\boldsymbol{B_{eff}} = \boldsymbol{B} - \frac{\omega_0}{\gamma}\boldsymbol{e_z}$. The result for the equation of motion within the rotating frame is $\frac{d'}{dt}\boldsymbol{M}(t) = \gamma \boldsymbol{M}(t) \times \boldsymbol{B_{eff}}$.

The effect of an oscillating RF magnetic field $\mathbf{B}_1$ is considered that is linearly polarized in the x-direction (Jin 1999) and can be expressed by two time dependent components in the xy-plane

$$\boldsymbol{B}_1 = A \cdot \boldsymbol{e}_x \cos \omega t = \boldsymbol{B}_{cw} + \boldsymbol{B}_{ccw}$$

where $\boldsymbol{B}_{cw} = \frac{1}{2}A\left(e_x \cdot \cos \omega t - e_y \cdot \sin \omega t\right)$ and $\boldsymbol{B}_{ccw} = \frac{1}{2}A\left(e_x \cdot \cos \omega t + e_y \cdot \sin \omega t\right)$ are rotating magnetic fields: one, $\boldsymbol{B}_{cw}$ is rotating clockwise and $\boldsymbol{B}_{ccw}$ is rotating counterclockwise.

When the $\mathbf{B}_1$-field oscillates with the same frequency $\omega = \omega_0$, than the two components of the so-called circularly polarized fields is given by

$$\boldsymbol{B}_{cw} = \frac{1}{2}A \cdot e_x \quad \text{and} \quad \boldsymbol{B}_{ccw} = \frac{1}{2}A\left(e_x \cdot \cos 2\omega_0 t + e_y \cdot \sin \omega_0 t\right) \tag{6}$$

When applying the $\mathbf{B}_1$-field then the value of the effective magnetic field $\boldsymbol{B_{eff}}$ in the rotating frame becomes

$$\boldsymbol{B_{eff}} = \left(B_0 - \frac{\omega_0}{\gamma}\right)\boldsymbol{e_z} + \frac{1}{2}A \cdot e_{x'} \tag{7}$$

Due to the rotation $\omega = \omega_0$ of the laboratory frame (the system is on-resonance) the external magnetic field $\boldsymbol{B}_0$ has no influence on the motion. The z-component disappears and the magnetization is precessing on a torque aligned to $\boldsymbol{B_{eff}} = \frac{1}{2}A \cdot e_{x'}$ and the resulting rotation has the angular frequency $\boldsymbol{\omega}_0 = \gamma \cdot \boldsymbol{B}_{eff} = \gamma \cdot \boldsymbol{B}_1$. It should be noted: the magnetization vector rotates to any desired angle α away from its equilibrium position (Fig. 3 right) when a magnetic field $\mathbf{B}_1$ oscillating with an angular frequency $\omega = \omega_0$ is applied that is perpendicular to the static field $\boldsymbol{B}_0$. In practice this$\mathbf{B}_1$-field is applied by RF coils using a RF-pulse with a frequency of $\nu_0 = \frac{2\pi}{\gamma}$. The final angle is dependent on the duration τ of the RF-pulse and the Amplitude of the applied transversal $\mathbf{B}_1$-field and is given by the integral of the envelope of the RF-pulse

DOI 10.1007/978-981-4560-75-7_121-1

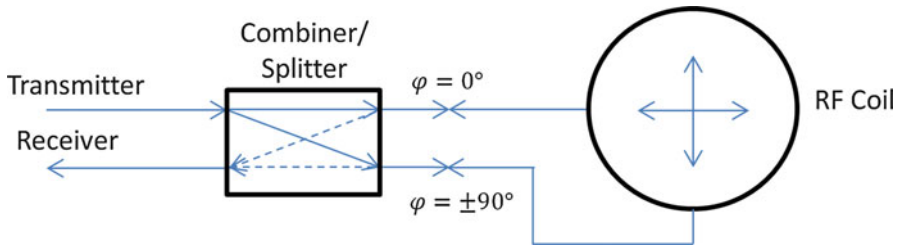


Fig. 4 Combiner/Splitter: The splitter separates the transmit power into two RF-channels and introduses a 90° phase shift in one of the channels to provide a circularly polarized $\mathbf{B}_1$ RF-field. During reception the Combiner adds the two signals with a -90°-phase shift

$$\alpha = \int_0^\tau \gamma \cdot B_1(t)dt \tag{8}$$

Quadrature Detection and Reception

The counterclockwise rotating component $\boldsymbol{B}_{ccw}$ has no effect on the magnetization vector **M**. This can be demonstrated by the same argumentation of the previous section where the laboratory frame rotates counterclockwise and results in an effective magnetic field

$$\boldsymbol{B}_{\boldsymbol{eff}} = \left(B_0 + \frac{\omega_0}{\gamma}\right)\boldsymbol{e}_{\boldsymbol{z}} + \frac{1}{2}A \cdot e_{x'} = 2B_0\,\boldsymbol{e}_{\boldsymbol{z}} + \frac{1}{2}A \cdot e_{x'}. \tag{9}$$

While the magnitude of $\mathbf{B}_0$ (in a range of few T) is much larger than the magnitude of $\mathbf{B}_1$ (in a range of few μT) the effect of the counterclockwise component is negligible and only the right hand side component of the transversal RF-field $\mathbf{B}_1$ has a significant effect on the magnetization vector. But for a linear polarized RF-field, only half of the energy will be used for reversing the magnetization vector.

To feed the whole energy of a transversal RF-field to the magnetization vector **M** a clockwise rotation $\mathbf{B}_1$-field $\boldsymbol{B}_1^{cw} = \mathbf{A}\left(e_x \cdot \cos\omega t + e_y \cdot \sin\omega t\right)$ can be used. Then the effective $\boldsymbol{B}_{\boldsymbol{eff}}$ becomes twice the amplitude compared to the linear RF-field $\boldsymbol{B}_{\boldsymbol{eff}} = A \cdot e_{x'}$ and takes twice as much energy into the spin-system to deflect the magnetization to a given angle α.

The components of the $\mathbf{B}_1$-field $\mathbf{B}_1 = \boldsymbol{B}_1^{cw} + \boldsymbol{B}_1^{ccw} = \boldsymbol{B}_1^{(+)} + \boldsymbol{B}_1^{(-)}$ in the field of MRI are commonly called left-hand and right-hand side component and are expressed in notation of complex numbers $\left(i = \sqrt{-1}\right)$ by

$$\boldsymbol{B}_1^{(+)} = B_{1x}^{(+)} + iB_{1y}^{(+)} = \frac{B_{1x} + i\,B_{1y}}{2} \quad \text{and} \quad \boldsymbol{B}_1^{(-)} = B_{1x}^{(-)} + iB_{1y}^{(-)} = \frac{\left(B_{1x} - i\,B_{1y}\right)^*}{2} \tag{10}$$

Knowing that the right hand side component is responsible for flipping the magnetization vector in the positively rotating frame it should be noted (Hould 2000) that the signal received is proportional to the left-hand side component (section Bloch Equations and Reciprocity). In practice a circular rotation RF-field is synthesized by two geometrically orthogonal $\mathbf{B}_1$ RF-fields; each oscillating linearly with a phase shift of 90° phase difference.

Commonly well-known devices from microwave technology, such as branch lines or ring hybrids (3 dB coupler) are used for quadrature splitting or combining of the RF-signals (Fig. 4). During transmission the input power is split into two RF-signals with a phase shift of 90°, while during reception the RF-signals from the coil are combined with the opposite phase shift for detecting the left-hand side polarized

DOI 10.1007/978-981-4560-75-7_121-1

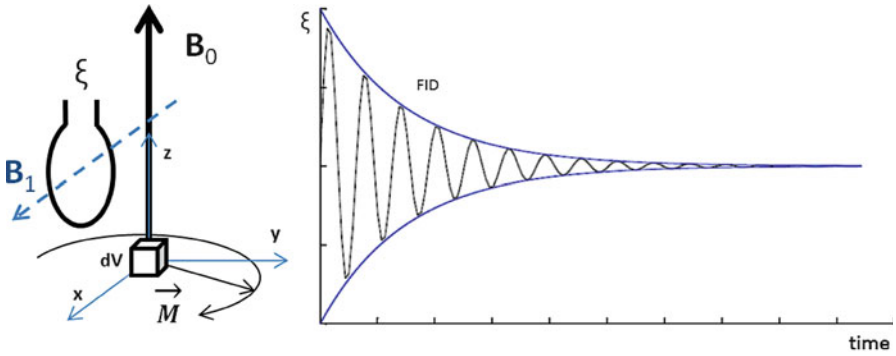


Fig. 5 *Left*: Schematic coil configuration for detection of the NMR signal; *Right*: Graph of the Time depending MRI signal, the so-called free induction decay (FID)

RF-signal from the sample. These two signals from the combiner/splitter are combined in a manner that the two RF-fields are perpendicular to each other and generate a circular $\mathbf{B}_1$-field. In this case the entire input power will be used for excitation of the nuclear spins or the whole energy from the reception process of the spin system can be detected from the object under investigation. Simply following the rules for signal theory, when the two phase correlated signals of a quadrature detected RF signal are combined, the resulting signal is doubled. The associated noise voltage, being uncorrelated, only increases by a factor of $\sqrt{2}$. This results in an increase of the SNR by a factor $\sqrt{2}$ compared to the detection of linearly polarized RF-fields. A good isolation of at least 20dB or more of the two quadrature channels is required.

Bloch Equations and Reciprocity

After application of the $\mathbf{B}_1$ RF pulse to the nuclear spin system the magnetization vectors are flipped and are precessing in the xy-plane. But the magnetization does not precess infinitely due to the losses and turns back to the equilibrium state aligned with $\mathbf{B}_0$. This process is called relaxation and is qualified by two different time constants T_1 and T_2. The time-constant T_1 describes the so-called longitudinal relaxation and is a result of the spin interactions with the lattice. The second relaxation process called transversal relaxation described by T_2 concerns the fact that an energy transfer beneath the spins occurs and the phase coherent movement of the magnetization vectors of the spins gets lost. These time constants are strongly dependent on the material of investigations (sample) and on the magnetic field strength.

The loss of coherent magnetization M_z is called Free Induction Decay (FID – Fig. 5 right). The so-called Bloch equations are a phenomenological model and provide one of the simplest models for the behavior of a nuclear spin system after a RF-pulse excitation. The Bloch equations in the laboratory frame (Bloch 1946) may be written:

$$\frac{dM_{x,y}}{dt} = \gamma(\boldsymbol{M} \times \boldsymbol{B})_{x,y} - \frac{M_{x,y}}{T_2} \tag{11}$$

$$\frac{dM_z}{dt} = \gamma(\boldsymbol{M} \times \boldsymbol{B})_{x,y} - \frac{M_z - M_0}{T_1} \tag{12}$$

For the case of the on-resonance condition $\Delta\omega = 0$ in the rotating frame the vector-product results in a time depending signal with the basic frequency ω_0 that is generated by the magnetization rotating with ω_0. The exponential envelope of the FID based on the two time constants T1 and T2 come from dephasing and longitudinal relaxation mechanisms (Fig. 5 right)

DOI 10.1007/978-981-4560-75-7_121-1

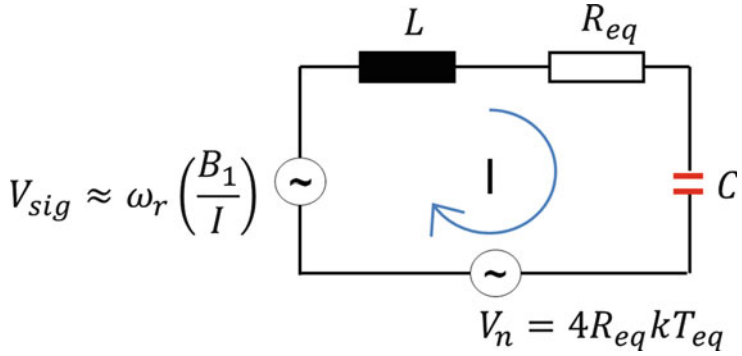


Fig. 6 Equivalent circuit for a MRI coil; including the noise source V_n

$$M_z(t) = M_0\left(1 - e^{t/T_1}\right) + M_z(0)e^{-t/T_1}$$
$$M_{x/y}(t) = M_{x/y}(0)e^{-t/T_2} \tag{13}$$

Consider the case that the $\mathbf{B}_1$-field is applied in the x-direction considering Eq. 11. A unit current is introduced that would be produced by a DC unit current flowing through the RF coil (Fig. 5 left). For times less than the transversal relaxation time T_2 the two Bloch-equations can be written in right hand notation (Hould 2000) $\frac{dM_x^+}{dt} \approx \gamma B_{1y}^+ M_z$ and $\frac{dM_y^+}{dt} \approx -\gamma B_{1y}^+ M_z$. For small flip angle approximation the case is considered that the pulse duration τ is short and only a small flip-angle ($<10°$) is applied. For this case $M_z = M_0$ is the result. When this short pulse with the duration τ is applied, the magnetization follows $M_x^+ \approx \gamma\tau M_0 B_{1y}^+$ and $M_y^+ \approx \gamma\tau M_0 B_{1x}^+$ or in complex notation for the positive rotating frame

$$M^+ = M_x^+ + iM_y^+ \approx \gamma\tau M_0 B_1^+ \tag{14}$$

This relation called *principle of reciprocity* make clears that the (complex) transversal magnetization following a RF-pulse is proportional to the right hand side component of the $\mathbf{B}_1$-field strength (10). Following the *principle of reciprocity* it can be shown that the receive signal ξ induced in a pickup coil (Fig. 5 left) by a magnetization in a voxel dV is given in the negatively rotating frame by

$$\xi = 2\omega M^+ B_1^{-*} = \omega M^+\left(B_{1x} - iB_{1y}\right) \tag{15}$$

and corresponds to the left hand side component of the $\mathbf{B}_1$-field strength (Hould 2000).

It should be noted for the transmit/receive case: the transmit/receive signal is only proportional to squared B_1-field strength for small flip angle and at low external magnetic field strength B_0 where the phase of the transmit/receive $\mathbf{B}_1$-field is independent from the position.

If the dielectric constant and/or conductivity of the sample causes the $\mathbf{B}_1$-fields to be complex and leads to a position dependent phase change, e.g. at high field applications, then the complex notation of the right and left hand side of the $\mathbf{B}_1$-field has to be taken into account. The *principle of reciprocity* is more than the statement that a good transmit coil is also an efficient receive coil!

Efficiency, Losses and Signal to Noise Ratio

In most cases the small voxel sizes of the detected NMR/MRI-signals are correlated to the detection of small signals. Consider a RLC circuit illustrated in Fig. 6 and in accordance with Kirchhoff's yields to equation

DOI 10.1007/978-981-4560-75-7_121-1

$$V = R_{eq}I + \frac{i}{\omega C}I - i\omega L \tag{16}$$

where ω denotes the angular frequency. For the case R $= 0$ this reduces to $I = V\left[\frac{L}{i\omega}\left(\omega^2 - \frac{1}{LC}\right)\right]^{-1}$ and for the boundary limiting case $I \to \infty$ the resonance condition $\omega = \omega_r = \frac{1}{\sqrt{LC}}$ is obtained where ω_r is the resonance frequency and where the RF coils must operate. In practice the resistance R cannot vanish completely and the magnitude of I cannot be infinity but in any case the current I would be at maximum at the resonance frequency ω_r. While the resistance R cannot be zero, some energy will be dissipated in the RLC- circuit and the efficiency is decreased. To provide a quantitative value that can be measured, a so-called quality factor is defined by $Q = 2\pi \frac{\text{Maximum stored energy}}{\text{Total dissipated energy per period}}$. For the RLC-circuit the quality factor can be found as

$$Q = \frac{1}{R_{eq}}\sqrt{\frac{L}{C}} \tag{17}$$

In principle the determination of the values of R, L and C can be difficult and per convention the Q-value is obtained as

$$Q = \frac{\omega_r}{\Delta\omega} \tag{18}$$

where $\Delta\omega$ denotes the so-called bandwidth. It must be noted: for comparison reasons it must be considered if the RF-coil is match to 50 Ω or not and that corresponds e.g., to a 3 dB or 6 dB bandwidth and yields to a change of the Q-value by a factor of 2.

The power signal-to noise ratio is given by $\text{SNR}^2 = \frac{V_{sig}^2}{V_n^2}$. With the introduction of so-called Johnson noise a very general formulation of SNR (neglecting the bandwidth of the measurement) is given by (Darrasse and Ginefri 2003)

$$\text{SNR} \approx \sqrt{F}\frac{\omega_0\left(\frac{B_1}{I}\right)}{\sqrt{4K_B T_{eq} R_{eq}}}\Delta V M_T \sqrt{\frac{t_0}{t_R}}\sqrt{t_{\text{scan}}} \tag{19}$$

We assume that all terms that are not strongly related to the RF-coil sensitivity can be neglected; this includes the noise factor F (which accounts for the additional noise power from the MR system), transversal magnetization M_{T}, the total scan time t_{scan}, the repetition time t_{R}, the sampling window duration t_0 and the elementary encoded volume ΔV. In addition, the bandwidth of the measurement is also neglected. For simplicity the coil current **I** is set to unity and to unit current, respectively. In this case the formula 19 is simplified and the SNR is expressed as follows:

$$\text{SNR} \approx \frac{\omega_0 B_1}{\sqrt{4K_B T_{eq} R_{eq}}} \tag{20}$$

A gain in SNR can be achieved either by increasing the effective RF-field strength B_1 per unit current in the sample volume or by reducing the noise sources (loss mechanism) represented by the term $R_{\text{eq}}T_{\text{eq}}$. Several different noise sources interact with the RF coil and are represented by four major terms:

DOI 10.1007/978-981-4560-75-7_121-1

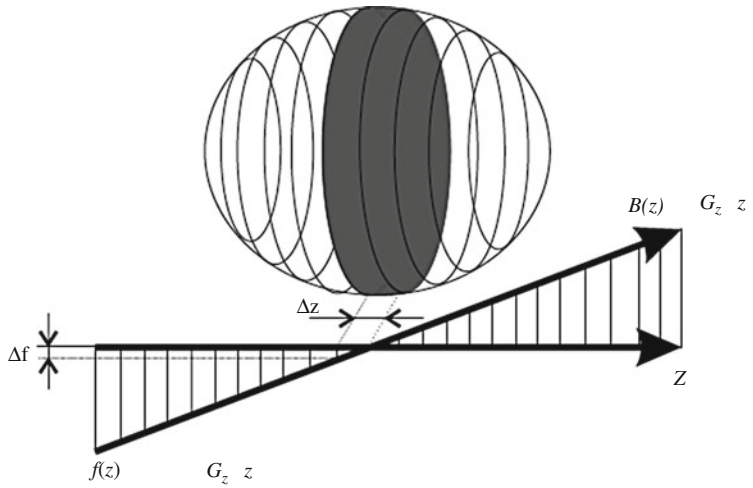


Fig. 7 The principle of slice selective excitation for MR imaging

$$R_{\text{eq}} = (R_{\text{M}} + R_{\text{E}}) + R_{\text{c}} + R_{\Omega} = R_{\text{s}} + R_{\text{c}} + R_{\Omega} \tag{21}$$

where R_{M} is the magnetic loss by magnetic field induced currents in the sample, R_{E} is the electric loss by stray capacitive coupling between the sample and the RF coil, R_{c} is the resistive loss of the RF coil, and R_{Ω} is the radiative loss. The magnetic loss R_{M} and electric loss R_{E} are combined into the so-called samples loss R_{s}. Simplifying these four loss mechanism to the two dominant parts of losses for RF detection that are most important and are temperature dependent, the SNR can be expressed as

$$\text{SNR} \approx \frac{\omega_0 B_1}{\sqrt{R_s T_s + R_c T_c}} \tag{22}$$

Basics of MRI imaging

There are numerous textbooks about MRI and a brief description can be found in the Encyclopedia of NMR (Lawrence 1996) or other basic textbooks about MRI (Canet 1996; Haacke et al. 1999). For MR-imaging (tomography) a spatially resolved signal is required which can be achieved with a spatially dependent Larmor-frequency $\omega_L = \gamma \cdot |\boldsymbol{B}_0|$ when an additional field gradient is present.

A constant field gradient $G = \begin{bmatrix} \frac{\partial B_x}{\partial x} & \frac{\partial B_x}{\partial y} & \frac{\partial B_x}{\partial z} \\ \frac{\partial B_y}{\partial x} & \frac{\partial B_y}{\partial y} & \frac{\partial B_y}{\partial z} \\ \frac{\partial B_z}{\partial x} & \frac{\partial B_z}{\partial y} & \frac{\partial B_z}{\partial z} \end{bmatrix} = \{g_{ik}\}$ is used which is aligned with the external magnetic field at z-axis. Only the 3rd row of G is from interest when $B_0 \gg \|G\| \cdot a$ is assumed, where a is the length of the imaging region. The field gradient G can be summarized to a vector $\boldsymbol{g} = (\frac{\partial B_z}{\partial x}, \frac{\partial B_z}{\partial x}, \frac{\partial B_z}{\partial z}) := \left(g_x, g_y, g_z\right)$ and commonly termed as field-gradient. With the use of this field gradient in addition to the external static magnetic field an allocation of a position x is provided. The Larmor-frequency for a voxel dV at this position is given by $\omega_L = \gamma \cdot (B_0 + \boldsymbol{g} \cdot \boldsymbol{r})$. It should be mentioned that with only one component of $\boldsymbol{g}$, e.g., g_x, a clear allocation of the frequency and position is impossible. It is required to measure ω_L for numerous field-gradients $\boldsymbol{g}$ to allocate the position in a three dimensional

DOI 10.1007/978-981-4560-75-7_121-1

volume. Today multiple different methods exist and in the following the principle of MRI is explained by one simple method – the spin warp gradient echo.

First a slice selective excitation is used when a RF-pulse with a carrier-frequency f_c and a bandwidth df is applied in presence of a field-gradient $\mathbf{g} = (0,0,g_z)$ (Fig. 7). The nuclear spins only in the slice $z = z_f = \text{const}$ corresponding to the carrier-frequency $f_c = \frac{\gamma}{2\pi} \cdot \left(B_0 + g_z \cdot z_f\right)$ are excited in the way explained in section Basic Introduction to the Principle of NMR and MRI. The slice thickness Δz is given by the bandwidth df corresponding to $df = \frac{\gamma}{2\pi} \cdot g_z \cdot \Delta z$ and can be chosen freely within technology limited boundaries.

The signal $\boldsymbol{S(t)}$ picked up by the receiver from a voxel at position $\boldsymbol{r} = \left(x, y, z_f\right)$ is (neglecting all relaxation effects) the sum of all spins in the selected slice with the spin density $\rho(\boldsymbol{r})$

$$\boldsymbol{S(t)} \propto \int \rho(\boldsymbol{r}) \cdot \boldsymbol{e}^{\,i\int_0^t \boldsymbol{\omega}(\boldsymbol{r},\boldsymbol{t'})\boldsymbol{dt'}} \boldsymbol{d^3r} = \boldsymbol{e^{i\omega_0 t}} \int \rho(\boldsymbol{r}) \boldsymbol{e}^{\,i\boldsymbol{\gamma}\int_0^t \boldsymbol{g}(\boldsymbol{t'})\boldsymbol{dt'}\boldsymbol{r}} \boldsymbol{d^3r} \tag{23}$$

When substituting the exponent by $\boldsymbol{k} = \boldsymbol{\gamma}\int_0^t \boldsymbol{g}(\boldsymbol{t'})\boldsymbol{dt'}$ where t denotes the time when the field gradient is applied and omitting the spatially independent frequency ω_L yields to

$$\boldsymbol{S(k)} \propto \int \rho(\boldsymbol{r}) \cdot \boldsymbol{e^{ikr}} \boldsymbol{d^3r} \tag{24}$$

And in this form the signal $\boldsymbol{S(t)}$ can be recognized as Fourier transform of the spin density $\rho(\boldsymbol{r})$ and the spin density itself can be determined by the inverse Fourier transform to the receiver signal

$$\rho(\boldsymbol{r}) \propto \int \boldsymbol{S(k)} \cdot \boldsymbol{e^{-ikr}} \boldsymbol{d^3r} \tag{25}$$

The so-called k-space is defined by the reciprocal space vector $\boldsymbol{k} = \left(k_x, k_y, k_z\right)$. After slice selective excitation the encoding is done in the two dimensional plane and one spatial dimension is encoded by a spatial frequency modulation by applying a constant gradient-field, e.g., g_x in x-direction. The received one-dimensional signal is sampled at N_x discrete points in time intervals of dt_x:

$$\boldsymbol{S}(k_{\boldsymbol{x}}) = \int \rho(\boldsymbol{r}) \cdot \boldsymbol{e}^{ik_x x} \boldsymbol{d^3r} \quad \text{with} \quad k_x = \gamma G_x \cdot t_x \tag{26}$$

The second spatial dimension is encoded by the phase of the signal (so-called phase encoding) and is carried out between the slice selective excitation and the signal acquisition by the gradient g_y for a duration of t_y. This gradient g_y is perpendicular to the gradient g_x in x-direction. This modulation of the spin phase persists through the following read-out procedure of the signal and results in a spatially dependent phase $\varphi(y) = \gamma g_y t_y y$. This phase encoding has to be repeated several times by changing the gradient strength of g_y in order to cover the entire two dimensional k-space.

With $k_y = \gamma G_y \cdot t_y$ the signal originating from the selected slice can be written as:

$$S\left(k_x, k_y\right) = dz \cdot \int \rho(x,y) \cdot e^{i\left(k_x x + k_y y\right)} dxdy \tag{27}$$

DOI 10.1007/978-981-4560-75-7_121-1

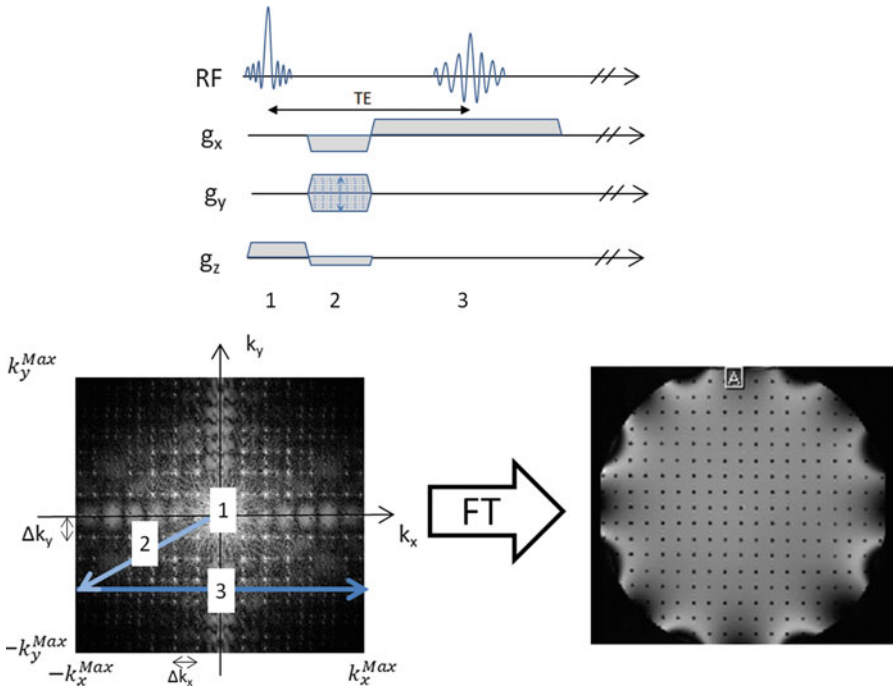


Fig. 8 Special encoding in MRI – *Top*: The principle diagram of a typical spin-warp gradient echo sequence; *Bottom*: The corresponding so-called k-space trajectory in the so-called Fourier-space for spatial encoding and the generation of the final image after Fourier-transform of the sampled data

In Fig. 8 a simple gradient echo experiment using a so-called spin-warp trajectory for phase encoding is shown. In the first step (1) the slice selection is carried out by applying the RF-pulse during the constant field gradient g_z. In the second step (2) a negative gradient g_x is applied to move the trajectory in the k-space to $-k_x^{\max}$ and at the same time the phase encoding gradient g_y is applied to move the trajectory in k-space to a specific k_y position. In addition a so-called rephase lobe of the slice selective gradient g_z is applied for alignment of the phase of the spin-system. In the last phase (3) a read out gradient g_x switches on that corresponds to move at a certain k_y -position in k-space from $-k_x^{\max}$ to $k_x^{\max}$. To cover the whole k-space in y-direction this procedure must be repeated with N_y steps with multiple phase encoding gradients with values running from $-g_y^{\max}$ to $g_y^{\max}$.

Apart from the SNR the imaging speed is one of the most important MR sequence parameters in MRI. The total time T_{tot} for acquiring an image is given by the number of required RF-excitations N_z, the number of phase encoding steps N_y, the number of repetition times N_{rep} and the repetition time TR. The repetition time TR is defined as time interval between the RF-excitations and is a fundamental limiting factor in terms of imaging speed and sensitivity predominant effects caused by the relaxation time of the different tissues. The total time for MRI scans is calculated by $T_{tot} = N_z\, N_y N_{rep} \cdot TR$.

In the past multiple imaging strategies such as FLASH have been developed for imaging time reduction allowing very short TR. Single shot or multi-shot sequences were developed, such as EPI or TSE. These sequences use multiple phase encoding steps during one repetition time TR. However with the introduction of so-called parallel imaging a decrease of the imaging time can be achieved without further need of enhanced gradient performance.

Parallel imaging is based on the principle of partially replacing the time consuming spatial encoding field-gradients. The spatial sensitivity information of the coil-elements of so-called array-coils

DOI 10.1007/978-981-4560-75-7_121-1

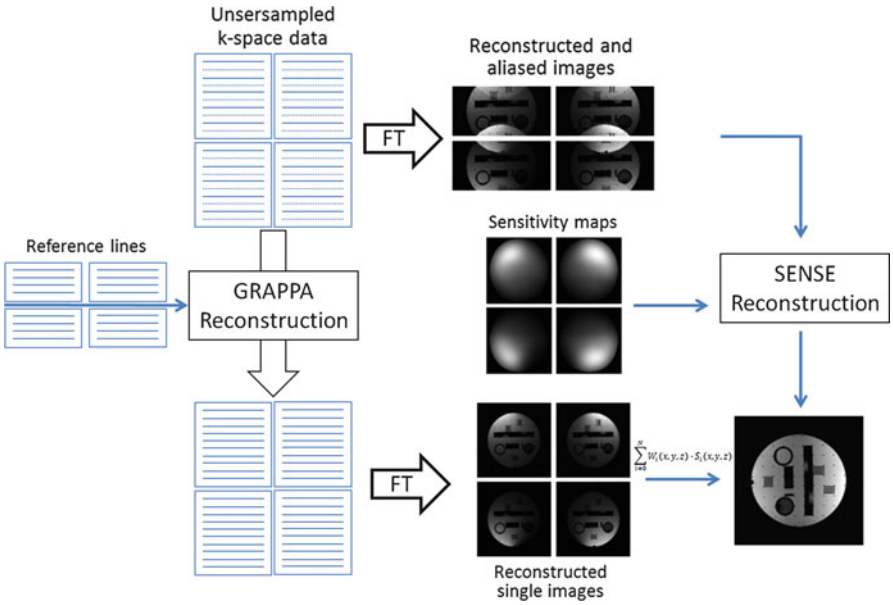


Fig. 9 Principle description of parallel imaging techniques in k-space or image domain

(section Cryogenically Cooled Coils, Double Tuned RF Coils) is used in a manner that only a fraction of the phase encoding steps has to be acquired. This results directly in an accelerated image acquisition.

The most prominent parallel imaging techniques are GRAPPA (Griswold et al. 2002) based on a *k*-space reconstruction approach and SENSE (Pruessmann et al. 1999) based on an approach in the image domain (Fig. 9). For parallel imaging the k-space is sampled with a rate of phase encoding steps that is below the Nyquist theorem, e.g., for an under sampling by a factor $R = 2$ only every second k-space line will be acquired using the phase encoding by the field gradients. The parallel image reconstruction can be directly performed in the k-space by solving the missing k-space lines. For that purpose every line in the center of the k-space is aquired so-called auto-calibration lines (ACS-lines). Using the ACS-lines for each coil element the missing k-space lines of the under sampled image are reconstructed and the finally the reconstructed image is the result of an inverse Fourier transform. For parallel imaging techniques using SENSE- algorithms the under sampled image is reconstructed via inverse Fourier transform and results in so-called aliased images for each coil-element. These images with a reduced FOV contain information from multiple equidistantly distributed pixels in the desired fully sampled image. With usage of the complex coil sensitivities of the individual coil-elements of the array, the superimposed pixels can be unfolded for a fully resolved image (see section Receive Array Coils for Parallel Imaging).

Hardware of MRI Systems

In this section a brief introduction of the general hardware for MR topography is presented. The RF coils are only one part but nevertheless play an important role in the transmit/receive chain of the MR -imaging system.

In Fig. 10 a sketch of an MRI system is shown and thereby explains the basic function of such an instrument. The largest component and in most cases also the most expensive component of a MRI-system is the magnetic field coil. For high and ultra-high field magnet systems superconducting magnets are commonly used. On one hand a higher magnetic field strength provides in general better SNR that can be used for better special or temporal resolution, but on the other hand the costs should be economically feasible.

DOI 10.1007/978-981-4560-75-7_121-1

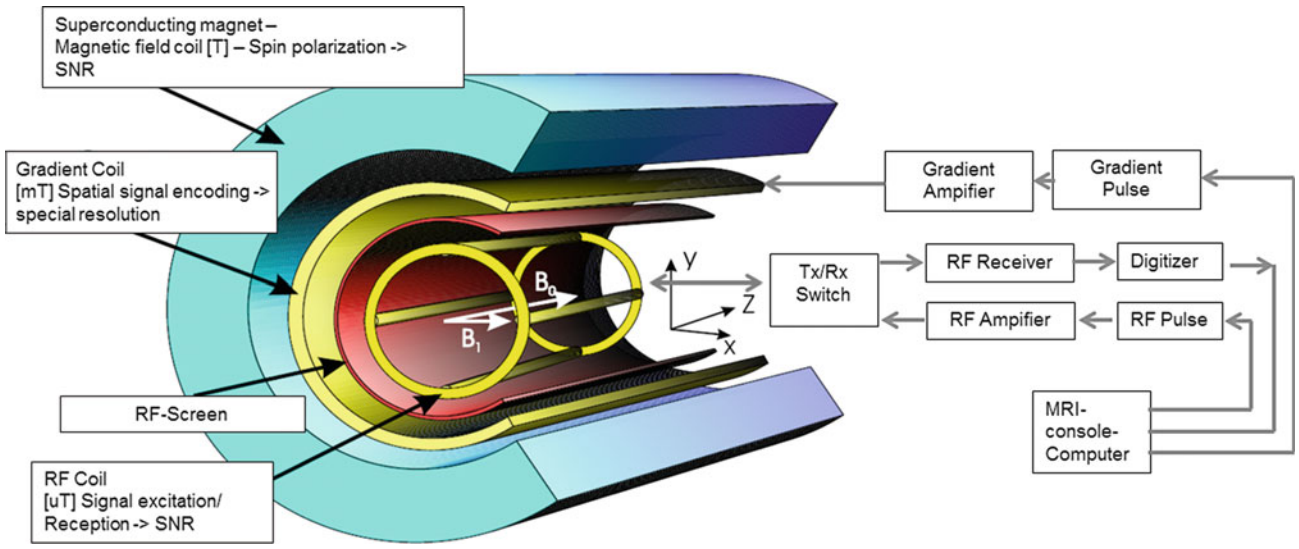


Fig. 10 Principle sketch of an MR-imaging system

The main magnet generates the external magnetic field $\boldsymbol{B}_0 = (0, 0, \mathrm{B0})$ used for spin polarization which is most as possible uniform over the volume of interest (VOI). For imaging applications the homogeneity is in order of a few parts per million (ppm) over a spherical volume up to 50–80 cm in diameter for clinical MRI systems. Increased demand with respect to the homogeneity of 1 part per billion over a spherical volume of a few cm is required for high resolution spectroscopy. For these reasons so-called shim coils are commonly employed as part of the gradient system or part of the main magnet for adjusting the uniformity of the main magnetic field $\boldsymbol{B}_0$ in the VOI locally.

The gradient coils are used for flexible switching of the linearly independent gradient fields $\boldsymbol{g} = \left(g_x, g_y, g_z\right)$ and these are temporarily superimposed to the main magnetic field $\boldsymbol{B}_0$. In this manner the spatial origin of the signal contributions are actively encoded either in its frequency or phase. Fast and so-called real-time imaging applications are limited in most cases by the gradient performance.

At the inner wall of the gradient coil or as outer part of the RF coil, a RF screen is included. This screen protects the inner volume from the outer space so that no additional noise is introduced into the received signal and prevents undesired RF-coupling of the RF-coil to the gradient system.

For applying the $\mathbf{B}_1$RF-field into the sample and receiving the RF-electromagnetic field from the sample the so-called RF coils, also known as RF resonators or RF-probes are used. This RF coil operates at a resonance frequency $v_0 = \frac{\omega_0}{2 \cdot \pi}$ in the general frequency band that is fixed by the desired Larmor-frequency and corresponds to the main magnetic field $B_0 = |\boldsymbol{B}_0|$ (section Basic Introduction to the Principle of NMR and MRI).

In principle three main RF systems of RF coils are distinguished:

- RF systems for excitation and reception (Tx/Rx) of the $\mathbf{B}_1$ electromagnetic RF field; commonly volume resonators and local coils and so-called RF-array coils are used.
- RF systems of excitation-only (Tx-only) of the $\mathbf{B}_1$ electromagnetic RF field; commonly volume resonators and dedicated transmit RF-array coils are used.
- RF systems for reception-only (Rx-only) of the $\mathbf{B}_1$ electromagnetic RF field; commonly local RF coils and so-called RF-array coils are used.

DOI 10.1007/978-981-4560-75-7_121-1

Until today, so-called volume coils were used for Tx/Rx operation. These RF coils generate a homogenous RF-field $\boldsymbol{B}_1$ inside a given volume of interest (VOI). Today these volume coils are generally used for imaging whole-bodies and deeper structures inside the VOI and for MRI systems where the wavelength of the RF-field $\boldsymbol{B}_1$ is much larger than the dimensions of the object under investigation. The $\boldsymbol{B}_1$ RF-field is generated by a suitable current distribution (see section The General Principle of Volumetric RF Antennas with Homogenous RF-Field Based on Magneto Static Theory) on conductive wires and structures. In most cases the RF coil is equipped with a RF screen. The distance of the conducting wires and the RF screen is an important parameter with respect to the efficiency of the RF-coil (see section Shielding of Birdcage Coils). While the magnetic field strength $\boldsymbol{B}_0$ increased rapidly over the years, the wavelength of the RF-field became comparable to the dimensions of the object under investigation. New imaging techniques were needed and so-called Tx/Rx RF-array coils were introduced. These coils consist of multiple independent small RF-structures (surface coils) that are arranged around the VOI and driven by their own RF-chain independently (see section Receive Array Coils for Parallel Imaging). Since the early days of MR imaging small loop Tx/Rx surface coils were used. Today this kind of coils is commonly used for applications where only a small VOI is of interest, inhomogeneous RF-fields can be tolerated and a high signal to noise ratio (SNR) is required. This is the case, e.g., for double resonance spectroscopy experiments (MRS) where the RF-coil operates on two different frequencies for detection of multiple nucleus at the same time or same application e.g., $^{1}H/^{13}C$, $^{1}H/^{31}P$.

For Tx-only operation commonly volume coils are used and for clinical MRI-systems these RF coils are often called “body-coils.” The key point is that these coils are only used for excitation of a homogenous RF-field $\boldsymbol{B}_1$ inside a well-defined VOI. These volume coils are geometrically optimized for a large diameter as possible of the accessible volume and have in most cases a less optimum RF-efficiency. In addition, these RF-coils are equipped with so-called active decoupling circuits for actively and temporally switching the active RF-elements out of the resonance frequency ω_0 . For the detection of the MR-signal a separate RF-system is used –a receive-only RF coil system.

The RX-only coil systems are subdivided in three main receive RF-systems (volume coils, surface coils and array coils): The RF- circuits of all these coils are actively or passively decoupled from the transmitter RF coil system in a manner that the transmitting and receiving RF-system is mutually switched off from the Larmor-frequency ω_0 for minimized RF coupling of the Tx and Rx subsystems. This so-called decoupling prevents the splitting of the resonance modes during the transmit or receive operation. The Rx-only volume coils work on the principle of the transmit-only counterparts but the geometrical dimensions are optimized for the highest possible efficiency. In most cases these RF coils are designed in a manner that the RF screen is missing and the inner volume of the RF coil fits as tight as possible to the objects under investigation. In most cases the geometry of these coils is a compromise between RF-efficiency, homogeneity of the RF-field $\boldsymbol{B}_1$, usability and user-handling. In contrast to Rx-only volume coils so-called Rx-only surface-coils provide inhomogeneous electro-magnetic RF-fields $\boldsymbol{B}_1$ and are used for dedicated imaging applications where the VOI is localized to a small area. Because of the dedicated localization of the RF-field and additional preamplifiers located directly at the RF-structure, these coils provide a high SNR which is in most cases much higher than that of volume coils.

Several surface coils can be arranged into a set of independent antennas around the VOI; this arrangement is called RF array coil (section Receive Array Coils, Transmit Sense/RF Shimming). For imaging with array-coils the MR signal is detected by the multiple independent RF array-elements and each of them have their own RF receiver-chain. The signals are combined by the software (section Receive Array Coils). In most cases the result image is more inhomogeneous compared to volume coils, but has the advantage of a very high SNR based on the pre-amplification of the MR signal and the close proximity of the array coil elements to the VOI.

DOI 10.1007/978-981-4560-75-7_121-1

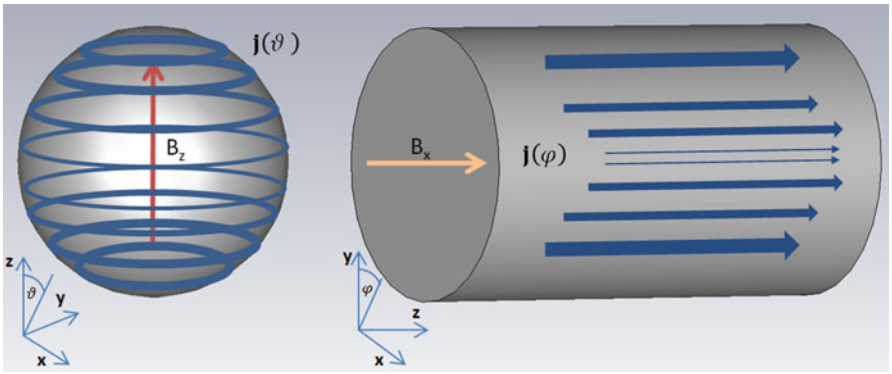


Fig. 11 Sinusoidal surface current distribution for generation of a homogenous transversal magnetic field

Design of MRI Coils

Design Principles of Volume RF Antennas for MRI

The General Principle of Volumetric RF Antennas with Homogenous RF-Field Based on Magneto Static Theory

The basic principle of volume RF coils is based on magneto static theory and on the general solution of the magnetic field in a source-free region, e.g., in spherical coordinates. Under the consideration that the current density **j** is zero outside of the surface of a sphere than **j** is defined in spherical coordinates at the surface with the radius $r = R$ (Fig. 11) by

$$\mathbf{j}(r, \vartheta, \varphi) = \left(\mathrm{j_r}, \mathrm{j}_\vartheta, \mathrm{j}_\varphi\right) = (0, 0, \mathrm{j}_0 \cdot \sin\vartheta) \tag{28}$$

Based on the Gauss's-law and Ampere's law in a source free region the following conditions apply

$$\nabla \cdot \mathbf{B} = 0 \tag{29}$$

$$\nabla \times \mathbf{B} = 0 \tag{30}$$

The scalar magnetic potential $\Phi^i(r, \vartheta, \vartheta)$ and $\Phi^a(r, \vartheta, \vartheta)$ at the inside and outside of the sphere $\mathbf{B}^{ia} = -\nabla\Phi^{ia}$ can be defined and this yields with Eq. 29 to the well-known Laplace equation $-\Delta\Phi^{ia} = 0$. The general solution can be written in spherical coordinates by

$$\Phi(r, \vartheta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left(C_{l,m} \cdot r^l + D_{l,m} \cdot r^{-(l+1)}\right) Y_{l,m}(\vartheta, \varphi) \tag{31}$$

and the current density at the equator is defined by the total current I by $\mathrm{j}_0 = \frac{I}{2R}$. Since j is independent from φ only terms of $\mathrm{m} = 0$ of the Eq. 31 are considered

$$\Phi(r, \vartheta) = \sum_{l=0}^{\infty} \left(C_{l,0} \cdot r^l + D_{l,0} \cdot r^{-(l+1)}\right) Y_{l,0}(\vartheta) \tag{32}$$

While the term $C_{l,0} \cdot r^l$ is regular at $\mathrm{r} = 0$ and goes to infinity for $r = \infty$ and the term $D_{l,0} \cdot r^{-(l+1)}$ is undefined at $\mathrm{r} = 0$ the scalar potential inside and outside of the sphere is defined by

DOI 10.1007/978-981-4560-75-7_121-1

$$\Phi^i(r,\vartheta) = \sum_{l=0}^{\infty} C_{l,0} \cdot r^l \cdot Y_{l,0}(\vartheta) \tag{33}$$

$$\Phi^a(r,\vartheta) = \sum_{l=0}^{\infty} D_{l,0} \cdot r^{-(l+1)} \cdot Y_{l,0}(\vartheta) \tag{34}$$

Based on $\nabla \cdot \mathbf{B} \equiv 0$, the normal component of B must be continuous at r = R and this yields to

$$\left(\frac{\partial \Phi^a}{\partial r} - \frac{\partial \Phi^i}{\partial r}\right)\Big|_{r=R} = 0 \tag{35}$$

$$\sum_{l=0}^{\infty} \left(l \cdot C_{l,0} \cdot R^{l-1} + (l+1) \cdot D_{l,0} \cdot R^{-l-2}\right) \cdot Y_{l,0}(\vartheta) = 0 \tag{36}$$

While the Eq. 36 is valid for every angle ϑ every member of the sum over l must be zero and the result is

$$D_{l,0} = -\frac{l}{l+1} R^{2l+1} \cdot C_{l,0} \tag{37}$$

The finite surface current **j** at the surface of the sphere results in a discontinuity of the tangential component of **B**. This can be expressed by the normal vector **n** of the spherical surface by

$$\left(\boldsymbol{B}^a - \boldsymbol{B}^i\right) \times \mathbf{n} = \mu_0 \cdot \mathbf{j} \tag{38}$$

By definition of Eq. 28 the current density has only nonzero components for the φ component and the discontinuity is only valid for components of ϑ

$$\left(\mathbf{B}^a_\vartheta - \mathbf{B}^i_\vartheta\right) = \mu_0 \cdot \mathrm{j}_0 \cdot \sin\vartheta \; at \; r = R \tag{39}$$

This yields

$$\mathbf{B}^a_\vartheta(R,\vartheta) = -\frac{1}{R}\frac{\partial \Phi^a(r,\vartheta)}{\partial \vartheta}\Big|_{r=R} = \sum_{l=0}^{\infty}\left(\frac{l}{l+1} \cdot C_{l,0} \cdot R^{l-1}\right) \cdot \frac{\partial Y_{l,0}(\vartheta)}{\partial \vartheta} \tag{40}$$

$$\mathbf{B}^i_\vartheta(R,\vartheta) = -\frac{1}{R}\frac{\partial \Phi^i(r,\vartheta)}{\partial \vartheta}\Big|_{r=R} = -\sum_{l=0}^{\infty}\left(C_{l,0} \cdot R^{l-1}\right) \cdot \frac{\partial Y_{l,0}(\vartheta)}{\partial \vartheta} \tag{41}$$

and substituting Eqs. 40 and 41 into 39 leads to

$$\sum_{l=0}^{\infty} C_{l,0} \cdot \left(\frac{2l+1}{l+1}\right) \cdot R^{l-1} \cdot \frac{\partial Y_{l,0}(\vartheta)}{\partial \vartheta} = \mu_0 \cdot \mathrm{j}_0 \cdot \sin\vartheta \tag{42}$$

The only valid derivate $Y_{l,0}(\vartheta)$ that is proportional to sin ϑ is $Y_{1,0}(\vartheta)$. With Eq. 42 and

$$\frac{\partial Y_{l,0}(\vartheta)}{\partial \vartheta} = -\sqrt{\frac{3}{4\pi}}\sin\vartheta \tag{43}$$

DOI 10.1007/978-981-4560-75-7_121-1

it can be concluded that all $C_{l,0} = 0$ for all $l \neq 1$ and $C_{1,0} = -\frac{2}{3}\sqrt{\frac{4\pi}{3}\mu_0 \cdot j_0}$. This yields for the scalar potential inside the sphere $\Phi^i(r,\vartheta) = \quad = -\frac{2}{3}\mu_0 j_0 \cdot r \cdot \cos\vartheta = \quad = -\frac{2}{3}\mu_0 j_0 \cdot z$.

In conclusion: A constant **B**-field $B\left(0,0,\frac{2}{3}\mu_0 j_0\right)$ inside the sphere in the z-direction is generated when a sinusoidal current distribution $\mathbf{j}(\vartheta,\varphi) = \mathbf{j}(0,0,j_0 \cdot \sin\vartheta)$ is applied. Using $j_0 = \frac{I}{2R}$ and $\mu_0 = 4\pi \cdot 10^{-7}\frac{Vs}{Am}$ a magnetic field **B** inside the sphere aligned with the z-direction of $B_z = \frac{4\pi}{30}\frac{I}{R}[\mu T]$ is the result with current I [A] and the radius R in [m].

This approach of sinusoidal current density modulation is the basic approach for generating a homogenous transversal **B**-field inside the volume of RF coils. From experience it is generally known that this approach is only valid for static currents and field distributions. But a static current distribution as an approach can be successfully used as long as the wave-length of the electro-magnetic fields is much larger than the dimensions of the object under investigation.

While access is needed to the inner volume of the resonator, the approach using a sphere is useless in practice. In general, a cylindrical structure is used to provide this access. The approach of applying a sinusoidal current distribution on a cylindrical surface will lead also to a homogenous **B**-field at the middle of an infinite long cylinder but requires also a defined current distribution at the fronts of a finite cylinder.

A homogenous magnetized finite long cylinder is considered. The material equation describes the relation between $\boldsymbol{B}$ and **H**-field and the magnetization $\boldsymbol{M}$

$$\boldsymbol{B}(x) = \mu(x)\,\mu_0 \boldsymbol{H}(x) + \boldsymbol{M}(x) \tag{44}$$

The current density **J**(**x**) in the areas inside and outside the cylindrical surface is zero and based on the 1. Maxwell equation the Eq. 44 can be written

$$0 = \nabla \times \mathbf{H}(\mathbf{x}) = \nabla \times 1/(\mu(x)\mu_0)(\boldsymbol{B}(x) - \boldsymbol{M}(x)) \tag{45}$$

It is considered that the permittivity in the inner cylinder is constant and is independent from position **x**. With the introduction of the vector potential **A** by $\boldsymbol{B} = \nabla \times \boldsymbol{A}$ the Eq. 45 yields to the so-called Poising equation $-\nabla^2 \boldsymbol{A} = \nabla \times \mathbf{M}$. This equation can be derived as volume integral (Kohlrausch 1956) while the magnetization **M**(**x**) is unequal zero only in a finite volume and the permittivity is constant:

$$\mathbf{A}(x) = \frac{1}{4\pi}\int \frac{\nabla' \times \mathbf{M}(\mathbf{x}')}{|x - x'|} dV' \tag{46}$$

For a finite volume where the magnetization **M** switches to zero at the surface of the body, the Eq. 46 can be written (Becker 1973) using the surface element $\mathbf{df}' = \mathbf{n}' \cdot \mathrm{d}f'$

$$\mathbf{A}(x) = \frac{1}{4\pi}\int \frac{\nabla' \times \mathbf{M}(\mathbf{x}')}{|x - x'|} dV' + \frac{1}{4\pi}\int \frac{\mathbf{M}(\mathbf{x}') \times \boldsymbol{n}'}{|x - x'|} df' \tag{47}$$

The vector potential A can be written by using the current density **J** as

$$\boldsymbol{A}(\boldsymbol{x}) = \frac{1}{4\pi}\int \frac{\mathbf{J}(\mathbf{x}')}{|x - x'|}\, dV' \tag{48}$$

DOI 10.1007/978-981-4560-75-7_121-1

On one hand it is considered that the magnetization of the body is constant and from Eq. 45 follows that the field **B** is also constant. On the other hand from Eq. 47 the volume integral disappears under this consideration.

The conclusion is: there are only surface currents on the body and the current density has no component in direction of the surface normal **n**. For that case Eq. 48 can be written by using the surface current density **j** to

$$\boldsymbol{A}(\boldsymbol{x}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{j}(\mathbf{x}')}{|x - x'|} df' \tag{49}$$

While the volume integral of Eq. 47 disappears than the surface integrals of Eqs. 47 and 49 must be equal and result in the surface current density of

$$\mathbf{j}(\mathbf{x}') = \frac{1}{\mu_0} \boldsymbol{M}(\boldsymbol{x}) \times \boldsymbol{n}' \tag{50}$$

This result means that a surface current density **J**(**x**) generates a constant **B**-field **B**(**x**') under consideration of a constant magnetization **M**(**x**) defined at a surface S' with a surface normal **n**'. This is an exact result that can be used now for a finite cylinder (Fig. 11 right).

The magnetization $\boldsymbol{M}$ is aligned to the x-axis: $\mathbf{M}(\boldsymbol{x}) = \boldsymbol{M}_0 \cdot \boldsymbol{n}_x$. The surface-normal on a cylindrical surface is defined by $\mathbf{n}' = (\sin\varphi, \cos\varphi, 0)$. With Eq. 50 the surface current density is given by

$$\mathbf{j}(\mathbf{x}') = \frac{1}{\mu_0} \boldsymbol{M}_0 \cdot \cos\varphi \cdot \boldsymbol{n}_z \tag{51}$$

This is the current density that is expected: a cosine or sinusoidal surface current distribution. While the cylinder has a finite length the surface current distribution on the fronts of the cylinder is required for a homogenous magnetic field. The surface-normal of the front side is defined by $\boldsymbol{n}' = (0, 0, \pm 1)$ and with Eq. 50 a constant surface current at the front sides of the cylinder $\mathbf{j}(\mathbf{x}') = \pm\frac{1}{\mu_0}\boldsymbol{M}_0 \cdot \boldsymbol{n}_y$ is the result. While access to the inner area of the cylinder is needed for the object under investigation this current distribution can be practically realized only at one side of the cylinder. The RF coils that are closed at one side are called end-caped RF coils.

Saddle- and Birdcage- Coils – An Approximation of an Ideal Current Distribution

A continuous sinusoidal or cosine current distribution is only a theoretical model and in principle this current distribution is approximated by two basic principles (see Fig. 12). (1) the current distribution can be approximated by conductive wires placed on nonequidistant angles using a fixed current $I = I_0$. (2) the current distribution is approximated on radially equidistant positions of the wires on the cylindrical structure and the current I is modulated by a sinusoidal or cosine current I_0.

The simplest case is the approximation by two currents I_0 and $-I_0$ at opposite positions at the cylindrical surface (Fig. 12a). A special case of this approach is the "Aldeman Grant" or "Slotted tube" resonator (Alderman and Grant 1979; Schneider and Dullenkopf 1977) where the current distribution is approximated by two rectangular conducting surfaces. These resonators were used in the early days of MRI. Still today they are common solutions for other applications such as electron paramagnetic resonance (EPR).

The approximation of the current distribution on four points leads to so-called saddle coils (Hoult 1979) (Fig. 12b) and consist of two 120° rotated saddle based on two conductive wires with the current I_0 that are

DOI 10.1007/978-981-4560-75-7_121-1

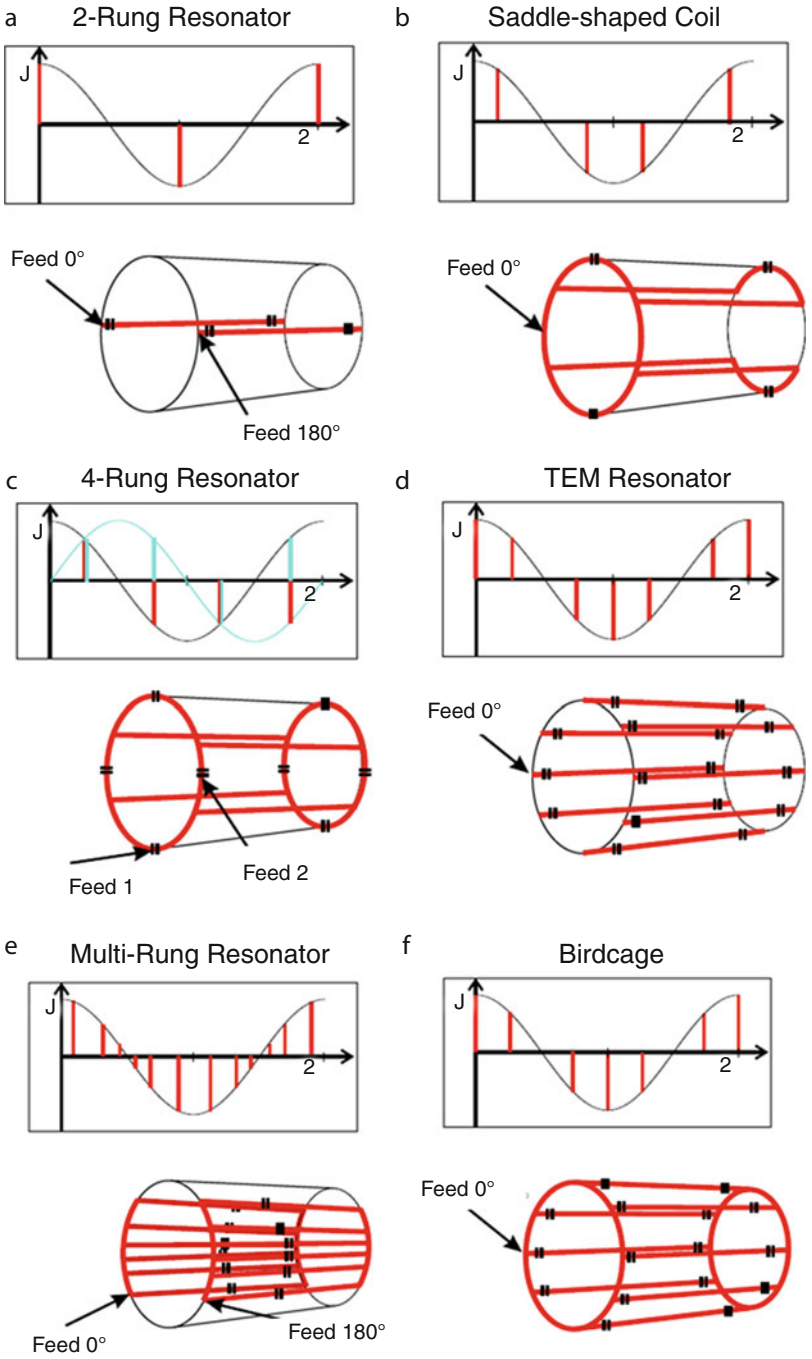


Fig. 12 Approximation of the ideal current distribution by RF coils (Link 1992. With permission of Springer)

rotated by 60° to each other. The value of this angle follows by the demand of the optimal bisection of the Integral of the cosine function in the area of 0° to 90°. This type of RF coil is very popular for the usage in low-field MR systems and for RF coils where the diameter is small enough compared to the wave-length of the RF-field. One drawback is the increased inductivity of the coil wire and the practical usage of this principle is limited by the self-resonance of the coils structure. Below this frequency the coil's structure can be adjusted to every resonance frequency that is required by the introduction of capacitors and can be optimized by the introduction of multiple capacitors that are placed symmetrically on the coil structure. Another draw-back is that saddle coils could not be easily built for quadrature operation

DOI 10.1007/978-981-4560-75-7_121-1

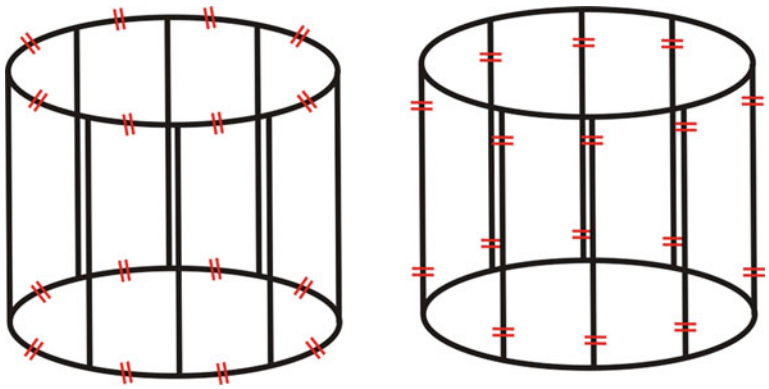


Fig. 13 Principle sketch of a Highpass birdcage RF coil (*left*); Lowpass RF coil (*right*)

(see section Quadrature Detection and Reception). Today saddle coils are commonly replaced by so-called Birdcage-coils (Hayes et al. 1985) since the diameter and Larmor frequency of the MR-system increased in the course of the time. The birdcage coil has become the "golden standard" volume RF coil in MRI systems because of its intrinsic excellent field homogeneity and ability to operate easily in quadrature excitation/reception based on the radial symmetry.

The birdcage coils are based on the approximation of the ideal current distribution on many equidistant points and the principle structure is based (see Fig. 13) on two conductive rings at the front side of the cylindrical surface connected by multiple conductive wires.

There are two basic types of birdcage coils: on one hand the high-pass birdcage on the other hand the low pass birdcage. The name is based on the structure of the latter network (see Fig. 14). Each mesh is based on a low-pass or high-pass filter and either the capacitors C_1 on the end-rings or C_2 at the rungs (legs) between the end-rings are being used. For a cylindrical structure the values of C_1 or C_2 for a low-pass or high-pass birdcage respectively have in principle the same values. When high- and low-pass birdcage structures are combined into one structure, this leads to so-called band-pass or hybrid birdcages.

A birdcage ladder network can be analyzed with the use of the nodal or mesh method based on the Kirchhoff law. In strict terms these rules are only valid if the maximum length of each conducting element is kept below app. 1/20 of a wavelength and if the structure is planar. But in practice the matrix treatment of Leifer (1997) based on the ladder network theory and the Kirchhoff's law (Tropp 1989, 1997) explains the Eigen-modes of the birdcages in a sufficient manner and can be used successfully in many cases (Chin et al. 2002a).

First a low-pass birdcage consists of N legs and meshes is analyzed. Following the rule of the Kirchhoff's voltage law, the sum of all voltages in a mesh equals zero. On the nth mesh (Fig. 14) the voltage of C_2 is on the left hand side $\frac{I_n}{j\omega C_2} - \frac{I_{n-1}}{j\omega C_2}$ and on the right hand side $\frac{I_n}{j\omega C_2} - \frac{I_{n+1}}{j\omega C_2}$. The voltage with respect to the self-inductance of the mesh is $j\omega L_n \cdot I_n$ and the voltages corresponding to the mutual inductances are added $M_{mn}\, m \neq n$ of the mesh cell n to all other meshes m: $j\omega \sum_{m \neq n} M_{mn} \cdot I_m$. Multiplying all terms with $j\omega$ yields to

$$-\omega^2 \sum_{m \neq n}^{N-1} (M_{mn} \cdot I_m + L_n \cdot I_n) + \frac{1}{C_2}(2I_n - I_{n+1} - I_{n-1}) = 0 \tag{52}$$

This equation rules for all meshes n = 1…N of the birdcage under the assumption that all designed meshes are equal and have the same self-inductance $L_n = L =: M_0$. Because of the cylindrical symmetry with N legs the mesh n = N + 1 corresponds n = 1 and n = 0 corresponds n = N. The Eq. 52 can be rewritten for all meshes n = 1..N

DOI 10.1007/978-981-4560-75-7_121-1

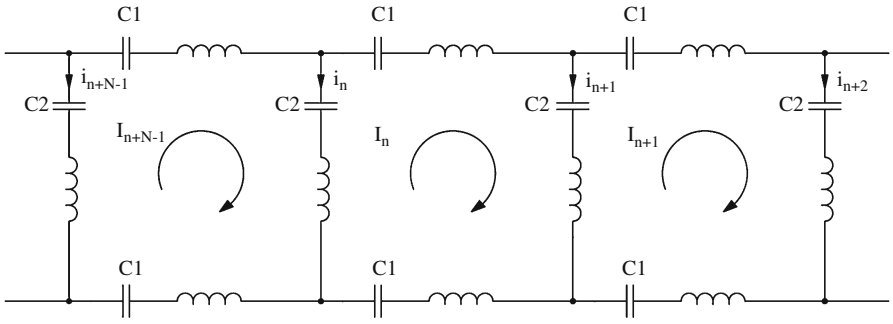


Fig. 14 Ladder network showing reactances and currents of a birdcage RF coil

$$-\omega^2 \sum_{m=0}^{N-1} M_m \cdot I_{m+n} + \frac{1}{C_2}(2I_n - I_{n+1} - I_{n-1}) = 0 \tag{53}$$

and these N equations are yield with the self-, mutual inductances (M_0, M_m) and the capacities C_2 to a linear system of equations (Leifer 1997)

$$\tilde{E}I - \omega^2 \cdot \tilde{M}I = 0 \tag{54}$$

with

$$\tilde{M} = \begin{pmatrix} M_0 & M_1 & \cdots & M_{N-1} \\ M_{N-1} & M_0 & \cdots & M_{N-2} \\ \vdots & \vdots & \vdots & \vdots \\ M_1 & M_2 & \cdots & M_0 \end{pmatrix} \tilde{E} = \begin{pmatrix} -\frac{2}{C_2} & \frac{1}{C_2} & 0 & \cdots & \frac{1}{C_2} \\ \frac{1}{C_2} & -\frac{2}{C_2} & \frac{1}{C_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{C_2} & \cdots & 0 & \frac{1}{C_2} & -\frac{2}{C_2} \end{pmatrix}$$ and the Column vector

$$I = \begin{bmatrix} I_0 \\ \vdots \\ I_N \end{bmatrix}.$$

The Eq. 54 has only solutions for discrete values of ω and can be rewritten to an eigen value equation

$$\tilde{M}^{-1}\,\tilde{E}I = \lambda I \tag{55}$$

The Eigen-frequencies given by the eigen values $\lambda = (i\omega)^2$ and are calculated from the properties of circulant matrices with respect to periodic functions. The Eigen-frequencies are given for a

$$\text{Lowpass birdcage } \omega_k = -\sqrt{\lambda} = \sqrt{\frac{\widetilde{E}}{\widetilde{M}}} = \sqrt{\frac{2}{\widetilde{M_k}}}\left[\frac{1}{C_2} \cdot \left(1 - \cos\frac{2\pi k}{N}\right)\right] \tag{56}$$

Using this approach in the same manner for high pass or band-pass birdcage the results are:

DOI 10.1007/978-981-4560-75-7_121-1

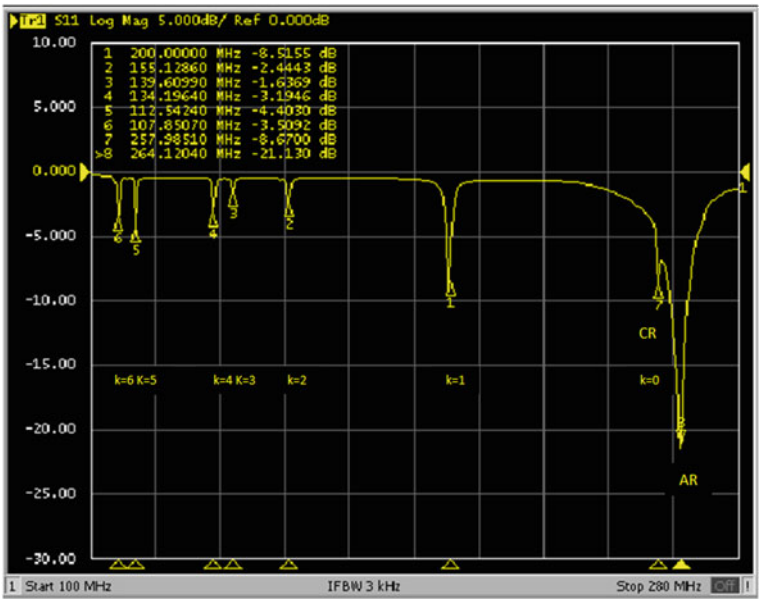


Fig. 15 Meassured eigenmode spectrum of a 12 rung high-pass-birdcage operating for MRI applications at 200 MHz. The modes are labeled with the wave number k and the ring-modes are labeled with AR and CR for the anti- and co-rotating ring-modes. The operating mode for high-pass-birdcages is the secound highestmode with k = 1

$$\text{Highpass birdcage} : \omega_k = \sqrt{\frac{2}{\widetilde{M}_k} \cdot C_1} \tag{57}$$

$$\text{Band-pass birdcage} : \omega_k = -\sqrt{\lambda} = \sqrt{\frac{\tilde{E}}{\tilde{M}}} = \sqrt{\frac{2}{\widetilde{M}_k}\left[\frac{1}{C_1} + \frac{1}{C_2} \cdot \left(1 - \cos\frac{2\pi k}{N}\right)\right]} \tag{58}$$

$$\text{where} \quad \widetilde{M}_k = \sum_{n=0}^{N-1} M_n \cdot e^{-i2\pi kn/N} \tag{59}$$

Based on the symmetry of the matrixes $\tilde{M}$ and $\tilde{E}$ and its periodic function the number of degenerated Eigen-modes pairs are limited to N/2-1.

With this approach all Eigen-modes and their current distributions can be calculated by using the given inductance values and capacity values. For the high-pass and bandpass birdcage two additional modes are present. These modes are a result from the coupling of the end-rings forming a co-rotating or counter-rotating mode. The frequency of the co-rotating mode can be calculated by the Eqs. 57 and 58 using the coefficient k = 0. The counter-rotating mode is calculated (Leifer, 1997) by the number of mesh-cells, the ring capacitors C_1, the self-inductance L_r of the end-rings and their magnetic coupling

$$M_{\text{ring}} : \omega_{cro} = \sqrt{\frac{N}{C_1\left(L_r - M_{\text{ring}}\right)}} \tag{60}$$

The magnetic **B**-field both of these modes are aligned in direction of the cylindrical axis (z-axis). These modes are at the two highest frequencies of the Eigen-mode spectrum. In most cases these modes are uninteresting for MRI due to the fact that the external **B**-field $\mathbf{B}_0$ points in the same direction. Among the number of N/2-1 degenerated Eigen-mode pairs of a birdcage with N rungs there exists only one

DOI 10.1007/978-981-4560-75-7_121-1

Eigen-mode that fits the requirement of Eq. 28, a cosine or sinusoidal current distribution. Apart from the ring-modes the frequency of this Eigen-mode for the high-pass and low-pass birdcage is at the highest frequency (k = 1) and at the lowest frequency of the Eigen-mode spectrum, respectively (Fig. 15).

In principle the structure and dimensions of a RF coil and the operating frequency ω_0 are predefined and the values of the resonance capacitors C_1 and C_2 are unknown. With knowledge of the inductances and mutual inductances these capacitor values can be calculated that are needed for an Eigen-mode at ω_0 and for the required sinusoidal current distribution on the legs of the birdcage. Based on this approach the necessary capacitance values can be calculated by approximation of the inductances and mutual inductance values (Grover 1946; Chin et al. 2002b)..

In most cases, especially when the wavelength of ω_0 is much larger than the dimensions of the RF-coil, this approach is useful for the identification of the desired frequency of the Eigenmode spectrum. The final tuning of the resonant frequency ω_0 can be done by fine adjusting the capacitor values C_1 or C_2 respectively (Fig. 14).

With introduction of numerical field calculation basically based on the Method-of-moments (e.g., commercially provided by Concept II, TET, TU Hamburg-Harburg; FEKO, Altair), Finite Difference –Time domain or Finite element Methods (e.g., commercially provided by MicroWave Studio, CST Darmstadt; Semcad, SPEAG; ANSOFT, ANSYS; XFdtd, RemCom) the calculation of the required capacitor values is much more precise and finally the designer gets the entire field-distributions of the **B** and **E**-fields. In addition the coil structures can be optimized easily by so-called parameter sweeps.

Following the sinusoidal current distribution on the legs or their corresponding cosine voltage distribution (Fig. 17) on the positions of resonance capacitors the desired field distribution on the coils structure can be forced in numerical field simulations by applying the corresponding voltage- or current-sources at the positions of the resonance capacitors. The capacitor value C_n can be calculated from the imaginary part of the input impedance of the multiple sources $Z_{C_n} = \frac{1}{i\omega C_n} = \frac{U_n}{\hat{I}_n}$ where $\hat{I}_n$ is the complex current at the position of the capacitor.

As outlined in section Quadrature Detection and Reception for the calculation of the MRI relevant RF-fields the circularly polarized RF-field has been taken into account. On one hand the counter clockwise rotating component B_1^+ is responsible for excitation anon the other hand the counter clockwise component B_1^- – is the component of the RF-field that is received. In order to qualify the performance of MRI RF coils different $\mathbf{B}_1$-field maps have to be taken into account, depending on the operation mode of the RF-coil for transmission/reception-only or as transmit/receive coil. A typically filed pattern of the transverse $\mathbf{B}_1$-field is presented in Fig. 16 for a high pass birdcage that consists of 12 rungs for the approximation of the required sinusoidal current distribution. The RF-field is linearly polarized and homogenous with respect to the amplitude and phase over a large area of the RF-resonator.

Feeding and Tuning of Birdcages RF Coils

The feeding of birdcage coils is based on the basic principles for feeding of high frequency RF-circuits. Generally all known varieties of feeding networks can be applied for power matching the RF coils to a 50 Ω source. Apart from the inductive coupling of a RF-coil, the capacitive coupled feeding is the most common method connecting RF coils to the RF-chain of an MRI-system. In the following, the high-pass birdcage will be used for a brief introduction on the basic principles.

The diagram below shows the theoretical current- and voltage-distribution on the endrings based on quasi static theory for a balanced (symmetric) feeding. For an unbalanced (asymmetric) feeding the Eigen mode and the corresponding transversal **B**-field points in the same direction as the feeding and the voltage and current distribution change their characteristic distribution vice versa.

DOI 10.1007/978-981-4560-75-7_121-1

Based on the symmetry of birdcages and the cosine or sinusoidal current distribution, the direct feeding on the position at a rung is not preferred while the current on the 90° positioned rung becomes zero and will not contribute to the $\mathbf{B}_1$-field of the coil. In general the feeding is done at one of the end-rings in the geometrical middle of birdcage mesh cell by a symmetric or asymmetric feeding of the resonance capacitor. For asymmetric feeding any known balun-transformation can be applied. The simplest way for feeding a quadrature birdcage is the symmetric feeding by splitting the resonance capacitor C_1 by two capacitors of the value $C_3 = 2 * C_1$ that creates a so-called virtual ground that is required for the perpendicular mode. The RF-shield is used as a common ground of the RF-circuit if the RF structure and the RF screen are in close proximity. The symmetry of the Eigen mode on the birdcage structure is dependent from the feeding and also changes the orientation of the transversal $\mathbf{B}_1$-field inside the structure. Balanced (symmetric) feeding generates a standing wave that is perpendicular to the feeding point and yields to a transversal $\mathbf{B}_1$-field that is perpendicular as well. For an unbalanced (asymmetric) feeding the standing wave is symmetric to the direction of the feeding and yields to a transversal $\mathbf{B}_1$-field that is aligned in same direction.

With the introduction of a serial capacitor C_m to the feeding point the RF-coil is matched to a source impedance of 50 Ω as long as the input impedance of the RF-coil is above 50 Ω. While the RF-coil is loaded with different subjects, additional losses and stray capacitance (section Efficiency, Losses and Signal to Noise Ratio) are introduced. This will change on one hand the input impedance to lower values compared to the empty case and on the other hand the stray-capacities shift the resonance frequency down to lower values. For this reason variable capacitors are required for adjusting the match and tuning of the RF coil. Following the principle of section Quadrature Detection and Reception for quadrature excitation, the coil is fed at two ports that are rotated by 90° from each other.

In principle the frequency of the Eigen-mode of the coil can be affected by changing the capacitance of each of the mesh capacitors, but the optimal position for tuning the frequency of each Eigen mode is at the position of the capacitor with the highest voltage drop (Fig. 17). Using a virtual ground for symmetric feeding the RF-coil (Fig. 18) the optimal tuning-capacitor C_4 + C_1 is 90° rotated with respect to the feeding port V_1. This position is at the opposite side to the port V_2 of the second channel for quadrature feeding of the birdcage. Because of the rotational symmetry and the cosine or sinusoidal voltage distribution, the effect of this capacitor for the second channel is minimal (the smallest voltage drop for the second channel). This assures an independent tuning and matching of the two quadrature channels of the birdcage.

Based on the theory (section Saddle- and Birdcage- Coils – An Approximation of An Ideal Current Distribution) the resonance-capacitors C_1 or C_2 of the mesh cells have the same capacitance value (Fig. 14) if an ideal cylindrical symmetry is provided. The capacitance value C_4 + $C_{T1/2}$ for the tuning is optimal at approximately the same capacitance value of C_1 (Fig. 18). To keep the rotational symmetry this capacitance value needs to be chosen carefully because of the ideal cosine or sinusoidal current distribution that is needed for a perfect homogenous RF-field inside the birdcage. A perfect rotational symmetry is in practice not the case and leads to an inhomogeneity of the transversal RF-field in any case. But this effect can be minimized by building the coil as symmetrically as possible. Different loss loadings will introduce different stray capacities to the coil structure including additional losses to the RF-circuit and for that case the tuning capacities must cover a tuning-range from unloaded to heavy loaded conditions.

In case that the asymmetry of the capacitor value introduced by capacitance C_4 + $C_{T1/2}$ varies widely from the ideal capacitor value of the resonance-capacitor C_1, specific balanced feedings have to be implemented (section Balancing of Birdcage Coils). The Eigen modes of the two channels must be perpendicular for quadrature operation. Any asymmetry leads to a so-called coupling of the modes and can be readjusted by the capacitors (C_5 + C_b) that are positioned diametrically to the feeding ports. At these points the effect for slightly shifting the zero-point of the voltage at the opposite feeding ports of

DOI 10.1007/978-981-4560-75-7_121-1

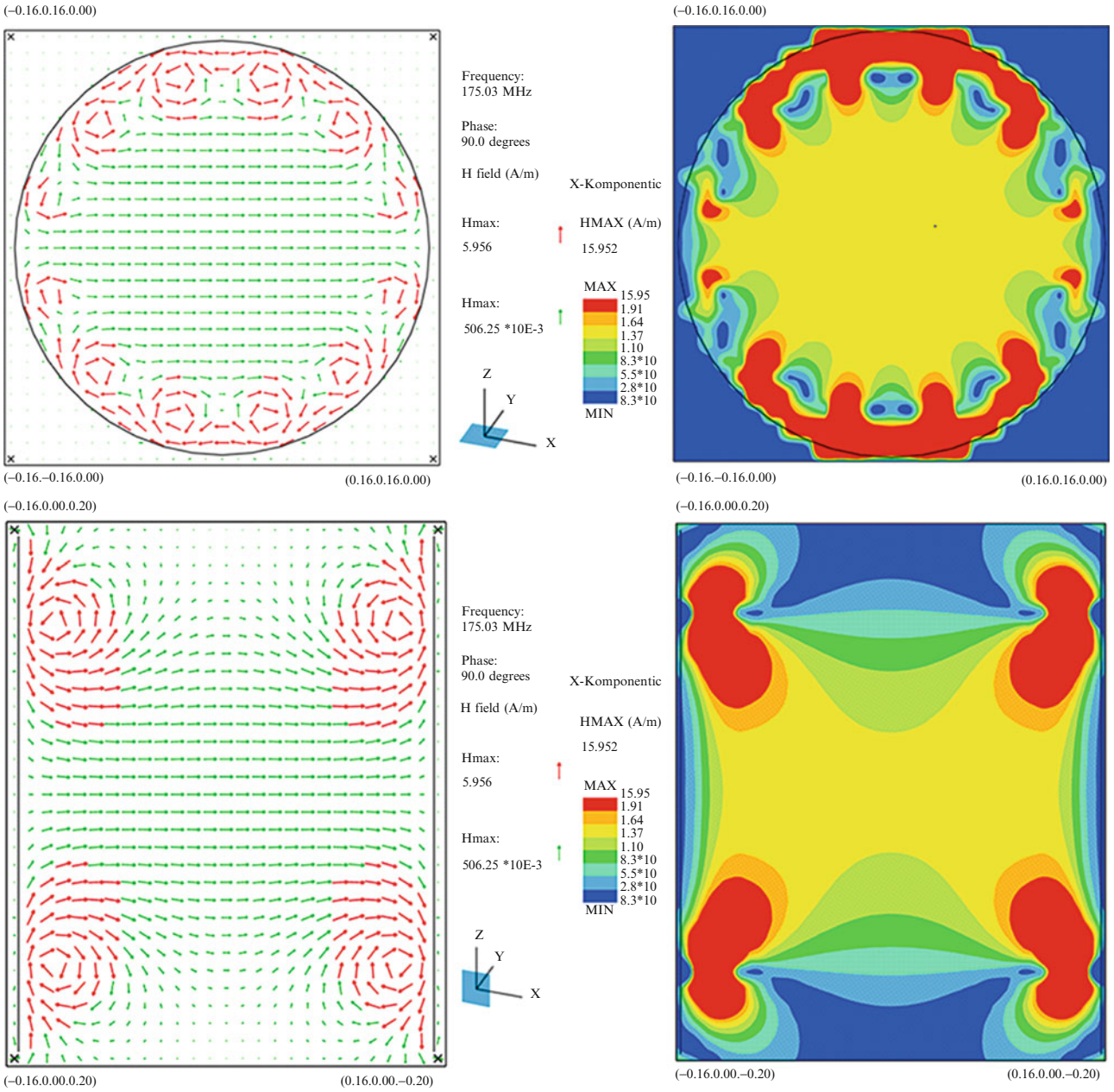


Fig. 16 Simulated $\mathbf{B_1}$-field distribution (MoM using Concept II, TUHH Hamburg, Germany) inside an empty high-pass birdcage operating at 175 MHz fitting the size of a human head. The transversal $\mathbf{B_1}$-field is homogenous over at least 2/3 of the diameter of the RF-coil. The longitudinal homogeneity is limited by the length of the birdcage for a specific operating frequency

both Eigen modes works in the same manner and is done by minimizing the power transfer S12 or S21 of the two quadrature RF channels.

In practice the capacitance values for a real designed birdcage are a compromise to fit the homogeneity of the RF-field, the independency of the quadrature modes and the value range of the variable capacitors to fit the requirements for shifting the resonance frequency of the homogenous Eigen mode for different loading conditions. When the birdcage is used as transmit-only coil (so-called body-coil) in some cases it is sufficient that the body-coil is adjusted once. This adjustment fits an optimal power-matching of a maximum load and the matching is not changed with respect to the different loading conditions. In case of

DOI 10.1007/978-981-4560-75-7_121-1

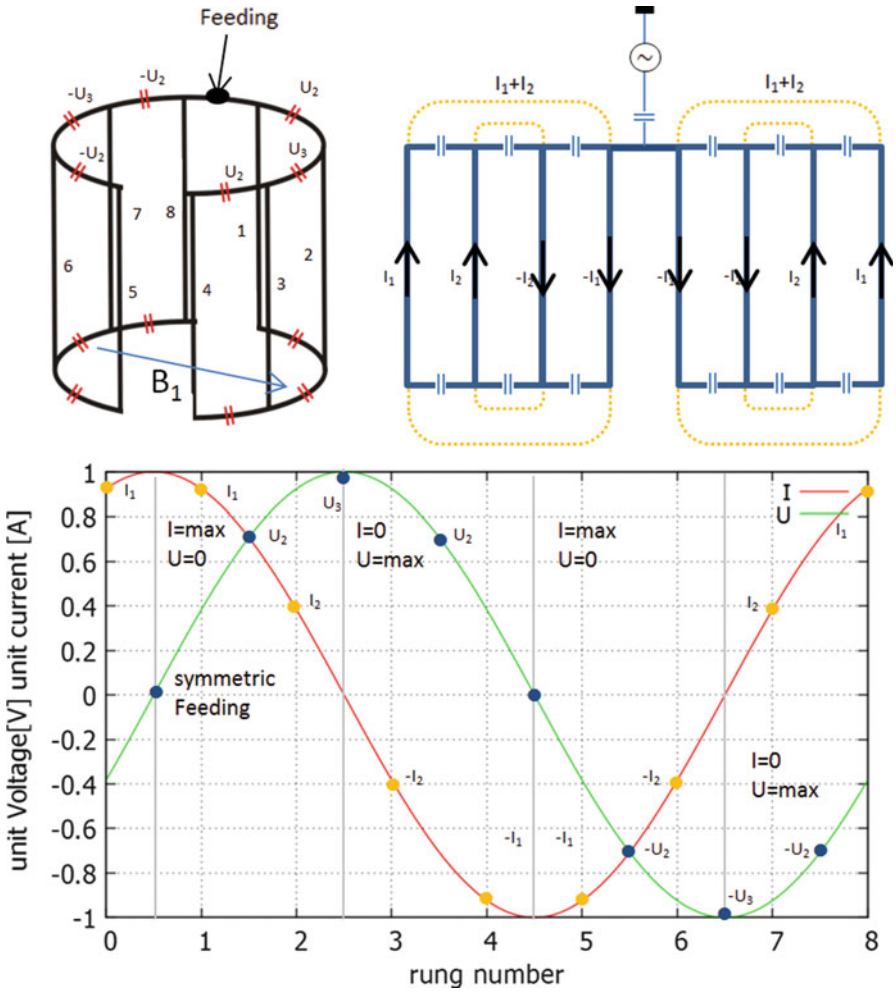


Fig. 17 Voltage and current distribution on a birdcage driven by linear excitation. A standing wave perpendicular to the feeding point is generated that yields to a transversal $\mathbf{B_1}$-field inside the structure

light loading the reflected power will be absorbed in a dummy load, e.g., on one port of the quadrature hybrid. However, in most cases birdcages are tuned and matched for optimal performance with respect to high SNR and low power losses.

Shielding of Birdcage coils

When the birdcage is used as transmit/receive coil, the outer shell of the coil includes in most cases a so-called RF-screen or RF-shield that is most important when the RF coil structure is close to the gradient coil. If the space between the RF-coil and the gradient coil is limited, this screen is glued or molded on the inner surface of the gradient coil. The screen performs three key tasks:

- It minimizes the RF power losses while the screen prevents the coupling of the RF magnetic fields to the gradient coil windings.
- It suppresses the residual RF noise radiated from the gradient coil and shim coil.
- It establishes a well-defined environment for the RF coil in order to generate a homogenous RF-field.

DOI 10.1007/978-981-4560-75-7_121-1

To achieve these tasks the RF shield should perform as a low-pass filter which should be transparent to the magnetic field of the gradient coil that operates at audio-frequency (couple of kHz). The RF shield should also be opaque with low loss to the $\mathbf{B}_1$RF magnetic field of the RF coil. In general the RF screen is designed as a conductive closed cylindrical screen with the thickness of a several RF skin-depth in a manner that the thickness is much smaller than the skin-depth of the audio-frequency of the gradient coil. But such a closed conductive RF screen can support eddy currents induced by the gradient field and thus increases the rise time of the switching gradients and results in imaging artifacts and distortions of the required homogenous $\boldsymbol{B}_0$ -field.

Even if the screen is thin compared to the audio-frequency skin-depth, it is not completely transparent to the high frequency components of the switch gradient field. This results in addition in an interference of the high frequency response and distortions of the gradient waveforms.

The presence of a cylindrical conductive shield can be quantitatively analyzed by the method of images (Smythe 1968) with some approximation to its finite length. Each element of current that is surrounded by a conductive cylinder creates a mirror image with the same current in the opposite direction and located outside of the conductive cylinder. A shield of a diameter D reduces the RF-field at the center compared to an unshielded coil carrying the same currents (Mispelter et al. 2006) by $B_1^{\text{shielded}} = B_1^{\text{unshielded}}\left[1 - \left(\frac{d}{d_s}\right)^2\right]$. The RF shield has also an overall effect in a reduction of the transverse homogeneity of the RF-field compared to unshielded coils. All these effects are strongest for closely fitting RF screens, but in any case the dimensions shield- versus coil-diameter are parameters that can be optimized if the overall geometrical dimensions of the MR-system support this. The current distribution produced by a linear excited birdcage is shown in Fig. 19. The currents on the RF shield are the mirror currents resulting from the currents on the birdcage-structure and forming saddle-shaped loops (Fig. 19). For linear excitation of the birdcage this saddle-shaped current distribution is stationary on the screen but rotates with the resonance frequency when the birdcage is fed in quadrature operation.

To prevent eddy currents on the shield and fit the requirements of a closed RF-screen, double sided copper sheets (Jin 1999) based on a thin dielectric layer (e.g., FR4 fiberglass or polyimide circuit board material) are commonly used.

The general design is based on the principle that each copper sheet is divided into multiple rectangular elements by etching or cutting narrow gaps (Fig. 19 bottom). The pattern on the other side is designed in the same manner and includes an offset of half of the elements-width of the rectangular elements. The overlapping elements separated by the dielectric layer serve as capacitor and the shield itself can be considered as a number of capacitors in series. The capacitance is proportional to the overlapping area and is inversely proportional to the thickness of the dielectric layer. The main impedance of the shield is inversely proportional to the frequency and capacitance.

For RF the conductive elements separated by the gaps of one side are capacitively coupled to the ones of the other side and therefore when the overlapping areas are sufficient in size, the conductive layers are electrically shorted. The shield serves as a single continuous conductor for the high frequency $\mathbf{B}_1$RF field. The eddy currents caused by the gradient coils are suppressed if the rectangular elements are sufficiently small. The gaps between the elements interrupt the current path of the eddy currents induced by the gradient fields. For example a shield having an impedance of 1 Ω at 64 MHz would have an impedance of 6.4 k Ohm at 10 kHz – this means that the shield will be opaque for RF and transparent for the low frequency gradient fields. The element-width and -length and the number of gaps are different for each RF-coil/gradient coil design and have to be optimized for optimal performance of the RF-coil e.g., by introduction of additional discrete capacitors in areas of high RF- current densities.

DOI 10.1007/978-981-4560-75-7_121-1

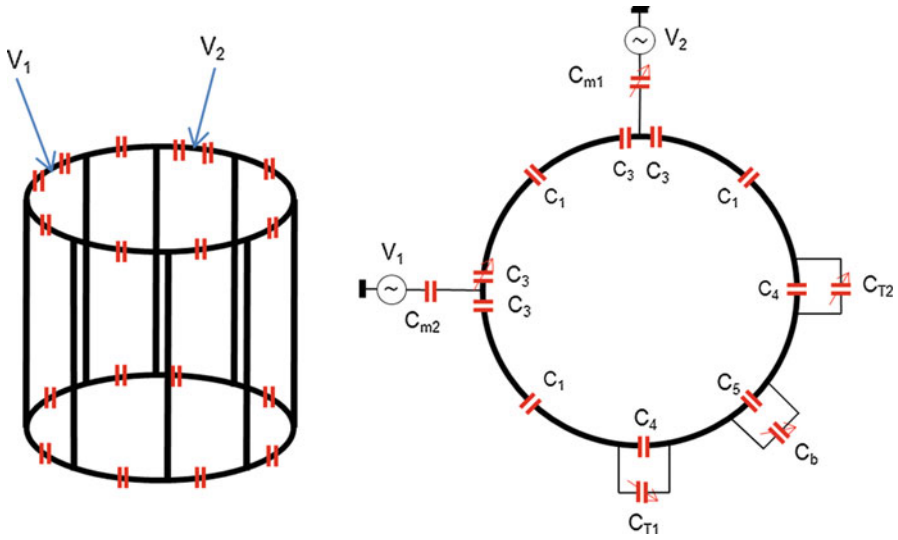


Fig. 18 Symmetric feeding of a high-pass birdcage and the corresponding end-ring configuration with matching-, tuning- and balancing- capacitors

Balancing of Birdcage coils

Many birdcage coil parameters can be optimized but in most cases the freedom for optimization is for practical reasons (unfortunately) limited. The minimum coil length is given by the homogeneity that is required by the application that will be performed and the maximum coil-length is limited by the gradient-field linearity. Ideally the RF coils should be built as short as possible while the SNR decreases with an increased RF-coil volume.

The number of rungs can be optimized with respect to the transversal homogeneity and the coil losses introduced by the mesh capacitors and find its limits with respect to the number of rungs by the so-called "Millipede Coils" (Wong 2001) including a couple of dozen rungs. The inner and outer diameter is in most cases limited by the inner diameter of the gradient-coil and the subject under investigations. The distribution of the electrical field, parasitically high electrical fields caused by local capacitors, can be optimized using distributed capacitors and locations as far away as possible from the loading.

But, beyond this optimization the most challenging part for coil optimization is the balancing with respect to the homogeneity of the transverse $\mathbf{B}_1$-field that the birdcage operates optimally for all loading conditions. The tuning presented in section Feeding and Tuning of Birdcages RF Coils has its limit when the tuning capacitor values are too far from the optimum value to generate an ideal current distribution. In most cases the acceptable deviation of the current from the ideal one have to be less than 10 % and larger deviations typically lead to inhomogeneous $\mathbf{B}_1$-field distributions shown in Fig. 20. Within the last decades many clever methods have been developed to improve the tuning range of birdcages so as to maintain its symmetry.

Multi-port feeding for the excitation of birdcages to keep the symmetry for a broad tunable range of different loading conditions is explained a later section. The use of so-called tuning-rings for balancing the meshes of a birdcage is another option. This principle is based on the capacitive connection on two diametrically opposed points of the tuning-rings to the birdcage (Fig. 21). The tuning-ring must be capacitively shorted to prevent eddy currents and connected to the end-ring of the birdcage via a tuning capacitor and a short connection or by two tuning capacitors, respectively. The currents from the feeding port split equally to the two tuning-ring segments and the capacitive coupling to the tune-ring moves the resonance-frequency of the desired homogenous mode without change of any capacitor values C_1 of the mesh cells of the birdcage. The principle of balancing by tuning-rings can be extended to both ends of the

DOI 10.1007/978-981-4560-75-7_121-1

end-rings (Mispelter et al. 2006) and is only limited by the mechanical effort for driving the tuning capacitors in equal and symmetric manner.

Volume Coils for High- and Ultra-High-Field MRI Applications – TEM Coils

For dimensions where the wavelength of the RF-resonance frequency becomes comparable with the dimensions of a volume coil the use of low and high-pass birdcage coils comes to a limit. The resonance frequency and inductance are dependent and limited by the dimensions of the end rings that are restricted by the inner diameter. The so-called transmission line (TEM) resonator is a good choice if volume resonator designs are associated with problems regarding coil-efficiency at high magnetic fields $\mathbf{B}_0$. The basic TEM design derives from the first coaxial cavity resonators for NMR of Pound and Purcell (Purcell et al. 1946) and further developed by Roeschmann (Röschmann, 1995) and Vaughan et al. (1994).

In principle when the "floating" end rings of a low-pass resonator are connected to the ground on the RF-shield, this design leads to a transmission line resonator (TEM coil) (Fig. 22). The TEM coils return path of the rung-currents follows now on the RF-shield rather than on the end-rings. The transversal homogeneity is dependent from the number of transmission-lines and provides in principle at the center of the RF coil the same transversal homogeneity as standard low and high pass birdcages with the same number of rungs and dimensions. While the end rings are missing, this design can be easily adapted to split-coil designs when the access to the inner volume is limited, e.g., for dedicated small volume coils for extremities of humans. For this case only the RF-shield has to be split in two parts and that can be realized much easier than for birdcages while the end-rings of the birdcages need an electrical connection in addition to the split RF-shield.

The longitudinal homogeneity is different from that of birdcages while the effects of interferences of the end ring currents are missing. For that reason the TEM-coils can be built shorter in z-length than birdcages for the same transaxial homogeneity. In Fig. 23 the transaxial homogeneity for two RF coils (in dimensions identical) are simulated and the normalized magnitude of the transversal RF-field is plotted along the central axis aligned with the z-axis. The RF-field of the TEM-coils has much less drop and provides at the end of the FOV approximately ± 30 cm, a 3 dB higher transversal RF-field. For same cases this effect can yield in a higher efficiency of the TEM-coil compared to standard birdcage coils. However the effort to build a homogenous TEM coil is in most cases more difficult than compared to standard birdcages. Especially for heavily loaded cases the excitation of the RF-coil at only one feeding port results in an inhomogeneous transversal RF-field. The Balancing by tuning rings in the same manner as for low- and high-pass birdcages could not be implemented and multiport excitation of the ring-structure is required either with the use of lambed elements circuits or a power-splitter/combiner (Fig. 24).

A common method for driving a TEM-coil is the multiport excitation that requires additional power splitter/combiner (Vaughan and Griffiths 2012, pp. 158). In Fig. 24 left the feeding of a quadrature driven transmit/receive TEM-coil is presented. The first transmit/receive switch separates the incoming power to the TEM-coil from the receive path of the MR-system. The power-splitter/combiner splits the power in two 180° phase shifted signals. These signals are split by two additional quadrature hybrids in four in amplitude identical signals that are phase shifted by 0°, 90°, 180°, 270°. For the reception the receive signal is combined in the reverse direction. Instead of using discrete power-splitters/combiners, cable baluns and cable networks can be used, e.g., half-wavelength baluns. The driving of a TEM-coil by multiport excitation can be easily extended to all TEM-rungs (Vaughan and Griffiths 2012, p. 169ff). This yields to TEM-arrays became popular within the last years for ultra-high field applications. For this purpose neighboring transmission line-elements are capacitive decoupled and each TEM-element is matched to 50 Ω and driven independently (Fig. 24 bottom). For receive RF-arrays commonly so-called local surface RF coils are used.

DOI 10.1007/978-981-4560-75-7_121-1

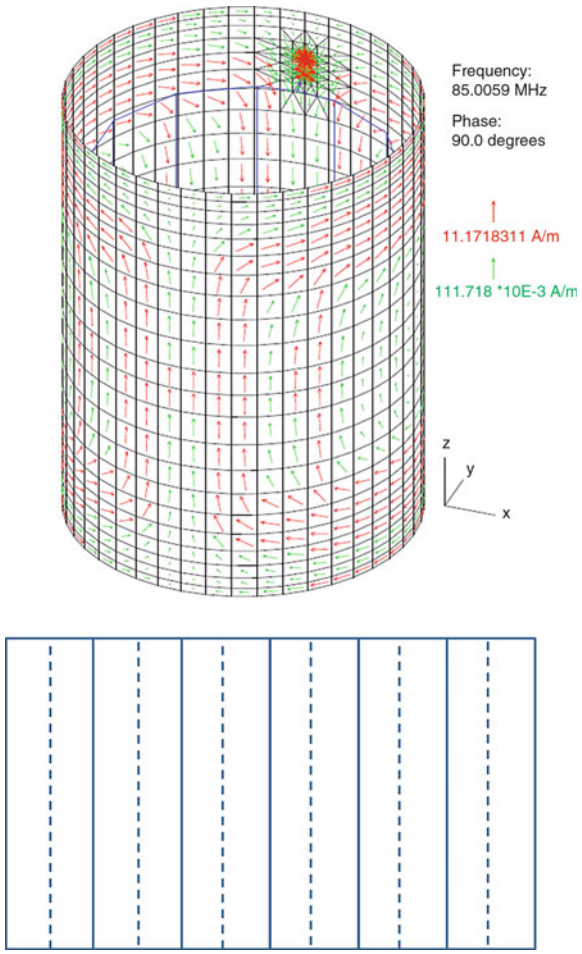


Fig. 19 *Top*: Simulated RF-current distribution (MoM using Concept II) on a RF-shield for linear excitation of a Birdcage that fits the size of a human head. *Bottom*: Example of an unwrapped RF-shield made of double layer copper sheets

Local RF Coils

From the early days of MRI to today local RF coils play an important role when high sensitivity is required, e.g., for enhanced spatial resolution. This includes single-loop RF coils and multi-loop coils of various shapes (Fig. 27). These coils are generally much smaller than volume coils and have therefore a higher SNR because they receive only noise from nearby regions and provide a much higher sensitivity when they are carefully designed. These local coils are optimized for high SNR within a certain volume of interest (VOI) and typically do not provide a homogenous sensitivity like a birdcage volume-coil. Local RF coils are commonly used either as transmit-receive coils for dedicated MRI and magnetic resonance spectroscopy (MRS) applications or used in receive-only RF coils in combination with a volume coil providing the required homogenous $\mathbf{B}_1$-field for excitation.

Single-loop coils can be arranged as elements of a large-scale RF-array (section Receive Array Coils) for simultaneous parallel reception (section Receive Array Coils for Parallel Imaging) of the MR receive signals. Dedicated MRI applications can be performed by so-called under sampling of the k-space with the result of a significant acceleration of image acquisition time (section Receive Array Coils for Parallel Imaging). This imaging technique based on the prior information of the sensitivity profiles for each of these array-elements. When the local surface coils operate as transmit/receive or transmit-only array coils these coils can be used for parallel excitation while the generated $\mathbf{B}_1$-field of the RF coil array is applied for shaping the transmit $\mathbf{B}_1$-field in a specific manner (section Transmit Array Coils).

DOI 10.1007/978-981-4560-75-7_121-1

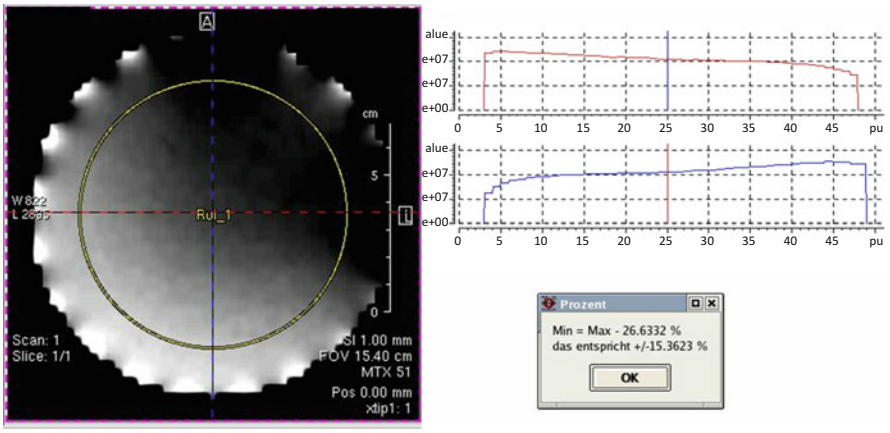


Fig. 20 Inhomogeneities of the transversal RF-field of a birdcage resonator caused by adapting the frequency-shift and impedance match to 50 Ω when the coil is loaded by an conductive, lossy load; *Right*: the diagrams plotted the sensitivity along the central axis though the center of the RF-coil. *Left*: nonoptimized RF-coil – inhomogeneities caused by change the RF-coil tuning and matching from unloaded to loaded conditions

The most commonly used and simplest surface coil is the planar loop coil. For the design of the loop coil the coil dimensions should be chosen carefully to maximize the SNR within the desired VOI. In the following we consider quasi-magneto static conditions and that the planar loop coil is placed directly onto a semi-infinite conducting load aligned in the z-x-plane. Then the magnetic field on the central axis of the loop coil is given by (Mispelter et al. 2006)

$$B_1 = \frac{\mu_0}{2} I \cdot \frac{R^2}{\left(y^2 + R^2\right)^{\frac{3}{2}}} \tag{61}$$

R is the loop radius and y the distance on the coil axis to the sample. If the case is considered that the coil resistance is smaller than the sample losses (sample loss dominates regime) then the coil losses could be neglected. Under the boundary condition of a semi-infinite sample, ignoring the gradient of scalar potential and considering only the vector potential A, the sample resistance R_s can be approximated (Vaughan and Griffiths 2012, pp. 42) as

$$R_s \approx \frac{1}{3} \mu_0^2 \sigma \omega^2 R^3 \tag{62}$$

Where R is the average radius of the coil and σ is the tissue conductivity. Finally for the SNR setting the coil current I to unity (unit current) the result for the SNR is:

$$\mathrm{SNR} \approx \frac{\omega_0 B_1}{\sqrt{R_s T_s}} \propto \sqrt{\frac{R}{\left(y^2 + R^2\right)^3}}. \tag{63}$$

The required coil radius R_0 for maximum SNR at a predetermined depth y is found to be

$$R_0 = \frac{y}{\sqrt{5}} \tag{64}$$

DOI 10.1007/978-981-4560-75-7_121-1

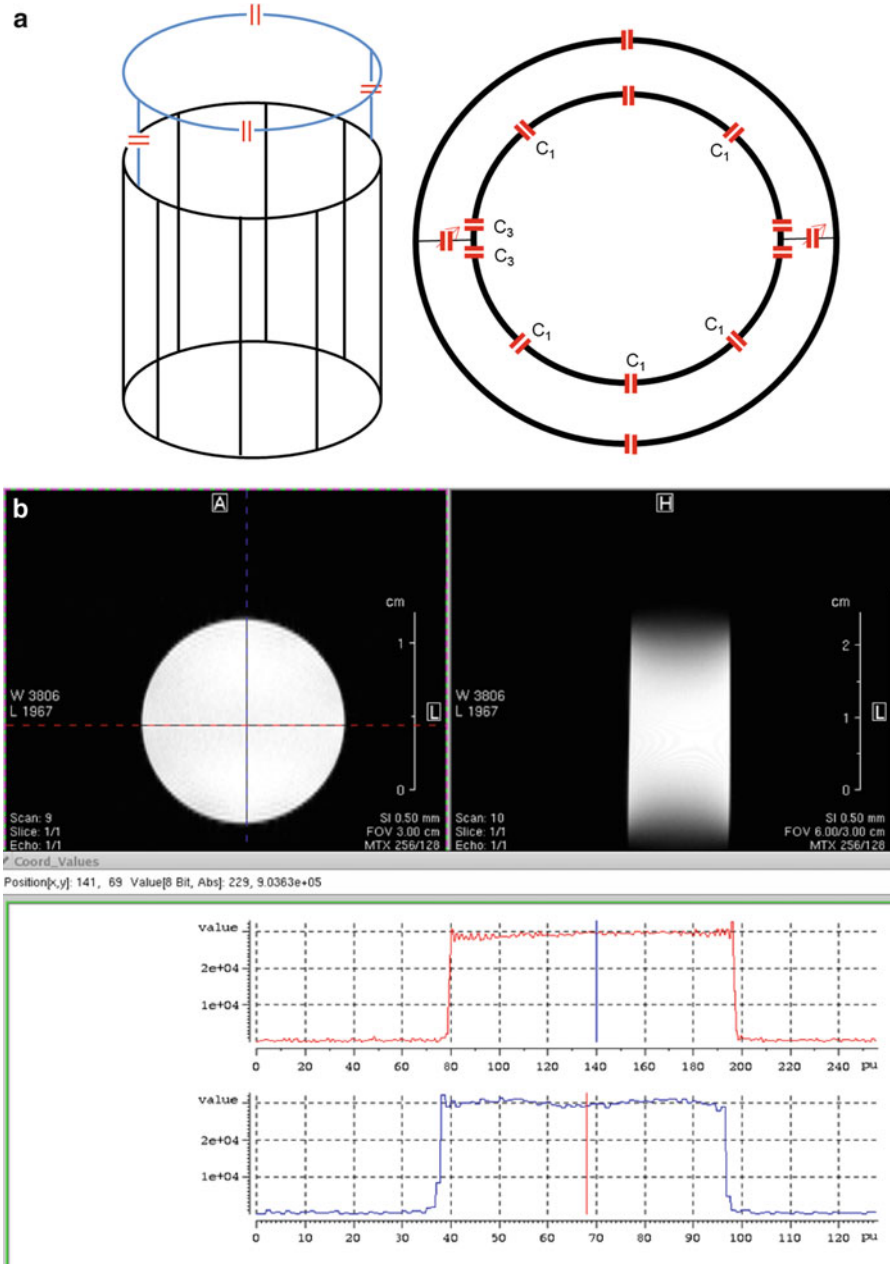


Fig. 21 *Top*: Balancing by a guide ring for homogenous $\mathbf{B}_1$-field distributions inside birdcages for the tuning of different loading conditions; *Bottom*: optimized birdcage – "quasi" perfect homogeneity of the transversal RF-field also for heavily loading conditions

by the differentiation of the Eq. 63 with respect to R. This is the simplest main design rule for planar loop coils. A loop coil with a radius smaller than R_0 is insufficient with respect to the sensitivity at the required depth. A loop coil with a radius larger than R_0 is too sensitive to sample noise. This is clearly a rough estimation only and the approximation considers only the sample noise and that the RF wavelength is large compared to the sample- and coil- dimensions. But this result gives a good starting point for more detailed investigations with respect to optimization e.g., by numerical field simulations.

Another common operation mode of local RF coils is the receive-only operation in conjunction with a homogenous so-called actively decoupled volume RF-coil. These receive RF coils are optimized for high

DOI 10.1007/978-981-4560-75-7_121-1

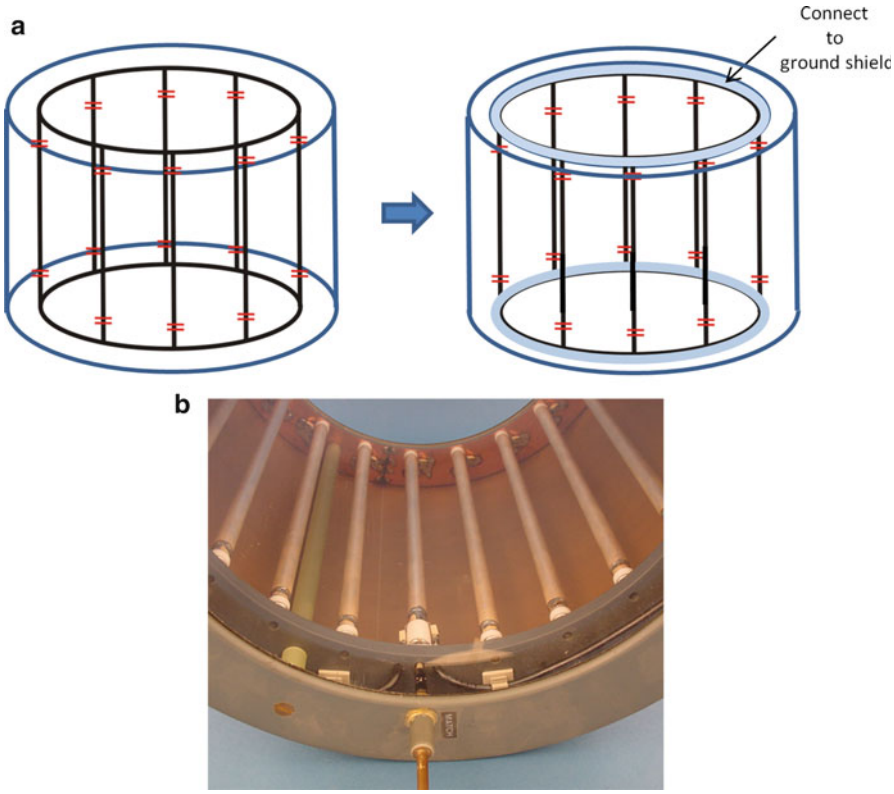


Fig. 22 *Left*: Basic sketch of a transmission line (TEM) resonator: connecting the end-rings to a common ground shield yields to a set of transmission lines that generate a sinusoidal current distribution for a homogenous transversal RF-field; *Right*: 16 element TEM-coil for human head applications at 3 T – the transmission-line elements connected via capacitors to the slotted ground shield

SNR within a certain volume of interest. For this purpose the RF circuit consists of the following components (Fig. 25):

1. The RF-coil structure: the size is optimized to fit the maximum SNR inside the required VOI (MRI coil).
2. Active decoupling/detuning circuit: suppression of induced currents during transmission of the RF from the transmitter coil.
3. Tuning and matching network of the RF-coil: for power matching the Rf coil to 50 Ω.
4. Impedance transformation network: minimization of the noise contribution of the preamplifier.
5. Low noise preamplifier with low or high input impedance: for so-called preamplifier decoupling.
6. Shielded RF-cable.
7. Balun or cable traps; minimization of so-called shield currents on the RF-cable and thus prohibit the cable itself from action as RF-antenna.

The Receive-only coils must be decoupled from the surrounding RF-Coil during the transmission of the RF-power. The transmission RF-power must never penetrate the receive chain. In addition induced currents must be avoided at the receive coil structure that disrupt the homogenous transmit RF-field. The currents induced in the receive coil during transmission create a magnetic field that adds or subtracts from the homogenous RF-field of the transmission RF coil. There are a few of different designs for such decoupling circuits but mostly based on the principle of actively switching PIN diodes for introduction of

DOI 10.1007/978-981-4560-75-7_121-1

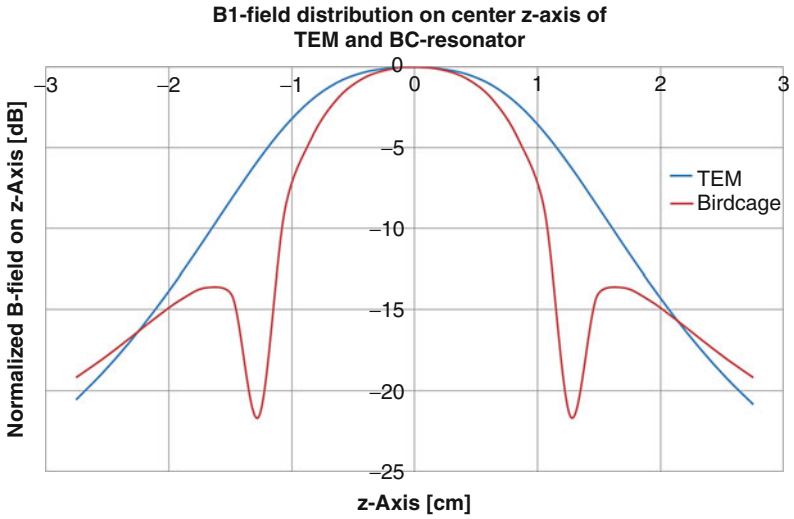


Fig. 23 The simulated and normalized *B*1 field magnitude along the center axis aligned with the z-direction of a birdcage- and TEM-coil at 11.7 T; the coil elements (rungs) for the birdcages and the TEM resonator for small animal applications are kept identical to 5 cm

an additional circuit that generates high impedance. Very common is the introduction of so-called pole circuits (Fig. 25 top). During the transmission the diode D (Fig. 25 bottom) is forward biased and forms with one of the resonance capacitors or the symmetry matching capacitor C_s a blocking resonant circuit using an additional inductance L_D.

During the receive operation the diode is reversed biased, thus removes the blocking circuit from the resonance structure. The DC-connections to the diode are in general made by so-called chock inductances, a high inductance having a large impedance value at the resonance frequency ω_0. These inductances block the RF from the DC-path and isolate the RF coils structure from additional noise from the DC-connections. Some other common used decoupling techniques for decoupling receive-only coils from a transmitter RF-coil are briefly mentioned:

1. Transmission lines –quarter wavelength cable
2. Passive decoupling by crossed diodes
3. Diodes in the RF coil circuit

For optimum performance with respect to SNR it is advantageous to place the preamplifier as close as possible to the feeding port of the RF-coil. The preamplifier is assumed to be calibrated to $R_{opt} = 50\,\Omega$ so that the noise performance optimum is at this impedance. Then, the coil is tuned and matched to R_{opt}. However in the reverse direction the input impedance Z_{in} of the preamplifier is usually considerably larger or smaller than R_{opt}. The preamplifier impedance Z_{in} is transformed via impedance transformation circuits (e.g., Γ-transformation) to a quasi-open circuit at the feeding port of the RF-coil (Fig. 25 top). This procedure is called preamplifier decoupling (Roemer et al. 1990; Reykowski et al. 2002). This blocks the currents even though the RF coil is receiving the MRI-signal, but if very little current flows in the RF coil, no noise or signal is coupled to other RF coils. This technique is commonly used for the decoupling of RF-structure that cannot be decoupled by traditional decoupling techniques such as inductive or capacitive decoupling e.g., when the RF coils are nonadjacent and too far away from each other.

The basic principle for inductive decoupling is based on the reduction of the mutual inductance of adjacent RF coils. A pair of identical resonant loops are considered with their self-inductance L_1 resonating at the same frequency ω_0. If these loops are brought into close proximity, the resonance will

DOI 10.1007/978-981-4560-75-7_121-1

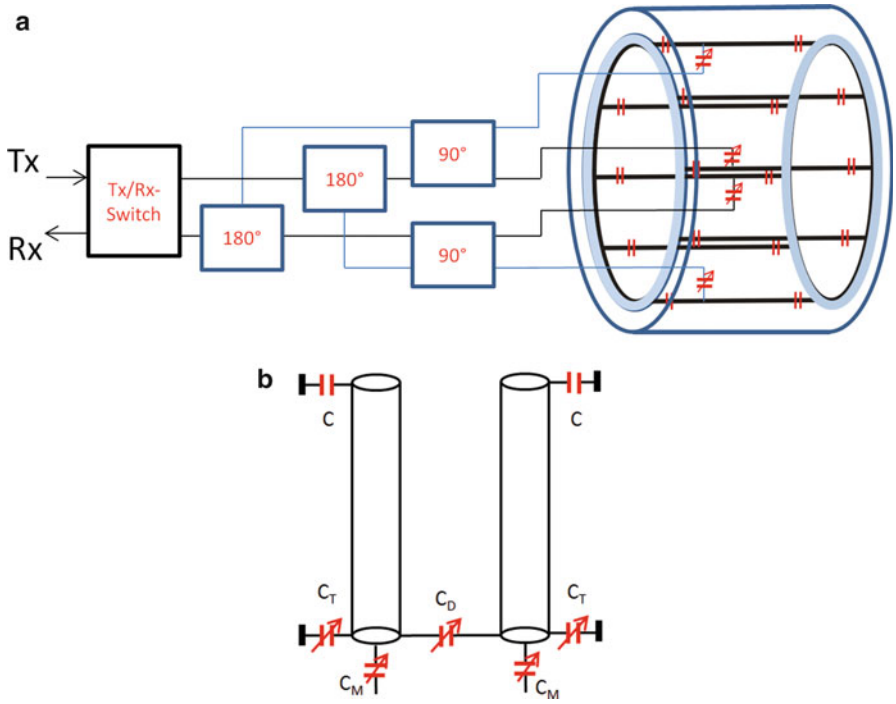


Fig. 24 *Left* – Four-port drive for the excitation of a RF-coil; the RF coil is feeded at 4 ports that are rotated by 90°. The power at the ports have almost the same amplitude and phase shifts of 0°, 90°, 180° and 270° right – Capacitive decoupling of neighboring elements by C_D; the elements tuned and matched by C_T and C_M respectively, the capacitors C and C_T connected to the ground shield.

split into two frequencies, one shifted above and the other shifted below ω_0, and results in a loss of sensitivity at ω_0. This coupling of the RF coils can be characterized by the coupling coefficient (Mispelter et al. 2006)

$$k = \frac{L_{12}}{L_1 \cdot L_2} \tag{65}$$

and the two frequencies can be calculated as

$$\begin{aligned} f_{\text{upper}} &= f_0(1+k) \\ f_{\text{lower}} &= f_0(1-k) \end{aligned} \tag{66}$$

Since the circuits now resonate away from their original resonance frequency ω_0, the efficiency of the combined coil is highly reduced. In addition, signal and noise are also transferred though the mutual inductance L_{12} from one coil to the other. The simplest method to reduce this coupling is the relative positioning by an overlap of the adjacent coils to force the mutual inductance to zero (Fig. 27) when the mutual inductance is predominant. For a set of two circular loop-coils the optimum distance of the centers of the coils is approx. 0.78 times of the loop diameter. For two identical rectangular loop-coils the optimal distance increases to 0.86 times of the linear dimension of the coils. For combinations like butterfly and single loop coils (Fig. 27. bottom right) the maximum decoupling can be achieved when the coil centers overlap exactly.

Here are other popular decoupling methods:

DOI 10.1007/978-981-4560-75-7_121-1

- Inductive decoupling by transformers uses two small inductors coils each connected to one of the RF coils and placed very close to each other so that their mutual coupling neglects the mutual coupling of the two RF coils. This method is useful if nonadjacent RF coils can't be decoupled by an overlap. In some cases, where overlapped RF coils are not suitable (e.g., for parallel imaging) the use of transformers is a good choice.
- Capacitive decoupling uses capacitors and for adjutant coils. So-called shared capacitors are used that replaces the inductive decoupling of the both RF coils.
- Extending the capacitive decoupling to multiple surface coils (e.g., array coils) yields to so-called external latter networks (so-called open-circuit voltage method). The capacitive or inductive elements needed for the decoupling are calculated as part of a 2n-port circuit network connecting the n input ports of the array with the n ports of the receiver or transmitter channels. In theory it is possible to transform a coupled n-port-system characterized by an impedance matrix Z into a decoupled system by a latter network (Lee et al. 2002).

Receive Array Coils

A major drawback of local RF coils is the limit with respect to the field of view (FOV). But this limitation can be overcome by combination of multiple surface coils. The signals detected by the individual coil-elements of the array coil are fed into the separate receiver channels and the final signal combination is done on a pixel-by-pixel basis on the individual processed images of the array coil elements (Roemer et al. 1990). This approach preserves the high SNR of each coil-element and offers the advantage so that the array covers a much larger FOV (Figs. 28, 32).

The performance of array coils highly depends on minimal coil coupling and thus on the signal distribution and noise correlation among the coil-elements in the array. Array coil-elements that are directly positioned to each other can be decoupled by inductive/capacitive decoupling as described in section Local RF coils. The coil-elements that are not direct neighbors usually have a fairly low mutual coupling that can be sufficiently reduced by preamplifier decoupling as described in the previous section.

After simultaneous acquisition of the RF-signals, the MR-data are reconstructed and the result are N individual images for each of the N each array-elements. The performance of the individual coil elements is not the only one design requirement. The overall performance of the array coil is dependent both on the sensitivity-distribution of the local coil-elements and on the coupling between the individual coil-elements. For optimal signal-to-noise ratio (SNR) the pixel signals $S_i(x,y,z)$ from each reconstructed image are weighted by a complex weighting factor $W_i(x,y,z)$ and added up to a combined signal (Roemer et al. 1990; Reykowski et al. 1995)

$$S_c(x,y,z) = \sum_{i=0}^{N} W_i^*(x,y,z) \cdot S_i(x,y,z) = W^H S \tag{67}$$

where W^H is the conjugate transpose of the column vector W including all weights W_i and S is the vector including all signals S_i. The noise correlation between the coil-elements can be expressed in a noise correlation matrix $\Psi = \begin{bmatrix} n_{11} & n_{21} & n_{j1} \\ \vdots & \ddots & \vdots \\ n_{1j} & n_{11} & n_{ij} \end{bmatrix}$ where n_{ij} is the noise power added by coil i to the coil j and all

DOI 10.1007/978-981-4560-75-7_121-1

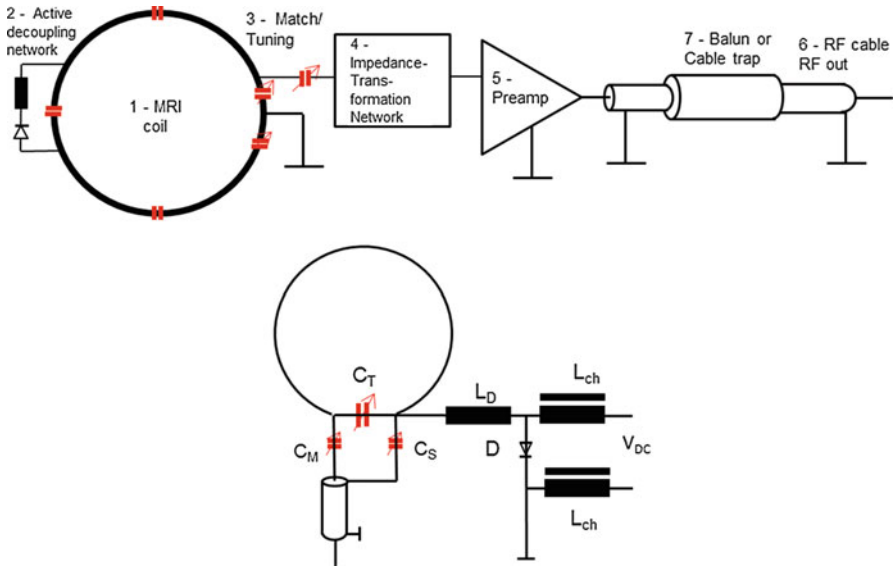


Fig. 25 *Top* – Brief description of a local receive-only coil system; *Bottom* – simple circuit for symmetric feeding and decoupling of a receive-only coil

diagonal elements n_{ii} are the noise power of the individual coils in absence of the other coil-elements. The squared SNR^2 can be expressed now by

$$\mathrm{SNR}^2 = \frac{\left(W^H S\right)^2}{W^H \Psi W} \tag{68}$$

The differentiation of the Eq. 68 with respect to the weights W and sets the result to zero, this yields in a set of optimal weights

$$W^{\mathrm{opt}} = C \ \ \Psi^{-1} S \tag{69}$$

Including a normalization factor C. The best obtainable SNR is dependent on the signals of the individual array coil-elements and the noise correlation between the coils-elements

$$\mathrm{SNR}^2_{\mathrm{opt}} = S^H \Psi^{-1} S \tag{70}$$

The information of the noise correlation is required to achieve the optimal SNR in the combined image. If the noise correlation between the individual coil-elements is sufficiently low, the noise correlation Matrix Ψ can be reduced to a diagonal matrix where the off-diagonal elements are zero and the squared SNR can be expressed by

$$\mathrm{SNR}^2_{\mathrm{opt}} = \sum\nolimits_i \frac{S_i^2(x,y,z)}{n_{ii}} = \sum\nolimits_i \mathrm{SNR}_i^2 \tag{71}$$

Thus the sum-of-squares combination is the simplest method for combining images of array coil-elements and yields in a SNR that is equal or larger than the individual coil-elements. Most important for achieving good results with this approach is an efficient inter coil-**de**coupling of all coil elements.

DOI 10.1007/978-981-4560-75-7_121-1

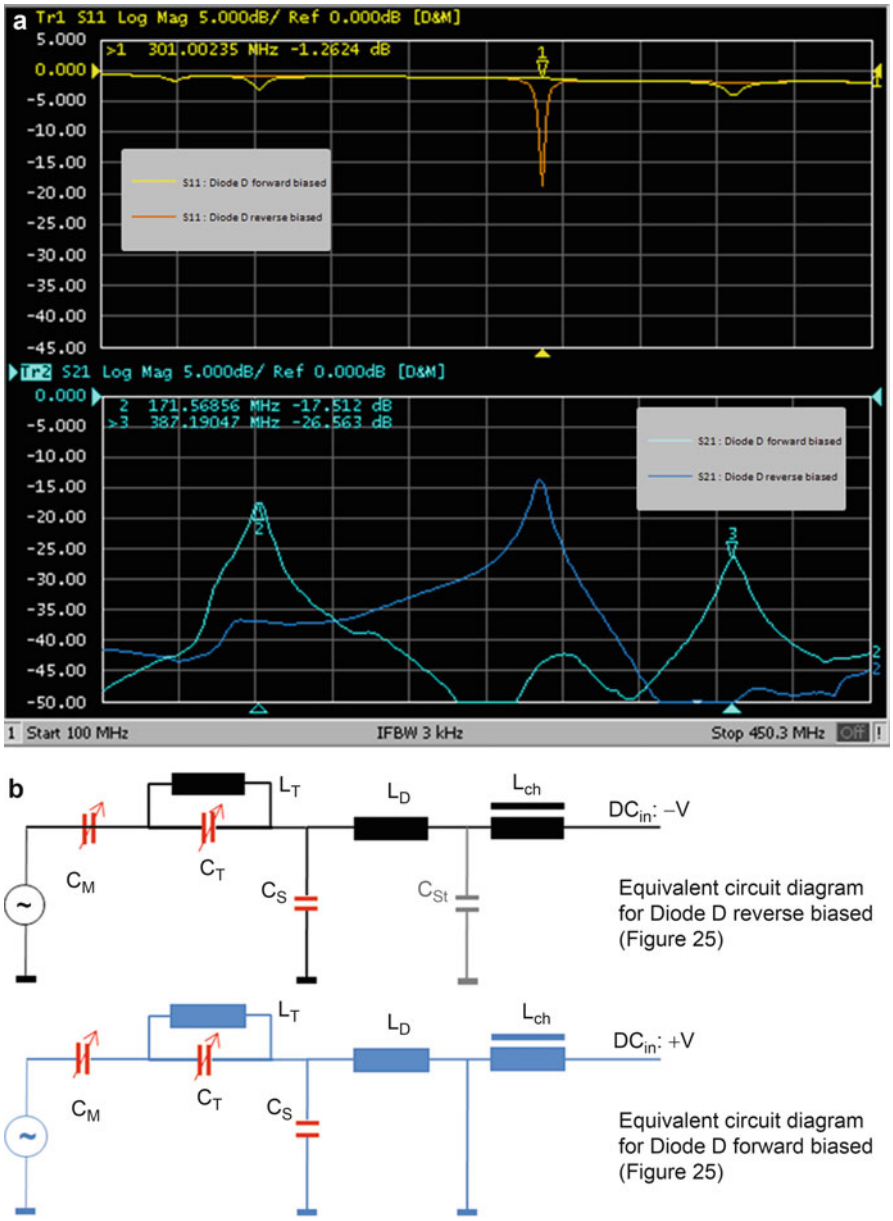


Fig. 26 Pole insertion method for active decoupling a receive-only coil: *Top*: Measured reflection coefficient S11 – is typically a minimum for a matched and tuned probe at the desired frequency ($f_0 = 301$ MHz) – *brown curve* and the corresponding *blue curve* shows the transmission S12 measured with a small so-called pick-up loop. The reversed biased diode behaves as a small capacitance C_{st} that can be neglected in most cases. When the diode is biased than the inductor L_D comes parallel to C_S resulting in a splitting of two resonances (*magenta line*) below and above the desired frequency ω_0 and the RF-coil is almost off-resonant (*yellow line*)

While an array coil consists of a set of local RF coils the field distribution and sensitivity is inhomogeneous over the VOI. There is no universally accepted or general design rule that fits all array coils. The optimal design of the array coils is highly dependent on the application that will be performed. Figure 29 demonstrates the basic principle of local sensitivity changes with an increased number of array elements under the boundary condition to cover a cylindrical FOV with a diameter of 60 mm and a z-length of 60 mm. Four rectangular coil loops (55 × 60 mm) surrounding the cylinder of 60 mm are the starting point. Remembering the Eq. 64 ($y = 30$ mm- $> R_0 = 13$ mm) the performance for loaded

DOI 10.1007/978-981-4560-75-7_121-1

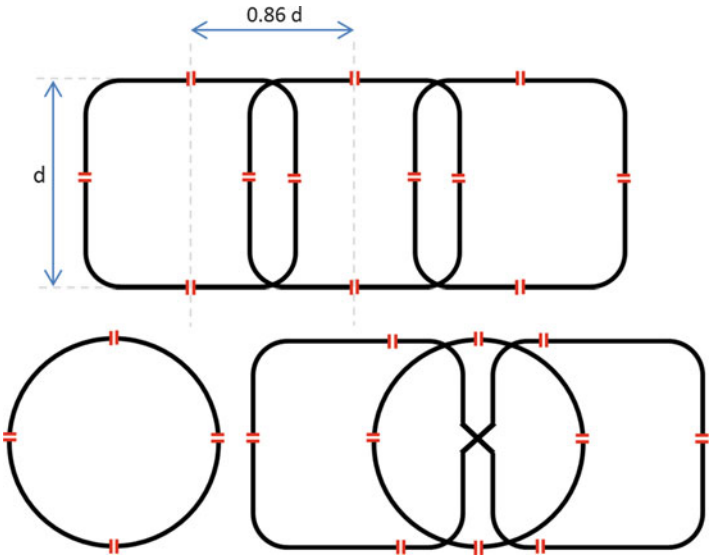


Fig. 27 Draft illustration of surface coils bottom left: single-loop surface coil; arrangement of surface coils that consists of three loop-coils working as separate coil-elements of an array-coil or as quadrature surface coil when driven in combination with fixed phase shifts. The adjacent coils are overlapped to set their mutual inductance to zero and prevent the splitting of the resonances for the nearest neighbor coils. For that case the center of the rectangular coils are separated by 0.86 times of the diameter of the coil elements.

conditions of these rather large surface coils is low because the coil-elements are much too sensitive to the sample noise. With an increased number of array elements the coil size in the radial direction is changed to 37 mm, 28 mm and 23 mm respectively. The length of the coil-elements in z-direction is kept constant to 60 mm to cover the whole FOV. The overall sensitivity increases, but the increase of sensitivity reaches its limit in the middle of the VOI. When enlarging the number of array coils the sensitivity benefits only at the edges near the surface of the array coil. For this example it is evident that an increase of number of array coil elements by more than 10 elements makes no sense if the main application requires a sufficient SNR over the whole area and not only at the edges of the FOV. One must remember that with an increased number of coil-elements the effort for decoupling of these array elements increases much in particular for opposite coil elements. For opposite array elements the method of choice is the decoupling by preamplifier decoupling or an external latter network. An example of the noise correlation for a 8-element volume array coil is shown in Fig. 30.

The values range from nearly zero up to 25–30 % when the elements are only decoupled by preamplifier decoupling. The next neighbor elements decoupled by overlap and preamplifier decoupling and the second next coil-elements decoupled by transformer (inductive decoupling). In principle the design of array coils is very challenging when the SNR of a comparable quadrature volume coil in the center of the VOI is targeted under the consideration that the dimensions of the RF-active elements of both coils are comparable in size and diameter. In most cases the SNR of array-coils are only higher than standard volume-coils for off-center positions.

Another example is shown in Fig. 31 and compares the SNR distribution of a quadrature surface coil (2 elements) with 4-element array coil. This array coil is used for applications where the target FOV is near the surface of the RF coils. The quadrature coil consists of two elements bent on a curved cylindrical surface with an area of 25 $\times$ 25 mm and the coil-elements are decoupled by a shared capacitor or overlap. The 2x2 array coil has been designed by four independent coil-elements with a size of 13 $\times$13 mm. All coil-elements are inductively decoupled by more than 20 dB. The highest gain in SNR to a factor greater than 2 is achieved at the edge of the surface of the array coil and decreases rapidly with the distance to the

DOI 10.1007/978-981-4560-75-7_121-1

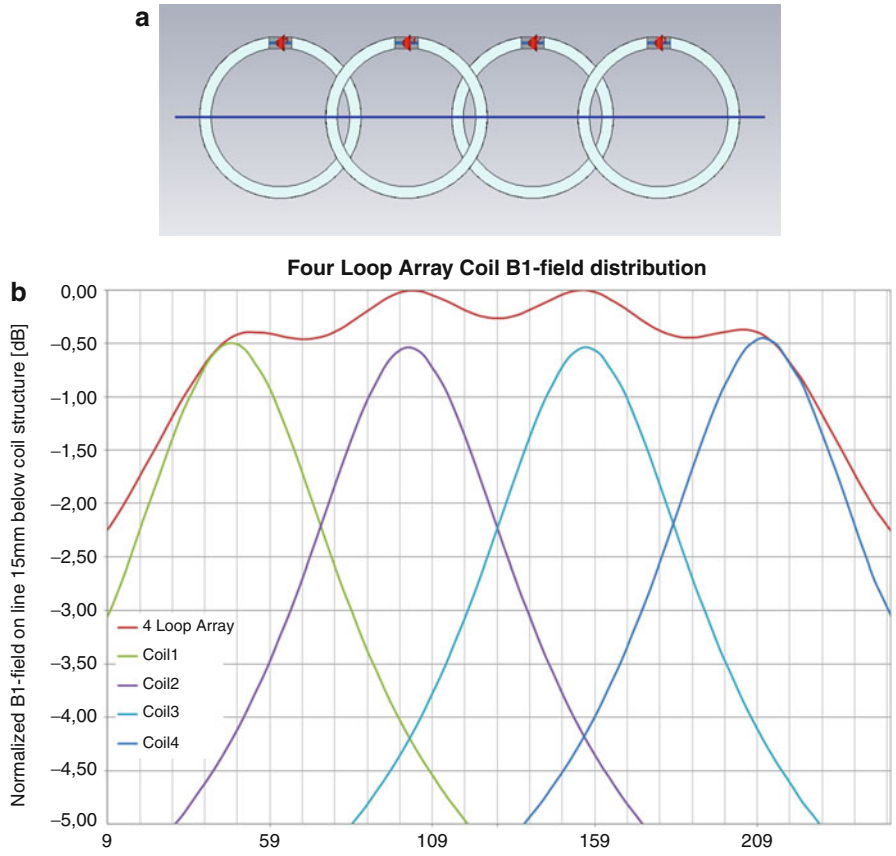


Fig. 28 Simulated $\mathbf{B}_1$-field-distribution (CST MWS) in the Field of View of a 4 channel loop array coil compared to single loop coils with an inner-diameter of 30 mm along a central line 15 mm below the coil structure for unloaded conditions

array-elements. The average SNR of the array coil is higher compared to the quadrature surface coil but is for a greater distance from the coil elements in the same magnitude.

Receive Array Coils for Parallel Imaging

In the last decades the array-technology was driven by the rapid developments of parallel imaging techniques. Most commonly used parallel imaging methods are SENSE, GRAPPA or SMASH. A detailed discussion of parallel MR image reconstruction (Pruessmann et al. 1999;Sodickson and Manning 1997) is beyond the scope of this chapter and therefore, only a very brief review is presented. These technologies pushed the envelope of RF-coil technology rapidly in the late nineties. The array coil design plays a major role in achieving the desired performance with respect to the SNR when image sequences are accelerated with respect to acquisition time (section Basics of MRI imaging).

Parallel MRI methods for accelerated image acquisition are based on extracting the spatial information from the sensitivity patterns of the RF-array coil elements and using this information to generate the data that normally is acquired by sequentially applied field-gradients $\boldsymbol{g} = \left(g_x, g_y, g_z\right)$.

The design considerations for arrays coil elements used for spatial encoding imply additional design requirements and these lead to an increased number of receiver channels and array coil-elements. As result commercial clinical MRI-systems with 32 channels or more have become routinely available and prototype systems with more than 128 channels in development.

In general a MR signal received from a coil-element l at a *k*-space position k_m can be written as (Vaughan and Griffiths 2012, p. 84 ff)

DOI 10.1007/978-981-4560-75-7_121-1

$$S_l(\boldsymbol{k}_m) = \sum\nolimits_{voxel\,j} C_l(\boldsymbol{r_j}) M(\boldsymbol{r_j}) e^{j\boldsymbol{k}_m \boldsymbol{r}_j} + \boldsymbol{n_l}(\boldsymbol{k_m}) \tag{72}$$

where n_l is the noise, $C_l(\boldsymbol{r_j})$ is the spatial sensitivity of element l and $M(\boldsymbol{r_j})$ is the precessing transverse magnetization (see Basic Introduction to the Principle of NMR and MRI). When the signals, noise and magnetization are written in column vectors this equation can be written in a general manner $\boldsymbol{S} = \boldsymbol{EM} + \boldsymbol{n}$, when the encoding matrix $\boldsymbol{E}$ is given by

$$\boldsymbol{E}_{(lm),j} \equiv C_l(\boldsymbol{r_j}) e^{j\boldsymbol{k}_m \boldsymbol{r}_j} \tag{73}$$

The matrix $\boldsymbol{E}$ is a rectangular matrix and within the formalism of parallel imaging the challenge is to find a Matrix $\boldsymbol{F}$ where the $\boldsymbol{EF}$ is equal to the identity matrix. There are many choices and it has been shown (Pruessmann et al. 1999) that the minimum norm solution to reconstruct an image with the least possible noise is given by

$$F_{\min norm} = \left(E^H \Psi^{-1} E\right)^{-1} E^H \Psi^{-1} \tag{74}$$

For parallel imaging the encoding functions 73 are generally spatially not orthogonal with the consequence that the reconstruction matrix 74 is typically not unitary. That yields to spatially varying noise increase. The image reconstruction used in Eq. 74 is a linear transformation. The noise power can be calculated from the encoding matrix and is given by

$$\left(\sigma_j\right)^2 \propto \left[\left(E^H \Psi^{-1} E\right)^{-1}\right]_{j,j} \tag{75}$$

The change in SNR that results from accelerated parallel imaging can be generally written as

$$\left(\frac{\mathrm{SNR}_j^R}{\mathrm{SNR}_j^{\mathrm{Full}}}\right)^2 = \left(\frac{\sigma_j^{\mathrm{Full}}}{\sigma_j^R}\right)^2 = \frac{\left[\left(E_{\mathrm{Full}}^H \Psi^{-1} E_{\mathrm{Full}}\right)^{-1}\right]_{j,j}}{\left[\left(E_R^H \Psi^{-1} E_R\right)^{-1}\right]_{j,j}} \tag{76}$$

For the special case of Cartesian sampling the Eq. 76 can be rewritten:

$$\mathrm{SNR}^R = \frac{\mathrm{SNR}^{\mathrm{Full}}}{\sqrt{R\left[\left(E_R^H \Psi^{-1} E_R\right)^{-1}\right]_{j,j}\left[\left(E_R^H \Psi^{-1} E_R\right)^{-1}\right]_{j,j}}} = \frac{\mathrm{SNR}^{\mathrm{Full}}}{\sqrt{R}g} \tag{77}$$

The so-called geometry factor is introduced and defined as:

$$g_j \equiv \sqrt{\left[\left(E_R^H \Psi^{-1} E_R\right)^{-1}\right]_{j,j}\left[\left(E_R^H \Psi^{-1} E_R\right)^{-1}\right]_{j,j}} \tag{78}$$

The factor $\sqrt{R}$ in the dominator of Eq. 77 is the result since the overall loss in SNR simply occurs by acquisition of fewer k-space points in the under-sampled image. The geometry factor g is always greater or equal than one and quantifies the fractional loss in spatially reduced SNR due to the nonorthogonality of the encoding functions based on the coil-element sensitivities. This factor g has become of general importance for qualifying the performance of coil arrays for parallel imaging.

DOI 10.1007/978-981-4560-75-7_121-1

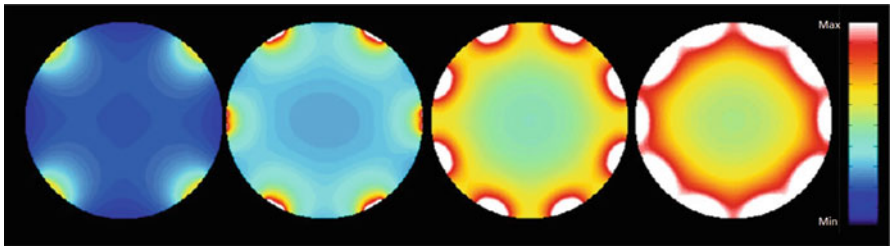


Fig. 29 Simulated field-distribution of the B_1^+ for an array coil with an inner diameter of 60 mm at 300 MHz (7 T) consists of 4,6,8 and 10 coil-elements under loaded conditions neglecting inter-coil coupling of the array-elements

The g-factor is a function of the coil sensitivities and the pattern of the acquired k-space lines. Therefore the pattern will be different for different undersampling factors, phase encoding directions and objects under investigations. Any change in the coil geometry and sample load, yields in a different g-factor and has an effect on the parallel imaging performance. In any case when the imaging acquisition is accelerated, the SNR will be decreased at least by a factor of $\sqrt{R}$. This factor will be even higher when the g-factor is larger than 1 (g>1). The imaging acceleration by factor of 2 leads to an SNR-decrease of more than 40 % compared to nonaccelerated imaging with the same imaging sequence parameters.

In addition, the noise correlation plays also an important role when drastic changes in the noise correlation affect the array's g-factor. The complex spatial sensitivity functions $C_l(\boldsymbol{r}_j)$ can be written based on the principle of reciprocity in terms of the two transverse components of the $\mathbf{B}_1$-field of the coil-elements (Vaughan and Griffiths 2012, p. 86)

$$C_l(\boldsymbol{r}) = \left[\widehat{B}_x(\boldsymbol{r})\right]_l - i\left[\widehat{B}_y(\boldsymbol{r})\right]_l \tag{79}$$

The noise correlation matrix elements can be computed in terms of the electrical fields of the individual coil-elements that are normalize to the unit current that flows around the conduction path

$$\Psi_{l,l} = 4kT\Delta f \int \sigma(\boldsymbol{r})\hat{\boldsymbol{E}}_l(\boldsymbol{r}) \cdot \hat{\boldsymbol{E}}_{l'}^*(\boldsymbol{r}) d^3\boldsymbol{r} \tag{80}$$

Where k is the Boltzmann constant, Δf the receiver bandwidth, T the temperature and σ is the conductivity of the sample. With these equations the theoretical evaluation of any prospective array coil can be performed, e.g., for optimization purposes. A typical pattern of the g-factor map is shown in Fig. 32 for reduction factors of 2, 3 and 4 for an under sampling in the left-right direction. For evaluation of this 8 channel array coil, a model in CST-microwave studio was designed. The model includes all conductive coil elements including the lumped elements, feeding ports and cylindrical conductive phantom as load. The sensitivity functions are calculated using the approach above. The g-factor was calculated for an under sampling with acceleration factors R = 2,3 and 4 that reduce the FOV by a factor 2,3 or 4. The pattern shown in the figures results in a folding from multiple equidistantly distributed data points (section Basics of MRI imaging).

The g-maps in Fig. 32 are normalized individually to the highest g-factor for the individual reduction factors g = 1.04, 1.50 and 3.88 and with Eq. 77 that yields a maximum spatial noise-increase compared to the unaccelerated image acquisition by a factor of 2.08, 2.6 and 7.76 including the factor of $\sqrt{R}$ for the intrinsically noise-increase for accelerated image acquisitions. As conclusion from the g-map pattern this coil will give reasonable results with respect to the noise introduction up to an acceleration factor of 3 (section Future Directions and Open Problems). While the structure of this coil and the loading has a

DOI 10.1007/978-981-4560-75-7_121-1

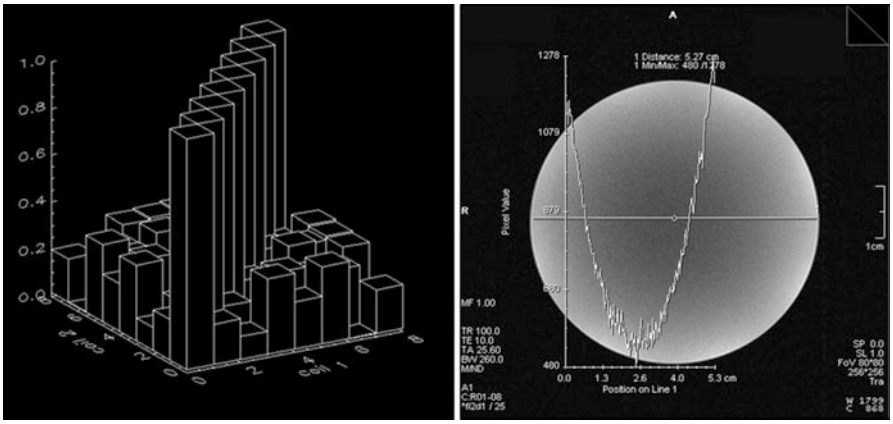


Fig. 30 *Left*: Visualization of the noise correlation for an 8 Channel loop coil; the noise correlation between opposite coil-elements increases up to 25–30 % when the elements are only decoupled by preamplifier decoupling; right: measured SNR-distribution in the center of an 8-element array –coil – the SNR from the surface to the center of the loaded array coil changes by a factor of 2.5

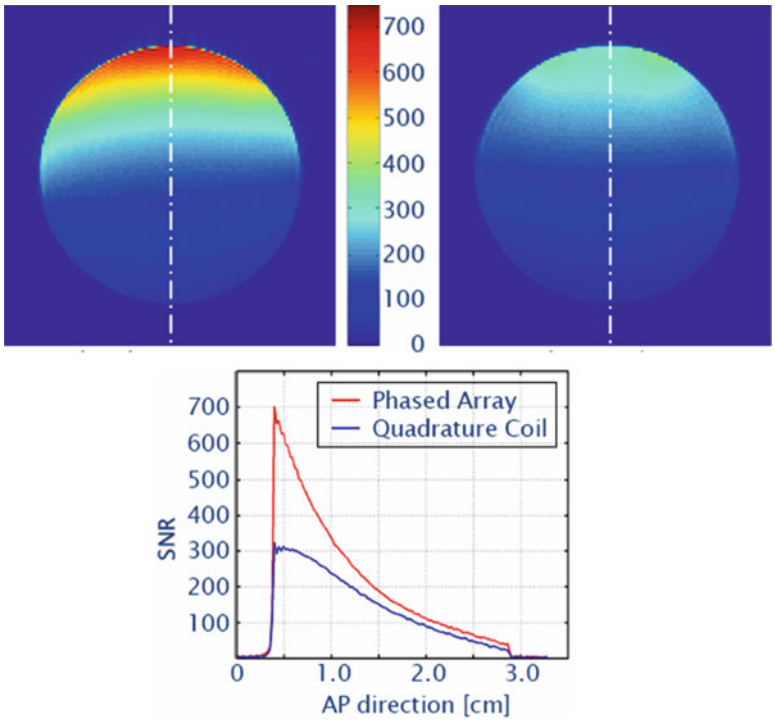


Fig. 31 *Left* – Comparison of the measured SNR-distribution for a quadrature and 2 × 2 array coil. The surface coils cover an area of 25 × 25 mm on the top of the load (diameter 30 mm); *Right* – The SNR is plotted along a line from the top to the bottom. The coils are loaded with a water phantom that represents the same loading conditions for the subject under investigation

rotational symmetry in the xy-plane, the g-factor-maps for left-right and top-down readout direction have the identical g-map pattern and are perpendicular to each other. Any asymmetry of the coil structure or loading will yield a different g-map pattern for the two readout directions.

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_121-1

Cryogenically Cooled Coils

Demands on high spatial or temporal resolutions require a high signal-to-noise ratio (SNR). For applications on small animals, spatial resolutions are required, that are up to a factor 10 higher than for human applications, which leads to a decrease of the sensitivity by a factor of the order of 1,000.

The SNR can be increased simply by increasing the static magnetic field $\mathbf{B}_0$. Today commercial animal-scanners up to 21 T are available. In some cases the decrease of the sample-noise and the noise contribution of the different receive-chain components of the MR-system are a common cost-effective option. The optimization of room-temperature (RT) coils is in most cases limited by theoretical and practical limitations. In most cases the design changes will lead to an increase of several percent in the sensitivity of the optimized coils. For dedicated MRI applications such as for small animals, the observed tissue volumes are of a size that renders the sample noise contributions comparable or smaller than the thermal noise contribution of the RF coil and associated components of the MRI-system receive-chain. The SNR can thus be enhanced by decreasing the noise of the sample. Noise contributions from the sample are the thermal noise, due to the thermal motion of the charge carriers and the so-called dielectric losses due to the electrical field of the RF-field.

Decreasing the temperature of living samples is usually not possible, but a reduction of the thermal noise and the resistive losses of the RF receive-chain will produce a significant increase of the SNR (section Efficiency, Losses and Signal to Noise Ratio – Eq. 22)

$$\mathrm{SNR} \approx \frac{\omega_0 B_1}{\sqrt{R_s T_s + R_c T_c}}.$$

The denominator of Eq. 22 describes the losses (noise) and their dependency upon temperature and frequency. Two domains can be distinguished. On one hand the case when the sample-losses dominate $R_s T_\mathrm{s} \gg R_c T_\mathrm{c}$; for this case the SNR is independent of the resistive coil-losses R_c and their temperature T_c. On the other hand, using better conductive coil conductors or lowering the coil temperature yields to an increased SNR when the coil-losses dominate $R_s T_\mathrm{s} \ll R_c T_\mathrm{c}$.

For the design of cooled RF coils it is required to know how the boundary between these two domains is dependent on the coil-size and frequency. In Fig. 33 a rough estimation regarding the limits of the coil-size for cryogenically cooled coils is presented. For this purpose the sample resistance R_s and coil resistance R_c for a single-loop copper coil under the boundary condition of a semi-infinite sample was examined (Darrasse and Ginefri 2003). The cross-over in a frequency band from 100 to 800 MHz between the two domains for room-temperature copper-coils is given by coil radii between 2.5 cm and 0.9 cm.

Above these radii the sample losses are dominating! For cooled copper coils at 77 K (LN2) will halve the respective coil radii. This will have an influence on $\mathbf{B}_1$-field distribution and the reduction of the sensitivity and so-called penetration depth (see Eq. 64). Even lower temperatures e.g., to 30 K for cooled copper coils or the usage of HTS-material can still improve the sensitivity. But with respect to frequencies from 100 to 800 MHz the coil radius has to be reduced to approximately 1.2 cm to 0.45 cm. Important to know is the SNR-gain that can be achieved by cryogenically cooled RF coils with respect to the frequency and the coil radius. The SNR gain for cryogenically cooled coils is evaluated from the SNR ratio at room temperature (RT) and low temperature (LT)

$$\mathrm{SNR}_\mathrm{gain} = \frac{\mathrm{SNR}_{LT}}{\mathrm{SNR}_{RT}} = \frac{\sqrt{R_s T_\mathrm{s} + R_{c,RT} T_{\mathrm{c,RT}}}}{\sqrt{R_s T_\mathrm{s} + R_{c,LT} T_{\mathrm{c,LT}}}} \tag{81}$$

For the quantification of the SNR gain for a circular loop coil in terms of parameters that can be measured the quality factor $Q = \omega L/R$ can be used. The inductance L of the RF coil is taken into account and for the

other expressions, an equivalent Q is defined by using the appropriate resistance term in (81). The term ωL cancels out, and leads to

$$\mathrm{SNR}_{\mathrm{gain}} = \frac{\mathrm{SNR}_{LT}}{\mathrm{SNR}_{RT}} = \sqrt{\frac{T_{RT} \cdot Q_{UL,RT}^{-1} + T_{RT} \cdot Q_{\mathrm{sample},RT}^{-1}}{T_{LT} \cdot Q_{UL,LT}^{-1} + T_{RT} \cdot Q_{\mathrm{sample},LT}^{-1}}} \quad Q_{\mathrm{sample}}^{-1} = Q_L^{-1} - Q_{UL}^{-1} \tag{82}$$

where the inverse of the quality factor of the sample (Q_{sample}) is calculated by the difference of the inverse of the quality factors for the loaded (Q_{L}) and unloaded ($\mathrm{Q_{UL}}$) coil condition. A very rough estimation of the SNR-gain for cooled RF coils can be used for the boundary limiting case $T_{LT} \rightarrow 0$. For this SNR-gain estimation the values only at room temperature are sufficient. For this case the Eq. 82 will change to

$$\mathrm{SNR}_{\mathrm{gain}=} \frac{\mathrm{SNR}_{LT}}{\mathrm{SNR}_{RT}} = \sqrt{\frac{1}{1 - \frac{Q_{L,RT}}{Q_{UL,RT}}}} \tag{83}$$

For room temperature surface coils commonly discrete elements are used to build the R-L-C-resonance circuit. Apart of possibly limited temperature operating specifications of commercially available capacitors, the soldering introduces several technological difficulties for cryogenic operation of resonance circuits built by discrete elements. Alternative approaches are the so-called monolithic designs and the use of distributed capacitors.

So-called transmission-line resonators (TLR: Fig. 34 top) or "Twin-Horseshoe resonator (Gonord and Kan 1994) and multi-turn transmission-line resonators (MTLR: Fig. 34 bottom) (Vaughan and Griffiths 2012, pp.236, Darrasse and Ginefri 2003) made either of copper or HTS material are commonly used for self-resonant high-Q structures. The TLR consists of a pair of planar, electric conductive rings facing each other and separated by a dielectric substrate. The gaps on the rings are diametrically opposed to each other. The MTLR based on the same principle consists of several concentric rings on different radii connected in series. Because of the increased inductance value of the MTLR the useful frequency domain is more limited compared to the TLR. The design for MTLR is best used for lower frequencies.

Copper material (conductivity $\sigma = 58{\cdot}10^6$S/m at 20 °C) exhibits a nearly linear dependency of the resistivity $\rho = 1/\sigma$ with respect to temperature down to approximately 30 K. A reduction of the thermal noise in the order of magnitude of about 20 at 77 K and even more at lower temperature (Ginefri et al. 2007) can be achieved. With the usage of HTS material cooled below their critical temperature even higher reduction factors of the thermal noise can be achieved. Very high Q-values in a range from 1,000 to 500,000 dependent on the coil-size, temperature, frequency and the static magnetic field can be attained. But it must be mentioned that HTS material loses its superconducting properties depending on the strength and orientation of the magnetic field. In any case a reduction of the high Q-values for HTS RF coils cannot be prevented inside MRI-systems.

There are several strategies for the cooling of superconducting coils. The simplest method is the direct immersion of the RF-coil in a bath of cryogenic liquid such as liquid nitrogen (LN2) placed inside a container for thermal isolation. The MR-measurements can only be performed after all the coolant has been consumed because of the mechanical vibrations caused by the boiling of the coolant. But the most common method is the use of closed-cycle cry coolers. The RF coil is cooled by a so-called cold finger by a cold-head either in the direct vicinity of the RF-coil or located at a certain distance from the cryogenic probe.

DOI 10.1007/978-981-4560-75-7_121-1

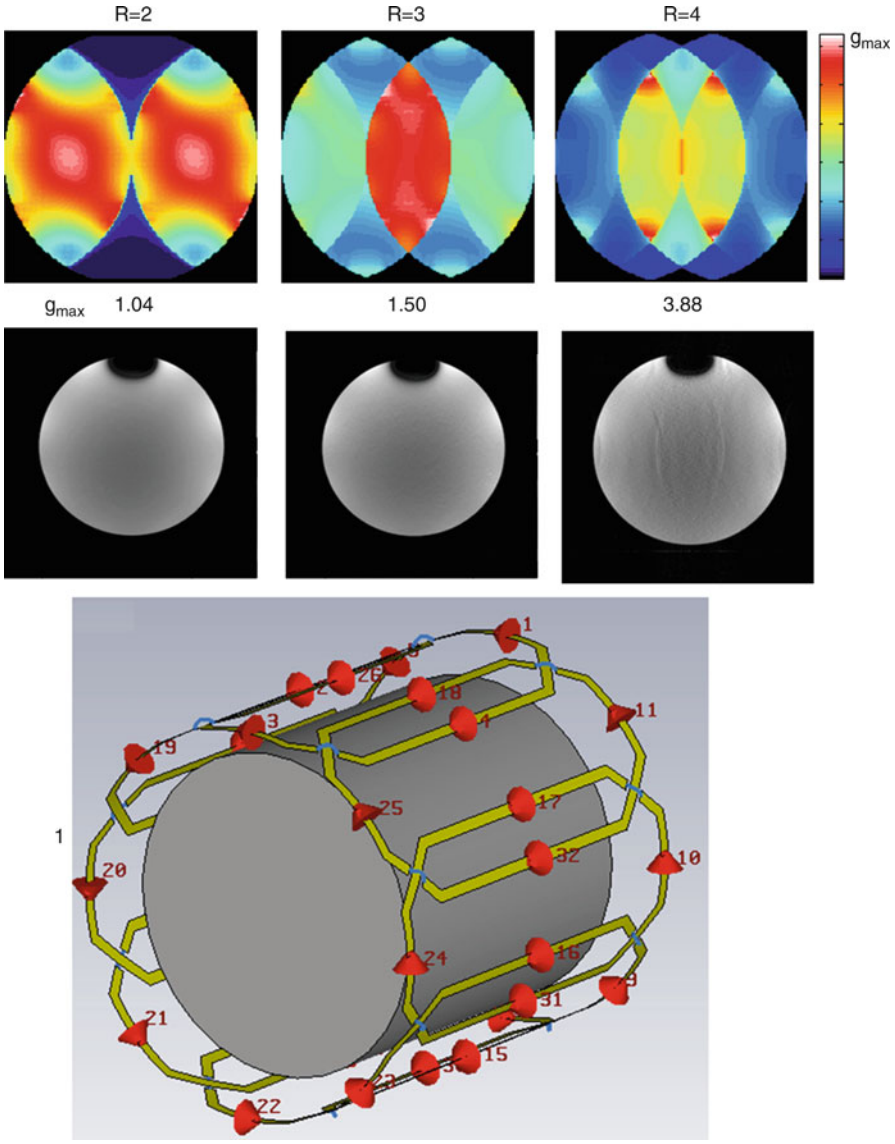


Fig. 32 *Left-top*: Maps of g-factor for an 8 element loop array coil – numerical calculated for acceleration left right direction; *left-bottom*: Accelerated MR-imaging of a cylindrical phantom performed with FLASH. The image acquisition can be accelerated upto a factor of 3 in one direction without limitations by imaging artifacts. *Right*: model for simulation of an 8 channel loop using CST Microwave Studio (CST-Darmstadt, Germany) coil modelconsisting the overlapped loop coil elements, all lumped elements and a lossy cylindrical phantom

The principle design (Haueisen et al. 2006) of a cryogenic probe-head is shown in Fig. 35. The RF-coil (1) is mounted on a cold finger (2) surrounded by housing (3) – dewar. The most efficient way for isolation is the use of vacuum to protect the object under investigations (5). This object has a temperature equal to or slightly higher than room temperature and needs to be isolated from the cryogenic temperatures of the RF-coil (20–30 K). The dewar-housing with a residual vacuum of about 10^{-6} mbar reduces the two dominant heat transmission processes (heat conduction and heat convection). For MR-imaging the conventional coating by mirror material cannot be used for additional reduction of the radiation. For this reason the dewar-wall would be cooled in an undesired fashion by heat radiation and this may have undesirable consequences, especially for living subjects under investigation. To protect the subject, the separating wall of the dewar is used as a heating device and is temperature-controlled for a suitable

DOI 10.1007/978-981-4560-75-7_121-1

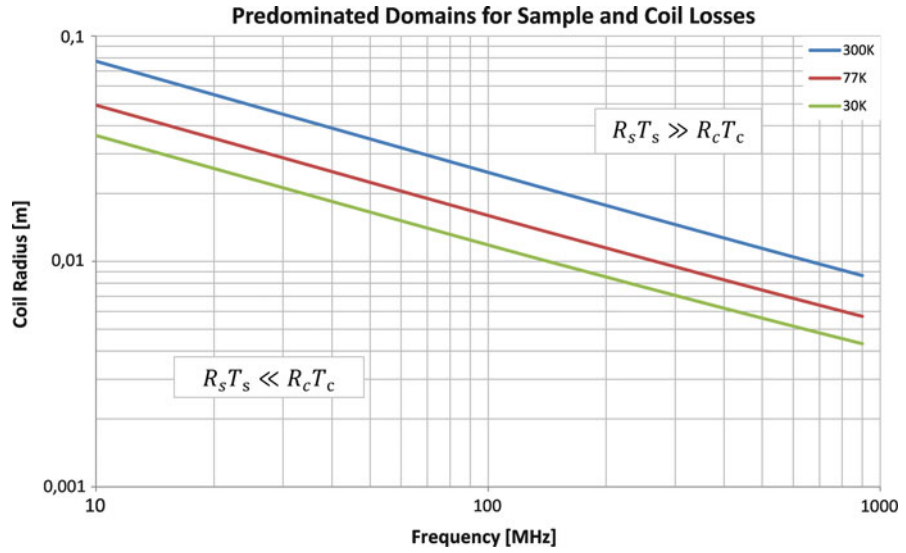


Fig. 33 Predominated domains (sample-losses dominate $R_sT_s \gg R_cT_c$ and coil-losses dominate $R_sT_s \ll R_cT_c$) in frequency and size coordinates for sample and coil losses. The tissue conductivity σ was chosen to an average value of 0.66 S/m (Darrasse and Ginefri 2003), the electrical resistivity ϱ_c to be $5.81 \cdot 10^7$ S/m for RT

temperature with respect to the subject (Haueisen et al. 2007). The heating device (4) for the dewar is conveniently positioned at the rear end of the dewar.

Double Tuned RF Coils

For the general study and treatment of diseases the combination of MR-imaging (MRI) and MR-spectroscopy (MRS) is an indispensable tool. The growing interest in applications of NMR spectroscopy for studies of biological systems has placed unique demands on the MR-systems and their instrumentation. In most cases MRI and MRS are combined into one study where MRI is used for so-called shimming (make the external **B**-field as most as possible homogenous) and for localization of the targeted region for the MRS-application. The MR-spectroscopy is used to obtain the metabolic information by acquiring spectra's of specific nucleus such as ^{31}P, ^{23}Na, ^{13}C and ^{19}F commonly so-called "X-nucleus" (corresponding frequencies see Table 1, section Introduction). This purpose demands that the RF coil has to be tuned to multiple frequencies.

In some cases, when two interacting nuclear spin-system are excited simultaneously, the received resonance-signal from the spin-systems can be enhanced or extenuated, e.g., by proton (^{1}H) decoupling (Luyten et al. 1989), Overhauser effect. For this purpose the RF coil has to operate for transmission and reception at two separate frequencies simultaneously. For proton-decoupling an excitation pulse is first applied at lower frequency (X-nucleus) and then simultaneously with the signal reception of the X-nucleus, the RF-power at the ^{1}H-channel is applied. A basic hardware scheme for such applications is shown in Fig. 36. On each transmit/receive path of the double resonance coil a RF-filter is introduced. For the higher frequency (^{1}H) a band-pass-filter is implemented and for the channel of the X-nucleus a low-pass-filter is used to suppress of out-of-band noise. On one hand the NMR-signal from the X-nucleus is in most cases very small (e.g., ^{13}C inverse NMR-spectroscopy) and every noise introduction from ^{1}H-transmission will over saturate the NMR-signal from the X-nucleus. On the other hand the broad-band noise level of the ^{1}H- transmitter can saturate the preamplifier of the X-channel easily if the filter and decoupling of both channels are not sufficient. A decoupling of the two channels at any point of the transmit/receive chain of more than 100 dB is commonly required for dedicated MRS- applications. These multiple frequency applications can be performed by double resonance surface coils, volume coils or by

DOI 10.1007/978-981-4560-75-7_121-1

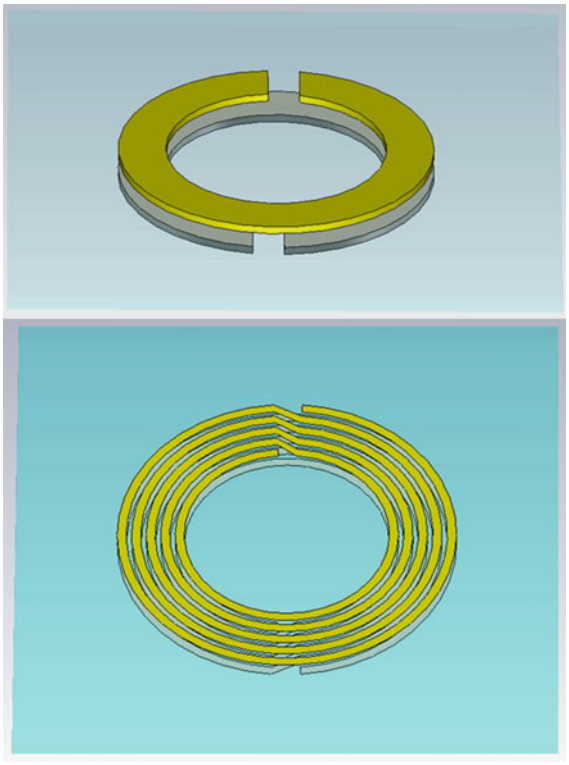


Fig. 34 *Top*: Transmission-Line resonator: Two planar split rings deposited on each side of a dielectric substrate. The gaps in the rings are diametrically opposed. – *bottom*: Multi-Transmission Line resonator: Two planar multi-turn split rings deposited on each side of a dielectric substrate. The gaps in the rings are diametrically opposed

combinations of the two. In this chapter only a brief description can be given while a wide number of various schemes have been introduced over the last decades.

The most common double-tuned circuits are based on two basic principles using so-called multi-pole or transformer-coupled RF-circuits. Multi-pole circuits have already been introduced in section Receive Array Coils for Parallel Imaging where these circuits are used for active decoupling of surface coils and the resonance-frequency ω_0 is split in two frequencies at a lower and at a higher frequency with respect toω_0.

For double resonance coils the multi-pole circuit inserts a discrete parallel resonance circuit (Fig. 36) into the coil and creates a pole in the reactance curve that creates a second capacitive reactance to resonate with the inductance L_s at a second frequency. The first and lower frequency is tuned by a serial capacitance C_L. While a second circuit is introduced into the RF-coil, special attention is required to the second inductor L_H that introduces a detectable resistive loss to the RF-circuit. The efficiency of a double resonance coil at the high- and low-frequency mode is given by (Schnall et al. 1985):

$$E_L = \sqrt{\frac{L_S}{L_H + L_S}} \tag{84}$$

$$E_H = \sqrt{\frac{L_H}{L_H + L_S}} \tag{85}$$

under the boundary condition of a loaded RF coil and that the resistance of the inductors is proportional to the inductance. The efficiency E_L for the low frequency will be the best if the Inductance L_H is small

DOI 10.1007/978-981-4560-75-7_121-1

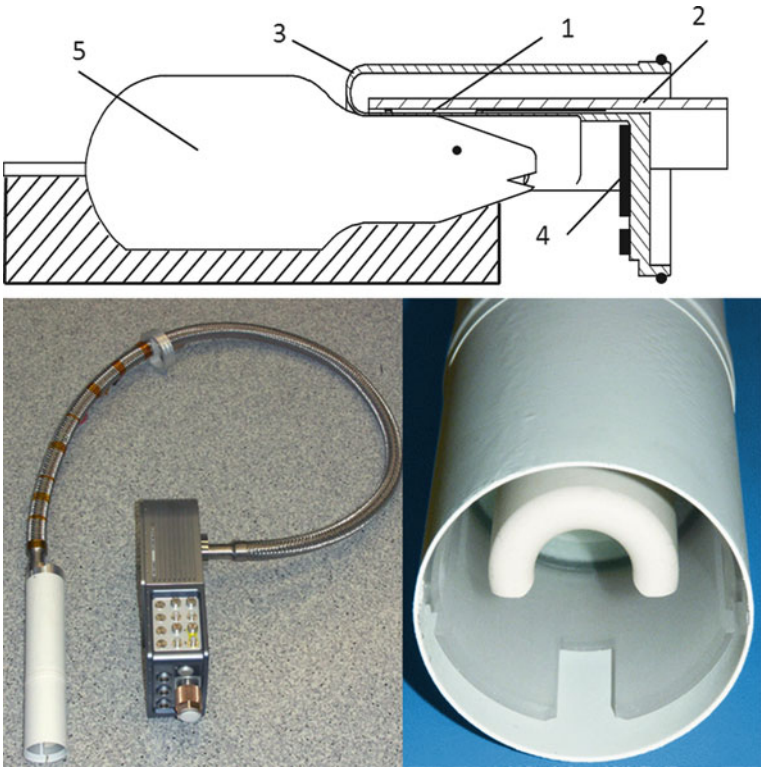


Fig. 35 *Top* – Principle design of a cryo-coil for MRI: *1* RF-coil; *2* cold finger; *3* ceramic Dewar; *4* heater *5* subject under investigation – bottom: the cryo-coil probe-head (*bottom right*) connected to the cooled preamplifiers via a evacuated, flexible line that supports the cooled probe head with cryogenic helium and RF and DC-lines

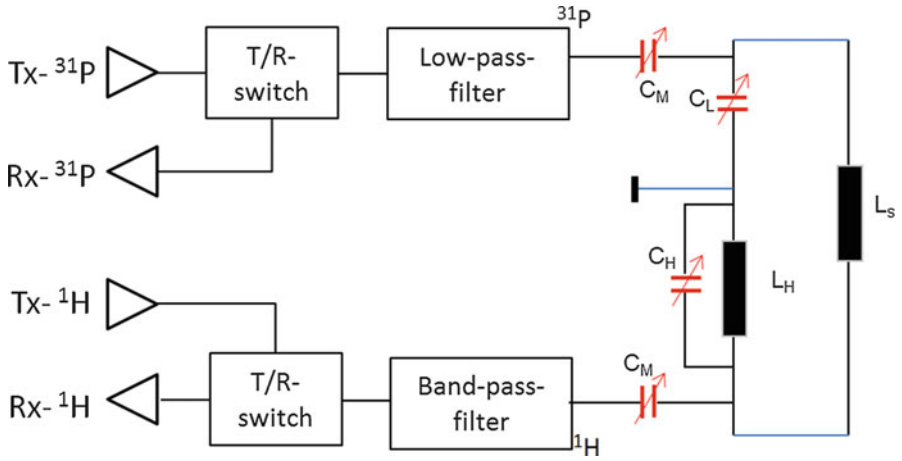


Fig. 36 One of the different configurations for the design of a double resonance RFcoil using multi-pole circuits. The surface RF-coil is interfaced via a transmit-receive-switch and a low- or band-pass-filter to the MR-system. The multiple channels are matched with C_M and tuned by C_L and C_H to the low- and higher frequency mode

compared to L_S and vice versa. In practice a good starting point for the ratio L_S/L_H is a factor of 4-5, yielding an efficiency of about 90 % for the low-frequency mode and about 45 % for the high frequency mode. For double resonance coils the focus is commonly on the performance of the low-frequency (X-nucleus) channel. Losses with respect to the higher efficiency on the ^{1}H-channel will be accepted. It must be clearly mentioned that the performance of the two frequencies of double resonance coils will

DOI 10.1007/978-981-4560-75-7_121-1

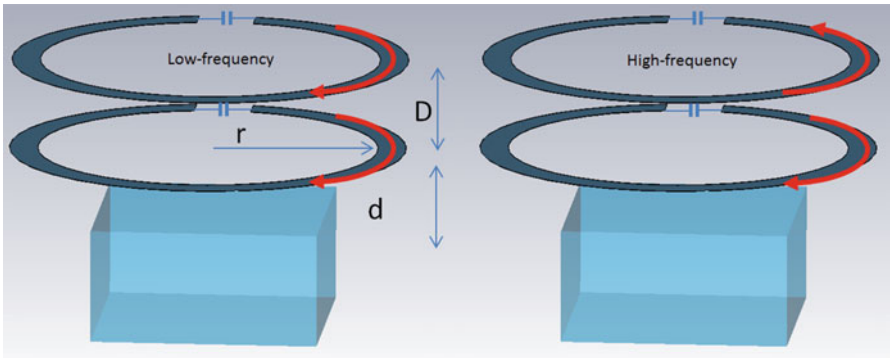


Fig. 37 Double resonance coil created by two identical inductive coupled and tuned circular loop coils. The final resonator has two modes, one with in-phase currents at low frequency and one with out-of-phase currents for the higher frequency. The design can be optimized with respect to the inter-coil distance D, the distance d to the sample and the radius r of the loops respectively

never achieve the performance of well-designed single resonance RF coils at the same frequencies. This fact should be taken into account for the chosen RF coil setup for a required application.

The basic principle of transformer coupled circuits (Fig. 37) based on the mutual coupling of tuned coils that was still introduced in section Local RF coils. The coupling constant and the frequencies are given by Eqs. 64 and 66 and for double tuned coils a large coupling constant k is required: The double resonance coils operate in the over coupled regime (dependent on the distance D) where the currents are split between the two frequencies (Fig. 37). As result a co-rotating mode at the lower frequency and a counter-rotating mode at the high frequency are generated. The generated RF fields are added at the lower frequency but subtracted at the higher frequency. As result the RF-efficiency at the low frequency is equivalent or slightly higher than a single coil but clearly underperform at the high frequency by approximately 50 %.

Multi-frequency birdcage coils have to be designed on the principle of preserving the sinusoidal /cosine current distribution around the cylindrical structure for both frequencies. The feeding is done in the same manner as for double resonance surface coils by introduction of band-pass- and low-pass-filters at the high-frequency (1H) and low-frequency channel (X-nucleus) respectively. For volume coils like bird-cages the principle of multi-pole circuits can be introduced as well and was one of the first methods that have been used for double resonance volume coils. The multi-pole circuit is introduced at the position of the resonance capacitors at the end rings for the high-pass birdcages or at the rungs for the low-pass birdcages (Fig. 38 left).

In the same manner as for surface coils the series capacitor C_L (Fig. 38) is used for the low frequency. The parallel capacitor C_H that has a smaller value as the series one is used for adjusting the high frequency. The adjustments of the circuit must be done very carefully on each position to maintain the coil symmetry and is a very time consuming procedure. Easier to build are double resonance birdcages that use alternatively the rungs for the both frequencies (Fig. 38 right). To maintain the coil-symmetry for the higher frequency (^{1}H) a trap for the lower frequency is introduced at the rungs for the 1H and prevents the destruction of the current distribution of the higher mode by the currents of the lower mode. The relatively small tuning capacitors at the ^{1}H-rungs generate at the low frequency high impedance and therefore the corresponding mode at the low-frequency is not disturbed by the high frequency mode. This method also relies on a very careful adjustment of the trap circuit on each position and needs a very careful and iterative adjustment procedure. These two designs-principles using the same birdcage-structure and both modes can be driven in quadrature because of the intrinsic symmetry of these structures.

DOI 10.1007/978-981-4560-75-7_121-1

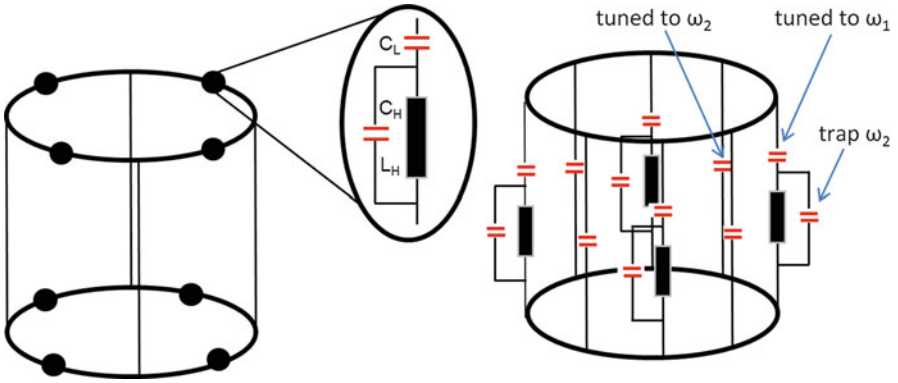


Fig. 38 *Right* – principle sketch for a low-pass double resonance birdcage. At the positions of the resonance capacitors the multi-pole circuits are introduced. *Left* – Double resonance Birdcage built on alternating legs that are tuned to the high-frequency ω_1 and low frequency ω_2. The parallel resonance circuits for the low-frequency-legs are tuned to the high-frequency ω_1 and operate as a trap and prevent the destruction of the current distribution of the lower mode by the currents of the higher one

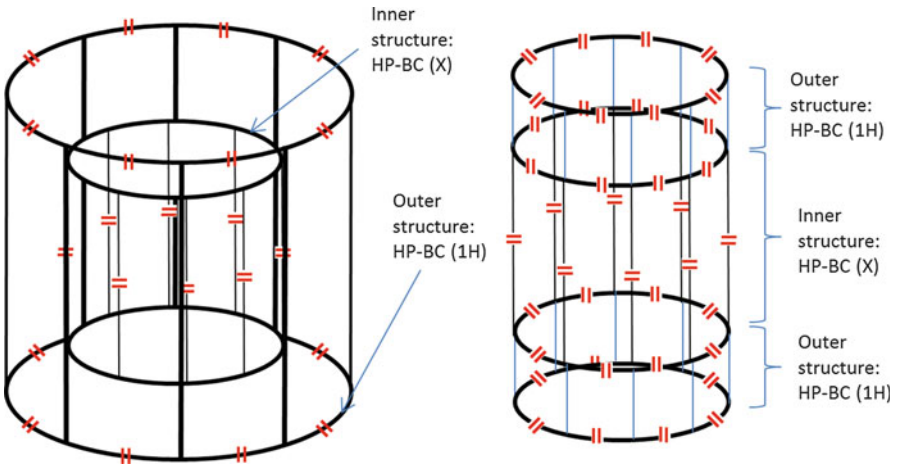


Fig. 39 *Left* – principle sketch of a double-resonance coil consisting of two imbricated birdcages. The inner birdcage is a low-pass-birdcage (LP-BC) that operates at the low-frequency (X-channel) and the outer structure designed as a high-pass-birdcage (HP-BC) that operates at the higher frequency (^{1}H). *Right* – lowpass-highpass birdcage combination of a Four-ring double resonance birdcage. The inner structure operates at the low-frequency while the two outer structures generate a transversal mode at the higher frequency. Due to the inductive coupling the co-rotating mode of the outer birdcages generates a common transverse, homogenous mode at the high frequency (^{1}H)

One of the easiest solutions for double resonance volume coils is the design principle based on TEM-coils. For this purpose the rungs are alternately tuned at the low and high frequency respectively. The adjustment procedure is in principle the same as for a single tuned TEM-coil and could be implemented in straight manner. The two modes can be driven in quadrature and fed in the same manner as birdcages via external low- and band-pass filters.

Another method to build double resonance-volume coils is the introduction of multiple cage-structures on different diameters (Fitzsimmons et al. 1993). In many cases these RF coils include a high-pass and a low-pass birdcage (Fig. 39 left). Also arrangements of high-pass or low-pass bird cages with saddle coils are commonly combined for multi-frequency volume coils. For such geometries the birdcage on the smaller diameter is used for the lower frequency while the larger diameter is tuned to the higher frequency (^{1}H). The strong inductive coupling effects between the coil structures have be to taken into account. To

DOI 10.1007/978-981-4560-75-7_121-1

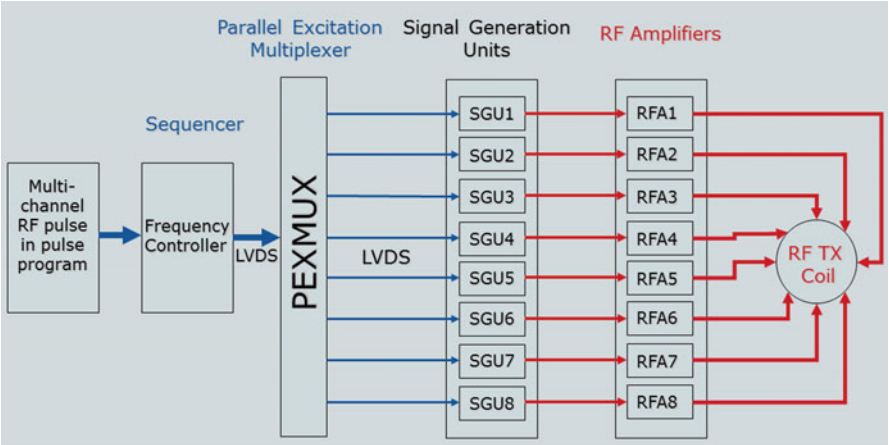


Fig. 40 Schematic of an example for parallel transmission systems based on a AVIII spectrometer platform (Bruker BioSpin, Rheinstetten, Germany)

achieve the optimal orientation to each other the proximity of the end rings needs special attention. The coupling of the modes of the two RF structures is minimized by slightly rotating one coil with respect to the other. In addition one of the coils can be built longer than the other but the implementation of quadrature driving of these structures can be quite cumbersome to build. One coil is usually a high-pass and the other one a low-pass birdcage structure according to the frequency and dimensions. It must be taken into account that the losses with respect to the performance of the two modes are strongly dependent on the mode coupling of the two separate electrical RF structures. In some cases it makes more sense with respect to the design-cost relationship to build the volume coils for linear excitation/reception than for quadrature-drive operation because of the reduced intrinsic coupling by adjusting the two modes perpendicular to each other. For some cases it is advantageous using the inner structure for the higher frequency (^{1}H) and the outer structure for the low-frequency e.g., for inverse ^{13}C-spectroscopy applications (Lanz 2000). In this special case the mode of the low-frequency that generates a linear RF field does not induce a significant current in the high frequency birdcage-structure due to its high impedance at the low-frequency. The sensitivity is recovered for both the coil-structures but this resonator cannot be driven in quadrature.

The issue of good circuit isolation and less coupling of the RF modes is intrinsically better solved by so-called four-ring birdcages (Murphy-Boesch et al. 1994). The principle structure is based on a birdcage that is merged together with two additional birdcage-structures and placed directly on the shared endrings of the inner structure (Fig. 39 right). The inner structure is built as a low-pass birdcage and the outer structures can be built as low or high-pass birdcage. The two identical outer structures are coupled inductively through the inner structure and generate a pair of two quadrature modes based on a co- and counter rotating current distribution of the outer structure. One of these modes produces a homogenous transversal RF-field and is operating at the higher frequency (^{1}H). The inner structure is tuned to the lower frequency. While the currents of the multiple frequencies flow substantially in different structures the circuit interactions between the frequencies are relatively small. This makes the construction and tuning of the four-ring birdcage relatively straightforward to construct.

DOI 10.1007/978-981-4560-75-7_121-1

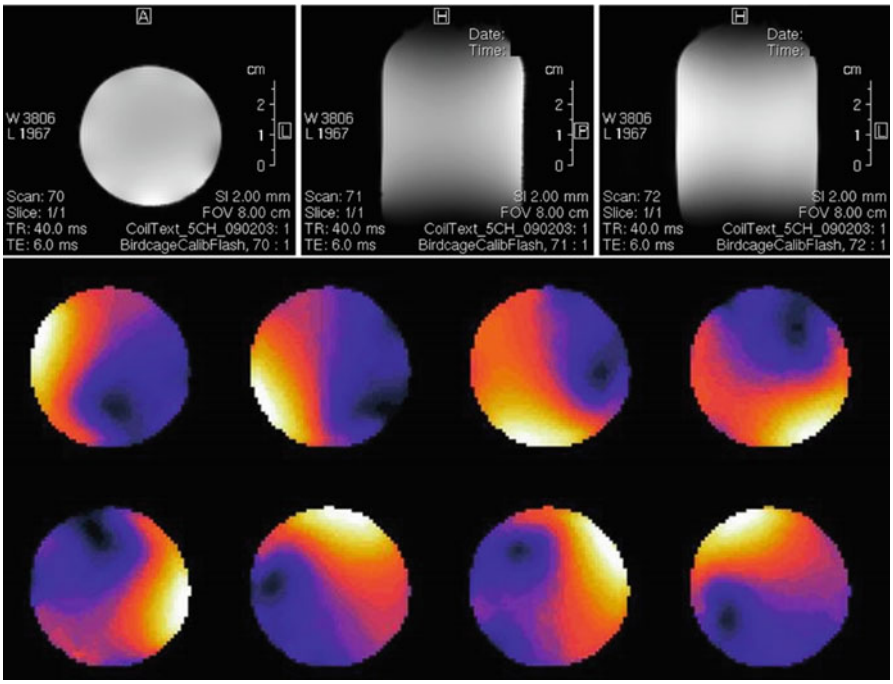


Fig. 41 *Top*: Image performed by a 8 channel coil driven by a birdcage-like excitation mode: *Bottom*: Transmit sensitivities of the 8-channelsof the transmit-coil (Fig. 32 right). The RF-coil was loaded with a water-phantom that represents the loading conditions of the object under investigation

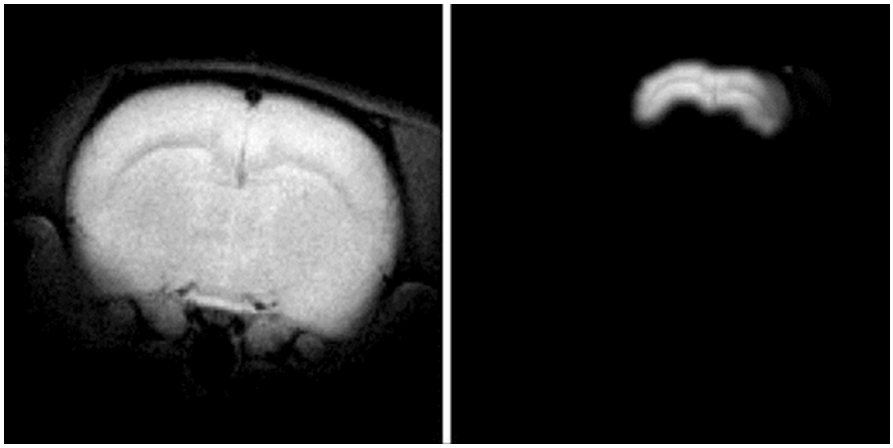


Fig. 42 Local excitation performed by PEX on a rat brain post-mortem; only a small portion of the brain of a rat was excited as shown on the right side

Future Directions and Open Problems

Since the early days of MR-imaging many different RF-coil concepts have been developed and today an almost unlimited amount of RF coil designs are in use. Birdcage- and local surface-coils are still commonly used in standard MRI-systems. The last two decades have been dominated by a rapid development of RF-array coils that are dedicated for parallel imaging. Many investigations have examined the upper bound on sensitivity of RF coil array with the use of the concept of the "ultimate" SNR (Lattanzi and Sodickson 2012; Wiesinger et al. 2004; Ohliger et al. 2003). The sensitivity profiles from

DOI 10.1007/978-981-4560-75-7_121-1

arbitrary coils outside a spherical sample based on a set of base functions satisfying the Maxwell equations were evaluated. A linear combination was found that maximizes the SNR at multiple locations in the sample (Ocali and Atalar 1998; Wiesinger et al. 2005). Based on the analysis of the ultimate SNR for accelerated imaging the number of 32 array-channels was found to be close to the theoretical limit at the center of the phantom for clinical MR-coils (Wiggins et al. 2005; Wiesinger et al. 2004). But there is substantial room for improvement towards the periphery of the array coil that is mainly influenced by the increased number of array coil elements. Array coils with 96-channel arrays have been built onto a helmet shaped former that fits closely to the head. The design of the individual circular surface coils was based on the combination of hexagonal and pentagonal symmetry of the soccer ball. Apart from the design effort of building these high count channel RF coil arrays the number of array coil elements reaches its limits today by the availability of receiver channels that are commercially available and economically feasible.

As seen from the theoretical investigations (Wiesinger et al. 2004) the sensitivity gain for accelerated imaging is a function of the number of array-elements and improves at higher field strengths. At higher field strengths where the wavelength is in the same range as the dimensions of the subject under investigations, a homogenous excitation performed by traditional birdcages or TEM-volume coils is impossible. For that reason the trend today goes to so-called transmit-arrays where the profile of the transversal $\mathbf{B}_1$-field can be tailored depending on the demands of the application, e.g., by RF-shimming for a more homogenous excitation in the subject under investigations. The trend to even higher magnet field strength for MR-imaging opens the door and creates new challenges for innovative RF coil designs and concepts.

Transmit Sense/RF Shimming

Transmit Array Coils

The basic idea of parallel-receive imaging techniques is the simultaneous use of multiple receiver coils. This principle can be adapted for parallel transmission of the spatially selective RF-pulses. A multi-dimensional RF-pulse follows a specific trajectory in k-space analogous to data acquisition and shorting this trajectory shortens the pulse duration. Using multiple transmitter channels, each has its individual time-dependent waveform and the spatial sensitivity distribution of the numerous array coil elements compensates for the missing parts of the k-space. Despite the shorting of the pulse duration, the reduction of $\mathbf{B}_1$-inhomogeneity's by tailoring the RF-pulses with Transmit SENSE (Katscher et al. 2003; Ullmann et al. 2005) has recently been successfully performed.

The design principle of receive-array coils can be extended to the transmit-only or the transmit/receive case and all commonly used methods for inductive or capacitive decoupling of the individual coil-elements can be used. When multiple array coil elements are fed by conventional voltage sources then the impedance of the mating network becomes a function of the current amplitude and phase of each of the RF elements. In order to have accurate control over the amplitudes and phases of the currents on each RF element it is necessary to have an accurate knowledge of the mutual coupling.

For proper handling of the mutual coupling so-called multi-port networks (Butler-matrix) have been developed for driving multiple ports with a defined phase and amplitude (Butler et al. 1963; Alagappan et al. 2008). The Butler matrix is a passive circuit network based on hybrid couplers, phase shifters and crossovers that corrects the shared impedances of the n-port coil interface by an $n \times n$ impedance network. The ports to the amplifier have possibly zero shared impedance and can be driven independently. While the impedance on the coil ports changes with different loadings the adjustments of the Butler-matrix network would be possible but are challenging for a high count channel system. The Cartesian feedback (Hoult et al. 2004a, p. 57; Hoult et al. 2004b, p. 64) can be a good choice. The output signal is fed back into the input before and after mixed-up and down. Alternatively the mutual coupling can be reduced by

DOI 10.1007/978-981-4560-75-7_121-1

so-called RF current sources using power- MOSFETS. The principle of using the high or low impedance of the amplifier (section Receive Array Coils for Parallel Imaging) is utilized to mitigate the effect of the mutual coupling (Kurpad et al. 2006).

An example of a multi-transmit MRI-system that consist of numerous independent transmit channels is given in Fig. 40. This system is based on a standard MRI platform that features up to 6 so-called frequency controllers. Each controller drives via a LVDS interface (Low voltage differential signaling – a high speed serial communication), one so-called signal generation unit (SGU). By using the digital frequency, amplitude and phase information issued by the frequency controller, an analogous RF signal is generated by the SGU. This analogous signal working on individual frequencies is fed into a high power RF amplifier (RFA) that generates the high power RF-pulse. For the extent of the 6 transmit channels a so-called Parallel Excitation Multiplexing Component (PEXMUX) is used in order to increase significantly the number up to 48 transmit channels for parallel excitation (PEX) applications e.g. Transmit SENSE. The multiplexing of the signals is performed entirely on the digital side. With this infrastructure standard pulse programs can be used by replacing the normal used RF-pulses by such multi-shape pulses. The individual RF waveforms are specific for each transmit channel and are generated simultaneously.

For PEX the generated transversal magnetization pattern M(**x**) can be expressed as a linear, sensitivity-weighted combination (Ullmann et al. 2005; Katscher et al. 2003) of virtual single-element magnetization patterns $M_n(\mathbf{x})$:

$$M(\mathbf{x}) = \sum_n S_n(\mathbf{x}) M_n(\mathbf{x}), \quad \text{where } M_n(\mathbf{x}) = i\gamma M_0 \int_0^T B_{1,n}(t) e^{-i\,\mathbf{x}\cdot\mathbf{k}(t)}\, dt \tag{86}$$

Where S_n is the sensitivity of the n-th array coil element in which the RF-waveform $B_{1,n}$ is incorporated and $\mathbf{k}(t)$ is the (reduced) k-space trajectory traversed during the RF-transmission. One of multiple approaches for the determination of the RF-waveforms $B_{1,n}$which are necessary to excite a given pattern $M(\mathbf{x})$ is the direct temporal and spatial discretization of the Eq. 86. That leads to the following linear system (u and v are the spatial and temporal indices respectively):

$$\mathbf{m} = \mathbf{A}\mathbf{b}, \quad \text{where } A_{u,(n,v)} = i\gamma M_0 S_n(\mathbf{x}_u) B_{1,n}(t_v) e^{-i\mathbf{x}_u\cdot\mathbf{k}(t_v)} \Delta t \tag{87}$$

For a given magnetization vector **m** the waveforms **b** can be calculated by expressing the linear system (87) as a minimization problem. An example of a PEX-experiment is given in Fig. 41. The transmit sensitivity of all eight coils were determined and converted to the sensitivity-map of the whole array-coil. With this information the individual transmitter channels of an 8 channel transmit-coil were independently driven to perform a homogenous transmit-RF-field with a birdcage-like excitation mode.

Apart from the flexibility for RF-shimming to overcome the problems of standing wave phenomena in the subjects under investigations any desired excitation pattern can be performed (Fig. 42). For this experiment an overview scan was carried out first. In this overview scan a region for excitation was selected (brain domain). Using parallel excitation (Ullmann et al. 2005) this region was then scanned with a full FOV and a reduced FOV.

Another strong motivation for the introduction of local transmitter arrays is an additional degree of freedom for steering E an B-fields, e.g. with respect to the specific absorption rate (SAR) at high-field MRI. The specific absorption rate (SAR) often limits the specific use of imaging methods with respect to the flip-angle or the number of slices at high and ultra-high fields. In addition the nonuniform excitation by traditional volume coils yields also to hot spots where the SAR-values exceed the permitted values locally. But with the usage of transmit excitation with multiple channels the user has more control about the **E** and

DOI 10.1007/978-981-4560-75-7_121-1

B-field distributions. But some practical issues for SAR-control and supervision are still under development and needs further investigations.

With the introduction of a multiple transmit/receive array especially at high and ultra-high magnetic field strengths, the development of new surface array coil elements advanced in new directions within the last years. At lower magnetic fields only the so-called near field of the RF coils is commonly of interest. The $\mathbf{B}_1$-field, which is close to the coil generated by the currents on the RF structure, is in most cases much higher than the electrical field that is generated simultaneously. Only a small fraction of the **B**-field is associated to a propagating wave. With the introduction of MR-systems >=7 T for human applications and increasing the field-strength up to 21 T for preclinical MR-systems, the so-called far-field is becoming more interesting for the design of MRI RF coils depending on whether the target region is located "far" away from the antenna or not. The so-called traveling wave imaging is under development. The waveguide modes supported by the scanner RF-shield are used to excite the MRI signal in the objects under investigations that maybe several meters away from the antenna.

By increasing up the magnetic field strength of MRI-magnets and expanding the resonance frequency of MR-coils to new horizons it becomes more important to investigate in directions that MRI RF coils perform as antennas. Potentially more concepts of commonly used antenna technology, e.g., from telecommunication, will have impact on the designs of antennas for MR-imaging in the future.

Cross-References

- ▶ Antennas in Medical Diagnosis and Treatment Systems
- ▶ Circularly Polarized Antennas
- ▶ Commercial Antenna Design Tools
- ▶ Impedance Matching and Baluns
- ▶ Loop Antennas
- ▶ Near-Field Antenna Measurement Techniques
- ▶ Phased Arrays
- ▶ Transmission Lines

References

Alagappan V et al (2008) A simplified 16-channel buttler matrix for parallel excitation with the birdcage modes at 7T. In: 16th annual meeting of ISMRM, Toronto

Alderman DW, Grant DM (1979) Efficient decoupler coil design which reduces heating in conductive samples in superconducting spectrometers. J Mag Reson 36:447–451

Becker/Sauter Theorie der Elektrizitätslehre Bd.1, Teubner, Stuttgart 1973

Bloch F (1946) Nuclear induction. Phys Rev 70:460–474

Butler J et al (1963) Beamforming matrix simplifies design of electronically scanned antennas. Electron Design 9:170–173

Canet D (1996) Nuclear magnetic resonance: concepts and methods. Wiley, New York

Chin C-L, Collins CM et al (2002a) Birdcage builder: design of specified-geometry birdcage coils with desired current pattern and resonant frequency. Concepts Magn Reson 15(2):156–163

Chin C-L, Collins CM, Li S, Dardzinski BJ, Smith MB (2002b) Birdcage builder: design of specified-geometry birdcage coils with desired current pattern and resonant frequency. Magn Reson Eng 15(2):156–163

DOI 10.1007/978-981-4560-75-7_121-1

Damadian R (1981) NMR in medicine, NMR basic principles and progress, vol 19. Springer, New York
Darrasse L, Ginefri JC (2003) Perspectives with cryogenic RF probes in biomedical MRI. Biochimie 85:915–937
Fitzsimmons JR, Beck B, Brooker HR (1993) Double resonant quadrature birdcage. Magn Reson Med 30:107–114
Ginefri JC, Quinot MP, Girard O, Darrasse L (2007) Technical aspects: development, manufacture and installation of a cryo-cooled HTS coil system for high resolution in-vivo imaging of the mouse at 1.5T. Methods 43:54–67
Gonord P, Kan S (1994) Twin horseshoe resonator. Rev Sci Instr 65:509–510
Griswold MA et al (2002) Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 47(6):1202–1210
Grover FW (1946) Inductance calculation. Van Nostrand, New York
Haacke EM et al (1996) Magnetic resonance imaging. Wiley, New York, 3ff
Haacke EM, Brown RW, Thompson MR, Venkatesan R (1999) Magnetic resonance imaging, physical principles and sequence design. Wiley, New York
Haueisen R, Marek D, Sacher M, Kong F, Ugurbil K, Junge S (2006) Flexible Cryoprobe-setup for mice brain imaging and spectroscopy at 9.4T. Magma 19(suppl 1):78
Haueisen et al. US Patent 2007/0139046
Hayes CE, Edelstein WA, Schenk JF, Mueller OM, Eash M (1985) An efficient, highly homogenouse radiofrequency coil for whole body NMR Imaging at 1.5T. J MagN Reson 63:622–628
Hould DI (2000) The principle of reciprocity in signal strength calculations – a mathematical guide. Concepts Magn Reson 12(4):173–187
Hoult DI (1979) Sensitivity optimization. In: Levy GC (ed) Experimental techniques in ^{13}C spectroscopy, vol 3. Wiley, New York
Hoult DI, Kolansky G, Kripiakevich D (2004a) A 'Hi-Fi' Cartesian feedback spectrometer for precise quantitation and superior performance. J Magn Reson 171(1):57–63
Hoult DI, Kolansky G, Kripiakevich D, King SB (2004b) The NMR multi-transmit phased array: a Cartesian feedback approach. J Magn Reson 171(1):64–70
Jin J (1999) Electromagnetic analysis and design. CRC Press, Boca Raton
Katscher U, Börnert P, Leussler C, van den Brink JS (2003) Transmit SENSE. Magn Reson Med 49:144–150
Kohlrausch F (1956) PraktischePhysikBd 2, TeubnerVerlagsgesellschaft
Kurpad K et al (2006) RF current element design for independent control of current amplitude and phase in transmit phased arrays. Concepts Magn Reson B 29B(2):75–83
Lanz T (2000) The double tuned 1H23Na Crosscage resonator for high field NMR spectroscopy. In: Proc. 8th ISMRM, p 1390
Lattanzi R, Sodickson DK (2012) Ideal current patterns yielding optimal signal-to-noise ratio and specific absorption rate in magnetic resonance imaging: computational methods and physical insights. Magn Reson Med 68(1):286–304
Lauterbur PC (1973) Image formation by induced local interactions: examples employing nuclear magnetic resonance. Nature 242:190–191
Lawrence EC (1996) Image formation methods. In: Harris RK (ed) Encylopedia of Nuclear Magnetic Resonance, Wiley, Chichester, pp 2439–1350
Lee RF et al (2002) Coupling and decoupling theory and its application to the MRI phased array. MRM 48:203–213
Leifer MC (1997) Resonant modes of the birdcage coil. J Magn Reson 124:51–60
Link J (1992) The design of resonator probes. NMR Basic Principle Progr 26:5–31
Luyten RP et al (1989) Broadband proton decoupling in human 31P NMR spectroscopy. NMR Biomed 1:177

DOI 10.1007/978-981-4560-75-7_121-1

Mansfield P (1996) Encyclopedia of nuclear magnetic resonance. Bd. 3, Wiley, New York, 183ff
Mispelter J et al (2006) NMR probeheads for biophysical and biomedical experiments. Imperial Colledge Press, Hackensack
Murphy-Boesch J, Srinivasan R, Caravajal L, Brown TR (1994) Two configurations of a four-ring coil for 1H human imaging and 1H-decoupled 31P spectroscopy of the human head. J Magn Res 103B:103–104
Ocali O, Atalar E (1998) Ultimate intrinsic signal-to noise ratio in MRI. MagnReson Med 39(3):462–473
Ohliger MA, Grant AK, Sodickson DK (2003) Ultimate intrinsic signal-to-noise ratio for parallel MRI: electromagnetic field considerations. Magn Reson Med 50(5):1018–1030
Pruessmann KP et al (1999) SENSE: sensitivity encoding for fast MRI. MRM 42:952–962
Purcell EM, Torrey HC, Pound RH (1946) Resonance absorption by nuclear magnetic moments in a solid. Phys Rev 69:37–38
Reykowski A et al (1995) Calculation of the signal-to-noise ratio for simple surface coils and arrays of coils. IEEE Trans Biomed Eng 42(9):908–917
Reykowski A et al (2002) Design of matching networks for low noise preamplifiers. Magn Reson Med 33(6):848–852
Roemer PB et al (1990) The NMR phased array. MRM 16:192–225
Röschmann P (1995) Analysis of mode spectra in cylindrical N-conductor transmission line resonators with expansion to low-, high- and bandpass birdcage structures. In: Proceedings 3rd annual meeting ISMRM, Nice, France, p 1000
Schnall MD, Subramaniam VH, Leight JS, Chance B (1985) A new double-tuned probe for concurrent 1H and31P NMR. J Magn Reson 85:122–129
Schneider H-J, Dullenkopf P (1977) Slotted tube resonator: a new NMR probe head at high observing frequencies. Rev Sci Instrum 48(1):68–73
Smythe WR (1968) Static and dynamic electricity', 3rd edn. McGraw-Hill, New York
Sodickson DK, Manning WJ (1997) Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. Magn Reson Med 38(4):591–603
Tropp J (1989) Theory of birdcage resonator. J Magn Res 82:51–62
Tropp J (1997) Mutual inductence in the birdcage resonator. J Magn Res 126:9–17
Ullmann P, Junge S, Wick M, Seifert F, Ruhm W, Hennig J (2005) Experimental analysis of parallel excitation using dedicated coil setups and simultaneous RF transmission on multiple channels. Magn Reson Med 54:994–1001
Vaughan JT, Griffiths JR (2012) RF coils for MRI. Wiley, Chichester
Vaughan JT, Hetherington HP, Otu JO, Pan JW, Pohost GM (1994) High frequency volume coils for clinical NMR imaging and spectroscopy. Magn Reson Med 32:206–218
Wiesinger F, Boesiger P, Pruessmann KP (2004) Electrodynamics and ultimate SNR in parallel MR imaging. Magn Reson Med 2(2):376–390
Wiesinger F, DeZanche N, Pruessmann KP (2005) Approaching ultimate SNR with finite coil arrays. In: Proceedings of the ISMRM, 2005. Miami, FL
Wiggins G et al (2005) A 96-channel MRI System with 23- and 90-channel Phase Array Head Coils at 1.5 Tesla. In: Proceedings of the International Society for Magnetic Resonance in Medicine, vol 13
Wong WH (2001) US Patent 6,285,189
Zeeman P (1882) Ueber einen Einfluss der Magnetisirung auf die Natur des von einer Substanz emittierten Lichtes, Verhandlungen der Physikalischen Gesellschaft zu Berlin, S. 127

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_123-2

Space Antennas including Terahertz Antennas

R. Gonzalo*, I. Ederra, J. C. Iriarte and J. Teniente
Electrical and Electronic Engineering Department, Public University of Navarra, Pamplona, Navarra, Spain

Abstract

This chapter deals with the analysis of several kinds of space antennae with a special section devoted to THz antennae. In particular, horn antennae, either corrugated as spline profiles for data downlink and uplink communications and TT&C applications are developed. More innovative antenna designs based on the use of Electromagnetic BandGap (EBG) or Metamaterial structures (MTM) are included. These ones exhibit very promising properties to be used in applications such as TT&C or Navigation. Finally, due to the increasing interest in scientific missions operating at THz frequencies, a section including the last results of using MTM technologies for implementing antennae at THz bands for imaging space applications is presented.

Keywords

Horn antennae; Corrugated antennae; TT&C; Space antennae; Electromagnetic band gap technology (EBG); Metamaterials; Terahertz antennae

Introduction

Following the IEEE standard definitions, an antenna is that part of a transmitting or receiving system, which is designed to radiate or to receive electromagnetic waves. In satellite communications, the use of antennae is fundamental in order to keep the satellite in operation or to be able to run the different applications such as TV broadcasting, data transmission, etc.

Each satellite needs different kind of antennae, and their requirements are very dependent on the applications: scientific, commercial, civil, etc. In general, the antennae on board of a satellite platform can be divided into three groups: antennas for telemetry, tracking, and control (TT&C) which control the satellite operation, antennae for high-capacity transmission, which are mainly in charge of the communications, and finally antennae for the different instruments mounted in the satellite which are fully related with the mission.

For each of the aforementioned groups, there are different antenna technologies, which can be applied in order to comply with the expected performances of the antenna. In the literature, it is possible to find from planar antennae to horn antennae, including arrays or reflectarrays. Anyway, this chapter will focus on two kinds of antenna developments. The first one will be horn antennae, either based on smooth or corrugated profiles. The second case will be devoted to more innovative antennae, which will follow Metamaterial concepts for their implementation.

Horn antennae are well known for many years. They exhibit very good radiation performances which make them very suitable to cover most of the needs for the radiation patterns in a satellite. They present very good pattern symmetry, very low cross-polar levels, very low sidelobe levels, and a wide bandwidth

*Email: ramon@unavarra.es

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_123-2

response. Even in the last years, new profiles have appeared making them very competitive versus other kind of antennae by reducing their overall size (length). This chapter will mainly present the performances of a new kind of corrugated horn antennae, named Gaussian Profile Horn Antennae (GPHA), which have been used in the last years to comply with the stringent requirements of the radiation performances of many different space missions or communication systems. The performances of different GPHA antennae for TT&C, high-capacity, or space instruments will be included.

On the other hand, Metamaterial-based antennae have been recently appeared as a new approach in the development of antennae. In particular and focusing on the space sector, there are different proposals which make use of Metamaterial concepts for improving antenna performances. This chapter will deal with the use of metasurfaces in satellite antennae for TT&C and GPS applications. This innovative concept is an alternative to conventional technologies which can be applied in those applications for improving overall performances, including mass and cost aspects.

A special section on THz space antennae is included. This one will mainly be centered in the progress of MTM antennae operating in this frequency range for space missions.

Data Downlink and Uplink Feed Horn Antennae for Spaceborne Communications

Introduction

Data feed horn antennae on today's communication satellite systems are of very different types. It is not possible to cover all feed horn antenna solutions for all types of spaceborne communications missions, but the following section provides an overview of the most common designs employed nowadays for geostationary communication satellites.

Not much information can be found from actual satellite programs since trade secrets in this highly competitive commercial field protects new developments from being disclosed. However, a lot of useful information from author's experience has been included, supported with several examples and highlighting the main constraints for each antenna type.

Spaceborne communications requirements for data feed horn antennae are always in evolution since next generation systems require higher capacity and flexibility complicating the feed horn design.

Geostationary satellites offer Fixed Satellite Services (FSS) and Broadcasting Satellite Services (BSS) that are often used for broadcasting to and from television networks and local stations. They can also be used for video conference, broadband internet, and general commercial telecommunications. The most common frequencies employed in actual commercial geostationary satellite communications are the Ku band from 12 to 18 GHz, and in K and Ka bands together (often referred simply as Ka band) from 18 to 40 GHz. Typically, Ku band uses the frequencies from 10.7 to 12.7 GHz for the downlink (Tx from the satellite) and 12.8–14.5 GHz for the uplink (Rx from the satellite's view). Note that in some cases, the uplink moves to the 17.3–18.2 GHz range. On the other hand, Ka band usually employs for the downlink (Tx) the 18.3–20.2 GHz range and the 27.2–31.2 GHz range for the uplink (Rx). It is common for FSS/BSS satellites to carry a multitude of feed horn antennae, which are operating at different frequency bands.

Metallic feed horns are the preferred feed elements for most spaceborne communications, because their size is reasonably small and because they have great power handling capability. Additionally, horn design techniques have evolved to the point where very good polarization purity, pattern symmetry, gain performance, reflector aperture illumination control, and frequency bandwidth are readily achievable through the synthesis of the internal horn wall profile. Several horn types are widely used for this type of antennae, but the most widely used nowadays are high-performance corrugated horns and high-

DOI 10.1007/978-981-4560-75-7_123-2

performance smooth-walled spline profile feed horns. Advanced synthesis techniques of rather complex horn wall profiles are able to optimize performance at a number of selected frequencies and therefore significantly widen the frequency band or achieve a very good dual-band performance covering both the Tx band and the Rx band.

Spaceborne Reflector Technology

Feed horn antennae for spaceborne communication missions usually employ reflectors to form contoured beams from the geostationary orbit. There are two main methods to cover the earth's surface from a geostationary satellite:

- A shaped reflector which consists in shaping the energy radiated by a single feed element (usually a high-performance corrugated feed horn) into a contoured beam to form the desired coverage area over the earth's surface (Viskum and Sorensen 1994). The shaping is done by appropriate reflector surface optimization.
- Multibeam reflector antennae where the coverage over the earth is divided in beams that cover the earth surface with cells. This technology employs several reflector surfaces (three or more) which are usually offset parabolic reflectors and a single feed horn per beam (usually high-performance smooth-walled spline profile feed horns) over the earth (Rao 1999). So all the beams must be covered by an array of feed horns.

One of the problems with shaped reflectors, as opposed to multifeed parabolic reflectors, is that only one beam can be generated per reflector profile. For higher rate communications and to allow more reuse of frequencies and more reconfigurability, multiple spot beams and contoured beams are actually the preferred solution.

These systems may use different antennae for Tx and Rx or common Tx/Rx antennae. Mass and cost-saving considerations increasingly drive the design into common Tx/Rx antenna solutions. One challenge in this case is to design the antenna for optimal performance in two widely spaced frequency bands over the same coverage spots on the ground. Designs have evolved to address this challenge. Clearly, the design antennae for Tx only or for Rx only is generally much simpler and often leads to slightly better gain performance across the coverage region. The mass and cost penalties are, however, quite high.

RF Performance Parameter Considerations in Spaceborne Antennae

RF performance parameters include gain (minimum gain across the coverage region as well as peak gain), isolation (gain levels on the predefined isolation areas on the ground) and cross-polarization discrimination (XPD) calculated as the copolarized gain minus the cross-polarized gain in dB.

Minimum gain across the coverage region must be optimized by means of proper reflector and feed horn design. Feed horn copolar radiation pattern plays a key role in this optimization since it should keep a pattern decay at the reflector edge as constant as possible over the whole frequency band as well as minimize spillover radiation to improve the isolation between different areas on the ground.

On the other hand, cross-polarization discrimination is also crucial since it must minimize the crosstalk interference between transmissions using alternative polarizations. Cross-polarization generation in spaceborne antennae is caused mainly by four factors: feed horn inherent cross-polarization, reflector surface curvature, feed horn location away from the focal point and inclination angle of the feed horn relative to the parabolic main axis. Other secondary order, but potentially important contributors to cross-polarization are scattering effects by support struts and feed lines, reflections of the horn primary pattern and the reflector radiated near fields by neighboring structures including the adjacent spacecraft panels.

DOI 10.1007/978-981-4560-75-7_123-2

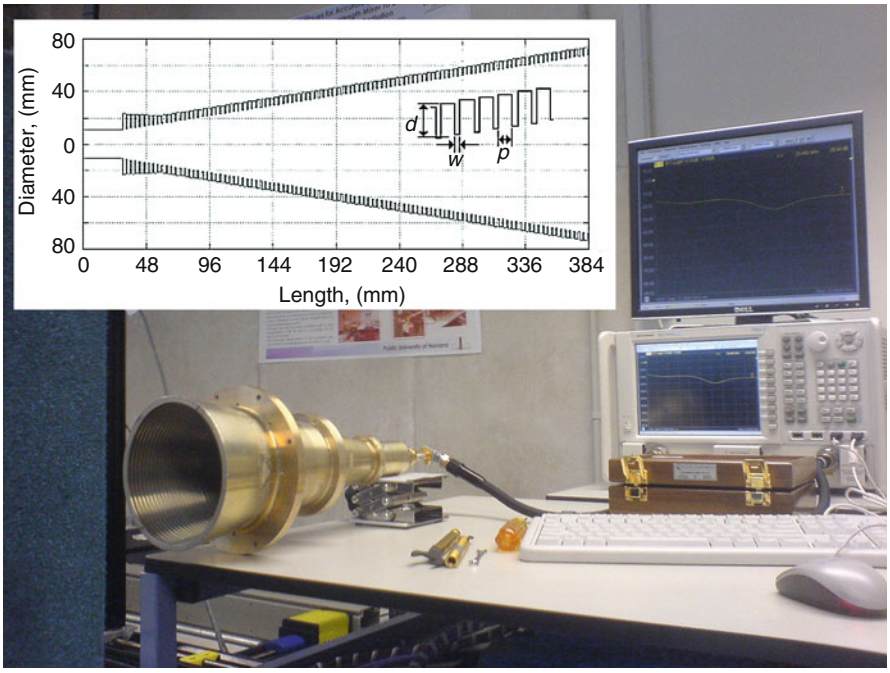


Fig. 1 Conical corrugated horn antenna

These effects should be quantified, analyzed and tested since they can potentially degrade performance significantly.

High Performance Corrugated Feed Horns

Spaceborne communications usually require the use of corrugated horn antennae to feed the shaped reflector surfaces designed to generate contoured beams that illuminate the desired coverage regions over the earth. The reason to use corrugated horn antennae and no other type of feeds is that these type of antennae offer, with improved performance over a quite wide bandwidth; low crosspolar levels, low sidelobe levels and high return losses as well as nice gaussian-like radiation patterns. Therefore, the performance of corrugated horns usually maintains high performance for both the Tx and Rx frequency bands.

A disadvantage of corrugated horns is that they are generally bulky with larger aperture diameter than smooth walled horns for a comparable gain. They are also heavier and more expensive than smooth walled horns. The basic corrugated horn profile is a linear corrugated flare angle leading to a long corrugated feed horn, see Fig. 1.

There is a very wide variety of corrugated horn profiles (Rudge et al. 1982; Olver et al. 1994; Teniente et al. 1999; Granet et al. 2000; Maffei et al. 2000; Hay et al. 2001; Teniente et al. 2002a; Granet and James 2005), some of them but not all of them are: linear, sinusoidal, sine-squared, exponential, gaussian, hyperbolic, polynomial, etc. However, new software tools are now available to facilitate and enhance the effectiveness of corrugated horn profile optimization, making it possible to design this type of horns to meet an even wider variety of performance requirements and volume constraints. Modern commercial software packages, use efficient full-wave analysis techniques such as mode matching, are usually employed in horn design (http://www.mician.com). The agreement between predictions and measured results has become excellent, such that horn prototyping is usually no longer required. Other analysis techniques, such as the finite element method and the method of moments, are also occasionally used, but they are not as computationally efficient as the mode matching method of analysis.

DOI 10.1007/978-981-4560-75-7_123-2

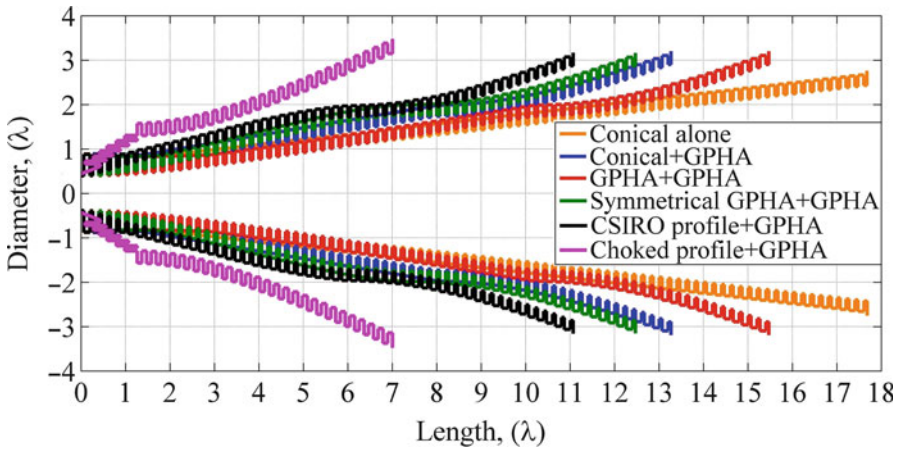


Fig. 2 Length comparison between different-profiled corrugated horn antennas (Teniente 2003) of the same directivity

After the simple conical corrugated profiles, different profiled corrugated profiled appeared, see Fig. 2. This new designs were shorter and the possibilities of optimization increased with the profile shape.

Nowadays, thanks to the available software tools and the speed of computers, the design of a corrugated feed horn for spaceborne communications is usually not any form of profile function. In fact, actual antenna designers leave every corrugation depth and step between corrugations free in the optimization process leading to efficient profiles that cannot be met by simpler profile formulas as the ones used previously.

The shortest corrugated horn profile and indeed very suitable for spaceborne communications as it presents reduced mass and volume is the horn antenna that combines horizontal corrugations (known also as axial corrugations (Bird 2008) or choke horns (Milligan 2005)) for the throat region and vertical corrugations (known also as radial corrugations) for the flare region, see Fig. 2. These horn antenna designs have been used in modern communication satellite feeders as well as in other applications, (Gonzalo et al. 2002; Teniente et al. 2005, 2006; Teniente et al. 2009).

This type of horn antenna achieves significant improvements in axial length, return loss over a wide bandwidth, computation complexity and manufacture complexity, when compared to a conventional corrugated horn with perpendicular corrugations. The only disadvantage is that, depending on the design, the manufacture could require splitting the horn in two pieces, one for the horizontal corrugations and another for the vertical corrugations. Both sections must be bolted together and aligned properly with special care to join them assuring a tight contact, to prevent narrow radiating slots (Granet et al. 2008).

Example of a Ku-band Corrugated Horn to Feed a Shaped Reflector for Spaceborne Communications

As an example, a Ku band communication satellite feedhorn whose specifications are given in Table 1, is presented here. Such feed horn have more than 50.2 % bandwidth and the partial bandwidths are 22.2 % for Tx and 4.9 % for Rx bands respectively. It can be observed that the overall requirements are extremely severe, since a return loss value above 32 dB and a crosspolar level below −45 dB for the whole frequency bands, are required. Note that this kind of profile is very short for a corrugated horn antenna.

Anyway, it is important to remark that every satellite has its own frequency plan that usually determines the feed horn design. However, all of the horn antennae for spaceborne communications that feed a shaped reflector need usually a simulated crosspolar maximum level below −45 dB and a copolar radiation pattern with a certain variation along the frequency band to assure the footprint of the radiation pattern from the satellite over the earth to be frequency independent, see Table 1.

DOI 10.1007/978-981-4560-75-7_123-2

Table 1 Example Ku band spaceborne corrugated feed horn specifications

Parameter	Value requested
Electrical specifications	
Frequency bands	Tx: f_{min1} to $1.25 \cdot f_{min1}$ GHz
	Rx: $1.62 \cdot f_{min1}$ to $1.67 \cdot f_{min1}$ GHz
Maximum crosspolar level	< -45 dB in a range of +/− 22 deg
Return loss	>32 dB in the whole frequency bands
Taper at 20°	~ −9.5 to −16 dB at Tx
	~ −19 dB at Rx
Maximum sidelobe level	<23 dB for angles above +/− 20 deg
Mechanical specifications	
Total length	$<10 \cdot \lambda(f_{min1}(GHz))$ mm
Maximum external diameter	$<3.55 \cdot \lambda(f_{min1}(GHz))$ mm

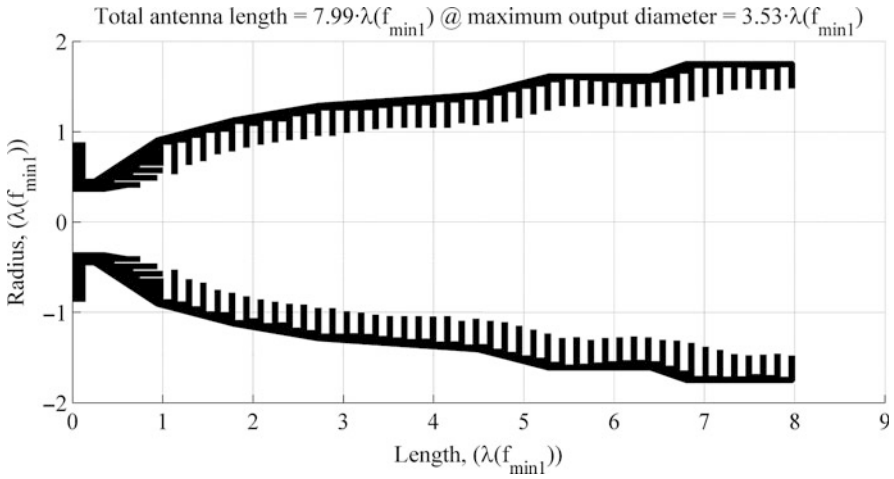


Fig. 3 Optimized corrugated feed horn for ku band

One of the best possible results to comply with the specifications given in Table 1 was a profile with 3 axial corrugations and 44 perpendicular corrugations, see Fig. 3. The resultant length was 20 % shorter ($8 \cdot \lambda(f_{min1}(GHz))$ mm), than the maximum specified, so the mass reduction is really remarkable. The rest of the specifications were met, see Fig. 4. The optimization process selected individually all the corrugation parameters (except thicknesses) and the length and angle of the initial linear taper at the throat. A total of 97 variables were individually optimized for the final design. Final results are shown in Figs. 4 and 5. The results are in good agreement with the specifications.

The resultant radiation patterns can be seen in Fig. 5. The results presented are mode matching simulations [1], but this method as it has been said previously is so accurate that the measurements are extremely similar if the feed horn is properly manufactured. This horn antenna was manufactured in a single piece, see Fig. 6, and is successfully operating onboard a geostationary satellite.

High-Performance Smooth-Walled Spline Profile Feed Horns

Applications requiring higher data rates, such as high-speed Internet services for consumers and small businesses, call for much higher antenna gains and wider frequency bandwidths. Large antenna gains are associated with a small coverage region on the ground and with large antenna aperture diameters in terms of wavelength. Realizing wider bandwidths is increasingly easier as the frequency of operation increases,

DOI 10.1007/978-981-4560-75-7_123-2

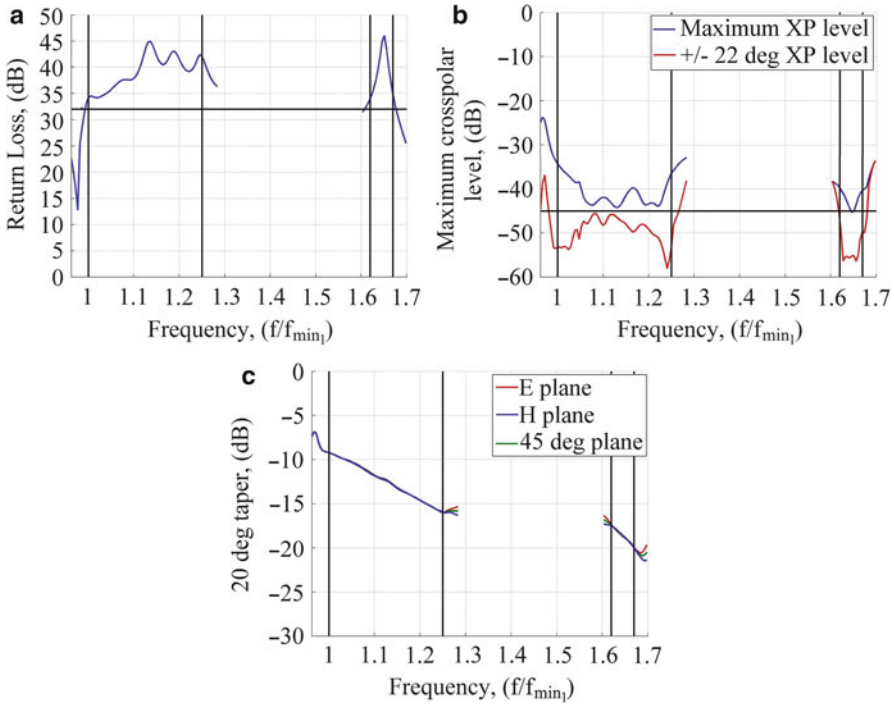


Fig. 4 Results of the optimized corrugated feed horn profile: (**a**) Return loss (**b**) Crosspolar level (**c**) Taper at 20°

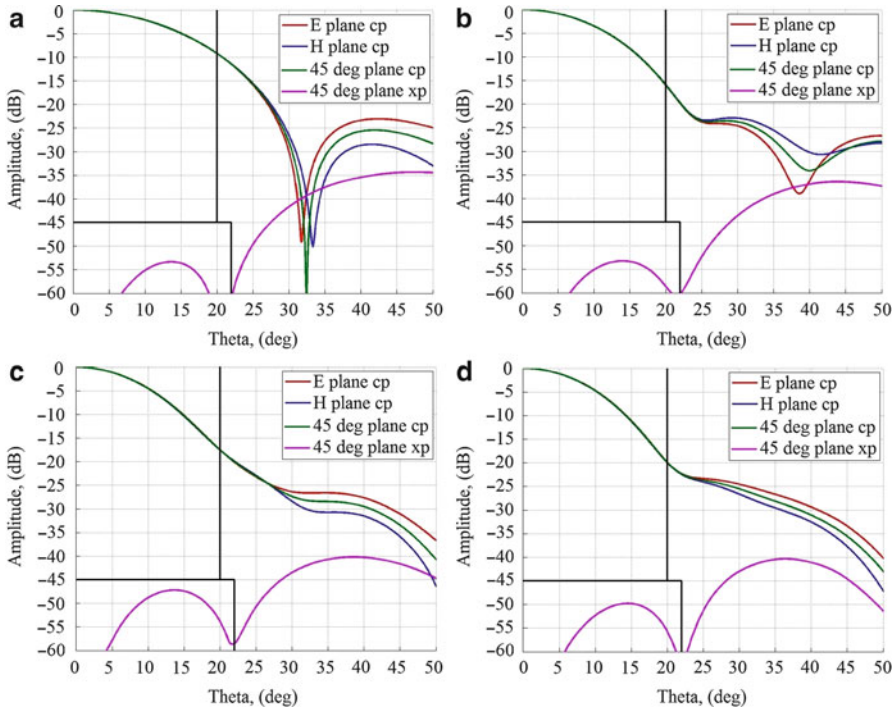


Fig. 5 Far field radiation patterns of the optimized corrugated feed horn profile: (**a**) At f_{min1} (**b**) At $1.25 \cdot f_{min1}$ (**c**) At $1.62 \cdot f_{min1}$ (**d**) At $1.67 \cdot f_{min1}$

since then the bandwidth as a percentage of the center frequency remains relatively small. For these reasons, this type of service has been typically provided with multiple spot beam antennae (with small

DOI 10.1007/978-981-4560-75-7_123-2

Fig. 6 Manufactured corrugated feed horn of the example (Reproduced with permission from ANTERAL company)

circular coverage spots) operating at K/Ka-band (Rao 1999). The coverage spots are typically arranged in a tight hexagonal grid on the Earth, where either all spots have the same diameter or different diameters are used to address different expected traffic density conditions.

These systems may use different antennae for Tx and Rx or common Tx/Rx antennae. Mass- and cost-saving considerations increasingly drive the design into common Tx/Rx antenna solutions and usually smooth wall horns are the preferred solution to reduce drastically the mass of the feed horn array since a lot of horns must be arranged together.

For these systems, known as spot beams, a high efficiency of the horn aperture is needed to ensure better edge of coverage gain (Amyotte and Martins-Camelo 2012; Bhattacharyya and Goyette 2013). This aspect also drives to the use of smooth wall feed horns since an aperture efficiency near 90 % can be met in Tx and Rx bands at the same time with modern spline profiles. On the other hand, if corrugated feed horns are selected, it is difficult to achieve better than 70 % aperture efficiency (aperture efficiency is the relation between effective area and physical area at the horn antenna aperture).

Nowadays, for these kinds of applications, the best solution for spot beams is the use of smooth-walled spline profile feed horns (Granet et al. 2004). Spline-profiled feed horns are longer than corrugated horns, their performance regarding crosspolar level and bandwidth is worse but they are low weight (suitable mass reduction in arrays for spaceborne communications) and aperture efficient. Crosspolar level around −30 dB can be achieved in Tx and Rx bands at the same time which for this kind of multiple spot beams where the reflectors are not shaped is considered enough to ensure the required crosspolar isolation or discrimination (XPI or XPD).

However, simulation of a smooth-walled horn is not as easy as a corrugated horn. A corrugated horn is discrete; their profile is well defined by steps, so the number of steps to analyze would be the number of corrugations multiplied by two. However, smooth-walled horns must be discretized to be analyzed and such discretization must be enough to assure good performance. This issue leads to have above 500 steps in a normal horn, much more steps than in a corrugated horn. In this situation, the designer has not such freedom and must rely on a certain formula to define the profile. Up to now, the best approach is the use of spline profiles. Some results can be found in for a 6 radii case in (Granet et al. 2004) and for higher number in (Zeng et al. 2010).

DOI 10.1007/978-981-4560-75-7_123-2

Table 2 Example Ka band spaceborne smooth-walled spline profile feed horn specifications

Parameter	Value requested
Electrical specifications	
Frequency bands	Tx: f_{min1} to $1.04 \cdot f_{min1}$ GHz
	Rx: $1.5 \cdot f_{min1}$ to $1.54 \cdot f_{min1}$ GHz
Maximum crosspolar level	< -26 dB in a range of +/− 15 deg
Return loss	>40 dB in the whole frequency bands
Taper at 13°	Spillover at Tx < 0.5 dB
	Aperture efficiency at Rx > 0.85
Mechanical specifications	
Total length	$<14 \cdot \lambda(f_{min1}(GHz))$ mm
Maximum external diameter	$<4.8 \cdot \lambda(f_{min1}(GHz))$ mm

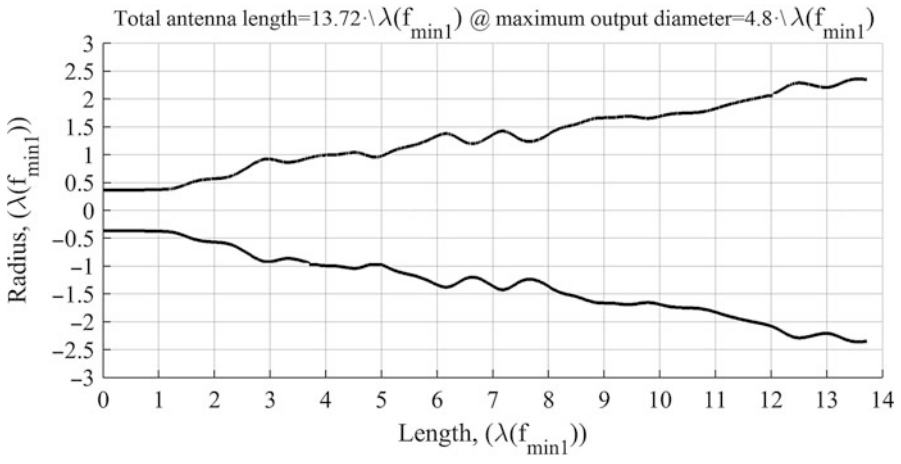


Fig. 7 Optimized smooth-walled spline profile feed horn

Example of a Ka-band Smooth-Walled Spline Profile Feed Horn to Feed Multibeam Reflector Antennae for Spaceborne Communications

As an example, a Ka band communication satellite feedhorn whose specifications are given in Table 2, is developed. Such feed horn has more than 42.3 % bandwidth and the partial bandwidths are 3.5 % for Tx and 2.4 % for Rx bands respectively. It corresponds with a very stringent requirement for a smooth-walled horn, since a return loss value above 40 dB and a crosspolar level below −26 dB for the whole frequency bands, are required. Furthermore, it is expected to present a reduced length.

The design employs 40 radiuses to define the profile, see Fig. 7. The result is rather curved inside for a smooth-walled horn, but with this profile, see Figs. 8 and 9, the requirements can be met.

The resultant radiation patterns are plotted in Fig. 9. This horn antenna was manufactured and it is successfully operating onboard a geostationary satellite where it forms part of a feed horn array for spot beam communications.

Navigation Antennae

Navigation systems are commonly used nowadays. They provide geospatial position with global coverage to users. There are different navigation systems as GPS from USA, GLONASS from Russia, Compass

DOI 10.1007/978-981-4560-75-7_123-2

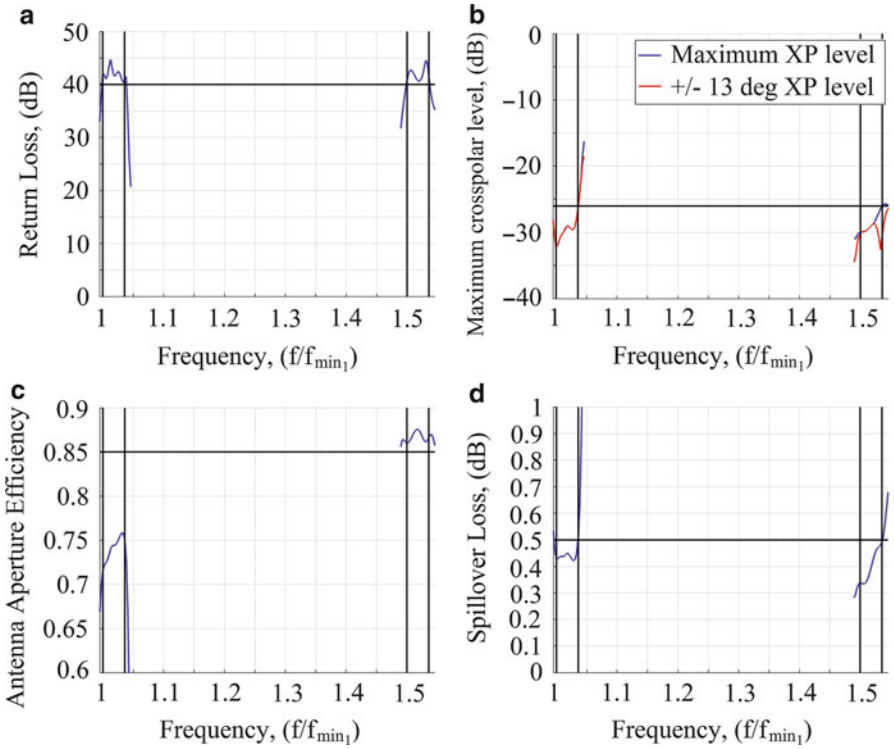


Fig. 8 Results of the optimized smooth-walled spline profile: (**a**) Return loss (**b**) Crosspolar level (**c**) Aperture efficiency (**d**) Spillover loss above 13°

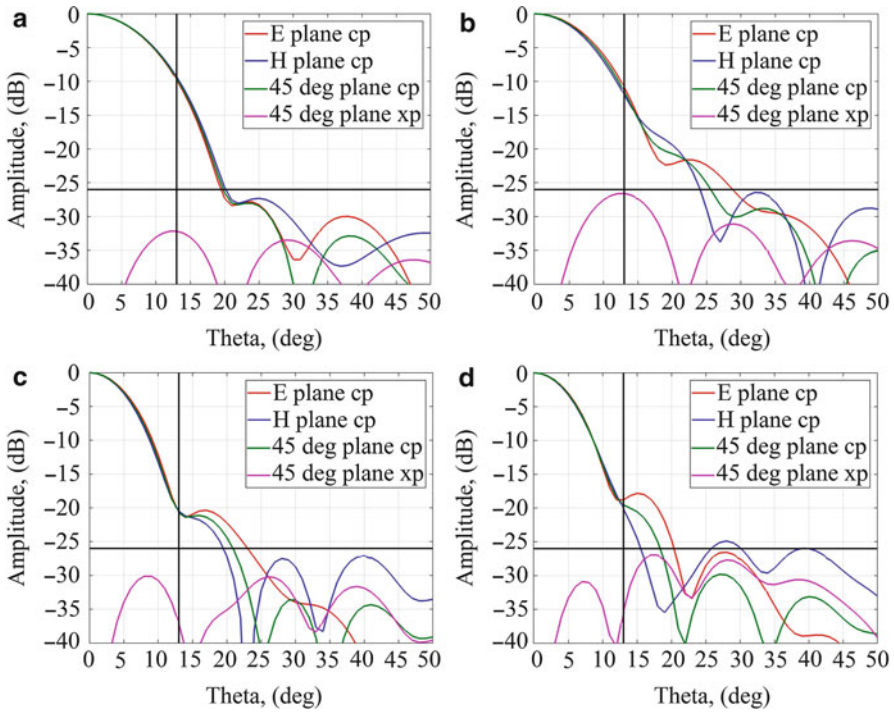


Fig. 9 Far field radiation patterns of the optimized smooth-walled spline profile: (**a**) At f_{min1} (**b**) At $1.04 \cdot f_{min1}$ (**c**) At $1.5 \cdot f_{min1}$ (**d**) At $1.54 \cdot f_{min1}$

DOI 10.1007/978-981-4560-75-7_123-2

Table 3 Frequency band for GPS, GLONASS, Galileo and Compass

System	Coding	Frequencies (MHz)	Country
		B1: 1559.052–1591.788	
Compass	CDMA	B2: 1162.220–1217.370	China
		B3: 1250.618–1286.423	
		E1: 1575.420	
Galileo	CDMA	E6: 1278.750	European Union
		E5b: 1207.140	
		E5a: 1176.450	
		L1: 1602 and 1575.420	
		L2: 1246 and 1242	
GLONASS	CDMA/FDMA	L3: 1202.025	Russia
		L5: 1176.450	
		L1: 1575.420	
GPS	CDMA	L2: 1227.6	USA
		L5: 1176.450	

from China, and Galileo from the European Union. Their operational frequencies cover two bands: 1150–1300 MHz and 1559–1611 MHz. The frequency band from 1300 to 1560 MHz is reserved for other radio users such as military, telemetry, radio astronomy, Earth-to-space operation, and satellite Internet services. There are frequency overlaps between systems, creating a compatibility between them and allowing users to obtain their position using several systems at the same time. The working frequencies of each system can be seen in Table 3.

The accuracy of these systems needs to be improved. Geosynchronous satellite Navigation antennae (GSNA) are used to provide GPS/GALILEO/GLONASS signal corrections for better position accuracy and to certify signal integrity. The accuracy of Navigation systems has been improved by the used of these complementary systems as the Wide Area Augmentation System (WAAS) in USA, EGNOS in Europe, and MSAS in Japan.

In this section, the Wide Area Augmentation System (WAAS) has been taken as GSNA reference application. WAAS is an extremely accurate navigation system developed to complement the Global Positioning System (GPS). One of the important components of this system is the Tx L-Band navigation antennae providing global Earth coverage from geostationary satellites. These on board antennae operate in circular polarization with Axial Ratio (AR) values lower than 1.5 dB and provide gain values around 17 dBi over the entire 18° beamwidth of the conical coverage area representing the antenna Field of View (FoV). The intended Navigation Antenna application requires also to achieve a precise and stable antenna phase center location (<20 mm) over both the antenna Field of View (FoV) and the operating frequency bandwidth, to allow for maximum overall system positioning accuracy. The required WAAS channel fractional bandwidth is relatively small and represents about 1.5 % of the positioning system L1 channel center frequency in L-Band. The main requirements from antenna point of view are shown in Table 4.

Conventional technology offers different solutions to comply with GSNA specifications. Array configurations with 9–19 elements size are necessary to obtain the desired gain values in the coverage area. Different solutions are given using patches or helixes elements to create the array. The number of elements in these arrays is not such a critical parameter. The main problem in these configurations is the Beam Forming Network (BFN) design, cost, and mass. The fabrication process of the helixes antennae could also be not either easy.

DOI 10.1007/978-981-4560-75-7_123-2

Table 4 WAAS specifications

Center frequency	1575.42 MHz
Bandwidth	24 MHz
Bandwidth at 4GHz	80 MHz
Gain over coverage	>16.9 dB
Edge of coverage	8.9°
Polarization	Circular RHCP
Axial ratio on coverage	<1.5 dB
Phase center collocation	<20 mm
3 Standard deviations over complete	
FoV and frequency band	
Diameter	<570 mm
Height	<350 mm

EBG GSNA Configuration

Electromagnetic Band Gap (EBG) technology offers new alternatives to design competitive solutions to conventional technology. EBG designs can obtain similar and even better, radiation performances (complying with all the radiation specifications) while the cost, dimensions, and manufacturing process complexity is minimized. The BFN can be simplified by reducing the number of radiating elements in the array configuration using the gain enhancement mechanism given by a superstrate EBG configuration. EBG technology can be applied to these applications due to the reduced required bandwidth (≅1.5 %).

In this section, an alternative to conventional technology based in EBG technology is presented. The EBG superstrate configuration complies with the RF requirements, while it reduces the cost and height and simplifies the BFN considerably. The comparison between the prediction and measurements will be shown.

The theory of the EBG gain enhancement phenomenon has its foundation in the pioneering work of Von Trentini (1956) on the gain enhancement properties of cavity antennae. It consisted in a partially reflective surface (PRS) located a quarter wavelength from a ground plane. The structure forms a Fabry-Perot cavity. Power is reflected and transmitted (radiated) in each bounce of the trapped waves. When the power radiated is in phase over the partially reflective surface, a directivity increase is achieved. Under this condition, the radiation aperture has been increased. Out of the resonance band of the cavity waves are radiated out of phase, creating destructive interference in the far field. This gain enhancement phenomenon can also be explained using leaky wave antennae theory (Lovat et al. 2006a, b).

Successively, more studies were done in 1985 (Jackson and Alexopoulos 1985) and 1988 (Jackson and Oliner 1988) where the reflective surface was substituted by a quarter wavelength dielectric. Higher directivity values can be achieved using several dielectric layers (Jackson et al. 1993; Thèvenot et al. 1999a; Feresidis and Vardaxoglou 2001).

Different techniques have been applied to create partially reflected surfaces. Frequency Selective Surfaces (FSS) or EBG with a defect in their periodicity can be used to produce the gain enhancement phenomenon (Biswas et al. 2001; Fehrembach et al. 2001; Cheype et al. 2002; Lee et al. 2005). The introduction of a defect in the stacked structure leads to the apparition of a narrow pass band, where the waves incident to the structure can suddenly be propagated in the now defective crystal. The quality factor of the defective crystal can be adjusted by playing with the permittivity of the slabs, and the quality factor decreases with the decrease of the dielectric permittivity. This quality factor is going to limit the maximum achievable gain and the operational bandwidth of the antenna configuration (Diblanc et al. 2005). The higher the Q-Factor, the smaller the operating relative bandwidth of the structure is.

DOI 10.1007/978-981-4560-75-7_123-2

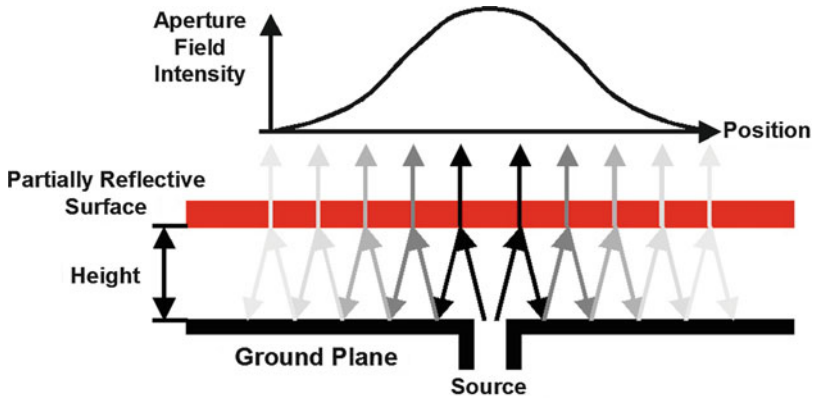


Fig. 10 Depiction of the spatial filtering concept

This technology is intrinsically limited to relatively narrow band applications where the operating relative frequency bandwidth is limited to a few percents. As the frequency increases from the optimum operating frequency providing the antenna peak gain, the directivity decreases and the pattern gets distorted. On the other hand, most of the rays are reflected back by the EBG when the frequency is lower than the optimum gain frequency, therefore trapping the RF energy emitted by the source between the ground plane and the reflecting slab, see Fig. 10.

EBG theory has provided a new and powerful framework for the understanding of the gain enhancement phenomena and has enabled significant advances in the design of antennae using partially reflecting surfaces. The gain enhancement phenomenon is now described in terms of wave propagation in 1D defective EBGs and it has been shown that the gain enhancement factor and the bandwidth are closely interrelated and essentially fixed by the quality factor of the defective 1D crystal resonance (Iriarte et al. 2006). The quality factor of the defective 1D crystal resonance can be adjusted by adjusting the reflectivity of the partially reflective surface and generally increases when the reflectivity of the surface is increased. Thin metallic sheets with holes or etched metallic patterns on thin dielectric sheets can be used to produce lightweight 1D defective EBGs for space antenna applications.

One of the main advantages that the EBG technology could bring to navigation antennae is the reduction of the antenna complexity by reducing the number of radiating elements to just a few or even to a single one. It has been shown in Thèvenot et al. (1999b), that a 1D defective crystal with a high quality factor and a single element source can meet the directivity and Axial Ratio requirements of the WAAS application. However, it has also been found that the phase center stability is significantly degraded with this large gain enhancement configuration. It therefore appears that the phase center stability needs to be traded-off against the quality factor of the defective 1D crystal and the intrinsic directivity of the source illuminating the EBG antenna. A quality factor of about 50 combined with a 2×2 element array exciting the antenna 1D defective crystal structure was found to be a good compromise in order to meet the desired gain and phase center stability requirements of the WAAS antenna application.

An EBG design, frequency scaled to C-Band is shown in Fig. 11. The model has been scaled due to measurement facilities in this frequency range. The result is completely scalable. It consists of a 2×2 array of circularly polarized two-port patches fed in quadrature and placed underneath an EBG superstrate consisting of an etched metallic pattern of circular holes on a thin dielectric sheet. The circular holes are positioned on a regular square matrix. The radiating elements are sequentially rotated in 90° steps to improve the Axial Ratio over the antenna Field of View (Iriarte et al. 2009).

At the edges of the ground plane, four dielectric threaded rods provide support to the EBG layer and allow for adjustment of the height of the layer above the radiating elements and antenna ground plane. The EBG superstrate is etched on a thin 50 μm thick Kapton substrate and tensioned on a fiberglass reinforced

DOI 10.1007/978-981-4560-75-7_123-2

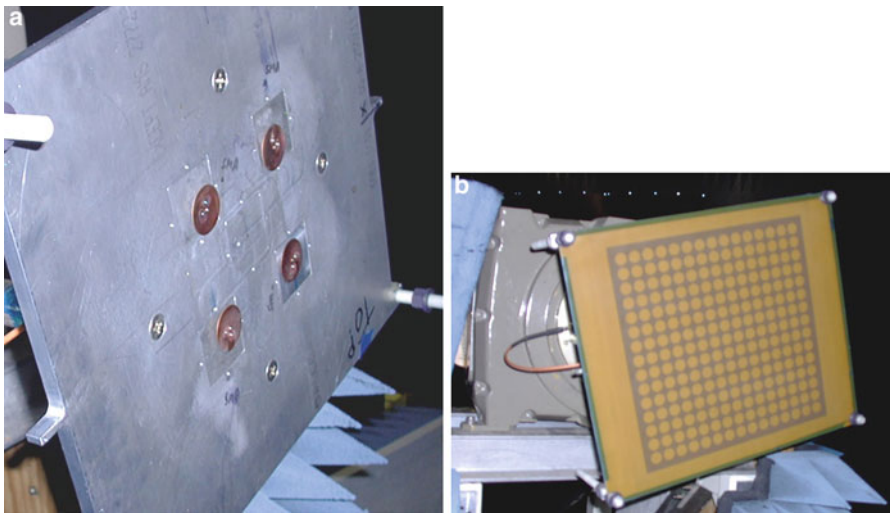


Fig. 11 (**a**) Patch array configuration and (**b**) EBG gain enhanced WAAS navigation antenna

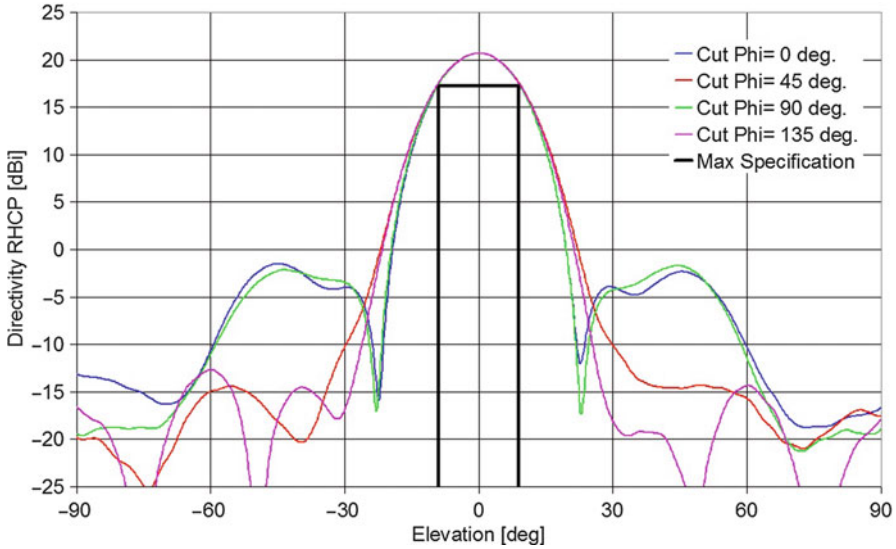


Fig. 12 RCHP directivity elevation cut at $\theta = 0°$, 45°, 90° and 135° azimuth at 3.74 GHz of GSNA EBG antenna

epoxy frame. Johannson Blocks were used to set the EBG layer height above the ground plane with an accuracy better than 50 μm. The height of the EBG layer above the ground plane is about 0.45 λ_0.

RF Performances

The optimum operating frequency band of the frequency scaled-up is from 3,700 to 3,780 MHz. The RHCP directivity elevation cuts at $\theta = 0°$, 45°, 90°, and 135° at 3,740 MHz are provided in Fig. 12, and the axial ratio cuts the same θ° cuts at 3,740 MHz are provided in Fig. 13.

The radiation pattern performance summary of the GSNA configuration is provided in Table 5. The antenna peak RHCP directivity ranges from 20.5 to 21.2 dBi. The worst RHCP directivity over the FoV ranges between 17.9 and 18.3 dBi.

0.47 dB of losses have been estimated for the design. Including the 0.47 dB of losses, the minimum RHCP gain over the FoV is 17.39 over the antenna operating bandwidth. Axial ratio over the FoV is better than 1.2 dB, therefore compliant to the targeted 1.5 dB value.

DOI 10.1007/978-981-4560-75-7_123-2

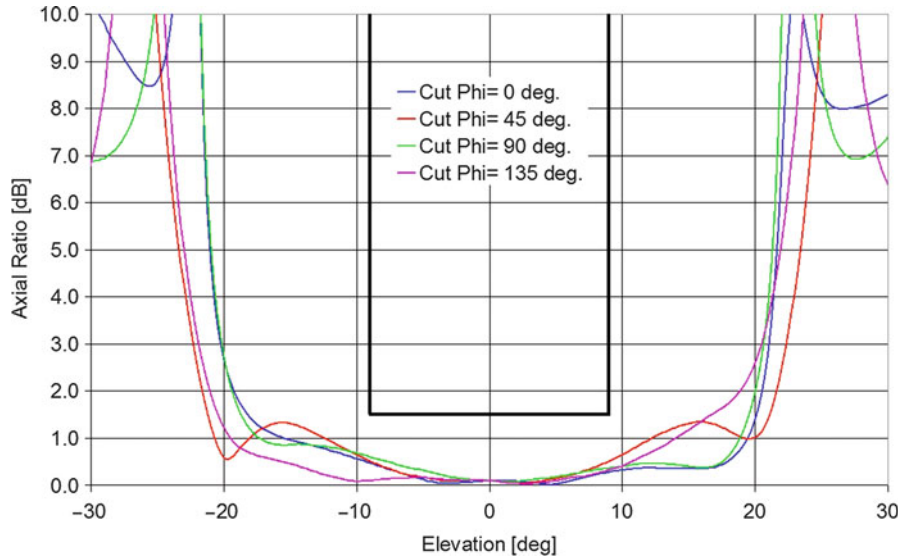


Fig. 13 Axial ratio elevation cut at $\theta = 0°$, 45°, 90° and 135° azimuth at 3.74 GHz of the GSNA EBG antenna

Table 5 Summary results for the GSNA EBG antenna

Frequency	RHCP Peak directivity (dBi)	Worst RHCP FoV directivity (dBi)	Worst FoV axial ratio (dB)
3700	20.51	17.86	1.11
3710	20.67	17.00	0.96
3720	20.80	18.11	0.84
3730	20.92	18.20	0.79
3740	21.01	18.26	0.69
3750	21.08	18.31	0.60
3760	21.14	18.33	0.58
3770	21.18	18.33	0.64
3780	21.21	18.32	0.60
Average over the channel bandwidth			
3740	20.95	18.19	0.76
Worst over the channel bandwidth			
3740	20.51	17.86	1.11

Different phase center calculation methods can be used to estimate the value from the antenna parameters. In this case, standard 3σ deviation and peak-to-peak methods are used to its calculation. The best phase center collocation properties of the GSNA EBG antenna over both the frequency band and the antenna FoV, provided in Table 6, are meeting the targeted specification presented in Table 4 using both methods.

Flight Implementation Considerations

In order to move towards a fully space qualified Navigation antenna product, several issues linked to both the mechanic and electrical environments specific to space applications must still be addressed. In this section, the focus will be on the impact of the thermomechanical distortions on the antenna frequency of operation, and on the evaluation of the available margin with respect to multipaction factor.

DOI 10.1007/978-981-4560-75-7_123-2

Table 6 Phase center variation over both the antenna field of view and the frequency bandwidth for the GSNA EBG antenna

Channel phase center data (scaled to L1) (mm)	
Radial variation 3σ dev.	17.55
Radial variation Pk-Pk.	19.83

Impact of Thermomechanical Distortions

The typical thermal environment of a low-power top-floor antenna with passive thermal control (thermal blankets and sunshields) is stringent when the antenna base is not radiatively coupled to the spacecraft Earth Deck. In this thermal configuration, the antenna can face minimum and maximum temperatures in the order of −130 °C to +120 °C and must remain operational.

The most critical dimension of the antenna is the separation between the ground plane and the superstrate. Any error in this dimension shifts the operational bandwidth of the antenna. In fact, the fractional frequency shift is proportional to the fractional variation of this dimension. Assuming that the manufacturing errors can be leveled out from the design by careful trimming of the antenna structure at ambient temperature (20 °C), the antenna will still suffer from some frequency shifts due to thermal distortions. Although that low Coefficient of Thermal Expansion (CTE) materials can be used in the fabrication of the antenna superstrate structure, such as Kevlar (CTE ~4 ppm/°C) or Invar (CTE ~2 ppm/°C), the fairly large temperature excursions seen by the antenna will lead to some dimensional changes as well as variations of the effective dielectric constant of the dielectric materials.

The separation between the ground plane and the superstrate is approximately $L = \lambda_0/2$ and the temperature variation will result in a variation of the electrical length between the ground plane and the superstrate. The fractional frequency shift Δf and the electrical length fractional variation ΔL_e are given approximately by:

$$\Delta f \approx \Delta L_e \approx \Delta L_p + \Delta k_{eff}$$

where $\Delta L_p = CTE\ \Delta T$ is the fractional variation of the physical separation between the ground plane and the superstrate and Δk_{eff} is the fractional variation of the effective propagation constant.

The dielectric constant of Kevlar varies by approximately 6 % over the −130 °C to +120 °C temperature range. In order to obtain a light low permittivity sandwich supporting the EBG layer, low density Kevlar honeycomb could be used to provide the structural stiffness. A conservative estimate of the filling factor of the cavity with Kevlar is 5 % resulting to an average low-density sandwich relative permittivity of $\varepsilon_{eff} \approx 1.1$. This leads to fractional variations of the propagation constant of:

$$\Delta k_{eff} = \sqrt{\Delta \varepsilon_{eff}}$$

The fractional bandwidth required in the antenna design is therefore given by:

$$\frac{BW}{f_0} = \Delta f + \frac{24}{1575.42} \cong \Delta f + 1.53\%$$

which leads to a required design bandwidth of 29.6 MHz at L1. This corresponds to the 80 MHz bandwidth at C-Band (4 GHz) used in the simulations and measurements of the breadboard.

Considerations on Multipactor

The strongest concern about the power handling capability of this type of antenna configurations in a space environment lies in the field buildup related to the 1D defective photonic crystal quality factor at its

DOI 10.1007/978-981-4560-75-7_123-2

resonant frequency. A multipaction analysis has therefore been conducted on the EBG gain enhanced antenna configuration to understand the power-handling capability of this structure.

Evaluation of the impact of thermomechanical distortions on the antenna operating frequency band shows that the proposed solution can be implemented using low CTE materials to make the critical dimensional parts of the antenna. As well, a preliminary multipaction analysis shows that RF power in the 100 W range can be used with this type of antenna in the space environment.

Telemetry and Telecommand Antennae

Telemetry, Tracking, and Control (TT&C) are vital functions for successful operation of all satellites, because they lead the spacecraft management. The main tasks of a TT&C system are, for instance, to monitor the performance of all satellite subsystems and transmit the monitored data to the satellite control center for routine operational and failure diagnostic purposes. Furthermore, it must represent a tracking platform to earth stations by accomplishing the determination of orbital parameters to maintain a satellite in its assigned orbital slot and provide look angle information to earth station in the network. Besides, the system receives commands from the control center to execute various function of the satellite, for example, power on/off subsystems, change subsystem operating modes, deploy booms, antennae, solar cell arrays, and so forth. Usually, uplink (UL) and downlink (DL) communications with TT&C satellites are carried out at S-band (UL: 1.6–2.2 GHz and DL: 2.2–2.3 GHz) such as in case of Inmarsat or Galileo (Galileo System Requirement 2002), C-band (UL: 5.9–6.5 GHz and DL: 3.7–4.2 GHz), and Ku-band (UP: 14–14.5 GHz and DL: 11.7–12.2 GHz).

For the purpose of providing robust communications with the established receivers, the TT&C antennae located onboard the satellites must provide appropriate earth coverage and work simultaneously with right hand (RHCP) and left hand circular polarization (LHCP). Particularly in the systems working at the C-band, one of the most stringent radiofrequency (RF) requirements regarding the antennae performance is the necessary gain; the typical value is 16.5 dBi at the edge of coverage (EOC), that is, $\pm 9^{\circ}$. Commonly, high bulky horn antennae (up to 4 kg for low C-band frequencies) and 2 kg for high C-band frequencies are implemented for TT&C applications. The reduction in the mass and dimensions of such antennae becomes a major concern for the spatial industry, especially when the carrier satellites are small. Due to this fact, the development of competitive and alternative antenna technologies complying with the overall RF requirements of conventional antennae but providing at the same time the advantage of lightness and compactness represents an attractive research area. TT&C onboard antenna requirements are shown in Table 7.

As it has been shown in the GSNA EBG antenna, the incorporation of EBG technology-based structures in different emerging communication frameworks consolidate them as feasible solutions for different applications, such as wireless and satellite industries (Galileo 2002; Rodes et al. 2006; Brand et al. 2009; Hajj et al. 2009; Iriarte et al. 2009; Oses et al. 2009; Kanso et al. 2011). Requirement of TT&C on board antennae applications can be covered by the use of EBG gain enhancement configuration antennae.

An EBG model which complies with the requirements of the selected TT&C application is presented in next section. These low-mass EBG superstrate antennae are capable to fulfill the given C-band system requirements. Characteristics such as the gain over coverage, dual circular polarization, return loss, isolation, and mass levels demanded by this kind of satellite systems are analyzed. The obtained results confirm the viability of the proposed EBG-based antennae for onboard TT&C antenna configurations.

Anyway, for Ku-band, horn antennae can be used for TT&C since the higher frequency reduces size and the mass. A very compact horn antenna for TT&C is presented afterwards. This horn antenna was

DOI 10.1007/978-981-4560-75-7_123-2

Table 7 TT&C antenna requirements

Center frequency	3.7GHz
Bandwidth	80 MHz (goal 120 MHz)
Gain over coverage	>16.5 dB
Edge of coverage	9°
Polarization	Circular (RHCP/LHCP)
Crosspolarization	< −25 dB (1 dB AR)
(CP/CX)	
Return loss	< −18 dB
Isolation	>25 dB
Power handling	10 Watt
Mass	<800 g
Envelope	To be minimized (height < 2.5λ)

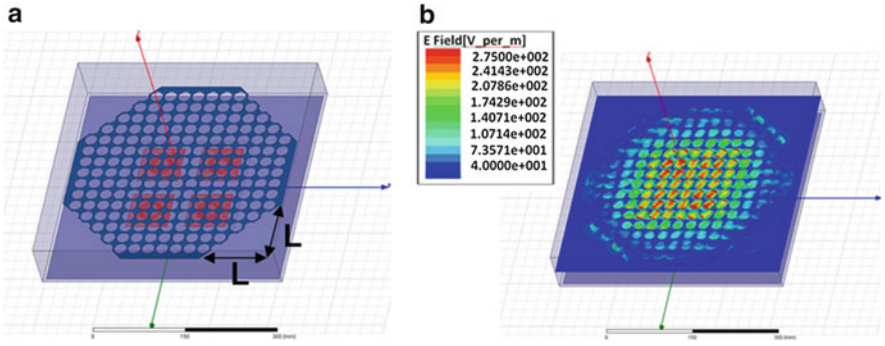


Fig. 14 Final truncated EBG superstrate configuration. (**a**) Isometric view. (**b**) Field confined in the center of the cavity

capable to cover simultaneously downlink and uplink frequencies and was connected to a very compact septum polarizer to produce the required circular polarization with enough isolation between ports.

EBG TT&C Antenna

As in the case of GSNA navigation antennae, gain increase can be accomplished by the resonance formed in a cavity, which, in the presented case, is created between a metallic ground plane and an EBG superstrate that is acting as a partially reflective surface. The latter consists of a metallic sheet with a square lattice of drilled holes, see Fig. 14.

The EBG superstrate fixes the operational frequency and the final directivity of the antenna. The radius of the holes determines both the magnitude and phase of its reflection coefficient. Meanwhile the phase determines the resonance frequency, the magnitude relates with the field spread on the superstrate surface and therefore determines the illumination, fixing the directivity.

The radiating system is composed by a 2 × 2 patch array feeding the resonant cavity created between the ground plane and the EBG superstrate (see Fig. 14a). The distance between patches was optimized in order to obtain an improved symmetric pattern with low sidelobe levels. Two feed ports are perpendicularly located at each radiant patch, which are fed with a phase difference of 90 °. In addition, each patch of the 2 × 2 array is sequentially rotated with respect to its adjacent radiators. Thus, not only circular polarization at each patch is ensured but also further improvement in the circular polarization of the complete feed system.

DOI 10.1007/978-981-4560-75-7_123-2

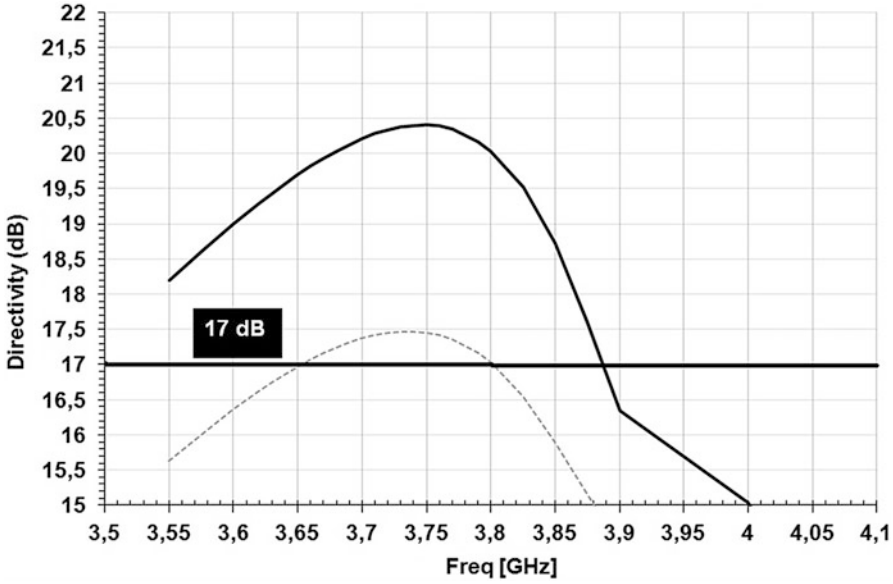


Fig. 15 Directivity vs. frequency at $\theta = 0°$ (*black solid line*) and $\theta = 9°$ (*gray dashed line*)

As the mass is critical in this application, the total dimension of the superstrate is also a critical parameter in the design process. It must be large enough to comply with the maximum directivity requirement stated in Table 7, but at the same time, it must be as light as possible. In a preliminary structure of the proposed configuration, the dimension of the superstrate was set to 365 × 365 mm. Since this area causes a significant increase in overall system weight, a strategy to overcome this problem consisting on truncating the corners of the superstrate is undertaken (see Fig. 14a). This solution does not produce any degradation of the antenna properties as the electromagnetic field is mostly confined in the center of the cavity (see Fig. 14b). After an optimization process trying to keep unaltered the radiation performances of the antenna, the cuts from the corners are stated to L = 110 mm so the resulting total area is 109.025 mm^2.

The EBG layer is placed 0.45 λ_0 at working frequency above the ground plane.

The losses of the radiators and BFN have been considered as 0.15 and 0.16 dB respectively, together with matching and connector losses (0.04 and 0.05 dB). An implementation margin including thermal effects of 0.10 dB has been added so that the total losses are predicted to be equal to 0.5 dB in the final system. Therefore, the gain of the proposed configuration should be above 16.66 dBi in the desired working bandwidth (see Fig. 15).

Figure 16 presents the RHCP and LHCP radiation patterns for 0°, 45°, and 90° phi cuts at 3.66, 3.73, and 3.79 GHz for the EBG-based antenna. Note that the low part of the black rectangular box embedded determines the −25 dB required crosspolar level. It is clear that the designed EBG antenna complies with this RF performance. Similarly, in Fig. 17, the axial ratio parameter is depicted. The AR value is under 1 dB at 3.74 GHz for all the phi cuts. At the same time, the AR for the worst phi cut at EOC is plotted in Fig. 17b with a gray dashed line. Analyzing this result, the AR operational bandwidth goes from 3.61 to 3.92 GHz, fulfilling the 80 or 120 MHz bandwidth stated at the specifications. The directivity and AR for the TT&C EBG truncated superstrate antenna are summarized in Table 8. Regarding the isolation between ports, the appropriate design of the BFN is the key issue. It must provide the optimized amplitude and phase coefficients to the array elements. In this case, uniform excitation has to be settled for both amplitude and phase. The BFN is composed by 1:2 T-dividers whereby supplied amplitudes and phases in each port guarantee RHCP.

DOI 10.1007/978-981-4560-75-7_123-2

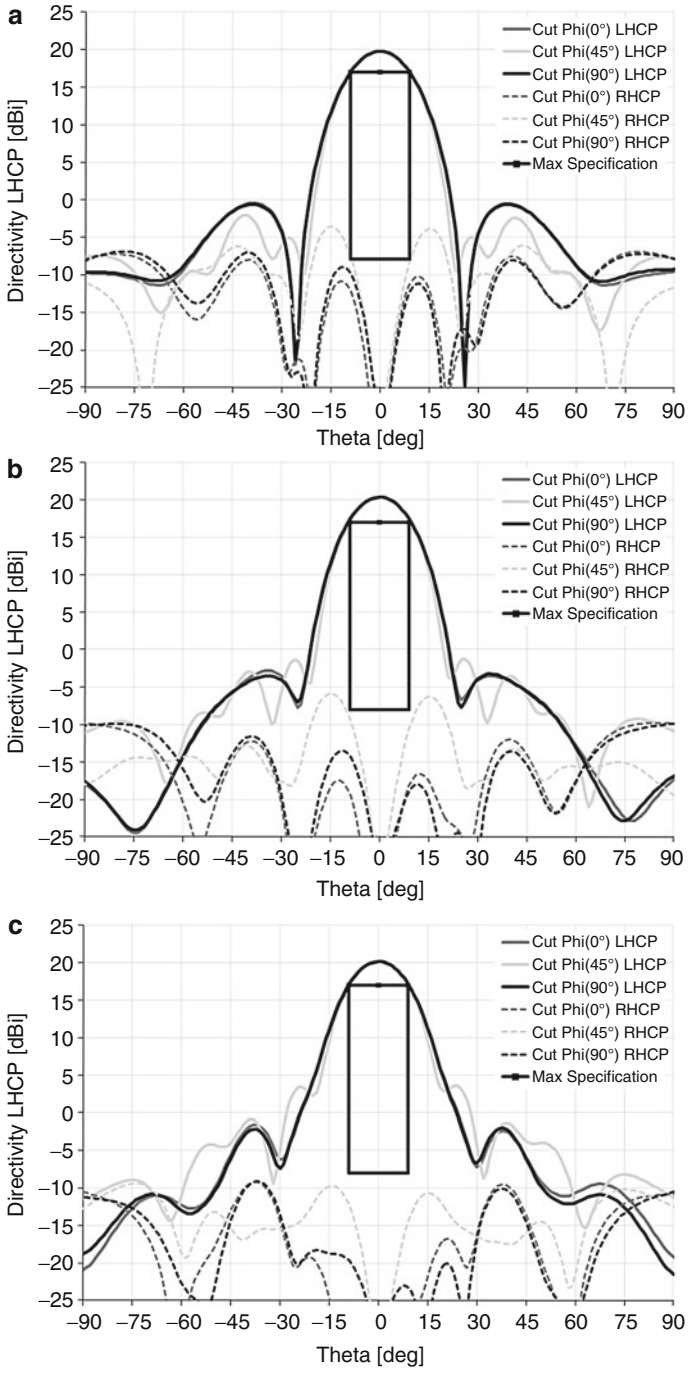


Fig. 16 RHCP and LHCP radiation pattern at $\theta(0°)$, $\theta(45°)$ and $\theta(90°)$. (**a**) 3.66 GHz, (**b**) 3.73 GHz and (**c**) 3.79 GHz

The directivity, axial ratio and requirements are met, together with the input matching below 18 dB and isolation level between ports below 25 dB within the operational bandwidth. The overall mass of the proposed EBG antenna system has been calculated taken into account all the constitutive antenna parts, such as the material used to build the brackets, holders, bolts and nuts, and a certain margin for each component. The superstrate is placed over a Kevlar layer to ensure its steadiness, which mass is estimated to be 222 g. The concluding outcome of the overall mass leads to estimation of 870 g. The overweight of 70 g is almost negligible.

DOI 10.1007/978-981-4560-75-7_123-2

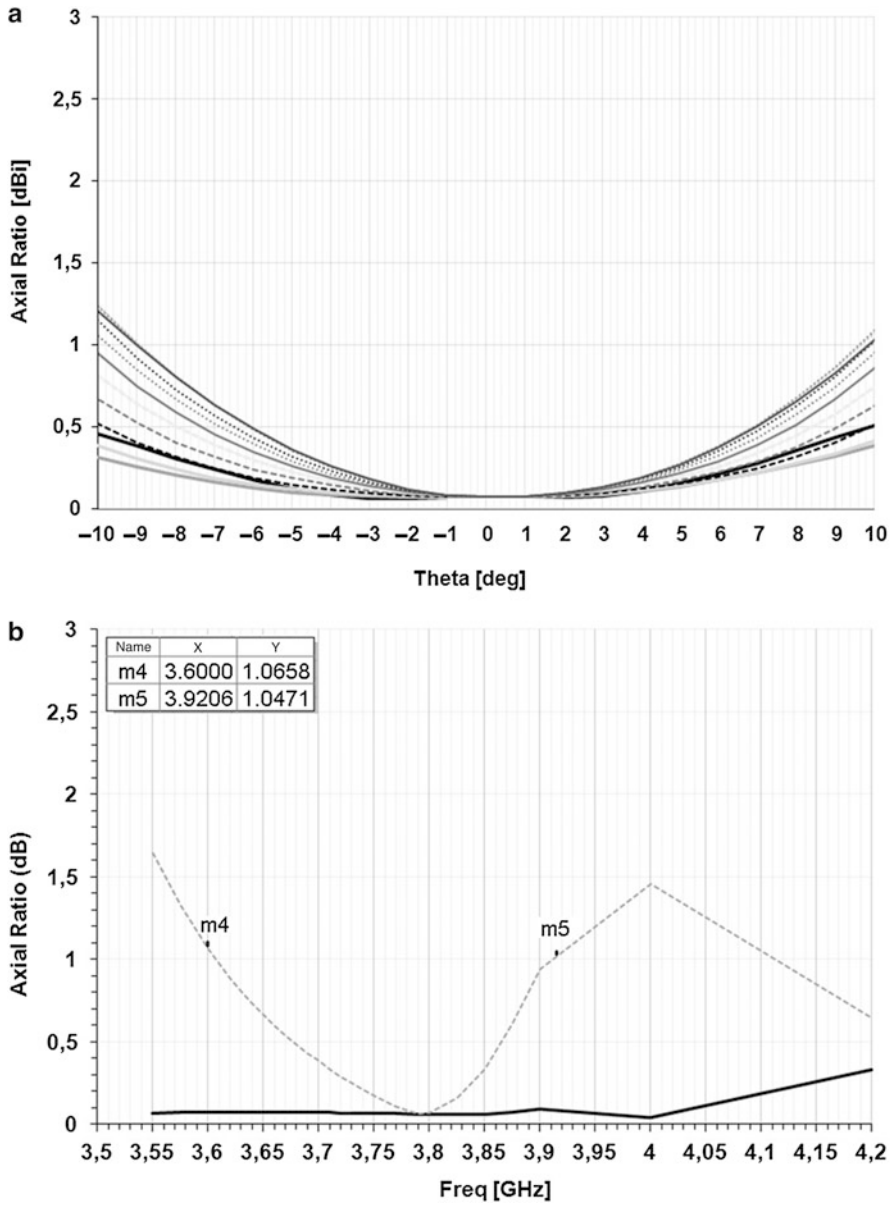


Fig. 17 Axial Ratio. (**a**) 0° to 180° phi cuts each 15° at 3.74 GHz. (**b**) AR for the worst phi cut (*gray dashed line*) in EOC

Table 8 Directivity and axial ratio summary for the TT&C EBG antenna

BW 80 MHz				BW 120 MHz			
Frequency	Peak directivity (dBi)	Worst directivity (dBi)	CP/XP (Fov) Worst case	Frequency	Peak directivity (dBi)	Worst directivity (dBi)	CP/XP (Fov) Worst case
3.69	20.13	17.31	25.35	3.67	19.93	17.16	25.12
3.74	20.40	17.47	27.13	3.74	20.40	17.47	27.13
3.77	20.34	17.36	28.44	3.79	20.17	17.16	29.34

EBG superstrate antennae are capable to fulfill the TT&C requirements of a global geostationary telecommunication system working at C-band (3.7 GHz). Other bands can also be covered with this technology as the design is completely scalable. The desired properties in the working bandwidth, as gain

DOI 10.1007/978-981-4560-75-7_123-2

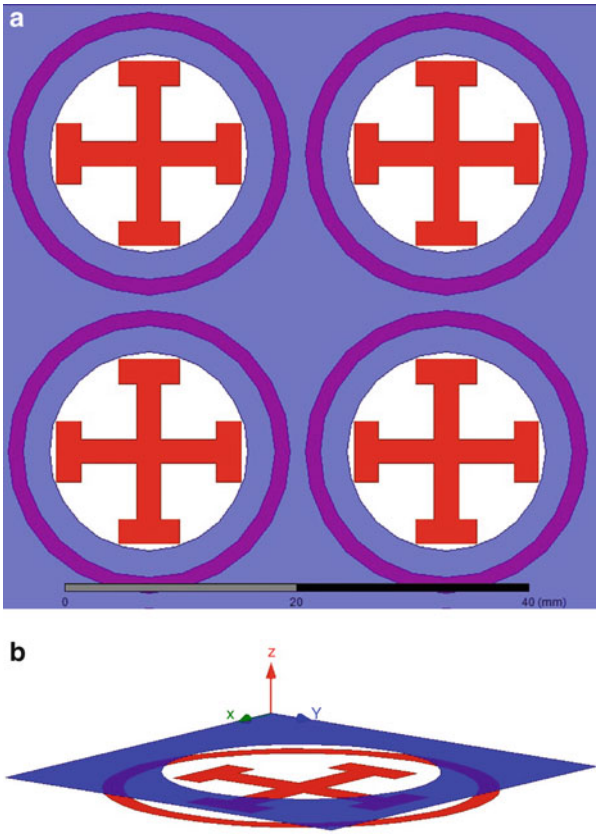


Fig. 18 Detail of the superstrate formed by two EBG layers

over coverage, dual circular polarization, return loss, and isolation values can be achieved with the presented structure. Although mass of the configuration is not under the established threshold, the overweight is almost negligible. Anyway the final mass is well below the typical horn antenna designs. In addition, the use of EBG-based technology antennae makes it possible to obtain gain enhancement with a reduced number of radiating elements; in this case just a 2×2 array of patches.

EBG-based systems could advantageously replace conventional horn antennae and arrays with comlicated BFN working at 3.7 GHz for TT&C applications, by reducing the antenna complexity, mass, and cost, similarly to the case of the GSNA application.

Dual Band EBG Antenna Configuration

TT&C frequency band requirements in C bands usually cover 3.7 GHz and 4.2 GHz bands at the same time. Two different antennae are used in conventional technology to cover these working bands or large arrays of broadband patches are also used to comply the frequency range specifications. Dual band antenna configurations are also being designed by the inclusion of another EBG superstrate in the configuration complying the TT&C requirement in dual band with a single EBG antenna.

The same configuration presented for the single frequency behavior can be used for the dual band with the inclusion of another EBG layer. The second layer is based on the combination of Frequency Selective Surfaces (FSS) metallic rings with Jerusalem crosses (see Fig. 18) inside them, placed just 1 mm below the EBG layer of circular holes in a metallic sheet.

The directivity versus frequency results are shown in Fig. 19. Maximum directivity of 21.74 dB at 3.78 GHz and 21.28 dB at 4.18 GHz is obtained. D(9°) is above 16 dB from 3.7 to 3.8G Hz (2.66 %) and

DOI 10.1007/978-981-4560-75-7_123-2

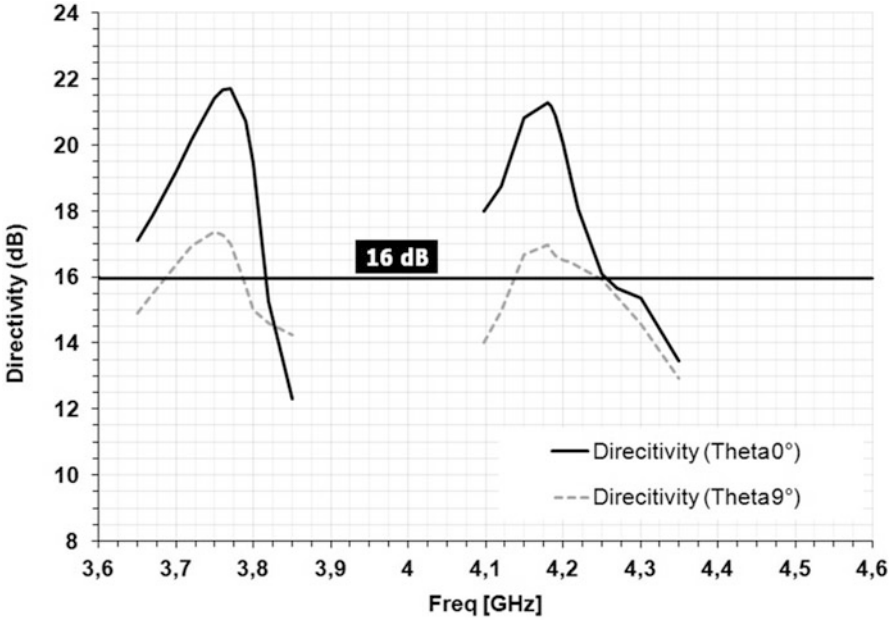


Fig. 19 Directivity vs. frequency for the 0° and 9° cuts of the 2 × 2 array EBG superestrate antenna

from 4.13 to 4.25 GHz (2.86 %). Therefore, 100 MHz working bandwidth is obtained at both frequency bands, complying requirements.

The directivity patterns at 3.76 and 4.32 GHz are shown in Fig. 20. Symmetric directivity patterns with cross polarization level below 25 dB are obtained, therefore fulfilling the specifications.

The overall mass of the proposed system is similar to the single frequency configuration, 870 g., as the second EBG layers is printed in the same sheet used to print the upper layer: one on a the top face of the sheet and the other one on the bottom face. The losses of the configuration have been estimated taking into account the EBG antenna and the feeder to be around 0.5 dB.

Horn Antenna for TT&C at Ku Band

From the geostationary orbit at about 36,000 km above the earth, the earth subtends an angle of 17.4°. With an increased number of satellites orbiting the earth, reducing the possibility of interference with other satellites is becoming more important than in the past. To minimize this interference, the amount of sidelobe energy should be as low as possible, both for the principal and cross-polarized signals. Taking these considerations into account, it is apparent that an ideal full-earth coverage antenna should be circularly symmetric and, therefore, most global coverage antennae are either reflectors or horns.

A corrugated horn can be made very low weight and also space qualified. Materials as carbon fiber reinforced plastic materials are suitable for fabrication of the final corrugated horn for TT&C (Bird and Granet 2008).

One of the first compact corrugated horn designs for low sidelobes and global earth coverage was described by (Granet et al. 2000). This horn exhibits quite low sidelobes (<36 dB) and it is quite compact $5.6{\cdot}\lambda_c$ but it presents two principal disadvantages: it operates in a narrow bandwidth, (less than 5 % for maximum −30 dB sidelobe level and −30 dB crosspolar level and −20 dB return loss) and it is very sensitive to manufacturing tolerances in the throat region (mode generator) (Bird et al. 2000).

Another wideband solution is the corrugated horn antenna that combines axial and perpendicular corrugations (Teniente et al. 2002b), shown in Fig. 21. This antenna presents a measured return loss better than 28 dB for above 22 % bandwidth. The measured sidelobe level is extremely low, below −35 dB for the overall operational bandwidth, and presents a crosspolar level below −35 dB (Table 9).

DOI 10.1007/978-981-4560-75-7_123-2

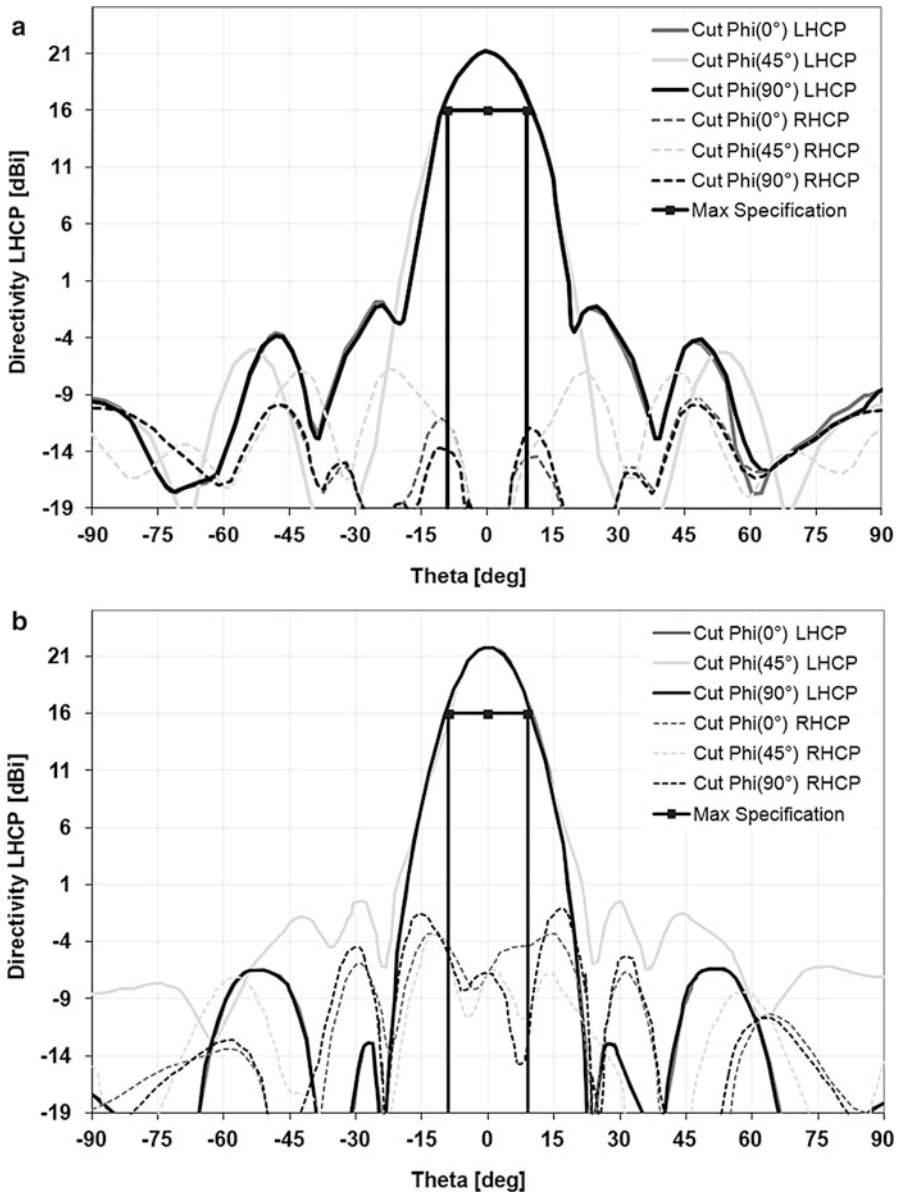


Fig. 20 RHCP and LHCP radiation patterns (**a**) at 3.78 GHz (**b**) at 4.12 GHz

THz Space Antennae

In the last years the number of space missions where submillimeter wave and Terahertz onboard instruments are operating is increasing. Some of the last missions which incorporate such instruments are PLANK and HERSCHEL with the HIFI instrument; in addition, in the near future scientific mission which are in need of high frequency instruments are expected to be launched, i.e., missions such as SPICA where the SAFARI European instrument will operate at THz frequencies (Goicoechea et al. 2009).

All of these instruments implement antennae which operate in bands covering frequencies between 100 and 900 GHz. At these THz frequencies there are different effects to be investigated. For example, understanding how the first galaxies were formed or the star formation process. Moreover, on a larger scale in this frequencies the development of galaxies can be monitored and studied; a wide range of molecules, including water, absorb radiation; and they can be used to determine exactly what the gases are

DOI 10.1007/978-981-4560-75-7_123-2

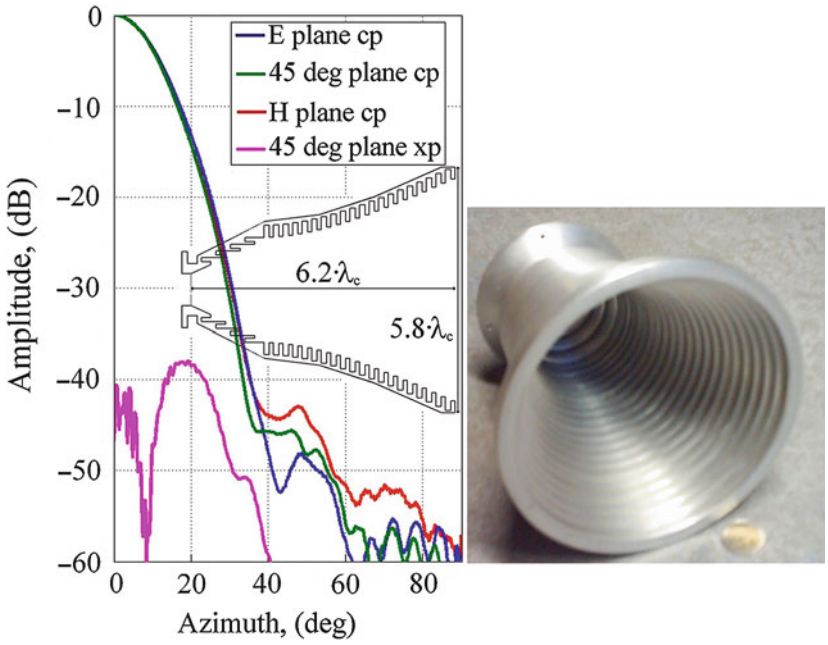


Fig. 21 Measured radiation pattern at f_c of the corrugated horn antenna that combines axial and perpendicular corrugations for TT&C (Reproduced by permission of ANTERAL)

Table 9 Radiation pattern details of the corrugated horn antenna that combines axial and perpendicular corrugations for TT&C

H plane taper @ 8.7°	45 deg plane taper @ 8.7°	E plane taper @ 8.7°	Maximum sidelobe level	Maximum crosspolar level	Directivity
−2.74 dB	−2.82 dB	−2.9 dB	−43 dB	−38 dB	20.5 dB

composed of and their temperature and pressure. Therefore, antenna technologies able to comply with the instrument requirements need to be developed. Apart from the conventional horn antennae, innovative concepts have been proposed, such as integrated antennae, Metamaterial (MTM) antennae or more recently antenna coupled Kinetic Inductance Detectors (KIDs).

Integrated antennae are the natural solution for the development of large imaging arrays operating in this frequency range. To this end, they would be placed in the focal plane of a collimating reflector or lens. However, they present several fundamental problems related to the use of electrically thick substrates such as the excitation of surface waves, which decrease their efficiency and degrade their performance. Solutions to this problem have been found by conforming the substrate creating a lens or by using a metamaterial substrate, which prevents surface wave excitation. Several antennae making use of substrate integrated extended hemispherical lenses fed by dual-slot antennae, log-periodic and spiral antennae, have resulted in excellent performance at millimeter-wave frequencies. In addition to these, new antennae such as integrated horn antennae, dielectric-filled parabolas or Fresnel plate antennae have been proposed.

A large effort has been dedicated to match these antennae to the detectors. These detectors range from Schottky diodes to bolometers. In principle it should be possible to build planar receivers with performance comparable to the best waveguide-based systems if efficient planar antennae and matching networks are developed. Recently, efforts have been made to integrate Kinetic inductance detectors (KIDs) (Day et al. 2003; Vayonakis et al. 2008; Baryshev et al. 2011) together with planar antennae. These are among the most sensitive cryogenic sensors available for detection of electromagnetic radiation

DOI 10.1007/978-981-4560-75-7_123-2

and probably the most promising sensors to be employed in large focal plane array configurations. They can be coupled to a single on-chip microwave transmission line enabling the read-out of thousands of detectors with just a single coaxial cable pair connecting the on-chip transmission line to room temperature read-out electronics. KIDs can be easily integrated with lens antennae and leaky lens antennae. In particular this last configuration, allows achieving a very large bandwidth, thanks to the phase stability of the leaky lens antenna.

Anyway, this section will focus on the MTM concepts applied to THz antennae. As mentioned, MTM substrates are one of the solutions for surface wave reduction and radiation pattern improvement. This chapter describes in detail the fundamentals behind their use and several applications.

One of the main goals for Terahertz technology is the realization of imaging arrays, which are of great interest for space astronomy and for atmospheric research. Particularly in astronomy most of the spectral line emitting regions are usually spatially extended over many observing beams in the sky and mapping is required to astrophysically understand the regions under study. In atmospheric research multiple beam observing systems allow push-broom measurements for limb scanning experiments. As the regions of interest in astronomy research are usually distributed over many observing beams, the time needed to build an image can be reduced in approximately inverse proportion to the number of elements included in the array.

THz imaging arrays have conventionally been based on feedhorns and waveguide technology and are usually assembled from discrete elements, some recent examples of missions using horn antennae can be found in PLANCK or HERSCHEL (Doyle et al. 2009). The costs, mass and volume associated with this approach may limit the maximum practicable number of pixels. In fact, PLANCK mission implements a focal plane formed by more than 40 antennae ranging from 30 to 857 GHz. On the other hand, HERSCHEL mission, although more focused at Infrared frequencies, also comprises 5 frequency bands covering 480–1,150 GHz with SIS mixers. The SPIRE detectors are spider-web bolometers using NTD germanium thermometers, which are coupled to the telescope by hexagonally close-packed diameter single mode conical feedhorns.

In the last time, the advances on lithography techniques and processes have open the possibilities to develop new THz antenna concepts which can be competitive versus conventional waveguide technologies. Being able to produce lithographically planar receivers with equivalent performance is allowing the realization of much larger two-dimensional arrays. Such an approach is dramatically reducing cost and greatly simplifying manufacture and assembly.

Although waveguide and feedhorn technology provides a natural isolation between adjoining pixels. In contrast, the performance of planar antenna arrays can be severely limited by the parasitic cross coupling of the antennae via surface waves (Pozar 1983). Electromagnetic bandgap (EBG) or Photonic Crystal (PC) structures, working as surface wave suppressing substrates, have consequently attracted much interest (Joannopoulos et al. 1995; Gonzalo et al. 1999; Sievenpiper et al. 1999; de Maagt et al. 2003; Yang and Rahmat-Samii 2003). EBG structures prevent the propagation of radiation, for frequencies within the bandgap, in 1, 2 or 3 spatial directions, the number usually corresponding to the number of dimensions in which the crystal is periodic (Joannopoulos et al. 1995). Whilst 2-D EBG materials have been proven to be an useful substrate for planar antennae (Gonzalo et al. 1999), a 3-D EBG appears to be more desirable because any antenna fundamentally radiates in 3-D and an antenna array by definition extends over at least one dimension perpendicular to the boresight.

Within this section, the performance of these structures as substrate for a dipole antenna will be studied. In particular two different single dipole antennae backed by a 3-D EBG structure are presented. This combination exhibited low back radiation and a higher directivity when compared to the equivalent dipole in free space. Two different EBG substrate shave been used for this purpose: the so-called Woodpile and FAN EBG structures (Ho et al. 1994; Sözuer and Dowling 1994; Martinez et al. 2007). In the case of the

DOI 10.1007/978-981-4560-75-7_123-2

Fan structure it was built with a very high dielectric constant material, Zirconium Tin Titanate (ZTT). This high dielectric material can be of interest for miniaturization. However, the excitation of surface would limit its performance as antenna substrate. This problem can be sort out if an EBG structure is implemented in the substrate. Once the performance of an element is studied a seven element 0.5 THz imaging array will be discussed.

Single Detector THz Configuration

In this section, the performance of a dipole antenna on top of an EBG structure operating at THz frequencies is described. Two cases are discussed. Both of them use as radiating element a printed dipole antenna and either a woodpile EBG or a FAN EBG structure as substrate.

When a dipole antenna is placed on top of a periodic structure, its radiation pattern and input impedance depend on the position and orientation of the dipole with respect to the features on its surface. Complete studies for both configurations can be found in (Gonzalo et al. 2001; Ederra et al. 2013). In summary, for each EBG structure it can be found a polarization dependence of the input impedance. The parallel to the main features (dielectric bars in the woodpile and air channels in Fan's) polarization presents very low input impedance, which can be ascribed to the close to 180° reflection coefficient phase. On the contrary, the perpendicular polarization exhibits higher impedance, easier to match to a CPS line, which would be the most natural feeding circuit for the dipole antennae. Regarding the radiation features, no significant differences are found between positions and polarizations. In all cases, there is an increase of directivity, thanks to the effect of the EBG substrate and coupling between neighboring antennae will be reduced.

For the design of the overall dipole antenna/EBG configuration as a detector, a subharmonically pumped heterodyne receiver was considered. This incorporates a flip chip mounted Schottky diode in the center of the dipole; and the local oscillator signal is delivered along a coplanar stripline (CPS). The low pass filter is a standard high–low impedance filter, which was constructed by widening the tracks of the CPS. Its design was carried out taking into account the loading of the underlying EBG. The filter dimensions were optimized in order to maximize the reflection of RF power at 0.5 THz. The down converted RF signals also passes along the same stripline. A low pass filter is incorporated to prevent the CPS from loading the dipole at frequencies around 0.5 THz. The design procedure was first to optimize the impedance matching when the dipole antenna was fed by a CPS and afterwards the RF power coupled to the diode was maximized.

In the first case, the dipole antenna is placed on the FAN EBG structure. The feeding CPS line and the dipole were printed on a 20 μm thick quartz substrate. Details of the dipole and feeding line dimensions can be found in Ederra et al. (2013). For these dimensions, the impedance of both CPS line and dipole were similar and close to 170 Ω, which assures a good impedance matching.

The Schottky diode was placed as close as possible to the dipole antenna in order to minimize its effect on the radiation pattern of the dipole. The distance between the diode and the low pass filter was optimized so that the power transferred from the incoming RF signal to the diode was maximized.

The circuits were patterned in a gold film, which had been deposited on a fused quartz substrate, using standard photolithographic techniques. The quartz was then cut to size and lapped to a final thickness of 20 μm.

The Fan structure was held to a low loss polymer, TPX high-density polyethylene, support by using cyanoacrylate adhesive. The quartz substrate was then aligned and fixed using a small quantity of the same adhesive at its edges. The output of the choke filter was connected to a coaxial line using gold ribbon and Epotek H20E. A photograph of the manufactured antenna and Fan EBG combination is shown in Fig. 22.

DOI 10.1007/978-981-4560-75-7_123-2

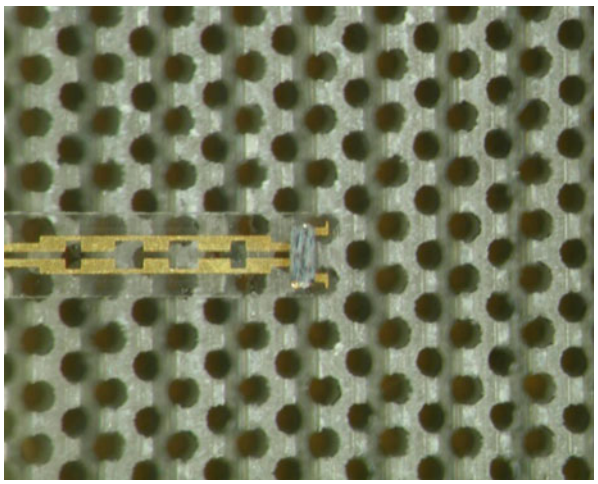


Fig. 22 Photograph of the dipole on top of the Fan EBG structure (From reference (Ederra et al. 2013) reproduced by permission of © 2013 IEEE)

Radiation Pattern Results

The predicted radiation pattern of the complete detector configuration, i.e., dipole, diode, and low pass filter, was computed using Ansoft HFSS. Since the diode and the feeding lines affect the radiation pattern some ripples in the E-Plane are appreciated. The computed directivity, 6.4 dBi, is larger than the directivity of the ideal dipole. This higher directivity corresponds to the peaks of the ripples and not to boresight direction.

The radiation pattern was measured at the University of Bern using an AB Millimeter submillimeter wave Vector Network Analyzer (VNA) from ABmm. Both E and H plane patterns were measured at 0.48, 0.5 and 0.52 THz. The comparison between the measured patterns and the predicted ones are depicted in Fig. 23, where normalization of measured data at boresight has been performed. The above directivity increase with respect to a dipole on a standard substrate can be noticed.

In the second case, a similar design procedure was followed, but in this design, the substrate was a woodpile EBG structure. More details about the woodpile properties and its manufacturing properties can be found in (Gonzalo et al. 2001). The dipole antenna was a lithographically patterned 2 μm microns thick gold structure. A similar receiver configuration was used, where a Schottky diode was used to convert the incoming RF signal at 500 GHz to DC.

The dipole antenna was fed using a transmission line on a quartz membrane and a simple quarter-wavelength, high-low impedance filter was used in order to isolate the RF from the DC signal. The membrane was kept thin enough so as not to disturb the performance of the EBG structure. The RF filter is adjoined to two bonding pads.

The fabricated structure of the dipole on top of the woodpile EBG is displayed in Fig. 24. The woodpile, the antenna, the filter, the membrane to support the feeding line, and the Schottky diode are clearly visible.

The measurement system included a Carcinotron operating at 500 GHz, a computer controlled positioning system with two axes of movement, onto which the configuration is placed and a lock-in amplifier to measure the detected DC signal. The results for the E and H planes are plotted in Fig. 25.

It should be noted that the obtained E-plane beamwidth is $\theta = 60^\circ$ while it is $\theta = 85^\circ$ for the H-plane. These values correspond to an estimated effective area much larger than that of a simple stand-alone dipole that has an effective area of $0.13\lambda_0^2$. This leads to higher directivity values that can be derived to be of the order of 8–10 dB for the EBG backed antenna compared with only 2.14 dB obtained for a simple dipole suspended in air.

DOI 10.1007/978-981-4560-75-7_123-2

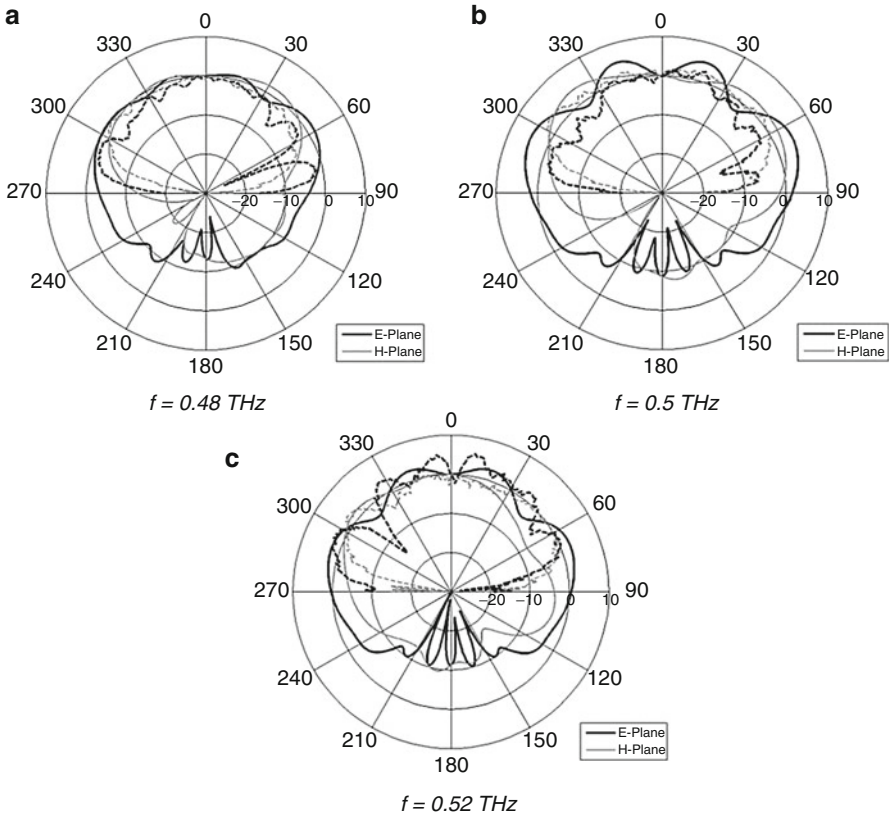


Fig. 23 Comparison between the measured (*dashed line*) and predicted (*solid line*) radiation pattern of a dipole antenna placed on top of the Fan structure at 0.48 THz (**a**), 0.5 THz (**b**) and 0.52 THz (**c**) (From reference (Ederra et al. 2013) reproduced by permission of © 2013 IEEE)

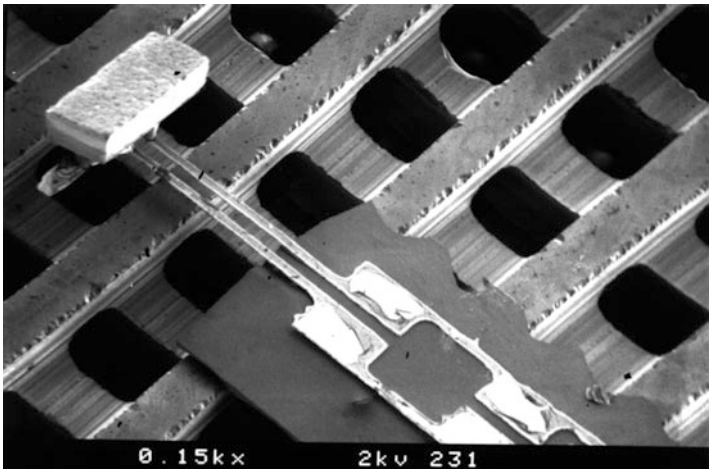


Fig. 24 Photographs taken with the electron microscope showing the fabricated dipole on top of the woodpile photonic crystal (From reference (Gonzalo et al. 2001) reproduced by permission of © 2001 IET)

In addition, both E and H plane patterns tend to zero at 90° and 270° confirming that the EBG substrate is suppressing the parasitic surface wave mode excitation.

DOI 10.1007/978-981-4560-75-7_123-2

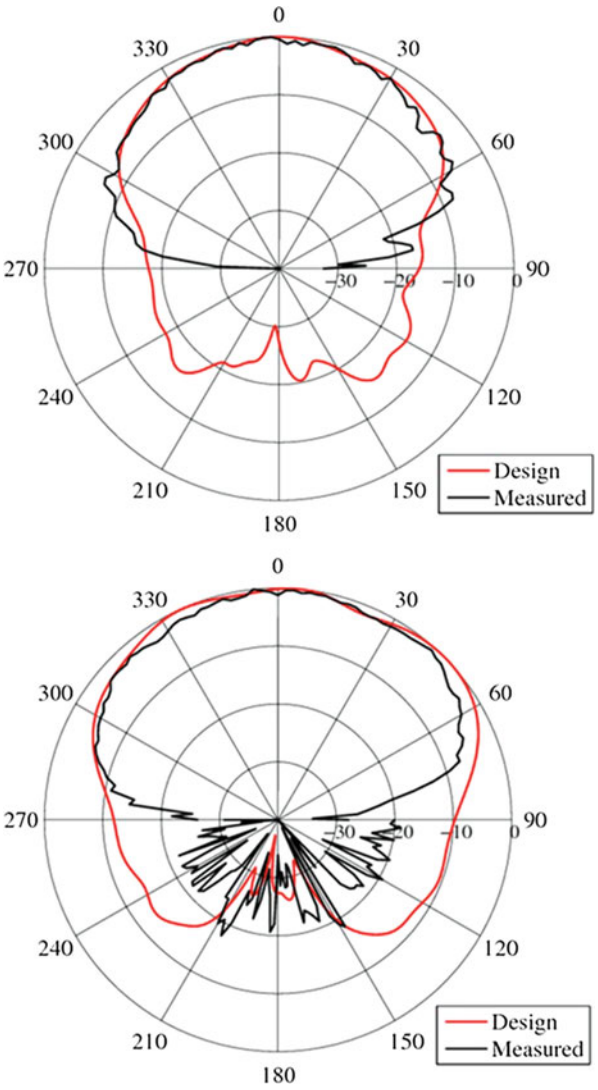


Fig. 25 Simulated and measured radiation patterns (*red* and *black* curves respectively) of a dipole on top of a woodpile EBG structure. 0° and 90° cuts are plotted

Array Configuration of Single Dipole THz Detectors

The enhanced performance of the single EBG backed element shown in previous sections is followed by the use of this configuration as base for an imaging array. In this case, a seven-element Schottky diode-based heterodyne array for 500 GHz which could be the base for future space instruments in imaging applications has been studied (Ederra et al. 2008). This design represents a significant step towards the goal of a planar technology imaging array. The array was designed to match the typical optical system of a submillimeter wave astronomical instrument: f/25 with a far field beam crossover level of −3 dB. In order to obtain these beam characteristics, a 90° paraboloidal mirror was used to modify the radiation pattern of the detector antennae. The design of the individual receivers was optimized for maximum sensitivity i.e., for best impedance matching between the Schottky diodes and the planar dipole antenna.

Dipole and Array Design

An optimized design of the 500 GHz dipole antenna together with the associated RF choke/ IF transmission line on top of the EBG structure is used for the array design (Fig. 26). Limitations on the

DOI 10.1007/978-981-4560-75-7_123-2

dipole antenna separations arise from the 500 GHz silicon woodpile, which has a 72 μm bar width and 232 μm period in the substrate plane. As the dipoles need to be placed above intersections of the bars in the top two layers, the separation between elements is consequently restricted to multiples of this period.

A previous study of the mutual coupling of dipole antennae on EBG substrates showed that a minimum separation of two periods is necessary to ensure coupling values below −20 dB in both E and H planes. One compact configuration of seven antennae that meets this requirement was developed. Although denser configurations are possible, the selected one offers the advantage that the six outer antennae lie between 3.6 and 4 woodpile periods from the central element. It will be shown later on that it is possible to design matching optics for this array that achieve the desired beam properties: the relatively narrow spread of radii assists in achieving a design that is compliant with the beam crossover point requirement.

The mutual coupling between the array elements was investigated. The results obtained with Ansoft HFSS shows that for the case under study this value is below 40 dB.

Mirror Design

Although in principle a lens could have been used, reflective optics exhibit lower losses and standing waves. The f# requirement on the array imposes the far-field power half- angle, i.e., the directivity of the resulting antenna–reflector combination. To meet the f/25 specification, the beamwidth of the output beams should be 2.2°. Therefore, the reflector must provide the required magnification, increasing the directivities of the radiating elements and simultaneously separating the beams. However, the further the antennae are from the optical axis of the system, the more degraded the directivity and beamwidths of their radiation patterns becomes. The central dipole is located at the optical focus of the mirror and the projection of the mirror on the x-z plane is an ellipse with its major axis in the z direction. To optimize the illumination of the mirror, the dipole array was also oriented with its major axis along the z-axis. Taking this into account, elevation (azimuth) corresponds to the H (E) -plane of the central beam respectively.

The mirror dimensions were optimized to match the array described above. The obtained optimum values were found to be D = 17.3 mm and f = 25.4 mm.

Physical optics GRASP (http://www.ticra.com/products/software/grasp) software was used to calculate the far-field radiation patterns. As an input for this program, the radiation of an isolated dipole on a finite silicon woodpile, containing 11 periods, was computed with Ansoft-HFSS. Predicted individual beamwidths of (2.1 ± 0.1)° were obtained, matching well the desired 2.2° value. GRASP simulations also showed that there was negligible impact on the far-field radiation pattern from mounting the woodpile on either a dielectric, TPX, or a metal plate.

Manufacturing of the Imaging EBG Array Configuration

A monolithic quartz substrate was used for the overall dipole array, this process ensures accurate alignment of all dipoles on the EBG substrate. Standard photolithographic techniques were used to pattern the array in a gold film that had been deposited on a fused quartz substrate. The quartz was then cut to size and lapped to a thickness of 20 μm. Separated and substrate thinned Schottky diodes were fixed in position on the metal using silver loaded epoxy (Epotek H20E). Electrical connections were made from the ends of the choke filters to coaxial lines using gold ribbon and silver loaded epoxy.

The characteristic curves of the Schottky diodes were monitored repeatedly during assembly. No deviation from the initial manufacturer's values was found, apart from for one diode. Here an increase in series resistance was localized to either the diode to substrate connection or to within the diode chip.

The mirror was made from a diamond turned, gold coated, parabolic reflector of 12.7 mm focal length and a projected diameter of 17.3 mm.

DOI 10.1007/978-981-4560-75-7_123-2

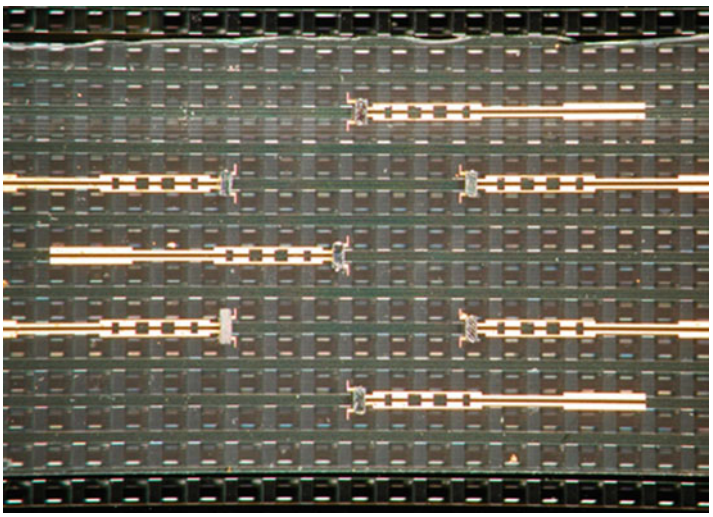


Fig. 26 Manufactured array configuration of 7 dipole antennas on top of an EBG structure (From reference (Ederra et al. 2008) reproduced by permission of © 2008 IEEE)

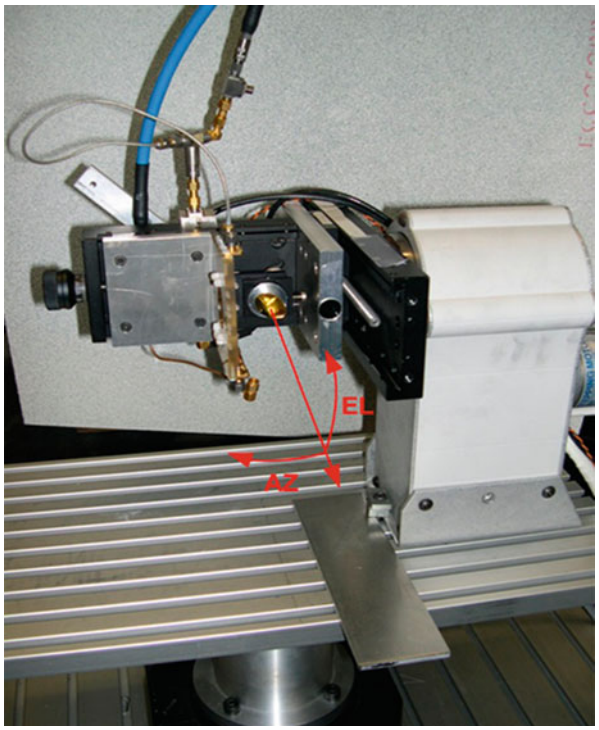


Fig. 27 Positioning system used to measure the array configuration

Experimental Results

A VNA based measurement technique and apparatus was used for the characterisation of the beam pattern of the array at 500 GHz. As a source, the 100 GHz output of a phase locked Gunn diode oscillator was quintupled in frequency and radiated by a corrugated horn antenna towards the EBG array. A stabilized DC bias voltage was connected to the diodes through a bias-T in the IF transmission line in order to set the optimum operating point of the receiver. In each case, the LO power and applied DC bias were adjusted to maximize the signal to noise ratio of the IF signal.

DOI 10.1007/978-981-4560-75-7_123-2

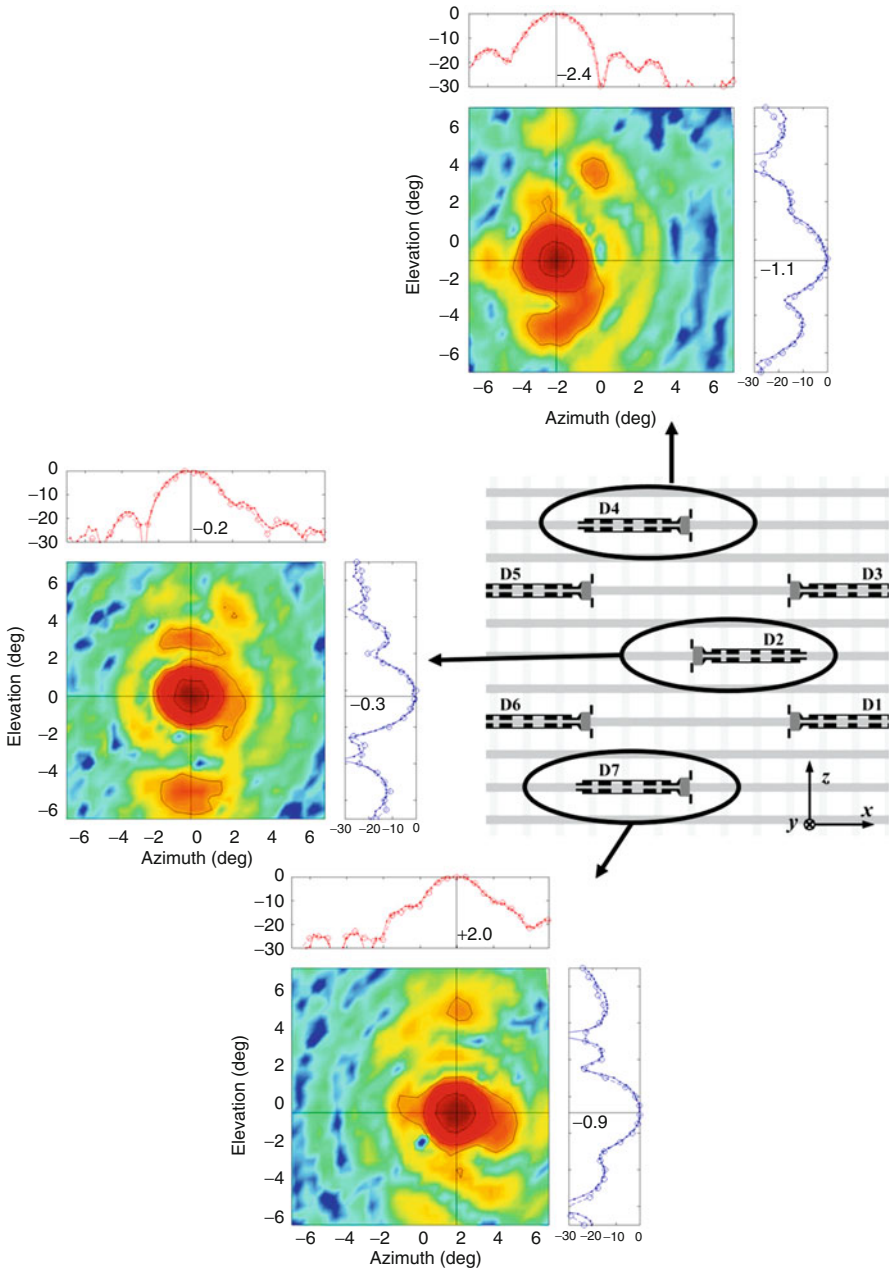


Fig. 28 Measured radiation patterns for the array elements. The −15 dB, −10 dB, −6 dB, and −3 dB contours are shown for beams 2, 4, and 7. Angular range of ±8° with a step size of 0.5°. Operating frequency 500 GHz

The imaging array was mounted so that it could be rotated about azimuth and elevation axes through the optical center of the parabolic mirror: Fig. 27. Beam pattern measurements were made at a feed horn to array separation of 0.8 m with a dynamic range above 25 dB.

Two dimensional beams scans, normalized to their maxima, are presented in Fig. 28 for some of the receivers over an angular range of ±8° with a step size of 0.5°. To improve the presentation quality, interpolation of the data has been performed and −3 dB, −6 dB, −10 dB and −15 dB contour lines added.

To represent better the overall behavior of the array, a contour plot of the −3 dB signal levels for all the beams is presented in Fig. 29. In the measurements the absolute pointing of the array had an offset of 1° in elevation and 1° in azimuth from the nominal optical axis, which has been subtracted for this figure. The

DOI 10.1007/978-981-4560-75-7_123-2

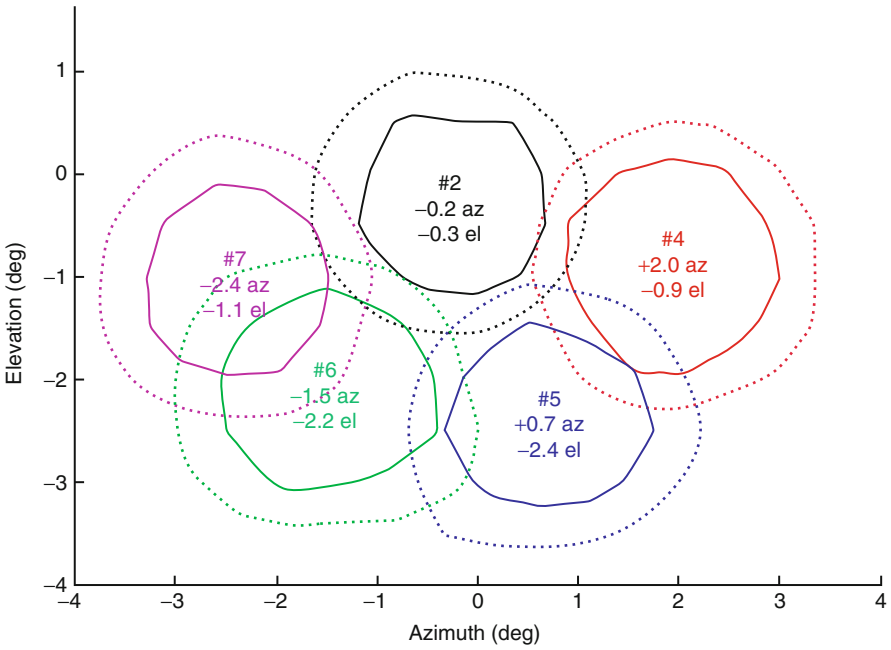


Fig. 29 Comparison of the 3-dB contours of the measured (*solid lines*) after correcting for mirror misalignment and the predicted patterns (*dashed lines*). In the measurements the absolute pointing of the array had an offset of 1° in elevation and 1° in azimuth from the nominal optical axis, which has been subtracted for this figure. Operating frequency 500 GHz

reason for this is probably a slight misalignment of the parabolic reflector and the test setup. Good agreement between patterns is observed, although the individual measured beams are slightly narrower than predicted, which leads to crossover points of around −4 dB. The predicted slightly compressed elevation plane characteristic is also exhibited by the measurements.

Finally, the radiation pattern cuts, azimuth and elevation, of different elements of the array are presented in Fig. 30. These measured pattern are compared with their counterparts obtained by simulation with HFSS software. In general, the agreement is good, some discrepancies can be found in the sidelobe levels, in particular, higher sidelobes are appreciated in the elevation cuts which are attributed to the effect of the mirror holder used in the measurement positioning system.

These results constitute the first demonstration of a planar imaging array based on EBG technology. Although the performance was still far from that obtained with more mature technologies it is the first step towards its application.

Conclusion

This chapter has demonstrated the use of advanced horn antenna profile to comply with the very stringent requirements of space antennae. Nowadays, the most of the space applications is requiring better electromagnetic performances to the antennae; lower cross-polar levels, low sidelobes, higher gains, higher efficiencies, larger operational bandwidths, etc. Conventional designs are not able to comply with these needs and more sophisticated antennae have to be developed. Several examples have demonstrated the possibilities of horn antennae to achieve those high performances with reduced lengths. New proposals based on the use of horizontal and vertical corrugations together with spline profiles have been analyzed and their radiation characteristics have been included.

DOI 10.1007/978-981-4560-75-7_123-2

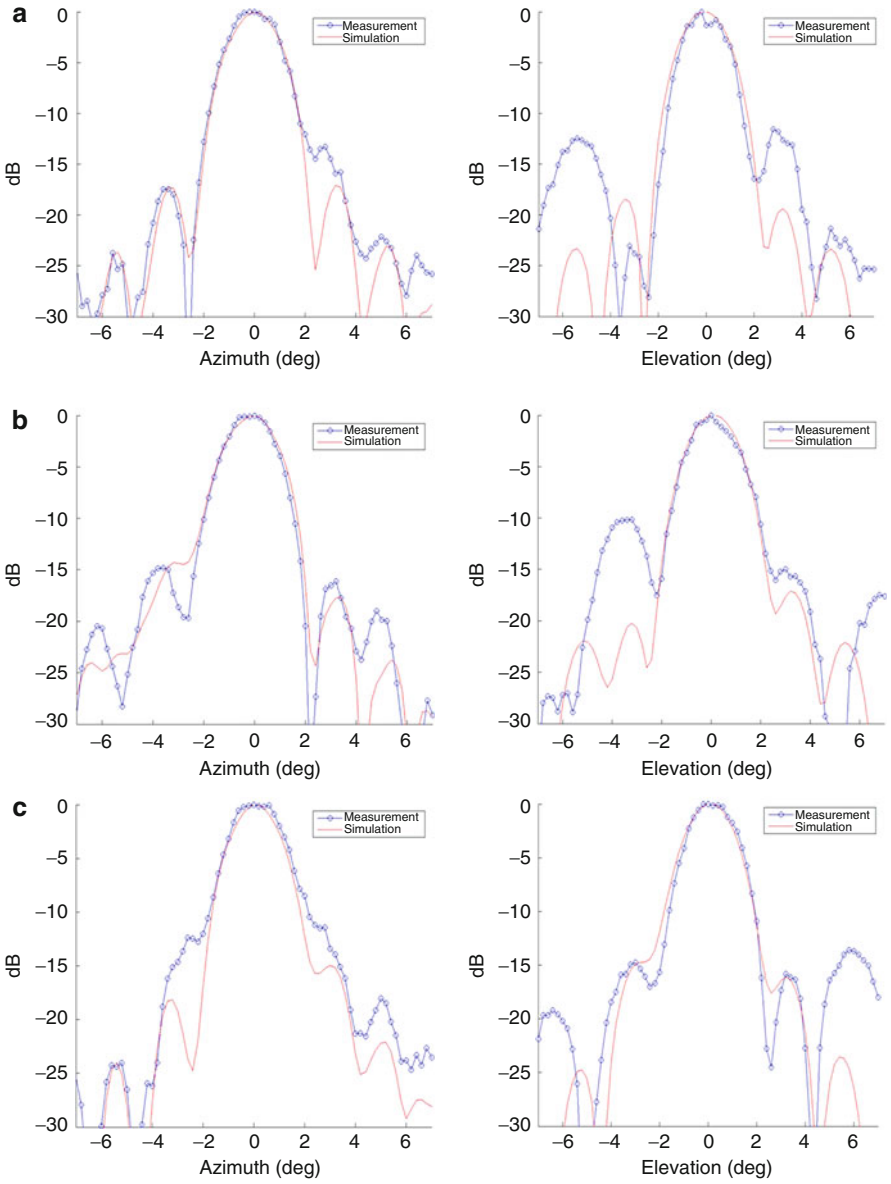


Fig. 30 2-D cuts of azimuth and elevation of beams 2 (**a**), 7 (**b**) and 4 (**c**). Operating frequency 500 GHz (From reference (Ederra et al. 2008) reproduced by permission of © 2008 IEEE)

On the other hand, there are space applications where conventional antennae suffer to obtain the desired radiation performances, mainly as they are heavy and bulky. In that case, this chapter has presented the possibilities of using EBG technology to develop innovative antennae. For instance, EBG-enhanced antennae could advantageously replace conventional array antenna in WAAS applications. This EBG technology simplifies the antenna configuration and leads to lighter and more cost effective solutions by simplifying the antenna BFN. The EBG GSN complies with all the requirements of a GSN application, resulting in a real alternative antenna for navigation on board antennae. Furthermore, EBG superstrate antennae fulfill the requirements TT&C application at C-Band in single and dual frequency performance. The system directivity is up to 17 dB in 100 MHz around 3.7GHz and 4.2GHz; the cross-polarization level is lower than 25d.B. Although mass of the configuration is not under the established threshold, the overweight is almost negligible. Consequently, EBG antennae technology system represents an

DOI 10.1007/978-981-4560-75-7_123-2

interesting alternative to the bulky horn antennae and complicated BFN of array usually used for TT&C applications in current satellites.

As the last part of this chapter, some results of EBG antennae operating at THz frequencies have been included. In particular, two cases are analyzed. In the first case, the enhanced performance of a dipole antenna placed on top of an EBG structure is discussed. The reduction of the back and side radiation and the improvement of the antenna directivity have been reported. These clear advantages have led to develop an array configuration suitable to be used in imaging THz application for scientific instruments. An array formed by seven elements has been presented. This configuration has been designed to comply with typical requirements in imaging applications by means of the integration together with a parabolic mirror. The radiation performances have been measured and good agreement with simulated results has been obtained. This constitutes the first demonstration of a planar imaging array based on EBG technology.

Cross-References

- ▶ Antennas in Radio Telescope Systems
- ▶ Commercial Antenna Design Tools
- ▶ Frequency Selective Surfaces
- ▶ Metamaterials and Antennas
- ▶ Mm-Wave Sub-mm-wave Antenna Measurement
- ▶ Radiometer Antennas
- ▶ Terahertz Antennas and Measurement

References

Amyotte E, Martins-Camelo L (2012) Antennas for satellite communications. In: Imbriale WA, Gao S, Boccia L (eds) Space antenna handbook, 1st edn. Wiley, Chichester, UK

Baryshev A, Baselmans J, Freni A, Gerini G, Hoevers H, Iacono A, Neto A (2011) Progress in antenna coupled kinetic inductance detectors. IEEE Trans Terahertz Sci Technol 1:112–123

Bhattacharyya A, Goyette G (2013) Smooth wall multimode horns for high aperture efficiency. Theory, design and applications. In: Rao S, Sharma S, Shafai L (eds) Handbook of reflector antennas and feed systems, vol II, 1st edn, Feed systems. Artech House, Boston

Bird T (2008) Feed antennas. In: Balanis C (ed) Modern antenna handbook. Wiley, Hoboken, pp 889–890

Bird T, Granet C (2008) Fabrication and space-qualifying a lightweight corrugated horn with low sidelobes for global-earth coverage. IEEE Antennas Propag Mag 50:80–86

Bird T, Granet C, James G (2000) Lightweight compact multi-mode corrugated horn with low sidelobes for global-earth coverage. In: AP2000 conference, pp 165

Biswas R, Ozbay E, Temelkuran B, Bayindir M, Sigalas M, Ho K (2001) Exceptionally directional sources with photonic-bandgap crystals. J Opt Soc Am B 18:1684–1689

Brand Y, Iriarte J, Cassivi Y, Fourmault A, Ederra I, Gonzalo R, De Maagt P (2009) WAAS space segment antenna based on EBG superstate gain enhancement technique. In: European conference on antennas and propagation, EuCAP 2009, pp 2613–2617

Cheype C, Serier C, Thèvenot M, Monediere T, Reineix A, Jecko B (2002) An electromagnetic bandgap resonator antenna. IEEE Trans Antennas Propag 50:1285–1290

Day P, LeDuc H, Mazin B, Vayonakis A, Zmuidzinas J (2003) A broadband superconducting detector suitable for use in large arrays. Nature 425:817–821

de Maagt P, Gonzalo R, Vardaxoglou Y, Baracco J (2003) Electromagnetic bandgap antennas and components for microwave (sub)millimeter wave applications. IEEE Trans Antennas Propag 51:2667–2677

Diblanc M, Rodes E, Arnaud E, Thevenot M, Monediere T, Jecko B (2005) Circularly polarized metallic EBG antenna. IEEE Microwave Wireless Compon Lett 15:638–640

Doyle D, Pilbratt G, Tauber J (2009) The herschel and planck space telescopes. Proc IEEE 97:1403–1411

Ederra I, Gonzalo R, Alderman B, Huggard P, de Hon B, van Beurden M, Murk A, Marchand L, de Maagt P (2008) Electromagnetic bandgap based planar imaging array for 500 GHz. IEEE Trans Microwave Theory Tech 56:2556–2565

Ederra I, Gonzalo R, Martínez B, Alderman B, Huggard P, Murk A, Marchand L, de Maagt P (2013) Design and test of a 0.5 THz dipole antenna with integrated schottky diode detector on a high dielectric constant ceramic electromagnetic bandgap substrate. IEEE Trans Terahertz Sci Technol 3:584–593

Fehrembach A, Enoch S, Sentenac A (2001) Highly directive light sources using two-dimensional photonic crystal slabs. Appl Phys Lett 79:4280–4282

Feresidis A, Vardaxoglou J (2001) High gain planar antenna using optimised partially reflective surfaces. IEE Proc Microwave Antennas Propag 148:345–350

Galileo System Requirement Document (2002) ESA APPNS-REQ-00011. Issue 2 - Rev 1

Goicoechea J, Isaak K, Swinyard B (2009) Exoplanet research with SAFARI: a far-IR imaging spectrometer for SPICA. arXiv:0901.3240

Gonzalo R, de Maagt P, Sorolla M (1999) Enhanced patch antenna performance by suppressing surface waves using photonic band-gap structures. IEEE Trans Microwave Theory Tech 47:2131–2138

Gonzalo R, Ederra I, Mann C, de Maagt P (2001) Radiation properties of terahertz dipole antenna mounted on photonic crystal. Electron Lett 37:613–614

Gonzalo R, del Río C, Goñi D, Teniente J (2002) Horn antenna combining horizontal and vertical ridges. International Patent WO03/100907

Granet C, James G (2005) Design of corrugated horns: a primer. IEEE Antennas Propag Mag 47:76–84

Granet C, Bird T, James G (2000) Compact multimode horn with low sidelobes for global earth coverage. IEEE Trans Antennas Propag 48:1125–1133

Granet C, James G, Bolton R, Moorey G (2004) A smooth-walled spline-profile horn as an alternative to the corrugated horn for wide band millimeter-wave applications. IEEE Trans Antennas Propag 52:848–854

Granet C, James G, Forsyth A (2008) Aperture antennas: waveguides and horns. In: Balanis C (ed) Modern antenna handbook, 1st edn. Wiley, Hoboken, pp 142–144

GRASP software. http://www.ticra.com/products/software/grasp. Accessed 13 April 2015

Hajj M, Rodes E, Monédière T (2009) Dual-band EBG sectoral antenna using a single-layer FSS for UMTS application. IEEE Antennas Wirel Propag Lett 8:161–164

Hay S, Barker S, Granet C, Forsyth A, Bird T, Sprey M, Greene K (2001) Multibeam earth station antenna for a European Teleport Application. In: IEEE AP-S international symposium and URSI radio science meeting, vol 2. pp 300–303

Ho KM, Chan CT, Soukoulis C, Biswas R, Sigalas M (1994) Photonic band gaps in three dimensions: new layer-by-layer periodic structure. Solid State Comm 89:413

Iriarte J, Ederra I, Gonzalo R, Gosh A, Laurin J, Caloz C, Brand Y, Gavrilovic M, Demers Y, de Maagt P (2006) EBG superstrate for gain enhancement of a circularly polarized patch antenna. In: IEEE antennas and propagation society international symposium, pp 2993–2996

DOI 10.1007/978-981-4560-75-7_123-2

Iriarte J, Ederra I, Gonzalo R, Brand Y, Fourmault A, Demers Y, de Maagt P (2009) EBG superstrate array configuration for the WAAS space segment. IEEE Trans Antennas Propag 57:81–93

Jackson D, Alexopoulos N (1985) Gain enhancement methods for printed circuit antennas. IEEE Trans Antennas Propag 33:976–987

Jackson D, Oliner A (1988) A leaky-wave analysis of the high-gain printed antenna configuration. IEEE Trans Antennas Propag 36:905–910

Jackson D, Oliner A, Ip A (1993) Leaky-wave propagation and radiation for a narrow-beam multiple-layer dielectric structure. IEEE Trans Antennas Propag 41:344–348

Joannopoulos J, Meade R, Winn J (1995) Photonic crystals; molding the flow of light. Princeton University Press, New York

Kanso A, Chantalat R, Naeem U, Chreim H, Thevenot M, Bila S, Monediere T (2011) Multifeed EBG dual-band antenna for spatial mission. Int J Antennas Propag 14p, Article ID 190358, doi:10.1155/2011/190358

Lee Y, Yeo J, Mittra R, Lee Y, Park W (2005) Application of Electromagnetic Bandgap (EBG) superstrates with controllable defect for a class of patch antennas as spatial angular filters. IEEE Trans Antennas Propag 53:224–235

Lovat G, Burghignoli P, Capolino F, Jackson D, Wilton D (2006a) Analysis of directive radiation from a line source in a metamaterial slab with low permittivity. IEEE Trans Antennas Propag 54:1017–1030

Lovat G, Burghignoli P, Capolino F, Jackson D (2006b) Highly-directive planar leaky-wave antennas: a comparison between metamaterial-based and conventional designs. In: EuMA (European Microwave Association) Proceedings, vol 2. pp 12–20

Maffei B, Ade P, Gannaway F, Wakui E, Wylde R, Murphy J, Colgan R, Dupuy J, Parini C (2000) Corrugated gaussian backto- back horns for cosmic microwave background continuum receivers, 24th QMW antenna symposium, pp 38–41

Martinez B, Ederra I, Gonzalo R, Alderman B, Azcona L, Huggard P, de Hon B, Hussain A, Andrews S, Marchand L, de Maagt P (2007) Manufacturing tolerance analysis, fabrication and characterisation of 3D submillimetre wave electromagnetic bandgap crystals. IEEE Trans Microwave Theory Tech 55:672–681

Mician Microwave Wizard software. http://www.mician.com. Accessed 08 April 2015

Milligan T (2005) Modern antenna design, 2nd edn. Wiley, New York, pp 358–359

Olver A, Clarricoats P, Kishk A, Shafai L (1994) Microwave horns and feeds, vol 39, IEE electromagnetic waves series. The Institution of Electrical Engineers, London

Oses A, Iriarte J, Ederra I, Gonzalo R, de Maagt P (2009) Multiband EBG navigation antenna. In: IEEE International workshop on antenna technology, iWAT, pp 1–4

Pozar D (1983) Considerations for millimetre wave printed antennas. IEEE Trans Antennas Propag 31:740–747

Rao S (1999) Design and analysis of multiple-beam reflector antennas. IEEE Antennas Propag Mag 41:53–59

Rodes E, Diblanc M, Drouet J, Thevenot M, Monediere T, Jecko B (2006) Design of a dual-band EBG resonator antenna using capacitive FSS. In: Antennas and propagation society international symposium, IEEE, pp 3009–3012

Rudge A, Milne K, Olver A, Knight P (1982) The handbook of antenna design, vol 15 and 16, IEE electromagnetic waves series. The Institution of Electrical Engineers, London

Sievenpiper D, Zhang L, Broas R, Alexopolus N, Yablonovitch E (1999) High-impedance electromagnetic surfaces with a forbidden frequency band. IEEE Trans Microwave Theory Tech 47:2059–2074

Sözuer H, Dowling J (1994) Photonic band calculations for woodpile structure. J Mod Opt 41:231–239

DOI 10.1007/978-981-4560-75-7_123-2

Teniente J (2003) Modern corrugated horn antennas. Ph.D. dissertation, Electric and Electronic Engineering Department, Public University of Navarra

Teniente J, Gonzalo R, del Río C (1999) Gaussian profiled horn antenna for hispasat 1C satellite. Int J Infrared Millimetr Waves 20:1809–1815

Teniente J, Gonzalo R, del Río C (2002a) Ultra wide band corrugated gaussian profiled horn antenna design. IEEE Microwave and Wireless Compon Lett 12:20–21

Teniente J, Goñi D, Gonzalo R, del Rio C (2002b) Choked gaussian antenna. Extremely low sidelobe compact antenna design. IEEE Antennas Wireless Propag Lett 1:200–202

Teniente J, Gonzalo R, del Rio C (2005) State of the art of corrugated horn antennas that combine horizontal and vertical corrugations. In: Proceeding of the 11th international symposium on antenna technology and applied electromagnetics, (ANTEM), pp 86–87

Teniente J, Gonzalo R, del Rio C (2006) Innovative high gain corrugated horn antenna combining horizontal and vertical corrugations. IEEE Antennas Wireless Propag Lett 5:380–383

Teniente J, Gonzalo R, del Rio C (2009) Satellite Horn Antennas Design. In: Proceeding of the 3rd European conference on antennas and propagation (EuCAP), pp 3081–3084

Thèvenot M, Denis M, Reineix A, Jecko B (1999a) Design of a new photonic cover to increase antenna directivity. Microwave Opt Technol Lett 22:136–139

Thèvenot M, Cheype C, Reineix A, Jecko B (1999b) Directive photonic-bandgap antennas. IEEE Trans Microwave Theory Tech 47:2115–2121

Trentini G (1956) Partially reflecting sheet arrays. IRE Trans Antennas Propag 4:666–671

Vayonakis A, Schlaerth J, Kumar S, Gao J, Day P, Mazin B, Ferry M, Noroozian O, Glenn J, Golwala S, LeDuc H, Zmuidzinas J (2008) Antenna-coupled Microwave Kinetic Inductance detectors (MKIDs) for mm and submm imaging arrays. In: 19th international symposium on space terahertz technology, pp 141

Viskum H, Sorensen S (1994) Dual offset shaped reflectors optimized for gain and XPD performance. Antennas Propag Soc Int Symp 2:894–897

Yang F, Rahmat-Samii Y (2003) Microstrip antennas integrated with electromagnetic band-gap (EBG) structures: a low mutual coupling design for array applications. IEEE Trans Antennas Propag 51:2936–2946

Zeng L, Bennett C, Chuss D, Wollack E (2010) A low cross-polarization smooth-walled horn with improved bandwidth. IEEE Trans Antennas Propag 58:1383–1387

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_124-1

Antennas in Radio Telescope Systems

S W Ellingson*
Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

Abstract

Radio astronomy is the study of the universe by measurement of radio frequency emission at frequencies ranging from a few MHz to the far infrared. Signals of interest are typically extraordinarily weak, necessitating large effective aperture and resulting in some of the world's largest antenna systems. Technologies now commonly employed include reflector antennas ("dishes") using horn-type feeds or feed arrays, beamforming arrays consisting of elements ranging from dipoles to large dishes, and interferometry. Many problems in radio astronomy also require very fine angular resolution, leading to aperture synthesis imaging instruments consisting of antennas distributed over apertures ranging from hundreds of meters to intercontinental distances. This chapter provides a brief review of antenna systems used in operational modern radio telescopes and in anticipated new radio telescopes.

Keywords

Radio astronomy; Reflector antenna; Phased array; Interferometry; Aperture synthesis imaging

Introduction

Radio astronomy is the study of the universe by measurement of radio frequency emission. Radio frequencies provide information which is difficult or impossible to obtain from emission in other regions of the electromagnetic spectrum (in particular infrared, optical, X-ray, and γ-ray) and which is complementary to observations in these wavebands. Ground-based radio astronomy is possible from frequencies of roughly 10 MHz, below which the Earth's ionosphere becomes a barrier, to about 10 GHz, where atmospheric absorption becomes a barrier. Radio astronomy at higher frequencies is possible in spectral "windows" within which atmospheric absorption is sufficiently low, up to the infrared frontier at roughly 1 THz. Radio astronomical measurements include the measurement of total power, power spectral density, and time–frequency distributions (i.e., dynamic spectrum) as well as imaging with resolution as fine as 60 microarcseconds (about 1.7×10^{-8} degrees, achieved by the EHT; see section "Very Long Baseline Interferometry (VLBI)") or fields of view as large as tens of degrees or the entire sky. Nearly all radio astronomical instruments employ antenna systems with extraordinarily large effective aperture, which is necessary because emissions of interest are typically extraordinarily weak, and very fine angular resolution, which is necessary to discriminate sources of interest from the backgrounds or for imaging. This chapter provides a review of modern radio telescopes and in particular of their antenna systems. The focus is on contemporary and emerging systems; for a historical perspective, Baars et al. (2009) (and references therein) is recommended.

*Email: ellingson@vt.edu
*Email: ellingson.1@vt.edu

DOI 10.1007/978-981-4560-75-7_124-1

To begin, a brief summary of problems of greatest interest in modern radio astronomy is in order, so as to provide context. Much of modern radio astronomy is devoted to the study of neutral hydrogen, the most common material in the universe and detectable at radio frequencies as a narrow spectral line at about 1,420 MHz (about 21 cm wavelength) for material at rest with respect to the Earth. The brightest regions of the Galaxy are observable through doppler-shifted hydrogen emission with brightness temperature on the order of 100 K, and other galaxies are detectable via emission which is orders of magnitude weaker and redshifted by 10–100 s of MHz. Some operational and planned telescopes are striving to detect hydrogen emission associated with the first stars (the "Epoch of Reionization," EoR), which is redshifted to frequencies in the range 100–200 MHz, and earlier (known as the "dark ages" and "cosmic dawn"), which entails observations of hydrogen redshifted to frequencies as low as 30 MHz (section "EoR and BAO Experiments"). Other molecules routinely observed at radio frequencies include the hydroxyl radical (OH, rest frequencies around 1,700 MHz), methyladyne (CH, around 3.3 GHz), formaldehyde (4.8–14.5 GHz), methanol (6.7 GHz), water vapor (22 GHz), ammonia (24 GHz), and many others. These molecules emit detectable emission at frequencies throughout the millimeter and submillimeter range of wavelengths. Observations of the associated spectral lines are used in the study of comets, astronomical masers, emerging solar systems, and astrobiology. Neutron stars emit beams of electromagnetic radiation which are routinely observed as broadband periodic pulsed emission as the beam sweeps past Earth with the rotation of the neutron star. Radio telescopes typically measure both orthogonal polarizations of the emission, both as a means to increase sensitivity as well as to explore the polarization of the emission. Radio astronomy plays an important role in cosmology via precision measurement of the intensity and polarization of the cosmic microwave background in the frequency and spatial domains. This is certainly not an exhaustive list of science topics but should suffice to demonstrate the scope and diversity of projects which are pursued using radio telescopes.

This chapter is organized as follows. The section "Contemporary Radio Telescope Systems" is a review of prominent radio telescopes currently in operation. This review is organized into sections according to the type of antenna system; categories include single-dish instruments (section "Single-Dish Instruments"), beamforming arrays (section "Beamforming Arrays"), interferometers (section "Interferometers"), and EoR and baryon acoustic oscillation (BAO) experiments (section "EoR and BAO Experiments"). The section "Future Directions and Open Problems" addresses trends in radio telescope antenna systems and describes instruments which are planned or in the early stages of development. Principal topics in the section "Future Directions and Open Problems" are (1) the evolution from "single-pixel" feeds toward phased array feeds ("PAFs") and from dishes toward "aperture arrays" (section "Evolution toward PAFs and Aperture Arrays") and (2) the Square Kilometre Array (SKA) project, which is working toward the next generation of large radio telescopes (section "Square Kilometre Array").

Two additional introductory comments: First, this chapter focuses on antenna systems specifically; that is, receiver electronics and signal processing are addressed only at the highest level. For additional reading on how radio telescopes are used as complete systems, Wilson et al. (2013, in particular, Chapters 1 and 10–16) is suggested. Second, it should be noted that this chapter does not attempt an exhaustive accounting of all operational or planned radio telescopes but rather focuses on those which are most prominent on a performance basis (e.g., among the most sensitive or among the highest angular resolution) and those with distinctive features that are relevant from an antenna engineering perspective.

Contemporary Radio Telescope Systems

This section provides a review of prominent currently operational radio telescope systems, categorized by type of antenna system: "single dish" instruments, beamforming arrays, interferometers, and EoR/BAO

DOI 10.1007/978-981-4560-75-7_124-1

experiments. It should be noted that some instruments fall into multiple categories; in particular, the distinction between beamforming arrays and interferometers is potentially ambiguous. In this case, instruments are categorized according to their "primary" use, and alternative uses are mentioned.

Single-Dish Instruments

The term "dish" is of course shorthand for an antenna system consisting of a paraboloidal reflector or a shaped reflector operating according to similar principles. The principles of steerable-dish radio telescope design, from a mechanical perspective, are summarized in von Hoerner (1967). In general, the largest possible dish is desired, since this yields the highest directivity and subsequently the highest sensitivity and spatial selectivity. However, the dish must also provide sufficient surface accuracy, both in terms of overall shape and in terms of local surface errors. Difficulties in achieving sufficient shape accuracy include mitigating deformation resulting from gravity and nonuniform solar heating, whereas local surface errors are due to nominally static features such as seams and rivets, surface misalignments, and surface roughness. Shape error degrades aperture efficiency, distorts beam shape, and introduces pointing error; local surface errors degrade aperture efficiency and tend to determine the highest frequency at which a dish is considered usable (Ruze 1966).

With a few important exceptions described later in this chapter, most existing single-dish instruments use axisymmetric paraboloidal (or nearly paraboloidal "shaped") main reflectors, cassegrain subreflectors, and horn-type feeds. Axisymmetric design is far from optimal due to aperture blockage and scattering associated with feed structures; however, compromise is motivated by design difficulty and cost of alternatives. The preference for Cassegrain optics over prime-focus feeds in axisymmetric systems is motivated primarily by the desire to minimize system temperature. This is because in the VHF band and above, noise from the sky is significantly less than noise originating from the ground. Therefore, mounting feed horns at the vertex of the dish, pointing skyward, tends to reduce the fraction of spillover that intercepts the ground.

There have been two important trends in modern radio telescope dish design that must be mentioned: offset-fed systems and feed arrays. Offset-fed systems are motivated by the limitations of axisymmetric systems (indicated above) as well as by the particular susceptibility of axisymmetric systems to ripple in the frequency response due with standing waves between feed and reflector (Morris 1978). Offset-fed systems also make practical the use of Gregorian (as opposed to Cassegrain) subreflectors, resulting in further reduction of system temperature through ground noise pickup. However, these advantages come at considerable cost and complexity and so have not been consistently adopted in new designs. Prominent examples of offset-fed reflector systems in radio astronomy include the GBT and ATA, the millimeter-wave instruments SPT and ACT, and the space telescopes WMAP and Planck; see later sections for additional information on these systems.

The other important trend has been the adoption of feed arrays. Feed arrays can be broadly classified as either focal plane arrays (FPAs) or phased array feeds (PAFs). (The radio astronomy community is not completely consistent in the use of this terminology, but the definitions employed here are relatively common.) An FPA is an array of "single-pixel" feed elements distributed in the focal plane, often implemented as a replacement for a single feed located exactly at the focus. Each feed element not exactly at the focus corresponds to a beam which points in a slightly different direction. In this way, a telescope with an N-element FPA achieves an effective field of view which is N times greater than that achieved by a single-feed system. This is an enormous advantage for science requiring observations over large areas, such as hydrogen intensity mapping, pulsar surveys, and certain cosmology experiments. However, N is limited to $\mathcal{O}(10)$ or so by the degradation of sensitivity and beam shape with increasing distance from the focal point, which constrains the area in the focal plane over which feeds may

DOI 10.1007/978-981-4560-75-7_124-1

effectively be positioned. Most of the largest single-dish instruments in the world now have FPAs, including GBT, Arecibo, and Parkes; see descriptions of these instruments below for additional details.

In contrast to an FPA, a PAF is a feed consisting of an array of a large number of low-gain antennas which are combined to form beams or cross-correlated to form images. Processing of the array elements is typically done in the digital domain, introducing the possibility of optimizing the main lobe or sidelobes (perhaps even dynamically) and performing interference mitigation by nullforming; see, e.g., Landon et al. (2010). Experimental phased array feeds have been demonstrated on small axisymmetric prime-fed systems (see in particular Elmer et al. 2012) and on Arecibo (Warnick et al. 2011) and will soon become the primary (permanent) feed systems for the WSRT and ASKAP (section "Other Prominent Interferometer-Type Radio Telescopes").

The Green Bank Telescope

The Green Bank Telescope (GBT), shown in Fig. 1, is the world's largest fully steerable filled-aperture radio telescope. (In fact, it is currently the world's largest land-based two-axis steerable structure of any kind.) It is located in Green Bank, West Virginia (USA) (Prestage et al. 2009). The main reflector of the GBT is an offset 100 m × 110 m paraboloidal section. The offset design positions the feed support structure outside the optical path, resulting in low sidelobe levels and avoiding the troublesome spectral ripple problem noted above. The main reflector consists of 2,004 separate panels employing 2,209 mechanical actuators used to independently adjust the panels to compensate for mechanical distortion due to gravity and nonuniform solar heating. The resulting surface deviation from ideal is $\leq$400 μm RMS, improving to $\leq$250 μm RMS after corrections for thermal effects. The associated loss of effective aperture due to surface efficiency is $<$1 % below 5 GHz, resulting in an effective area of about 5,500 m^2 and aperture efficiency of about 70 %. The main reflector is considered usable from 100 MHz to 115 GHz.

The GBT uses a 7.6 m × 8.0 m Gregorian subreflector, visible in Figs. 1 and 2, for frequencies above 1.15 GHz. The Gregorian design reduces main reflector spillover relative to a Cassegrain design, subsequently reducing the contribution of ground noise to the system temperature. The effective *f*/*D*

Fig. 1 The Green Bank Telescope. In this image, the prime focus feed is retracted (Image credit: NRAO/AUI/NSF)

DOI 10.1007/978-981-4560-75-7_124-1

Fig. 2 GBT subreflector and prime focus feed. In this image, the prime focus feed is deployed (Image credit: B. Saxton, NRAO/AUI/NSF)

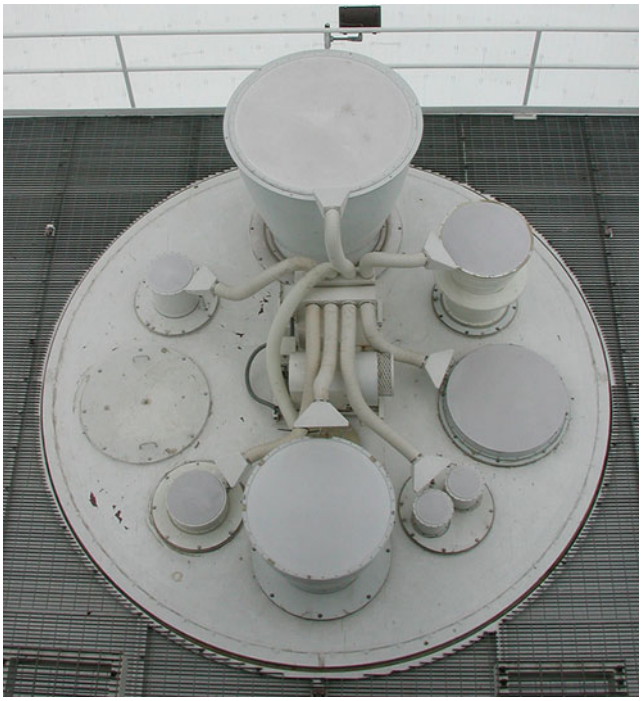


Fig. 3 GBT feed turret (Image credit: B. Saxton, NRAO/AUI/NSF)

when using a Gregorian subreflector is 0.29 referred to the 208 m circular "parent paraboloid." The feeds for this mode of operation are arranged in a turret located on the feed arm, visible in Figs. 1 and 3. The turret rotates the desired feed into the secondary focus. At any given time, the turret accommodates up to eight of the following feed/receiver combinations: Horns for L-band (1.15–1.73 GHz), S-band (1.73–2.60 GHz), C-band (3.95–5.85 GHz), X-band (8–10 GHz), Ku-band (12–15.4), K-band (18–22/22–26.5), Ka-band (26–40 GHz), Q-band (38–48 GHz), and W-band (68–92 GHz). Most of these feeds illuminate the subreflector with about 14 dB edge taper, resulting in about 1 % rear (sky-facing) spillover. The turret also accommodates a K-band FPA and MUSTANG, a 64-element bolometer FPA for 82–100 GHz.

DOI 10.1007/978-981-4560-75-7_124-1

Fig. 4 Arecibo (Image courtesy of the NAIC-Arecibo Observatory, a facility of the National Science Foundation)

Below 1 GHz, the GBT uses a prime focus feed system having an effective $f/D = 0.6$ with respect to the projected diameter of the paraboloid. One horn at a time may be mounted to a support structure which pivots outward to position the feed at the prime focus, as shown in Fig. 2. When extended, the prime focus feed blocks the subreflector. Frequency ranges available in prime focus configuration include 290–395 MHz, 385–520 MHz, and 510–690 MHz using a dipole-type feed and 680–920 MHz using a horn.

Arecibo

The world's largest filled-aperture radio telescope, shown in Fig. 4, is at the Arecibo Observatory, located near Arecibo, Puerto Rico (USA) (Salter 2012). The main reflector of the Arecibo telescope is a fixed spherical section 305 m in diameter, having a radius of curvature of 265 m and subtending a half-angle of 35° as seen from the center of curvature. Surface accuracy is about 2 mm RMS, limiting the instrument to frequencies below about 10 GHz. Feeds are located in a platform suspended above the main reflector by cables, as shown in Fig. 5. The platform rotates for pointing in azimuth, and feeds traverse the curved rail for pointing in elevation up to about 20° from the zenith.

Because Arecibo's main reflector is spherical, there is no prime focus, and feed systems must account for the particular optics of reflection from a spherical surface. The principal feed system consists of Gregorian secondary and tertiary subreflectors, located in the dome visible on the left side of the feed platform in Fig. 5. Illumination reflected by the main reflector arrives first at a Gregorian secondary reflector located in the upper part of the dome and curved downward and is then culminated by a Gregorian tertiary reflector located in the lower part of the dome and curved upward (Kildal et al. 1994). Signals are received by feeds located in a turret on the top side of the dome, opposite the secondary reflector. This system results in roughly uniform illumination of an ellipse on the main reflector approximately 213 m × 237 m in extent, resulting in aperture efficiency (including blockage) of about 75 % below 2 GHz and beamwidth of about 1′ at 5 GHz. Feed systems available via this system include

DOI 10.1007/978-981-4560-75-7_124-1

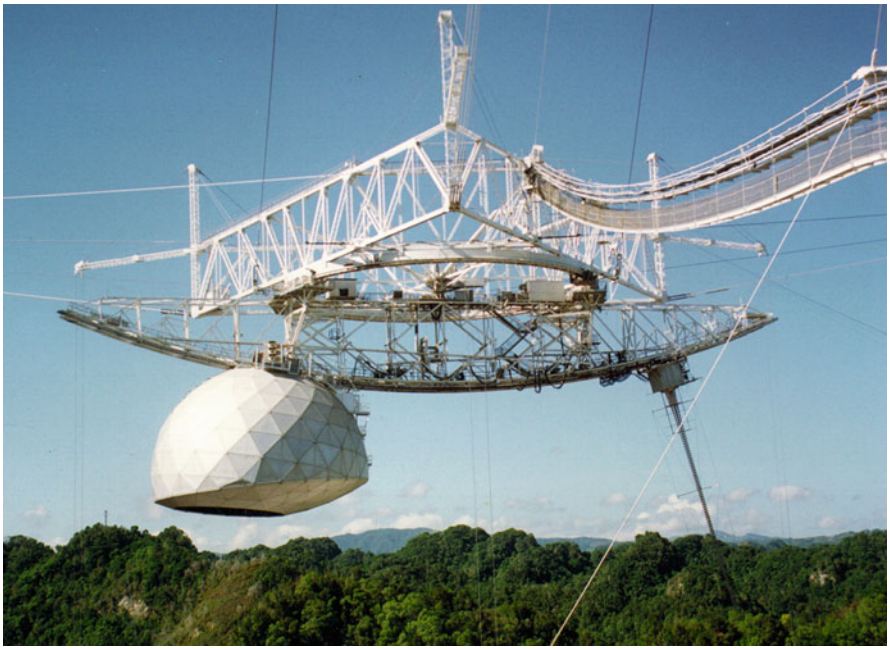


Fig. 5 Arecibo's feed platform (Image courtesy of the NAIC-Arecibo Observatory, a facility of the National Science Foundation)

horns for a 327 MHz band (312–342 MHz), a 430 MHz band (425–435 MHz), a 800 MHz band (705–800 MHz), L-band (1.12–1.73 GHz), S-band (separate feeds for 1.8–3.1 GHz and 3–4 GHz), C-band (separate feeds for 3.85–6.05 GHz and 6–8 GHz), and X-band (8–10 GHz). This system accommodates the Arecibo L-band Feed Array (ALFA), an FPA consisting of seven horns operating at 1,225–1,525 MHz. Arecibo also includes a "line feed" for 430 MHz (visible on the right side of the feed platform in Fig. 5) and a "point feed" for 47 MHz. The line feed is about 29 m long and lies along the caustic associated with the desired pointing direction, yielding a beamwidth of about 11′ over a bandwidth of about 10 MHz. The 47 MHz feed provides a beamwidth of roughly 100′.

Other Large Dishes for cm-Wavelengths and Longer

The world's other operational large single-dish radio telescopes are predominantly two-axis steerable axisymmetric paraboloidal systems. Prominent among these are Effelsburg (100 m, Germany; Hachenberg et al. 1973), Jodrell Bank (76 m, England), and Parkes (64 m, Australia). Parkes is equipped with a 13-beam L-band FPA located at its prime focus. A summary of the world's radio telescopes having dishes greater than 20 m diameter is provided in Appendix A.2 of Ekers and Wilson (2013) (also, Figure 8.7 of this reference shows the Parkes FPA being installed).

For a variety of reasons (see especially the first paragraph of section "Beamforming Arrays"), the construction of new cm-wavelength single-dish instruments on the scale of the GBT, Effelsburg, Jodrell Bank, and Parkes is quite unlikely; nearly all planned or imagined future radio telescopes are either beamforming arrays (section "Beamforming Arrays") or interferometers (section "Interferometers"). A notable exception is the Five Hundred Meter Aperture (FAST) telescope currently under construction in China (Nan 2011). FAST will be similar to Arecibo, only significantly larger – when completed, it will become the world's largest single filled-aperture radio telescope.

DOI 10.1007/978-981-4560-75-7_124-1

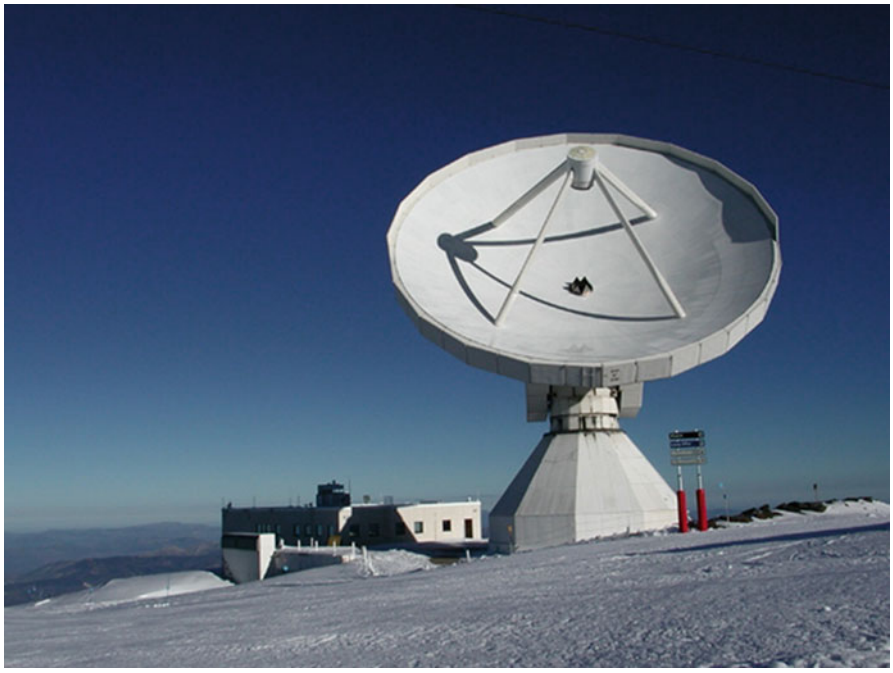


Fig. 6 The IRAM 30-m millimeter-wavelength radio telescope (Image credit: IRAM/Alexandre Beelen)

Large Single-Dish Instruments Operating at Higher Frequencies

What constitutes a "large dish" at millimeter and submillimeter wavelengths is physically smaller than what constitutes a large dish at centimeter wavelengths. This is because beamwidth is proportional to wavelength but also because surface accuracy emerges as a fundamental limitation on aperture efficiency, requiring more stringent and costly measures to achieve acceptable thresholds. The largest instruments currently capable of mm-wave operation (aside from GBT and Effelsburg, described above) include the Sardinia Radio Telescope (SRT) and the Large Millimeter Telescope (LMT). The SRT consists of a 64-m main reflector usable to 3 mm (115 GHz) with a Gregorian feed. The LMT, located atop Sierra Negra in Mexico, has a 32-m main reflector with plans to increase this to 50 m. The LMT is an axisymmetric cassegrain design and is intended to operate from 4 mm (75 GHz) to 0.85 mm (353 GHz). The instrument includes two FPAs: AzTEC, a 144-element FPA for 1.1 mm (Wilson et al. 2008), and SEQUOIA, a 32-element FPA for 3 mm. Other prominent large millimeter-wave single-dish instruments include the IRAM 30-m dish on Pico Veleta, Spain (Fig. 6), with two FPA systems for 1 mm (Baars et al. 1994), the Arizona Radio Observatory 12 m and 10 m dishes, and the 12 m dish of the Atacama Pathfinder Experiment (APEX) at Llano Chajnantor, Chile.

Prominent single-dish instruments designed to operate at submillimeter wavelengths (and into the infrared, depending on one's definition of wavebands) include the 15-m James Clerk Maxwell Telescope (JCMT) on Mauna Kea, Hawaii, the 10.4-m instrument of the Caltech Submillimeter Observatory (CSO) also on Mauna Kea, and the 10-m South Pole Telescope (SPT; Carlstom et al. 2011). A number of prominent space- and ground-based telescopes dedicated to cosmology (described in the next section) also operate in the submillimeter regime. For additional reading on submillimeter telescopes specifically, Phillips et al. 2013 is recommended.

Cosmic Microwave Background Experiments

There is currently intense interest in millimeter-/submillimeter-wave polarimetric observations of the cosmic microwave background with the aim of detecting large-scale (on the order of degrees) variations in

DOI 10.1007/978-981-4560-75-7_124-1

brightness, temperature, and polarization that would provide insight into the early history of the universe, extending backward to the epoch of inflation, just a tiny fraction of a second following the Big Bang. This is best achieved using an FPA operating as "radio camera" on a relatively small dish for a large field of view. Because the signal of interest is extraordinarily weak, very long observations (on the order of years) are required in order to achieve significant signal-to-noise ratio. Suitable instruments are typically purpose built, as existing instruments are not suitable and are typically not able to provide the necessary observing time. Examples of instruments in this class include the Atacama Cosmology Telescope (ACT; Swetz 2011), the Background Imaging of Cosmic Extragalactic Polarization (BICEP) series of instruments, the Cosmic Background Imager (CBI), and the Q/U Imaging ExperimenT (QUIET). For additional reading on these instruments, Hanany et al. (2013) is recommended; see also the discussion of space-based radio telescopes in the next section.

Space-Based Radio Telescopes

Some applications do not require high angular resolution but rather relatively large bandwidth and extreme sensitivity; for these applications, operation from space is preferable. The cosmology instruments WMAP (Page et al. 2003) and *Planck* (Doyle et al. 2009; Tauber et al. 2010) are small (0.7–1.5 m, respectively) offset Gregorian reflector systems at the L2 Lagrangian point, beyond the orbit of the Moon. Operation at L2 allows coverage of the entire celestial sphere while mitigating the limitations imposed by atmospheric losses and interference from terrestrial sources. *Hershel* (Doyle et al. 2009; Pilbratt et al. 2010) is a 1.5 m cassegrain submillimeter-wavelength radio telescope operating at L2, designed primarily to study star formation in the early universe. Also, Earth-orbiting satellites equipped with single dishes have been used as stations for very long baseline interferometry (VLBI); see the section "Very Long Baseline Interferometry (VLBI)."

Beamforming Arrays

The sensitivity and spatial selectivity of a single-dish instrument is fundamentally limited by the size of the dish, which in turn is limited by mechanical considerations. On the other hand, sensitivity is also limited by receiver noise temperature. Technological improvements over the last two decades have resulted in lower-cost receivers which perform nearly as well as state-of-the-art radio telescope receivers. Taken together, these trends make it quite likely that there will never be another steerable single-dish radio telescope as large as the GBT. Instead, the trend is in the direction of arrays of smaller dishes, each having receivers which are nearly as good as those on the best existing single-dish instruments and which can be distributed over a greater aperture to achieve improved spatial selectivity. These dishes may then be combined to form one or more beams within the beamwidth of the individual dishes. Within the radio astronomy community, this concept is often referred to as "large N small D" (LNSD). Two prominent instruments that are able to operate as beamforming arrays are the Allen Telescope Array (ATA; section "Allen Telescope Array (ATA)") and the Very Large Array (VLA; described in section "Very Large Array (VLA)").

For wavelengths longer than about 1 m (300 MHz), it becomes cost effective to switch from dishes to even larger numbers of lower-gain elements, such as dipoles. This is due in part to the poor directivity of a small dish at these wavelengths and in part due to the dramatically increasing antenna temperature with decreasing frequency associated with the Galactic synchrotron background, which facilitates the use of simpler, less expensive uncooled electronics without significant penalty in sensitivity (Ellingson 2005; 2011). The section "Long Wavelength Array Station 1 (LWA1)" describes the Long Wavelength Array Station 1 (LWA1) as an example of an operational large meter-wavelength beamforming instrument; other low-frequency instruments such as LOFAR (described as an interferometer in section "LOw Frequency ARray (LOFAR)") also have beamforming modes. Beamforming arrays consisting of

DOI 10.1007/978-981-4560-75-7_124-1

large numbers of low-gain elements have some unique issues, including evolution of beam shape and sidelobe structure as a function of pointing, degradation of sensitivity with increasing zenith angle, and high overall system complexity.

Before describing specific instruments, it is worth noting that beamforming arrays are not solely a response to mechanical challenges or engineering economics but are also of interest because of the distinct advantages that beamforming provides. Primary among these is multibeaming – that is, combining the signals from individual elements so that multiple beams, each pointed in a different direction, – can be simultaneously generated. Each additional beam that can be independently pointed is one additional observer that can be simultaneously accommodated – dramatically increasing the scientific productivity relative to a single-dish system. Another advantage of beamforming is the potential to form beams having custom shapes or having sidelobes that are tailored to reduce system temperature or null interference; see, e.g., Raza et al. (2002) and Ellingson and Cazemier (2003). It should be noted that PAFs (and, to a lesser extent, FPAs) offer similar capabilities for dishes; see, e.g., Jeffs et al. (2008).

Allen Telescope Array (ATA)

The ATA (Welch et al. 2009), located in Hat Creek, California (USA), is a seminal implementation of the LNSD concept. The instrument currently consists of 42 6.1-m dishes distributed over an aperture roughly 300 m in maximum dimension. Each dish is an offset-fed paraboloid using a Gregorian subreflector with a dual-polarization log-periodic feed, achieving aperture efficiency of about 60 %. The feed system has an instantaneous tuning range of 0.5–11.2 GHz; however optimum performance requires a frequency-dependent linear mechanical positioning of the feed with respect to the subreflector focus. The dish provides a 2.5° main lobe at 1.4 GHz; subsequent beamforming results reduce this to about 4.3′ per beam, with the ability to position multiple beams located within the field of view defined by the dish pattern.

Deuterium Array

The Deuterium Array, shown in Fig. 7, was an array of low-gain elements which successfully observed the extraordinarily weak and previously undetected 327 MHz spectral line of deuterium (Rogers et al. 2007). The array consisted of 24 “stations,” each consisting of 24 dual-polarization two-element (main + director) Yagi antennas positioned according to a 0.8-wavelength rectilinear grid above a wire

Fig. 7 The Deuterium Array (Image credit: MIT Haystack Observatory)

DOI 10.1007/978-981-4560-75-7_124-1

Fig. 8 LWA1 (Image credit: LWA)

mesh ground plane. Each polarization of every antenna was separately digitized, and beams were formed entirely in software. Although actually a short-lived (2004–2006) experiment, the instrument was significant as a demonstration of the efficacy of digital beamforming for high-sensitivity radio astronomy, and no comparable instrument currently exists with the ability to observe in this frequency range.

Long Wavelength Array Station 1 (LWA1)

LWA1 (Ellingson et al. 2013a, b) is a large modern operational general-purpose beamforming array operating in 10–88 MHz. The instrument, shown in Fig. 8, is collocated with the VLA in central New Mexico. LWA1 consists of about 256 dual-polarized dipole-type antennas. The output from each antenna is individually digitized, and beams are formed using a digital-domain delay-and-sum technique. The dipoles are distributed within a 100 m × 110 m elliptical aperture, resulting in a beamwidth of roughly 2° for zenith pointing at the highest frequencies, increasing to tens of degrees at the lowest frequencies and for pointing at low elevations. The major axis of the aperture is aligned in the north–south direction in order to improve the symmetry of the beam for Galactic center observations, which require pointing to low elevations toward the south. The distribution of antennas is pseudorandom with 5 m minimum spacing to facilitate maintenance. The pseudorandom distribution ameliorates grating lobes, which would otherwise become a problem for frequencies greater than about 30 MHz.

The LWA1 antenna element is a wire-grid bowtie-type dipole, having dipole arms that are angled downward into an inverted "V" configuration to improve uniformity of gain over the sky (Hicks et al. 2012). The useful bandwidth of the antenna element is considerably greater than the impedance bandwidth of the antenna element; this is attributable to the fact that the sky – in particular, the Galactic synchrotron radiation – is very bright in this frequency regime (Ellingson 2005). As a result, the ratio of irreducible external noise after the impedance mismatch to the internal (receiver) noise (about 300 K) is large, and thus the impact of the impedance mismatch on sensitivity is small. For LWA1, the system temperature is dominated by the Galactic synchrotron background by a factor of 4:1 from 24 to 87 MHz; thus improved impedance matching would result in only a very small improvement in sensitivity.

A *disadvantage* of having external noise-limited sensitivity is that the sensitivity of the array does not increase linearly with the number of antennas, due to the deleterious effect of correlation of external noise between closely spaced antennas (Ellingson 2011). Further, the brightness temperature distribution of the external noise varies with time of day, and the specific effect depends on beam pointing direction. So, whereas it is reasonable to characterize the sensitivity of large dishes in terms of effective aperture and

DOI 10.1007/978-981-4560-75-7_124-1

system temperature (i.e., A_e/T_{sys} or, in communications engineering nomenclature, G/T), such a characterization is not necessarily meaningful for modern low-frequency beamforming arrays employing antennas with large impedance mismatch, such as LWA1. In this case, system equivalent flux density (SEFD) is particularly convenient. SEFD is the flux density (i.e., W m^{-2} Hz^{-1}) of a source, unresolved by the beam, which results in a doubling of the apparent system temperature. LWA1 achieves an SEFD on the order of 10 kJy at the zenith. This sensitivity is approximately independent of frequency, because the effective aperture of dipoles increases in proportion to λ^2, whereas the antenna temperature due to the Galactic noise background increases roughly as $\lambda^{2.5}$, and sensitivity is proportional to the ratio.

Other Low-Frequency Beamforming Arrays

Another large modern operational low-frequency array which is sometimes used in a beamforming mode is LOFAR, which is designed primarily as an interferometer and so is described in more detail in the section "LOw Frequency ARray (LOFAR)." Before moving on, however, it is worth noting that essentially all currently operational radio telescopes designed to operate below 300 MHz consist of large numbers of low-gain elements combined to form beams, for reasons described earlier. LWA1 and LOFAR are relative newcomers to this frequency regime; other instruments operating below 100 MHz (and which have been operational for decades as analog beamforming arrays) include GEETEE (35–70 MHz), located in Gauribidanur, India, UTR-2 (Braude et al. 1978, 5–40 MHz), located in Ukraine, and the Nançay Decametric Array (10–100 MHz), located in France.

Interferometers

A distinct category of radio telescopes is *aperture synthesis imaging* instruments, often referred to as "interferometer arrays" or simply as "interferometers." Instruments in this category are designed primarily for imaging, in contrast to the "single pixel per feed" mode of operation associated with single-dish instruments, single-dish instruments using FPAs, and beamforming arrays. (To be absolutely clear: these instruments are *also* routinely used for imaging, e.g., by sweeping a beam across the desired field of view or by using an FPA as a "radio camera." This section addresses instruments which are designed primarily for aperture synthesis imaging.)

The typical method of operation for an interferometer consists of cross-correlating the signals received from widely separated antennas, followed by an application of the Fourier transform to obtain an image via the Van Cittert–Zernicke theorem. The actual processing is considerably more involved but beyond the scope of this chapter; for details, Thompson et al. (2001) is recommended. The principal characteristics of an interferometer are resolution, which is limited by maximum baseline length (i.e., maximum separation between two antennas in the array), and FOV, which is limited by the beamwidth of individual antennas. Useful FOV is also limited by practical considerations, such as imperfect coherence over bandwidth resulting from processing limitations. For this reason, the useful FOV is sometimes referred to as largest angular scale (LAS), which is inevitably less than the dish beamwidth.

Very Large Array (VLA)

The VLA (also known as the "Expanded VLA" (EVLA) and "Jansky VLA" (JVLA)) is perhaps the seminal example of an operational large radio interferometer (Napier et al. 1983; Perley et al. 2009). The array is located in central New Mexico (USA) and consists of 27 25-m dishes arranged in a "wye" geometry as shown in Fig. 9. The individual dishes are transportable along the 21-km-long arms of the wye, facilitating baselines up to 36 km. At 1.4 GHz, the configuration with the longest baselines ("A") facilitates imaging with resolution on the order of 2″, scaling inversely with frequency. The dishes are periodically moved into configurations which are factors of 3 smaller (known as "B," "C," and "D") in order to obtain improved sensitivity over larger angular scales.

DOI 10.1007/978-981-4560-75-7_124-1

Fig. 9 The VLA in its most compact ("D") configuration (Image credit: NRAO/AUI/NSF)

Fig. 10 VLA dishes showing details of cassegrain feed system (Image credit: B. Saxton, NRAO/AUI/NSF)

The individual dishes use a cassegrain-type system in which the feed horns for eight contiguous bands covering 1–50 GHz are located at the vertex of the dish, as shown in Fig. 10. The 2.3-m cassegrain subreflector is mounted with a slight tilt relative to the axis of revolution of the main reflector, allowing feed selection by rotating the subreflector around this axis.

The VLA also has "P-band" (230–470 MHz) and "4-band" (54–83 MHz) feed systems, which are not cassegrain but rather are implemented using dipole feeds mounted between the subreflector and main reflector, using the subreflector as a crude ground plane (Kassim et al. 2007). The 4-band dipoles reduce

DOI 10.1007/978-981-4560-75-7_124-1

Fig. 11 ALMA 12-m dishes (Image credit: W. Garnier, ALMA (ESO/NAOJ/NRAO))

the sensitivity of the higher-frequency cassegrain-mode feed systems and so are only intermittently deployed. An experimental/alternative 4-band feed array has recently been deployed to six dishes. This system consists of four end-fed dipoles positioned outside the optical path of the cassegrain feed system, significantly reducing interference to higher-frequency systems with a small reduction in effective aperture relative to the legacy on-axis dipole system (Harun and Ellingson 2011; Ellingson et al. 2013c).

As mentioned previously, VLA also has a "phased array" mode in which it operates as a beamforming array, which is particularly useful for observations of sources which cannot be resolved by imaging, such as pulsars. In this mode, the VLA achieves a beamwidth approximately inversely proportional to its maximum baseline with an effective aperture of roughly 10,000 m^2 – nearly 70 % greater than that of the GBT.

Atacama Large Millimeter/Submillimeter Array (ALMA)

ALMA (Mangum et al. 2006; Wooten and Thompson 2009) consists of 64 12-m dishes plus a compact array of 7-m dishes located near San Pedro de Atacama, Chile. ALMA just recently (2011) began science operations and represents the current state of the art in millimeter-wave astronomy in many respects. Figure 11 shows ALMA 12-m dishes, which operate in seven bands in 84–720 GHz (3.6–0.4 mm) and are transportable to form configurations with baselines up to 16 km. At 300 GHz (1 mm), this facilitates imaging with resolution of roughly 20 milliarcseconds. Like the VLA, ALMA may also be used as a beamforming array.

Giant Meterwave Radio Telescope (GMRT)

The GMRT consists of 30 45-m dishes located north of Pune, India. Whereas ALMA is essentially the millimeter/submillimeter-wave complement of the VLA, the GMRT addresses primarily lower frequencies, in bands from 50 MHz to 1.5 GHz. The relatively low upper bound on frequency facilitates the use of a wire-grid dish surface, which in turn makes the significantly larger dishes practical and cost effective. Eighteen dishes are fixed in a wye configuration similar to the VLA, with 14-km arms (baselines up to 20 km), and the remaining 12 dishes are concentrated in a 1 km × 1 km pseudorandom configuration at the center. At 610 MHz, this facilitates imaging with resolution on the order of 10″.

DOI 10.1007/978-981-4560-75-7_124-1

Fig. 12 An aerial view of the LOFAR "superterp," a region within the 2-km-diameter "core" region, within which 24 of the stations are located. Nine stations are visible here: 6 within the circular island in the *center* of the image, one at the *lower left*, and two more on the *upper right*. The *square dark-colored* panels are HBA tiles, organized here into 24-tile subarrays. The *smaller dark-colored* patches are the ground screens of individual LBA antennas (Credit: van Haarlem et al. (2013) A&A, 556: A2, reproduced with permission © ESO)

LOw-Frequency ARray (LOFAR)

LOFAR (van Haarlem et al. 2013) is an operational large low-frequency interferometer consisting of tens of "stations" (sites containing beamforming arrays; 48 planned) distributed over Northern Europe. Figure 12 shows several of these stations. Each station consists of a 10–90 MHz "low-band array" (LBA) component and 110–240 MHz "high-band array" (HBA) component. These components use different antenna elements and geometries. LBA arrays consist of 48 or 96 dual-polarized thin-dipole inverted V-shaped antennas over ground screens arranged in a pseudorandom geometry within a circular aperture about 80 m in diameter. (See section "Long Wavelength Array Station 1 (LWA1)" for discussion as to why narrowband antenna elements are effective over multiple octaves in this wavelength regime.) HBA arrays consist of 24, 48, or 96 "tiles" (subarrays), with each tile consisting of a $4 \times 4 = 16$ array of dual-polarized "fat" wire-grid dipoles in a compact rectilinear geometry, with the tiles themselves being arranged in a compact geometry. The HBA tiles use analog beamforming; the LBA dipole outputs and HBA tile outputs are digitized, and all subsequent processing is digital. The stations are distributed in a pseudorandom manner with baselines ranging from 68 m to 1,158 km, facilitating imaging with resolution ranging from 0.5° to subarcsecond scales, depending on frequency, ionospheric conditions, and interplanetary scintillation. Although intended primarily as an aperture synthesis imaging instrument, LOFAR may also be operated as a beamforming array.

Other Prominent Interferometer-Type Radio Telescopes

Other prominent large currently operational radio interferometers include the Australia Telescope Compact Array (ATCA) (6 transportable 22-m dishes, maximum baseline 6 km, recent description: Wilson et al. 2011) and the Westerbork Synthesis Radio Telescope (WSRT; 14 25-m dishes, maximum baseline 3 km, Baars et al. 1973). WSRT is being retrofitted with a PAF system known as APERTIF ("APERture Tiles In Focus," van Cappellen and Bakker 2010). Each APERTIF PAF is a compact rectilinear array of 112 Vivaldi-type broadband horns operating from 1.0 to 1.7 GHz. In this application, the PAF serves to

DOI 10.1007/978-981-4560-75-7_124-1

Fig. 13 Three ASKAP dishes with PAFs visible (Credit: Schinckel AET, Bunton JD, Cornwell TJ, Feain I, Hay SG (2012) The Australian SKA Pathfinder, Proc. of SPIE Vol. 8444:2A)

simultaneously increase the imaging FOV and facilitates sophisticated spatial filtering techniques for mitigation of interference.

Associated with the SKA project (see section "Square Kilometre Array") are two "pathfinder" instruments that are expected to become operational over the next few years. The Australian SKA Pathfinder (ASKAP, Schinckel et al. 2012), shown in Fig. 13, consists of 36 12-m prime-focus dish systems in Western Australia. Although not clearly conveyed by the name of the instrument, ASKAP will be a prominent centimeter-wavelength interferometer in its own right and will be particularly well suited to survey-type projects due to large FOV (facilitated by relatively small dish diameter) and an innovative PAF-type feed system consisting of a self-complementary array of 94 dual-polarized square patches (Hay et al. 2011). The South African instrument MeerKAT will consist of 64 13.5 m offset Gregorian dishes with single-pixel feeds for bands between 580 MHz and 14 GHz (Davidson and Jonas 2012).

In the millimeter/submillimeter regime, other prominent interferometers include the Combined ARray for Millimeter Astronomy (CARMA, California, U.S.A.), Plateau de Bure and its successor, NOEMA (France), and the Sub-Millimeter Array (SMA; Mauna Kea, Hawaii; Blundell 2007). At the long-wavelength end of the radio spectrum, the Murchison Widefield Array (MWA, Tingay et al. 2013) has recently begun science operations. MWA is an 80–300 MHz array located in western Australia, intended primarily for angular power spectrum detection of the EoR but well suited to a variety of other science in this frequency band. MWA consists of 128 tiles, each consisting of a rectilinear array of $4 \times 4 = 16$ dual-polarization elements, similar to LOFAR HBA tiles. The tiles are distributed in a pseudorandom fashion with baselines up to 3 km. This results in a relatively large FOV (due to the small size of the beamforming tiles) simultaneously with useful angular resolution.

It should be noted that LWA1 (section "Long Wavelength Array Station 1 (LWA1)") is also able to operate as an interferometer; in the case of LWA1, the FOV is equal to the dipole element pattern – i.e., essentially the entire sky – whereas the spatial resolution is not significantly better than the beamwidth, since both modes are limited by the 110-m $\times$ 100-m physical aperture. Recently, a second LWA station

DOI 10.1007/978-981-4560-75-7_124-1

has been constructed at the Owens Valley Radio Observatory in California (USA) which has been designed primarily for interferometry. The Owens Valley LWA station array is essentially identical to that of LWA1, except the array geometry has been scaled up by a factor of about 2.3 with a corresponding improvement in resolution while retaining essentially full-sky FOV.

Very Long Baseline Interferometry (VLBI)

VLBI is interferometry using baselines extending beyond the boundaries of a particular radio observatory, typically with baselines on continental scales. A prominent operational VLBI instrument is the Very Long Baseline Array (VLBA), which consists of 10 25-m dishes (very similar to VLA dishes) distributed throughout the United States, including dishes in Hawaii and St. Croix (U.S. Virgin Islands), yielding baselines up 8,000 km. This yields spatial resolution on the order of 5 milliarcseconds at 1.4 GHz. Another prominent VLBI instrument is the Multi-Element Radio Linked Interferometer Network (e-MERLIN), consisting of seven dishes distributed throughout the United Kingdom (including Jodrell Bank), with baselines up to 217 km. In the millimeter regime, existing radio telescopes – now including ALMA – may be operated jointly for VLBI.

A prominent example of millimeter-wavelength VLBI is the Event Horizon Telescope (EHT), which aims to image the event horizon of the supermassive black hole at the center of the Galaxy at 1.3 mm wavelength. Finally, it should be noted that existing telescopes in all wavebands – including GBT, Arecibo, and the VLA – are commonly operated together to perform VLBI observations. Because data paths linking these instruments are typically not suitable for real-time correlation, it is common to record the signals from each instrument separately and then to perform correlation at a central location using the recorded data. Using this technique, baselines have been successfully extended even to small-dish radio telescopes in Earth orbit, also known as "space VLBI." A contemporary space VLBI mission is the Russian *RadioAstron* mission, which employs a satellite equipped with a 10 m dish.

EoR and BAO Experiments

As explained at the beginning of this chapter, one of the principal science drivers in modern radio astronomy is the quest to observe the EoR via 21-cm hydrogen emission redshifted to frequencies between (roughly) 100 and 200 MHz. This signal should appear as a millikelvin-scale variation in the spectrum of the CMB or alternatively as a distinct signature in the angular (spatial) power spectrum. The former approach can, in principle, be implemented using a single well-calibrated wideband dipole, since the signal is "global," i.e., is incident from the entire sky. Experiments pursuing the global spectrum approach include EDGES (Rogers and Bowman 2012) and SARAS (Patra et al. 2013). The spatial power spectrum approach requires an interferometer and is being pursued by LOFAR (section "LOw Frequency ARray (LOFAR)"), MWA (section "Other Prominent Interferometer-Type Radio Telescopes") and the "Precision Array to Probe the Epoch of Reionization" (PAPER; Parsons et al. 2010). PAPER is a purpose-built instrument consisting of dual-polarized sleeve dipoles over ground planes. This instrument is transportable, reconfigurable (various array geometries have been used), and scalable with deployments from 4 elements to 128 elements so far.

Due to the extreme weakness of the EoR signal, a detection is expected to require months to years of observation using existing instruments. A much larger interferometer will be required to productively observe the EoR signal. One such instrument now in the early stages of planning and development is the Hydrogen EoR Array (HERA), which will consist of hundreds of 14-m dishes, all fixed to point to zenith. Another instrument which aims to observe the VHF-band EoR signal is SKA Phase-1 Low, discussed in the section "Square Kilometre Array."

A distinct science effort leading to conceptually similar instruments is the quest to observe BAO in the frequency range 400–900 MHz, which promises to constrain the nature of dark energy. This is the primary

DOI 10.1007/978-981-4560-75-7_124-1

science objective of the Canadian Hydrogen Intensity Mapping Experiment (CHIME; Bandura et al. 2014), which would be a transit (fixed pointing) instrument consisting of fixed parabolic cylinder reflectors with line feeds.

Future Directions and Open Problems

This chapter has classified radio telescopes as being primarily single-dish systems, beamforming arrays, interferometers, and EoR/BAO experiments. In each category, prominent existing/operational instruments have been identified as well as instruments which are on track to become operational within the next few years; these include FAST, ASKAP, MeerKAT, NOEMA, HERA, and CHIME. This chapter concludes by looking a little further down the road to instruments that are in the early stages of planning and development and the associated technical trends.

Evolution toward PAFs and Aperture Arrays

A technical trend that seems likely to influence future instruments in most applications is the continuing shift from large filled apertures to arrays of larger numbers of low-gain elements followed by beamforming or interferometry. The primary motivations are addressed in the first few paragraphs of the section "Beamforming Arrays" and include the decreasing cost of electronics compared to filled apertures; the ability to support multiple users or observing programs simultaneously through multibeaming, increased flexibility in beam shape and sidelobe control, and potential for more capable interference mitigation techniques. In applications typically addressed with large single dishes, this trend manifests as growing interest in FPAs, the emergence of PAFs, and the transition to LNSD architectures with larger "*N*" and smaller "*D*." Extrapolating to future decades, it seems likely that in many applications, dishes may be replaced by "aperture arrays," that is, compact planar arrays of low-gain elements with per-element digitization for maximum flexibility in subsequent processing.

At the low-frequency end of the radio spectrum, this architectural approach has already both arrived and matured, as demonstrated by operational instruments such as LWA1 (section "Long Wavelength Array Station 1 (LWA1)") and LOFAR (section "LOw Frequency ARray (LOFAR)") and at 327 MHz by the Deuterium Array experiment (section "Deuterium Array"). Realization of this paradigm is more difficult at higher frequencies, both because the collecting area per element is approximately proportional to λ^2 and because the sensitivity of instruments becomes inevitably internal noise limited above about 400 MHz where the antenna temperature due to sky noise drops below 50 K or so. Thus, the case for new aperture array instruments is most compelling at meter wavelengths, extending into the centimeter-wavelength regime in certain cases. Prominent among experiments exploring this approach is EMBRACE (Kant et al. 2011), a beamforming array of 10,368 dual-polarized Vivaldi-type antenna elements for 500–1,500 MHz collocated with the WSRT, with a smaller version located at Nançay, France.

Square Kilometre Array

Figuring prominently in the future of radio astronomy is the Square Kilometre Array (Dewdney et al. 2009). The SKA project emerged in the 1990s as a concept for a new interferometer that would have an effective aperture of 1 km^2 at 1.4 GHz – hence the name. SKA is now an "umbrella" project coordinating multiple national and international instrument development programs. "SKA Phase 1" comprises three instruments: (1) SKA Phase 1 Low, a new meter-wavelength array architecturally similar to LOFAR, consisting of 900 stations in Western Australia, each consisting of 300 dual-polarized antennas of the log-periodic yagi type, (2) An expansion of ASKAP from 36 to 96 dishes, and (3) an

expansion of MeerKAT from 64 to 254 dishes. SKA Phase 2 would follow and is presumed to include a large meter-/centimeter-wavelength aperture array along the lines described above.

Acknowledgments

The author acknowledges the helpful comments and suggestions of R.M. Prestage (NRAO), G.B. Taylor (U. New Mexico), and K.F. Warnick (Brigham Young U.).

Cross-References

▶ Applications of Phased Array Feeders in Reflector Antennas
▶ Millimeter-Wave Antennas and Arrays
▶ Multibeam Antenna Arrays
▶ Phased Arrays
▶ Radio Frequency Beamforming for Scanned and Multibeam Antenna Systems
▶ Reflector Antennas

References

Baars JWM et al (1973) The synthesis radio telescope at Westerbork. Proc IEEE 61:1258–1266
Baars JWM et al (1994) Design parameters and measured performance of the IRAM 30-m millimeter radio telescope. Proc IEEE 82:687–696
Baars JWM et al (2009) Radio astronomy in the early twenty-first century. Proc IEEE 97:1377–1381
Bandura K et al (2014) Canadian hydrogen intensity mapping experiment (CHIME) pathfinder. arXiv:1406.2288 [astro-ph.IM]
Blundell R (2007) The submillimeter array. In: IEEE international microwave symposium, Honolulu, pp 1857–1860
Braude SY et al (1978) Decametric survey of discrete sources in the northern sky: I. The UTR-2 radio telescope. Experimental techniques and data processing. Astrophys Space Sci 54:3–36
van Cappellen WA, Bakker L (2010) APERTIF: phased array feeds for the Westerbork synthesis radio telescope. In: IEEE international symposium on phased array systems and technology, Boston, pp 640–647
Carlstrom JE et al (2011) The 10 meter south pole telescope. Publ Astron Soc Pac 123:568–581
Davidson DB, Jonas J (2012) An overview of electromagnetics, signal processing, imaging and calibration for MeerKAT. In: International conference on electromagnetics in advanced applications, Cape Town, pp 308–311
Dewdney PE et al (2009) The square kilometre array. Proc IEEE 97:1482–1496
Doyle D et al (2009) The Hershel and Planck space telescopes. Proc IEEE 97:1403–1411
Ekers R, Wilson TL (2013) Radio telescopes. In: Planets, stars and stellar systems, vol 1. Springer, Dordrecht, pp 315–359
Ellingson SW (2005) Antennas for the next generation of low frequency radio telescopes. IEEE Trans Antennas Propag 53:2480–2489
Ellingson SW (2011) Sensitivity of antenna arrays for long-wavelength radio astronomy. IEEE Trans Antennas Propag 59:1855–1863

DOI 10.1007/978-981-4560-75-7_124-1

Ellingson SW, Cazemier W (2003) Efficient multibeam synthesis with interference nulling for large arrays. IEEE Trans Antennas Propag 51:503–511
Ellingson SW et al (2013a) The LWA1 radio telescope. IEEE Trans Antennas Propag 61:2540–2549
Ellingson SW et al (2013b) Design and commissioning of the LWA1 radio telescope. In: IEEE international symposium on phased array systems and technology, Boston, pp 776–783
Ellingson SW et al. (2013c) Strut-straddling arrays for the VLA 4-meter observing system. EVLA Memo 173. http://www.aoc.nrao.edu/evla/memolist.shtml. Accessed 6 Dec 2014
Elmer M et al (2012) Beamformer design methods for radio astronomical phased array feeds. IEEE Trans Antennas Propag 60:903–914
van Haarlem MP et al (2013) LOFAR: the low-frequency array. Astron Astrophys 556:A2
Hachenberg O et al (1973) The 100-meter radio telescope at Effelsberg. Proc IEEE 61:1288–1295
Hanany S et al (2013) CMB telescopes and optical systems. In: Planets, stars and stellar systems, vol 1. Springer, Dordrecht, pp 431–480
Harun M, Ellingson SW (2011) Design and analysis of low frequency strut-straddling feed arrays for EVLA reflector antennas. Radio Sci 46:RS0M04. doi:10.1029/2011RS004710
Hay S et al (2011) Connected patch array analysis using the characteristic basis function method. IEEE Trans Antennas Propag 59:1828–1837
Hicks BC et al (2012) A wide-band, active antenna system for long wavelength radio astronomy. Publ Astron Soc Pac 124:1090–1104
von Hoerner S (1967) Design of large steerable antennas. Astron J 72:35–47
Jeffs BD et al (2008) Signal processing for phased array feeds in radio astronomical telescopes. J Sel Top Sig Process 2:635–646
Kant G et al (2011) EMBRACE: a multi-beam 20,000-element radio astronomical phased array antenna demonstrator. IEEE Trans Antennas Propag 59:1990–2003
Kassim NE et al (2007) The 74 MHz system on the very large array. Astrophys J Suppl Ser 172:686–719
Kildal P-S et al (1994) The Arecibo upgrading: electrical design and expected performance of the dual-reflector feed system. Proc IEEE 82:714–724
Landon J et al (2010) Phased array feed calibration, beamforming, and imaging. Astron J 139:1154–1167
Mangum JG et al (2006) Evaluation of the ALMA prototype antennas. Publ Astron Soc Pac 118:1257–1301
Morris D (1978) Chromatism in radio telescopes due to blocking and feed scattering. Astron Astrophys 67:221–228
Nan R (2011) The five-hundred-meter aperture spherical radio telescope (FAST) project. Int J Mod Phys D 20:989–1024
Napier P et al (1983) The very large array: design and performance of a modern synthesis radio telescope. Proc IEEE 71:1295–1320
Page L et al (2003) The optical design and characterization of the microwave anisotropy probe. Astrophys J 585:566–586
Parsons AR et al (2010) The precision array for probing the epoch of reionization: 8 station results. Astron J 139:1468–1480
Patra N et al (2013) SARAS: a precision system for measurement of the cosmic radio background and signatures from the epoch of reionization. Exp Astron 36:319–370
Perley R et al (2009) The expanded very large array. Proc IEEE 97:1448–1462
Phillips TG et al (2013) Submillimeter telescopes. In: Planets, stars and stellar systems, vol 1. Springer, Dordrecht, pp 283–313
Pilbratt GL et al (2010) Herschel space observatory: an ESA facility for far-infrared and submillimetre astronomy. Astron Astrophys 518:1–6

DOI 10.1007/978-981-4560-75-7_124-1

Prestage RM et al (2009) The green bank telescope. Proc IEEE 97:1382–1389
Raza J et al (2002) Spatial filtering of RF interference in radio astronomy. IEEE Signal Process Lett 9:64–67
Rogers AEE, Bowman JD (2012) Absolute calibration of a wideband antenna and spectrometer for accurate sky noise temperature measurements. Radio Sci 47:RS0K06 doi:10.1029/2011RS004962
Rogers AEE et al (2007) Observations of the 327 MHz deuterium hyperfine transition. Astron J 133:1625–1632
Ruze J (1966) Antenna tolerance theory – a review. Proc IEEE 54:633–640
Salter C (2012) An astronomer's guide to the Arecibo 305-m telescope [online]. http://www.naic.edu/~astro/guide/. Accessed 6 Dec 2014
Schinckel AET, Bunton JD, Cornwell TJ, Feain I, Hay SG (2012) The Australian SKA Pathfinder, Proc. of SPIE, Amsterdam, Netherlands. Vol. 8444:2A
Swetz DS (2011) Overview of the Atacama cosmology telescope: receiver, instrumentation, and telescope systems. Astrophys J 194:41
Tauber JA et al (2010) Planck pre-launch status: the optical system. Astron Astrophys 520:A2
Thompson AR et al (2001) Interferometry and synthesis in radio astronomy, 2nd edn. Wiley, New York
Tingay SJ et al (2013) The Murchison widefield array: the square kilometre array precursor at low radio frequencies. Publ Astron Soc Aust 30:e007
Warnick KF et al (2011) Design and characterization of an active impedance matched low noise phased array feed. IEEE Trans Antennas Propag 59:1876–1885
Welch J et al (2009) The allen telescope array: the first widefield, panchromatic, snapshot radio camera for radio astronomy and SETI. Proc IEEE 97:1438–1447
Wilson GW et al (2008) The AzTEC mm-wavlength camera. Mon Not R Astron Soc 386:807–818
Wilson WE et al (2011) The Australia telescope compact array broadband backend (CABB): description & first results. Mon Not R Astron Soc 416:832–856
Wilson T et al (2013) Tools of radio astronomy, 6th edn. Springer, Berlin
Wooten A, Thompson AR (2009) The Atacama large millimeter/submillimeter array. Proc IEEE 97:1463–1471

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_125-1

Radiometer Antennas

Markus Peichl*
DLR, Microwaves and Radar Institute, Oberpfaffenhofen, Germany

Abstract

Microwave radiometry is concerned with purely passive sensing of naturally generated microwave radiation of thermal origin. Microwave radiometers are corresponding measuring devices typically designed and built as a very low-noise receiver followed by a signal recording unit. Usually, radiometers contain an antenna as the first reception component collecting the incoming radiation, and they measure radiation power expressed in an apparent temperature called brightness temperature. The observable brightness temperature of any object or surface depends on various chemical and physical quantities, whose concurrence is expressed by the objects' emission (absorption), reflection, and transmission properties and its true temperature. Since the Earth has a temperature typically close to 300 K and the universe close to 3 K, a nearly arbitrary mixture of these two extreme temperatures can be expected. Consequently, our environment can show quite different brightness temperature values depending on the direction of actual observation.

On the one hand, radiometer measurements are carried out stationary with respect to the antenna pointing direction in order to observe time-dependent variations of the brightness temperature. On the other hand, the brightness temperature of a whole scene is scanned in order to acquire locally changing one- or two-dimensional profiles, while the latter ones are assembled as a two-dimensional image comparable to a conventional photograph. Depending on the specific application, various antenna types are considered, where usually hard requirements with respect to beam width, side-lobe level, scan capability, and losses have to be addressed (► Transmission Lines). Radiometric measurements are performed for Earth or planetary observation in space (► Space Antennas including Terahertz Antennas), from aircraft platforms on the Earth's surface and the atmosphere, or on the ground, either sensing the environment or sensing the universe, the latter being performed in radio astronomy (► Antennas in Radio Telescope Systems). Usually, the brightness temperature is rarely used as the physical quantity of interest. More often, it is transferred via adequate physical models to other secondary or third quantities for more direct use in the case of Earth observation (e.g., soil moisture, ocean salinity, rain rate, snow cover, etc.), being performed already since the 1950s of the last century. However, in the last decades, microwave radiometry is as well used in many safety- and security-related applications, for which often only sufficient temperature contrast between an object and its surrounding is required besides spatial resolution for detection and recognition purposes.

In this chapter relevant fundamentals of microwave radiometry are outlined for better understanding of antenna requirements, followed by an overview of typical types of radiometer antenna systems. Some existing antenna systems are discussed in order to illustrate the variability with respect to applications. A section on basic antenna quantities addresses key figures for practical design and verification and illustrates the results exemplarily for selected cases. Finally, a brief summary and an outlook on possible future implementations and other frequency ranges are given.

*Email: Markus.peichl@dlr.de

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_125-1

Keywords

Microwave radiometry; Earth observation; Security; Antenna temperature; Antenna losses; Antenna reflection; Mechanical scanner; Focal plane array; Aperture synthesis; Frequency scanning; Digital beamforming; Sparse array; Spatial resolution; Sensitivity; Sampling

Introduction

Microwave radiometry addresses the domain of passive measurement of the natural thermally caused electromagnetic radiation of matter at a thermal temperature above 0 K. The thermal or true temperature of an object is usually called physical temperature in order to discriminate it from an apparent temperature, the latter being observable in microwave radiometry or even infrared measurements. This apparent temperature is called the brightness temperature. In the case of Earth observation, significant contrasts can be observed between reflective and absorbing materials due to the impact of reflected sky radiation of cosmic origin (► Radio Wave Propagation Models). Thus, an approximate brightness temperature range from 3 K to more than 300 K can be observed. In the microwave region, the spatial two-dimensional brightness temperature distribution can be used as a daytime and often weather-independent indicator for many different physical phenomena. Hence, interesting application areas incorporate geoscience, climatology, agriculture, pollution and disaster control, detection, reconnaissance, surveillance, and status registration in general. Since little more than a decade, safety and security applications like personnel screening and the monitoring of critical infrastructures are also of major interest. General reading on microwave radiometry and its applications can be found in Ulaby et al. (1981), Janssen (1993), Mätzler (2006), NAS (2007), and Peichl et al. (2007).

Fundamental Quantities

The origin of thermal radiation is given by quantum-mechanical exchange of energy with matter, i.e., in the case of radiation-based interaction, the energy can be either absorbed or emitted in very small but fixed portions. The size of those energy portions is exactly defined by the differences of single-energy levels of excitation states all atoms or molecules of matter can adopt. Since many different energy levels for single atoms or molecules exist and an object can be composed of many different materials, and since the event of a transition from one energy level to another occurs statistically in time, the character of the whole emitted radiation of an object or piece of matter is described by broadband noise. In the same way absorption of radiation energy occurs. Consequently, the capability of a certain matter to emit or absorb radiation energy is identical. These capabilities are denoted and quantified by the quantities of emissivity e and absorptivity a.

For a blackbody, the power distribution of thermally based noise radiation as a function of frequency and physical temperature is given by Planck's law (Ulaby et al. 1982). A blackbody is a hypothetical piece of matter which absorbs all received radiation totally and consequently emits the same amount in the case of thermodynamic equilibrium. So if the total energy exchange between the blackbody and the surrounding is 100 %, or equivalently denoted as 1, the relation $e = a = 1$ is valid. In nature and even our man-made world, a blackbody does not exist, but it can be sufficiently approximated by specific material compositions, as in the case of a microwave absorber used, for instance, in anechoic chambers. For that reason natural objects are often called gray bodies, meaning that their emissivity and absorptivity in fact are equal but lower than the value 1. The residual amount of interacting energy is related to reflected radiation on the surface of the gray body and transmitted radiation through the body. Due to the requirement of energy

DOI 10.1007/978-981-4560-75-7_125-1

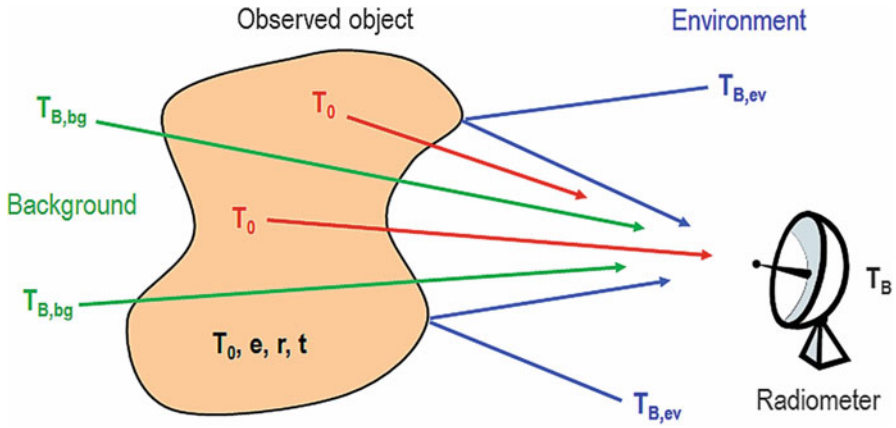


Fig. 1 Contributions to the observed brightness temperature. The object has a physical temperature T_0 and its interaction capability with radiation is described by emissivity e, reflectivity r, and transmissivity t. The radiation is received by the antenna of a radiometer

conservation and by introducing reflectivity *r* and transmissivity *t*, the following relations are valid for natural objects in thermodynamic equilibrium:

$$e + r + t = 1, \quad e, r, t \leq 1, \quad e = a. \tag{1}$$

In the microwave region, Planck's law can be approximated by the Rayleigh-Jeans law for frequencies up to 300 GHz and physical temperatures down to 1 K as given next:

$$B_f = \frac{2 k_B T f^2}{c^2}. \tag{2}$$

The quantity B_f is the so-called brightness (formerly radiation intensity), given in units of watts per steradian and hertz, which is the observed radiation power related to an infinitesimal segment of solid angle and an infinitesimal increment of bandwidth. The quantities T and f are the physical temperature in kelvins and the frequency in hertz; $c \approx 3 \times 10^8$ m/s is the speed of light and $k_B \approx 1.38 \times 10^{-23}$ J/K is Boltzmann's constant. The important fact in Eq. 2 is the linear dependence of the brightness from the physical temperature, which directly relates observed radiation power to temperature at a certain frequency in a simple way. This relationship can be as well applied to radiation originating from a gray body, and in this case, the given temperature is the brightness temperature and in general not identical to the physical temperature. The charm of using that relationship is that, if one uses a radiometer system in an outside area looking at a certain object and being calibrated for brightness temperatures, a low temperature value can indicate a highly reflective material and a high value can indicate a highly absorptive material, depending of course on the direction of observation with respect to sky or space.

Since we can use now the brightness temperature instead of radiation power to describe a certain observation, the brightness temperature can be considered as a composition of several contributions, as shown in Fig. 1, being expressed in a simple notation as

$$T_B = e T_0 + r T_{B,ev} + t T_{B,bg}. \tag{3}$$

The first summand represents the thermal energy content of the object, determined by its physical temperature T_0. The second contribution is the portion of reflected power from the ambience being represented by the brightness temperature $T_{B,ev}$. The third part is the radiation originating from directions

DOI 10.1007/978-981-4560-75-7_125-1

behind the object being able to penetrate it, and $T_{B,bg}$ is the corresponding brightness temperature of this background.

Note that $T_{B,ev}$ and $T_{B,bg}$ for their part can be generated as well according to Eq. 3. Furthermore (3) usually has to be an integral equation summing up all contributions from those directions, which the antenna can receive and which are deflected into that particular direction range. The material properties e, r, and t can depend strongly on frequency, observation angle, and polarization, and so does the observed brightness temperature T_B (▶ Radio Wave Propagation Models).

The equivalence of radiation power and temperature is expressed as well by Nyquist's law for ohmic resistors (Ulaby 1981):

$$P_N = k_B \Delta f T, \tag{4}$$

where k_b again is Boltzmann's constant, Δf is the noise bandwidth of observation, and T is the physical temperature. Since the noise being produced by an ohmic resistor cannot be discriminated from noise provided by a receiving antenna, the latter one assumed for the moment to have no losses, the temperature in Eq. 4 can be replaced by the brightness temperature. In case of radiometric measurements, the range of observable brightness temperatures is roughly in the order of 3–300 K for sensing in space or on Earth, and a typical number for the receiver bandwidth is 100 MHz. Using those values, the expected power range for natural brightness temperature observations is in the range of 10^{-15}–10^{-13} W, which is very small and in the order of the noise power the receivers generate for their own.

Sensing spatially the brightness temperature distribution requires an antenna and a moving sufficiently focused antenna beam to collect the radiation from a certain direction. The observed brightness temperature T_B is called antenna temperature T_A and can be expressed in a spherical coordinate system, using Θ and Φ as elevation and azimuth angles covering the ranges of $0 \le \Theta \le \pi$ and $0 \le \Phi \le 2\pi$:

$$T_A(\Theta', \Phi') = \frac{1}{\Omega_A} \int\limits_{\substack{antenna \\ pattern}} T_B(\Theta, \Phi)\ F_{Pn}(\Theta - \Theta', \Phi - \Phi')\ \sin(\Theta - \Theta')\ d\Theta\ d\Phi, \tag{5a}$$

$$\Omega_A = \int\limits_{\substack{antenna \\ pattern}} F_{Pn}(\Theta, \Phi)\ \sin\Theta\ d\Theta\ d\Phi. \tag{5b}$$

The quantity F_{Pn} denotes the normalized power pattern of the antenna and Ω_A is her effective solid angle from which the antenna would receive if the pattern had constant amplitude. The term $\sin(\Theta)\, d\Theta\, d\Phi$ of course represents the infinitesimal solid angle for the integration. The moving beam is represented by primed coordinates, Θ′ and Φ′, indicating the actual pointing direction of the antenna. Following Eq. 5a, it can be recognized that mathematically the antenna temperature results from a correlation or special convolution (i.e., exchange of $x - x'$ by $x' - x$) of the scene-dependent brightness temperature distribution with the antenna power pattern. The normalizing character of solid angle Ω_A becomes clear when a constant brightness temperature distribution is scanned by the antenna beam, i.e., $T_B(\Theta,\Phi) = T_0 = \text{constant}$. In that case the result for the antenna temperature is always $T_A = T_B = T_0$, and the antenna pattern has no blurring impact on the true brightness temperature distribution. In all other cases of a spatially varying distribution, the antenna pattern determines the obtainable spatial resolution

DOI 10.1007/978-981-4560-75-7_125-1

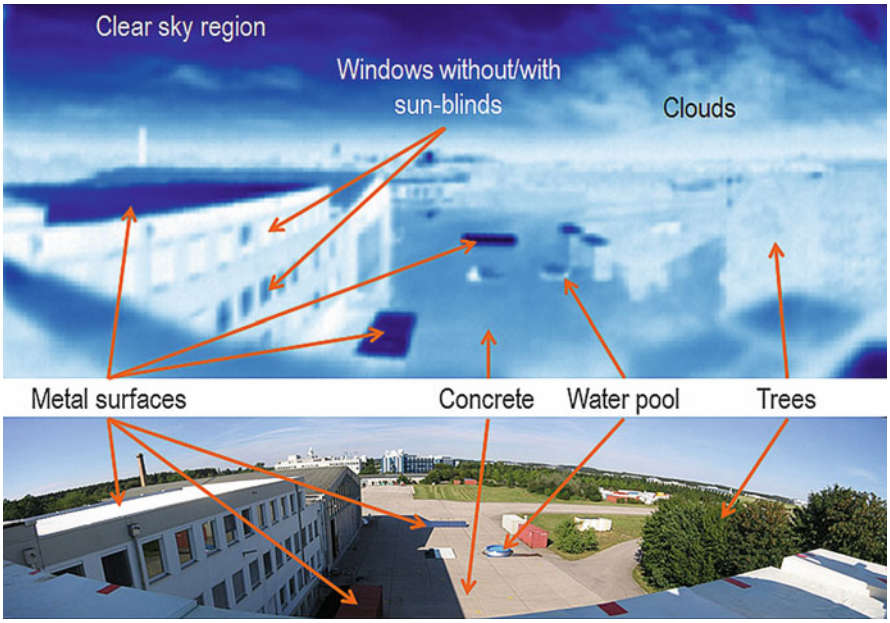


Fig. 2 Radiometer image (*top*) and photograph (*bottom*) of a complex scene for surveillance purposes. The radiometer image was recorded around 90 GHz using a mechanically scanning offset parabola antenna. Both images are geometrically not adapted. *Black* or *dark blue* color indicates low and *white* or *bright blue* indicates high brightness temperature (Peichl et al. 2007)

and contrast of the scanning measurement, as discussed later in section Basic Quantities of Radiometer Antennas.

Application Examples

At this point some practical examples of two-dimensional imaging shall be illustrated and briefly discussed for a few applications, where quite different antenna types have been used. Figure 2 shows an image of a ground-based measurement for surveillance applications.

First of all, the quasi-optical appearance of the radiometric millimeter-wave image and its complexity of details can be observed. Some special objects or materials are indicated in order to illustrate the brightness temperature variance of radiometric Earth observation. Note especially the radiometric "cold" appearance of the sky region and metallic objects reflecting sky radiation and the rather "warm" appearance of trees, the horizon region, and the vertical faces of buildings reflecting ambient radiation from horizon direction.

A second example in Fig. 3 shows a global map of brightness temperature measurements acquired by the SMOS (Soil Moisture and Ocean Salinity) satellite mission launched in November 2009. The SMOS instrument is an L-band two-dimensional aperture synthesis radiometer system using sparse interferometric array technology for imaging (► Space Antennas including Terahertz Antennas), as pointed out in a later section. The antenna array forms a "Y" shape and uses patch antennas (► Microstrip Patch Antennas) for the single elements (ESA 2009).

A last example Fig. 4 illustrates a two-dimensional map of buried objects, i.e., landmine simulants and other targets, measured by a multispectral radiometer. The radiometer uses a broadband Vivaldi antenna (► Ultra-Wideband Antennas) and is scanning a scene by two-dimensional raster motion close to the ground in the frequency range of about 1.5–7 GHz, divided in many different bands of 50 MHz bandwidth (Peichl et al. 2007). Note that the antenna was operated in the near field where usually the antenna pattern is different from far-field conditions.

DOI 10.1007/978-981-4560-75-7_125-1

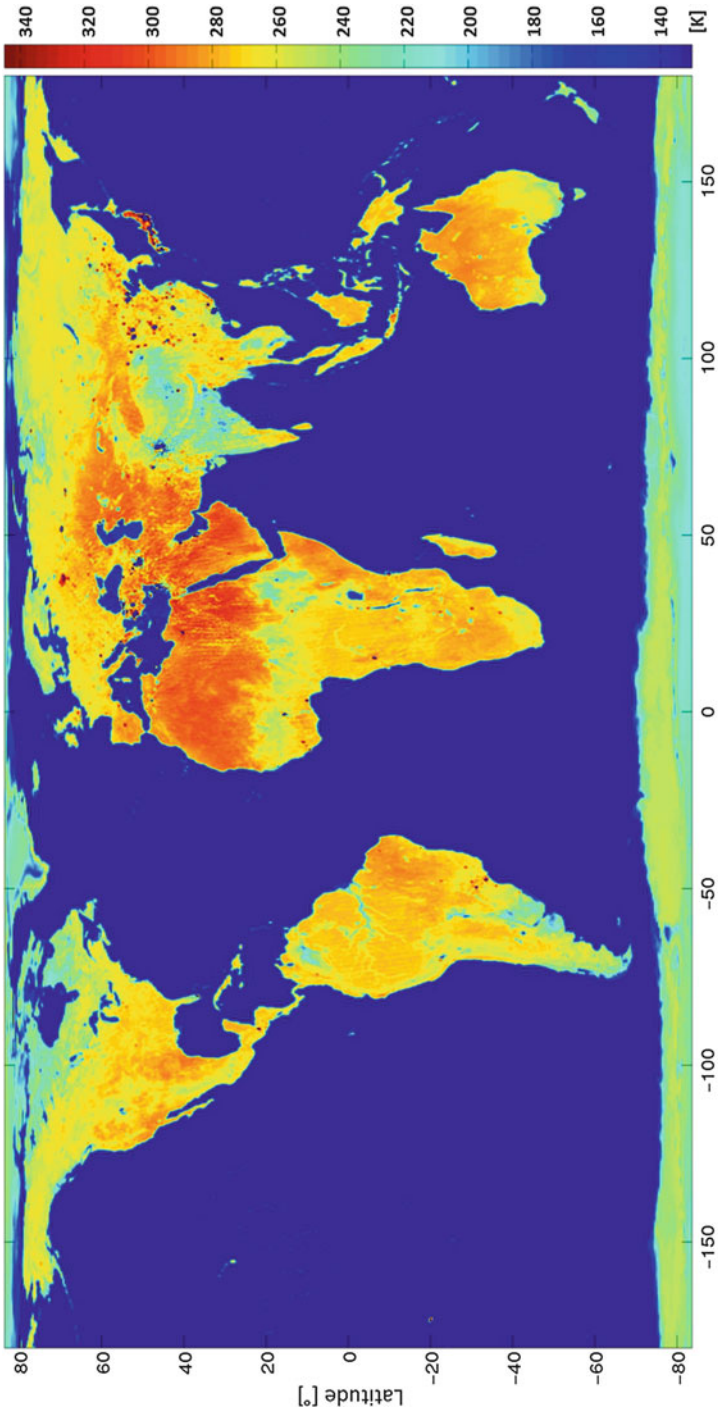


Fig. 3 Global mean brightness temperature map at L band (1.413 GHz center frequency) and vertical polarization acquired by the SMOS satellite. Data have been processed for 29 July to 24 August 2014 for an incidence angle of 42.5° (Reproduced courtesy of Kerr/Albitar from CESBIO/CATDS, France)

DOI 10.1007/978-981-4560-75-7_125-1

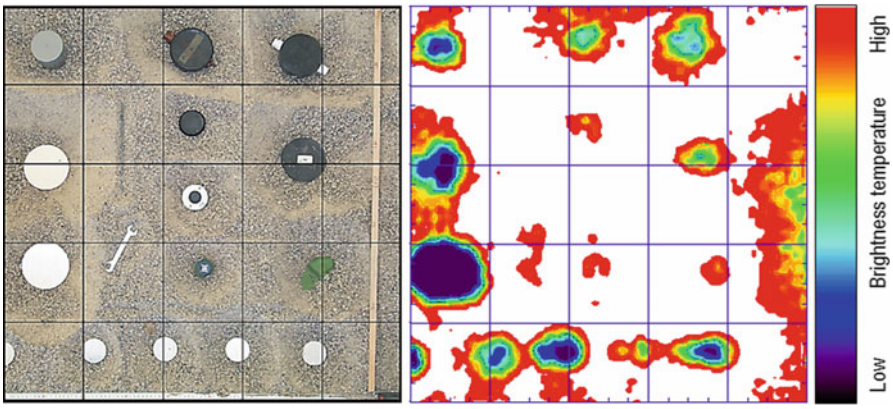


Fig. 4 Photograph (*left*) and brightness temperature image (*right*) of a buried object scene. The objects visible on the photograph (landmine simulants, metal disks, water can) have been buried in the sand in depths between 1 cm and 15 cm. The *right image* is a superposition of several single images at different center frequencies between 2 GHz and 6 GHz

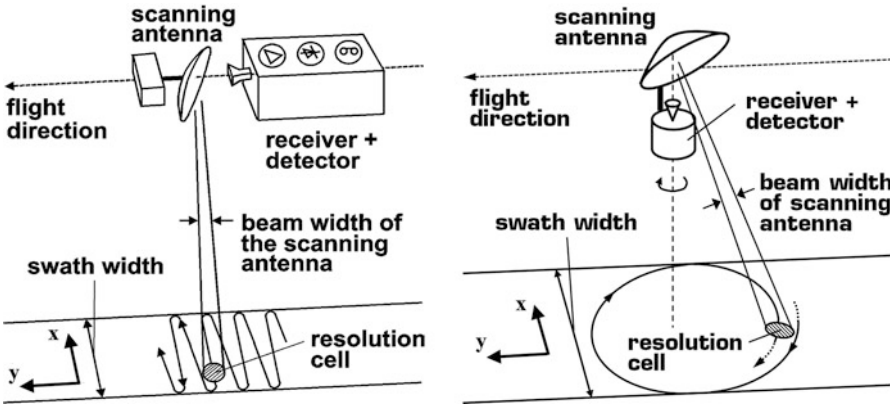


Fig. 5 Principles of fully mechanical scanning of the antenna beam for an airborne situation. In the example (*left*), the reflector of an offset parabola antenna and hence the antenna beam is swinging back and forth across flight direction. In the example (*right*), an offset reflector and hence the antenna beam perform a conical scan around the vertical axis. In both cases, a moving antenna footprint on the ground is produced in conjunction with the platform motion

Already these few examples have shown a high variability of applications and antenna types to be used for microwave radiometry. However, there is much more which cannot be covered or cited in this section, but fundamental insight in basic radiometric quantities and how they relate to the antenna or antenna system have been given. Next, a closer and more formal overview on different antenna types used in microwave radiometry is addressed.

Overview on Radiometer Antennas

Basically, each antenna type could be used for radiometric applications. However, in most cases, the choice of antenna type is driven by the peculiarity of the application and the basic requirement that the antenna should have very low losses for its own (► Transmission Lines). The latter condition again is based on the fact that the observable brightness temperature is in the order of the noise of the receiving system, in which the antenna is the first and radiometric most influential component. Since most applications in microwave radiometry deal with imaging or at least one-dimensional profiling of the

DOI 10.1007/978-981-4560-75-7_125-1

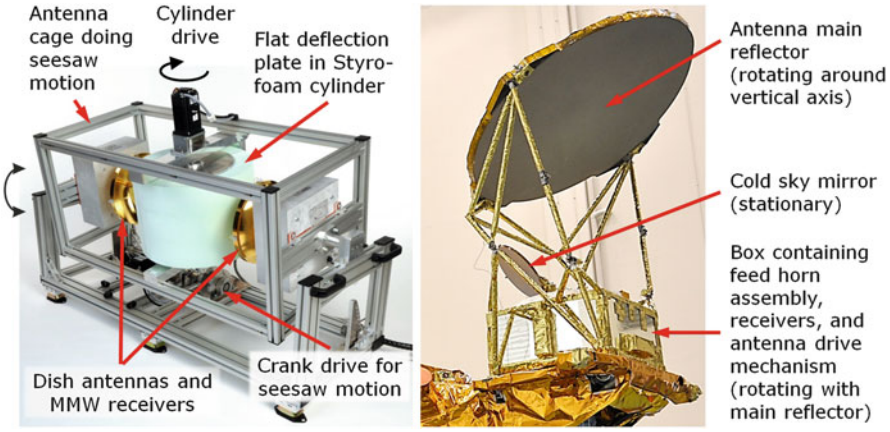


Fig. 6 Examples for fully mechanical scanning radiometer systems. The *left* system SUMIRAD is a vehicle-based imager for security applications like threat detection. The *right* scanner AMSR2 is a spaceborne imager for Earth observation and climate change monitoring (*Right* picture reproduced courtesy of JAXA, Japan)

environment, the intrinsic antenna usually is part of a scanning mechanism being either mechanic, electronic, or a hybrid of both. Consequently rather an antenna system has to be considered for performance characterization instead of the pure antenna element. Next, the most useful and applied scanning mechanisms for antenna beams are briefly discussed.

Mechanical Scanner

The operational principle for mechanical scanning of an antenna beam across a certain scene is illustrated in Fig. 5. In most cases, such a system is using a reflector antenna type realized as direct-fed parabola, offset parabola, or Cassegrain- or Gregorian-type antenna system (▶ Reflector Antennas). The focused beam is either moved by rotating or swinging back and forth the main or subreflector while keeping the feed antenna fixed or an additional flat metallic plate deflects the focused beam by a similar motion. This scanning principle can be applied in either two or only in one dimension. In the latter case the second beam motion is accomplished by use of the platform motion from an aircraft or satellite, for instance. Note that for the seesaw scanning shown left, the incidence angle varies for each sample, while for the conical scan on the right, the incidence angle is always constant.

Figure 6 illustrates two examples of existing fully mechanical scanners. The left system Surveillance in an Urban environment using Mobile sensors Imaging RADiometer (SUMIRAD) (Peichl et al. 2013) is intended for fast imaging of a scene in front of a vehicle during a drive. Two receivers are connected to Cassegrain antennas pointing toward a rotating metallic deflection plate, hence providing the horizontal scan of the antenna beam. The deflection plate is encased by a cylindrical structure made of Styrofoam in order to reduce considerably aerodynamic drag during rotation (up to 2,400 rpm). In vertical direction the whole unit of antennas, receivers, and deflection plate is forced to seesaw motion by a crank drive, providing the vertical scan of the antenna beam. The system works at a center frequency of 90 GHz and a bandwidth of 20 GHz. The angular resolution is about 0.75° (30 cm antenna diameter) and the temperature resolution (i.e., the sensitivity) is <2 K for a frame rate of up to 1 image per second. The system generates simultaneously two images of the identical scene for a field of view of $30° \times 80°$ (elevation $\times$ azimuth), and each linear polarization state can be chosen individually.

The right system AMSR2 (Advanced Microwave Scanning Radiometer 2) is part of GCOM (Global Change Observation Mission) intended for long-term global observation of various geophysical parameters for understanding climate variability and water cycle (Website JAXA; Imaoka et al. 2010). An assembly of feed horns is located in the focal plane of an offset parabola antenna providing tilted beams

DOI 10.1007/978-981-4560-75-7_125-1

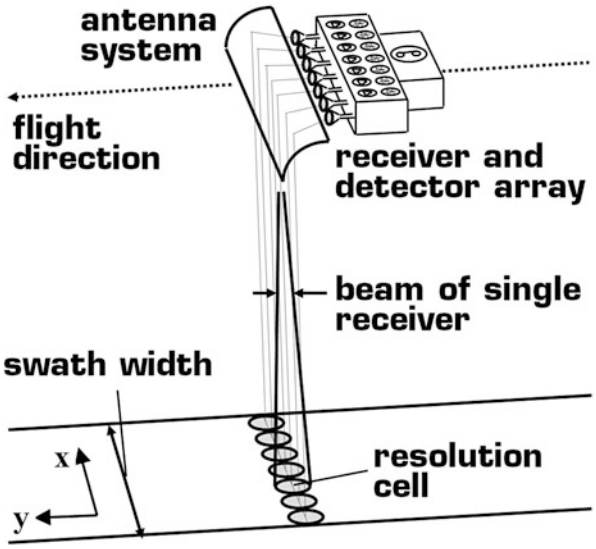


Fig. 7 Principle of focal plane array scanning for a one-dimensional case. Several feed antennas are arranged in the focal plane of a double-curved reflector in order to produce single beams across flight direction. In the illustrated case the second image dimension along flight direction is obtained again by platform motion

with respect to the vertical axis (► Space Antennas including Terahertz Antennas). By rotating the whole setup at 40 rpm (but not the whole satellite), a conical scan at around 47.5° with respect to nadir line is performed simultaneously for all frequency channels, and the satellite motion provides the second image dimension. The system operates at six frequency bands from about 7 GHz to 89 GHz center frequency providing an angular resolution between 1.8°–0.15° (2 m antenna diameter). The corresponding temperature resolution is within 0.4–1.4 K and two linear polarizations are provided. The satellite altitude is around 700 km and the swath width is about 1,450 km. AMSR2 was launched in May 2012.

Of course, many implementations of the mechanical scan can be realized, but the basic procedure of moving any reflecting device mechanically is always required. Consequently this approach has limitations with respect to antenna size and mass inertia when a very large antenna has to be moved very fast to achieve high-resolution and/or real-time imaging capabilities. However, mechanical scanning systems can be realized with relatively low expense and low losses of the antenna system. In addition, calibration of such systems can be performed relatively simply by applying a so-called hot-cold method, where two well-known reference targets on a known and well-separated brightness temperature are used. Finally, it should be noted that, by the mechanical scanning procedure, the antenna beam is always generated instantaneously in angular space. Thus, these antenna systems are called real-aperture antenna systems and they perform real-aperture imaging.

Focal Plane Array

Focal plane arrays (FPAs) are already known for a longer time in optical imaging systems like modern digital cameras (► Multi-Beam Antenna Arrays, Radio-Frequency Beam-Forming for Scanned and Multi-Beam Antenna Systems). The basic principle is to locate an array of small detectors or antennas in the focal plane of a much larger focusing device. The latter usually is a suitable dielectric lens (► Dielectric Lens Antennas) or any curved metallic reflector producing collimated radiation. The situation of a one-dimensional FPA for again airborne radiometric imaging is shown in Fig. 7. For a two-dimensional FPA, the one-dimensional row of N antennas has to be repeated M times to get an image of N × M pixels, where of course an adequate focusing reflector has to be designed to point the single

DOI 10.1007/978-981-4560-75-7_125-1

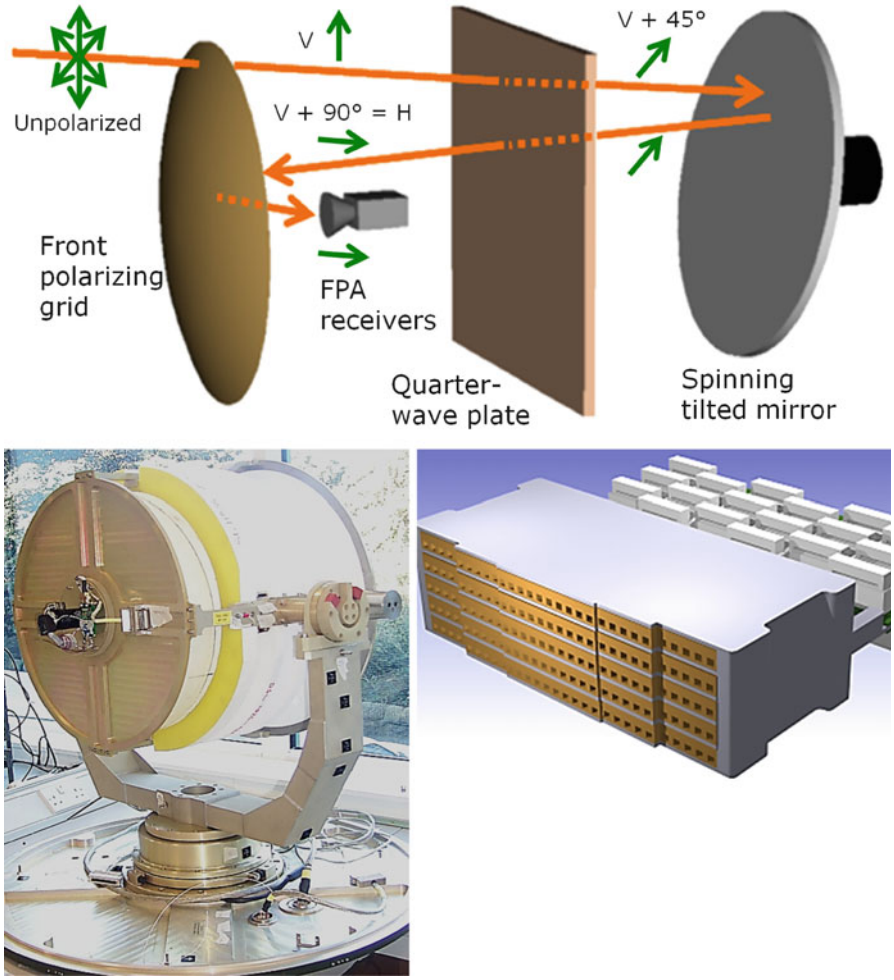


Fig. 8 Scanning scheme (*top*) and photograph (*bottom left*) of the QinetiQ PMMW real-time imager combining FPA technology and mechanical scanning for the antenna system. In the *lower-right* picture, the FPA block including feed antennas and receivers is shown (Reproduced courtesy of QinetiQ, UK)

beams in both directions. FPA technology-based radiometer systems are as well called push-broom systems in Earth observation (Skou and Le Vine 2006).

Figure 8 shows an example of a radiometric FPA system intended for security applications like enhanced vision for aircraft or concealed object detection under clothing or other dielectric obstacles (Appleby et al. 2004). In fact the scanning approach is a hybrid of FPA techniques and mechanical scanning in order to limit the required number of single receivers. The scan is performed by feeding unpolarized radiation from the scene through a polarizing grid to extract only vertical polarization (V), for instance. Then, the radiation passes a quarter-wave plate which rotates the linear polarization by 45° (V + 45°) before it is reflected by a metallic spinning mirror. That mirror is additionally tilted in order to perform a conical scan when rotated. After reflection, the radiation passes again the quarter-wave plate and the polarization is rotated by another 45° in the same direction as before, thus resulting in horizontal polarization (V + 90° = H). The H-polarized radiation now is fully reflected by the curved polarizer grid (transparent only for V polarization) and focused at the focal plane where the FPA is located. Since each element of the FPA produces its own conical scan displaced slightly from all others, a complete scene is scanned during each 360° rotation of the spinning mirror.

DOI 10.1007/978-981-4560-75-7_125-1

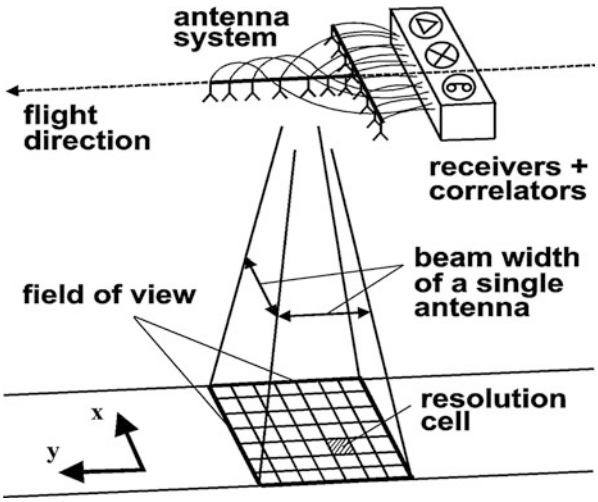


Fig. 9 Imaging principle for two-dimensional aperture synthesis. Each antenna/receiver pair of a sparse array forms an individual two-element interferometer measuring a single spatial frequency. The image is formed by processing the signals from all spatial frequencies. Note that for the shown configuration, a two-dimensional image is already obtained without additional platform motion. Aperture synthesis can be performed as well in only one dimension

The QinetiQ Passive MilliMeter-Wave (PMMW) system works at a center frequency of 94 GHz and a bandwidth of 20 GHz. The angular resolution is about 0.47° (60 cm antenna diameter) and the temperature resolution (i.e., the sensitivity) is about 2 K for a frame rate of up to 25 images per second. The field of view is 30° × 60° (elevation × azimuth), and the FPA consists of 150 single receivers, each one using a small rectangular horn antenna as feed.

FPA technology using many receiver elements is rather new in passive microwave remote sensing due to a former lack in sufficiently small and cheap microwave integrated circuits (MIC) or even monolithic MICs (MMIC). That instance can be noticed still at present in the realization of hybrids of FPAs and mechanical scanning to avoid large numbers of receivers (e.g., hundreds or even thousands), which is applied as well by Heinz et al. (2014) at submillimeter waves (▶ Mm-Wave Sub-mm-Wave Antenna Measurement). For Earth observation only studies have been carried out to date (Nielsen et al. 1989) and no operational system has been built yet. However, it can be expected that future antenna systems for high frame rate imaging will make more use of FPA technology, when the whole FPA unit can be even more miniaturized into a compact block. It should be noted that the physical distance of the single elements does not allow full Nyquist sampling of the scene. Consequently techniques of lateral adjustment of single FPA rows by half the element distance or crossing and overlapping scan patterns as in the case of the QinetiQ PMMW system are applied. Furthermore, calibration issues are more challenging since all receiver elements have to be adjusted to a common scale. It should be noted that by FPA imaging, all antenna beams are generated instantaneously in angular space. Thus, these antenna systems belong as well to real-aperture antennas and perform real-aperture imaging. Finally, the instance of measuring the brightness temperature of all single-antenna beams in parallel requires no mechanical motion. Hence, large antenna structures and high spatial resolution can be obtained.

Aperture Synthesis

Aperture synthesis is an imaging technique used in radio astronomy (▶ Antennas in Radio Telescope Systems) since the 1950s of the last century (Thompson et al. 2001). The goal of this technique is to make use of redundancy in an antenna aperture by only recording the required spatial frequencies of a scene with minimum replication but following the Nyquist sampling law. The spatial frequency is a measure for

DOI 10.1007/978-981-4560-75-7_125-1

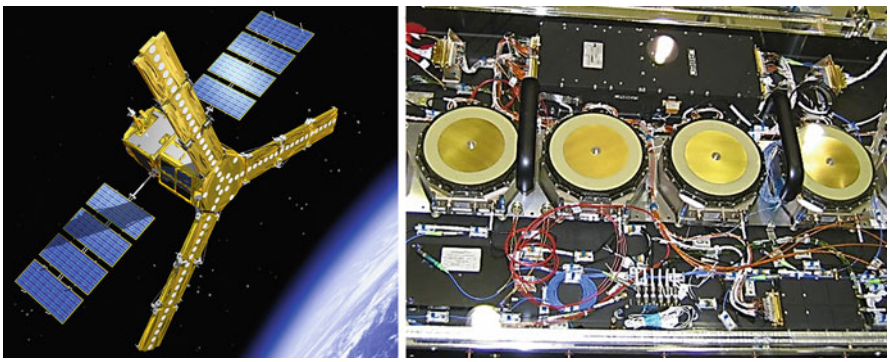


Fig. 10 Artist view of the SMOS instrument in space (*left*) and photograph of four receiver elements and corresponding electronics (*right*) of the real hardware. The single-element antenna type is a dual-polarized aperture-coupled microstrip patch (▶ Microstrip Patch Antennas) of 165 mm diameter with integrated feed network (Reproduced courtesy of CESBIO, France, and EADS CASA Espacio S.L., Spain)

a specific spatial variation of a quantity, e.g., the brightness temperature distribution, so that in general a scene consists of a certain number of differently weighted spatial frequencies, i.e., the spatial frequency spectrum. The measurement of a single spatial frequency is performed by cross-correlating the input signals of two spatially separated antennas, i.e., an interferometer. The spacing between single antennas is called a baseline. The imaging principle of aperture synthesis is illustrated in Fig. 9 again for an airborne situation.

Note that for the shown situation, only a sparse array consisting of two branches populated with antenna elements is necessary instead of a fully filled two-dimensional array. The imaging here is not performed by generating instantaneously the antenna beam through the whole array, but by Fourier inversion of the acquired spatial frequency spectrum for the case of an ideal system. Hence, aperture synthesis performs scanning of a scene in the spatial frequency domain, in contrast to the previously described methods which are scanning in the spatial domain. The sparse array represents a synthetic aperture simulating roughly a real-aperture antenna of similar dimensions. A synthetic antenna beam is only generated by processing which therefore does not exist instantaneously.

As an operational example for aperture synthesis, Fig. 10 shows the Microwave Imaging Radiometer with Aperture Synthesis (MIRAS) instrument of the spaceborne SMOS mission launched in 2009 by European Space Agency (ESA) (▶ Space Antennas including Terahertz Antennas). The SMOS mission is intended to predominantly provide global observations of soil moisture on land and sea surface salinity to improve modeling of weather and climate. It is also used for monitoring vegetation water content, snow cover, and ice structure (Barré et al. 2008; Font et al. 2010).

MIRAS is a Y-shaped aperture synthesis radiometer, the first operational instrument of this imaging type in space. It is operated in a protected band for radio-astronomical observations at a center frequency of 1,413 GHz using a bandwidth of about 20 MHz. Besides the protection, this L-band frequency range is used due to sufficient sensitivity of the brightness temperature with respect to soil moisture and ocean salinity, being harder obtained at higher frequencies. The sparse array consists of 69 receiver elements in total, where 18 elements are located in equal distances of 0.875 wavelengths on each arm of the Y, and the other 15 in the hub. The maximum spacing between two antenna elements is about 7.8 m (arm length 4.5 m) leading to a boresight angular resolution in the order of just under 2°, depending on the weighting function used for the sampled spatial frequency spectrum. The scene-dependent temperature resolution (e.g., for land or ocean) is in the order of about 2–4 K. For a satellite altitude of 758 km, the useful swath on the ground is about 1,050 km. The signal correlation of all antenna pairs is performed by 1-bit digital correlators operated at about 55 MHz clock rate.

DOI 10.1007/978-981-4560-75-7_125-1

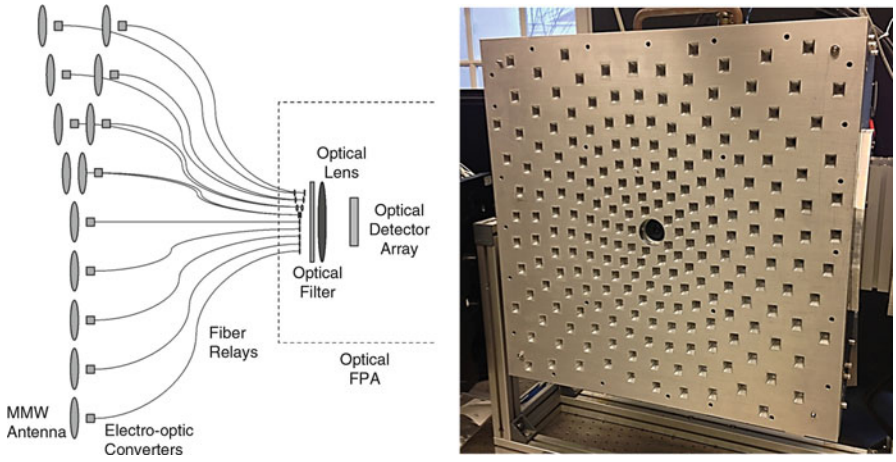


Fig. 11 Principle of aperture synthesis scanning using a Y-shaped array and an optical upconversion approach for optics-based image reconstruction (*left*). On the right the realized millimeter-wave demonstrator using a spiral antenna distribution is shown (▶ Millimeter-Wave Antennas and Arrays). Single antennas are realized as rectangular horns mounted in an aluminum plate (Courtesy of Phase Sensitive Inc., USA)

Another example for the application of aperture synthesis radiometry is realized by a real-time scanner for security screening to pilotage in degraded visual environments (Schuetz et al. 2013). Due to weight and volume restrictions on aircraft or helicopter, a very compact design is required. The realized approach uses aperture synthesis sampling of a scene and immediate upconversion of millimeter-wave signals to optical frequencies as shown in Fig. 11 (left), where directly the image reconstruction is performed by lenses and an optical camera, instead of correlation at Radio frequencies (RFs) and Fourier inversion by computer processing as it is done in the SMOS case (Schuetz et al. 2007).

The system is operated at 75 GHz center frequency using a bandwidth of a few 10 GHz. The antenna array is built in a five-arm spiral pattern consisting in total of 220 single-antenna and receiver elements as shown in Fig. 11 (right). A temperature resolution of around 3 K is achieved at a video rate of 25 frames per second. The field of view is about 20° and the angular resolution is reported for 0.4° in both azimuth and elevation direction for a square aperture size of 0.6 m. As indicated in Fig. 11 (left), the millimeter-wave radiation is collected by the single antennas, low-noise amplified and then directly upconverted to optical frequencies, where further signal configuration and especially full-image reconstruction is performed.

In microwave radiometry, two-dimensional aperture synthesis imaging is an attractive new method for high-resolution imaging at high frame rates without any mechanically moving parts. Due to the considerable thinning of the arrays, large apertures and thus very high spatial resolution can be realized by a reasonable number of single-antenna and receiver elements. However, as shown previously, a few hundred elements for present systems and probably thousands of them in future systems are required. Image reconstruction still is a challenge, since rarely a simple Fourier transform can be applied to the measured spatial frequency spectrum due to imperfections of the real system. Moreover the synthesized beam strongly depends on the specific array configuration (e.g., Y shape, spiral shape, etc.) and consequently how the corresponding sampling pattern of the spatial frequency spectrum fulfills the Nyquist criterion. The synthesis of the antenna aperture and the corresponding sampling character of that principle lead to grating lobes of the synthesized beam and a limited useful field of view without aliasing impacts. The calibration of an aperture synthesis system as well is rather complicated, since the classical hot-cold method as used typically for mechanical scanner systems cannot be applied, since many single elements with slightly different transfer functions influence not only a single pixel in the final image

DOI 10.1007/978-981-4560-75-7_125-1

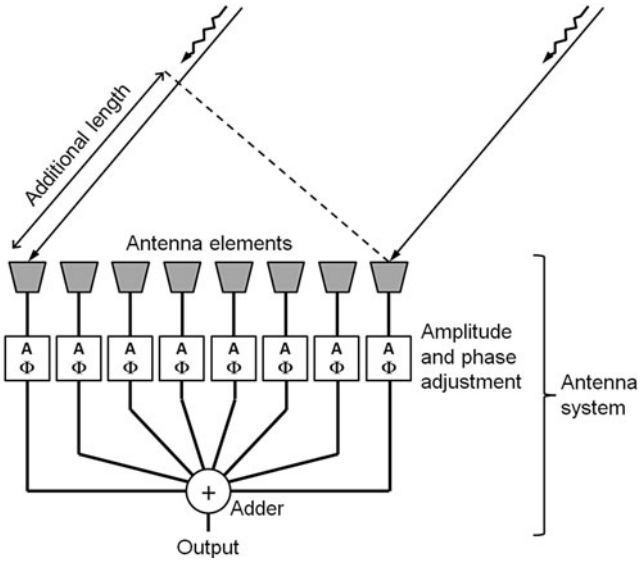


Fig. 12 Basic principle of phased-array scanning for radiometric applications. Brightness temperature radiation is received by single-antenna elements and shifted in phase and weighted in amplitude to add up coherently to the output for only one specific direction. Shown is only one-dimensional operation but two-dimensional scanning is possible as well

but altogether the whole image. Although thinned arrays can be designed, still a high number of single antennas and receivers compared to the mechanical scanners have to be used, making presently such systems rather expensive. An essay about the evolution of aperture synthesis technology for microwave radiometry in Earth observation and security applications for the last 25 years is given by Martin-Neira et al. (2014).

Other Scanning Principles

Previously discussed scanning principles and the corresponding antenna types are the mostly used ones or those with large potential for the future. However, several other approaches exist, at least in theory when up to date no practical system of such type has been realized. These principles are scanning by phased-array technology, frequency scanning, digital beamforming, and lens-based scanning. They are briefly discussed next.

Phased-Array Scanners

Phased-array technology has meanwhile large heritage from radar technology going back already to the 1950s of the last century (► Phased Array). The goal of this scanning principle is to replace a large antenna aperture by an array of small antennas and feed those with adequate phase and amplitude information in order to focus and steer the antenna beam fully electronically (see chapter XXX for further reading). The basic operation is illustrated for one-dimensional beamsteering in Fig. 12. Basically, for each antenna element, the additional path length with respect to a reference element (e.g., the right one) in the case of off-axis steering has to be compensated to produce maximum output signal for a certain pointing direction. This compensation is exclusively done by proper insertion of additional phase shift and time delay, while additional amplitude adjustment allows shaping the beam. Finally, all signals are added up coherently, usually implemented in hardware.

However, for microwave radiometry applications, the phased-array principle has been rarely used to date (Srinivasan et al. 2010; Sönmez et al. 1992). There are two major reasons. The first one is the typically large number of single-antenna elements and, moreover, of phase shifters for the steering and low-noise amplifiers for sufficient temperature resolution. It should be noted that proper beamsteering and

DOI 10.1007/978-981-4560-75-7_125-1

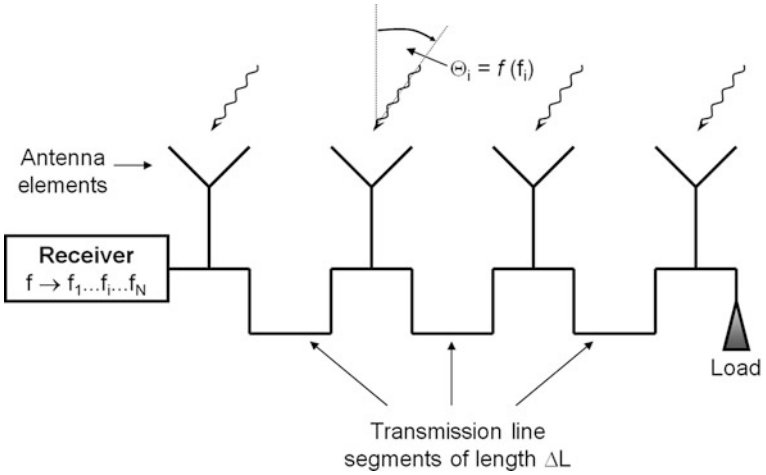


Fig. 13 Principle of antenna beamsteering by changing the operational frequency. The transmission line segments have identical and constant length, so that the amount of linear phase variation along the array is changed by varying frequency f_i, which steers the beam to a corresponding direction Θ_i

shaping requires a high degree of granulation, i.e., a very high number of single-antenna elements, located as close as possible together in a fully populated array. Secondly, the losses in the antenna system have to be kept as low as possible, which can be challenging when either no low-noise amplifiers are available or when such a large number of those are not affordable (► Transmission Lines). Decreasing the number of elements, by increasing the spacing and/or by increasing the single-element aperture size, results in the generation of strong grating lobes and a narrow field of view. In addition, large antenna structures require the implementation of true-time delays, which is performed to date by insertion of additional low-dispersion transmission lines. In conclusive summary, the expense for radiometric phased -array scanning is still very high to date. Nevertheless, it can be expected that increasing miniaturization of components using MMIC technology or a system-on-a-chip approach (► On-Chip Antennas), for instance, will allow the use of even thousands of combined amplifier–phase shifter circuits, possibly with already integrated antenna elements.

It should be further noted that, besides the capability of very fast fully electronic scanning, the beams to scan a scene are generated sequentially in time by common phased-array operation. For that reason the phased-array approach belongs to the classification of real-aperture antenna systems. Hence, the imaging of a large field of view at high resolution requires short dwell time of the beam at each scene location, resulting in reduced temperature resolution. As before in aperture synthesis, the calibration of such an imaging array approach is challenging due to the individual characteristics of the single elements, since all together contribute to a single beam. Here, again the classical hot-cold method is not applicable. Furthermore, the dense positioning of the single antennas in highly granulated arrays can lead to varying and bad antenna impedance matching in dependence of the scan angle, resulting in brightness temperature errors when not sufficiently known and compensated.

Frequency Scanning

In a one-dimensional antenna array, where the single-antenna elements are sequentially fed by equally long interconnecting pieces of transmission line, the phase shift between adjacent elements is equal and depends on frequency or wavelength, respectively. Hence, from the beginning to the end of the array, a linear phase shift between $1 \times \Delta\varphi$ and $(N-1) \times \Delta\varphi$ is obtained with respect to the first antenna with N being the number of antennas. Depending on $\Delta\varphi = 2\pi\,\Delta L/\lambda$, the array-generated beam is steered to a certain off-axis direction $0° \leq \Theta \leq \pm\Theta_{max}$ with respect to boresight, ΔL being the length of

DOI 10.1007/978-981-4560-75-7_125-1

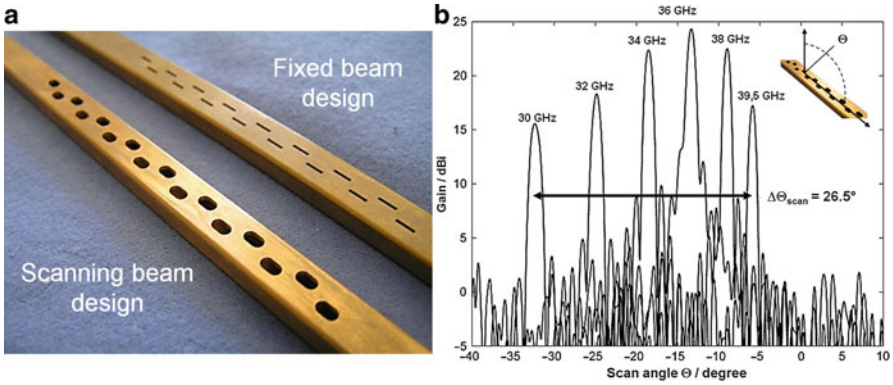


Fig. 14 Photograph of Ka-band slotted waveguide antennas for fixed and scanning beam design (*left*). On the right measured gain profiles of the scanning beam version for different frequencies are superimposed in one diagram to show the dependence on scan angle variation. The waveguide is of standard rectangular WR28 type

interconnecting transmission line and λ the wavelength according to Fig. 13. This beamsteering principle can be also understood as a special type of phased-array scanning (► Phased Array) where, instead of adjusting the phase for a fixed frequency, a frequency change produces a certain phase value (► Beam-Scanning Leaky Wave Antennas).

Such antenna types can be realized by any traveling wave configuration providing a repetitive arrangement of transmission line segment and antenna element (► Travelling-wave Antennas (Including Leaky Wave and Surface Wave Antennas), ► Transmission Lines). In microwave radiometry, it is most important that the antenna structure has lowest losses since it is the first component in the receiving chain to which the total receiver noise temperature is most sensitive. A suitable antenna type especially at higher frequencies, e.g., the millimeter-wave domain, is the slotted waveguide antenna (► Waveguide Slot Antennas and Arrays, ► Millimeter-Wave Antennas and Arrays, ► Tapered Slot Antennas, ► Travelling-wave Antennas (Including Leaky Wave and Surface Wave Antennas)). The waveguide itself is a low-loss transmission line and radiating slots can be relatively simple milled into its broad or narrow metallic walls.

A first approach on radiometric imaging using frequency scanning in combination with phased-array techniques is summarized in Clark et al. (2004). The system operating in W band (75–93 GHz) with 232 single channels using a parallel plate slotted waveguide antenna was intended for real-time security applications like concealed weapon detection. With a square antenna size of 60 cm side length, an angular resolution of about 0.4° and a field of view of around 20° in frequency scanning direction and 30° in the other direction were achieved.

Figure 14 shows another implementation and the scan patterns for Ka band in comparison to the classical slot geometry used for a fixed beam (Schreiber and Peichl 2013). The scanning-type antenna is intended to be used in conjunction with aperture synthesis (► Waveguide Slot Array Antennas) for a fully electronic scanning radiometer system to be used in security applications like enhanced vision for vehicle or aircraft. Noticeably, the slots for the scanning version have formed into rather elongated holes in order to serve the large bandwidth to be used for steering the antenna beam. The shown slot design allows a range of about 26.5° scan angle for a bandwidth of 9.5 GHz resulting in about 2.8°/GHz scan rate according to Fig. 14 (right). The aperture length of the slotted waveguide antenna used for that diagram is 0.5 m, housing 122 equidistant slots in total. Due to the large operational frequency range, the corresponding half-power beam widths are between 1.26° and 0.84° for the lowest and highest frequency. Note that the scan angle range is not centered at 0° boresight direction for mid-band frequency but at about −17°. That shift is caused by required optimization between antenna efficiency, i.e., lowest losses, and

DOI 10.1007/978-981-4560-75-7_125-1

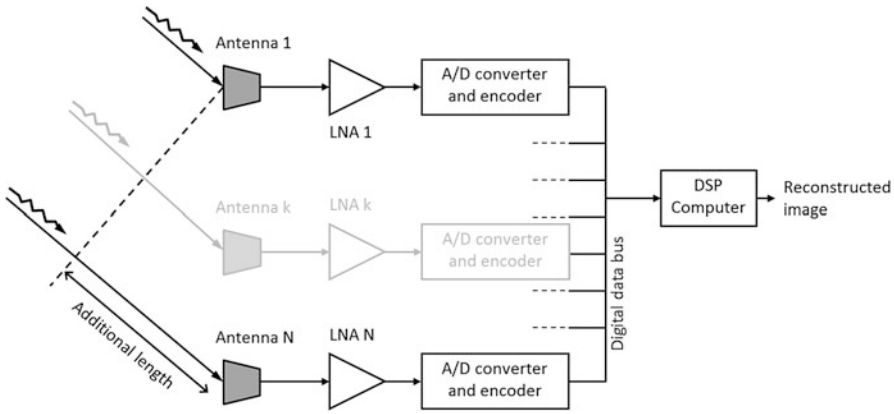


Fig. 15 Imaging principle for one-dimensional radiometric digital beamforming. Brightness temperature radiation is received by each antenna element and preprocessed for A/D conversion and storage. All channels are sampled coherently and the final image is generated by adequate common signal processing of all channels. Only the radiation from one direction is shown, but radiation from all directions is received simultaneously (LNA, low-noise amplifier; A/D, analog to digital; DSP, digital signal processor)

sufficient matching of input impedance (► Impedance Matching and Baluns), i.e., low reflections (Schreiber 2014). The obtained beam efficiency is at least 20 % at the band edges and 90 % in the central band region. Note that for the radiometric operation of this antenna, only an instantaneous bandwidth of around 100 MHz should be used in order to minimize beam blurring due to unintended beamsteering within the bandwidth.

Beamsteering by frequency shift is an attractive method for some radiometric scanning applications, since it can be realized purely electronic without additional expense in multi-receiver designs. However, it can be realized only in one dimension, so that the second scan direction for two-dimensional imaging has to be realized by any other suitable method. In addition, some degrading impacts like varying angular resolution and increased losses at the band or scan region edges have to be accepted, requiring careful design of this antenna type. In addition, only narrow-band operation for a certain scan position can be used, resulting in a decrease in temperature resolution if not compensable by other setting parameters. It should be further noted that applications requiring a certain fixed frequency band for observation cannot be served by this scanning principle due to the frequency dependence of beam pointing.

Digital Beamforming

Digital beamforming (DBF) is an imaging or scanning technique, which generally can be defined as a procedure performing adequate signal sampling by a special arrangement of antennas and the mathematical processing of the sampled signals in order to generate a synthetized focused antenna beam for each position in the image to be constructed (► Multi-Beam Antenna Arrays; ► Radio-Frequency Beam-Forming for Scanned and Multi-Beam Antenna Systems). Consequently the antenna beam of the antenna system does not exist instantaneously in reality but virtually within the data processing and the final image, as it is already the case for aperture synthesis. However, in this context DBF is intended to work according to Fig. 15.

Each channel of an adequately spatially distributed antenna/receiver array is sampled coherently and simultaneously and digitized for storage in a computer or the memory of a digital signal processor. So, no focusing occurs at this point. The digital data then are collectively phase processed such that for each direction the additional path length between the single-antenna elements is compensated, which performs focusing in a certain direction. Since the compensation can be done for each direction using the identical data set, multiple synthesized antenna beams weighted according to the received radiation power for each

DOI 10.1007/978-981-4560-75-7_125-1

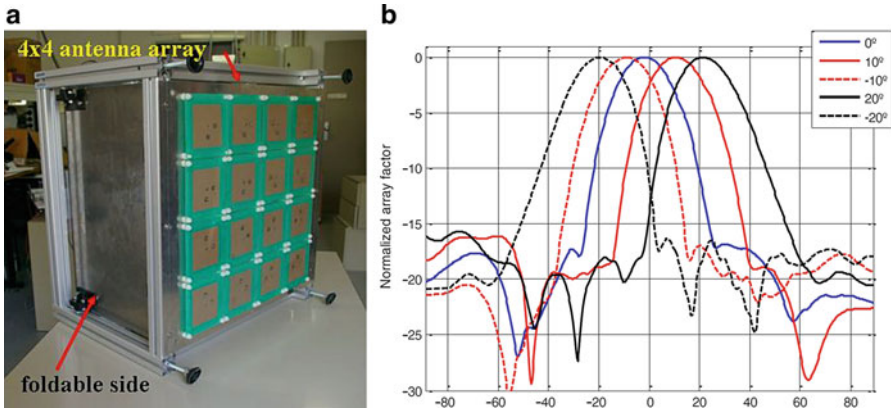


Fig. 16 Photograph (*left*) and measured synthesized antenna beams (*right*) of the PAU-RAD radiometer system working at L band. Multiple antenna beams are generated in one scan direction by DBF techniques. The single-antenna element is realized as microstrip patch (▶ Microstrip Patch Antennas) receiving two linear polarization states (Reproduced courtesy of Bosch-Lluis, UPC, Spain)

required direction are generated in parallel, thus reconstructing altogether the final image. Adequate choice of phase manipulation and amplitude weighting of the single channels allows a high degree in shaping the synthesized antenna beams, even individually, and performing error correction for the imperfectness of the real system.

DBF is a rather new scanning or imaging technique since the required sampling devices have to be provided in large numbers and at usually very high sampling rates according to the desired bandwidth of the radiometer system, which was not possible or too expensive in the past. Conventionally an array has to be fully populated to avoid or minimize grating lobes, which means a high number of single channels for a high spatial resolution and a large field of view, the latter being defined by the single-element antenna pattern. However, first attempts for radiometric applications have been undertaken as shown in Fig. 16 (Bosch-Lluis 2011).

The Passive Advanced Unit-Radiometer (PAU-RAD) system is a demonstrator device intended for ground-based sea surface salinity monitoring of the ocean from an elevated platform. It is a combination of a dual-polarized microwave radiometer and a Global Positioning System (GPS) reflectometer; hence, the center frequency was chosen to be 1,575 GHz for the benefit of using the identical front end, and the bandwidth is 2.2 MHz. DBF is only applied in one array direction and analog fixed beam generation in the other. So, there are 16 single-antenna elements grouped into four channels, each one digitized to 8 bits at 5.745 MHz sampling frequency for each polarization. The lateral length of the square array is about 2.6 wavelengths or 0.5 m (green area of Fig. 16 (left)), resulting in a half beam width of about 24°.

DBF can be considered as the most flexible and powerful imaging technique for future high-performance radiometer systems using fully electronic scanning. Hence, large apertures and corresponding high spatial resolution are possible without any mechanically moving parts. Note that even aperture synthesis imaging using sparse arrays can be performed via DBF, when the digital signal processing is done by complex correlation of all pairs of antennas in the array. However, the data revenue is considerable due to sampling of each channel at a usually high rate. As an example, consider a future array of 1,000 elements, operated at 1 GHz bandwidth for 10 ms integration time and sampled by 8 bit quantization. The corresponding data volume will be at least around 20 GB for tightly fulfilling Nyquist criterion, which has to be processed to the final image. This task is still a hard challenge for present memory and computer technology, especially when there is demand for real-time operation of the system.

DOI 10.1007/978-981-4560-75-7_125-1

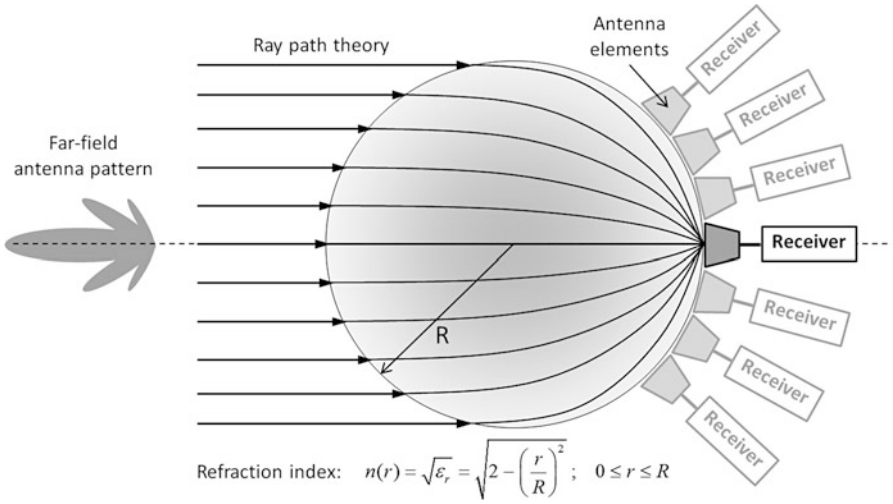


Fig. 17 Imaging principle for a Luneburg lens radiometer system. An antenna array can be located along the surface of a spherical lens, each one being in the focus of a bundle of parallel rays from another direction. The ray bundle is only shown for the central element. The refraction index n of the lens has to change according to the given equation as a function of radius r. The quantity ε_r is the real part of the permittivity of the sphere material

Lens-Based Scanning

Lenses are used in optical systems since centuries for focusing light to a passive detector, e.g., the human retina or a photographic film, for instance. Those detectors basically act as a focal plane array (FPA) in the way as described previously, in order to generate an image in an instant (► Dielectric Lens Antennas, ► Multi-Beam Antenna Arrays, ► Radio-Frequency Beam-Forming for Scanned and Multi-Beam Antenna Systems). For practical reasons the same principle can be applied to microwave frequencies. However, the classic optical lens design, i.e., a thin dielectric body of convex shape, was rather rarely used in the past for microwave radiometry. Main reasons are the required bulky profile and associated weight for a given frequency and aperture size, the impacts of reflection and dielectric losses, and the occurrence of typical imaging errors known from optical systems (e.g., aberrations, astigmatism, coma, etc.), requiring expensive measures for compensation or reduction. Furthermore, the problem of pure FPA systems requiring a large number of receivers/detectors for two-dimensional imaging of a large field of view at a given spatial resolution is still challenging. Nevertheless, some attempts of developing lens-based passive FPA systems in the millimeter-wave range have been undertaken in the last few decades (Richter et al. 2006; Kim et al. 2011, ► Millimeter-Wave Antennas and Arrays).

A special version of dielectric lens is the Luneburg type. It consists of a spherical dielectric medium whose refractive index n varies fully symmetric from the center to the surface of the sphere from a value of $2^{1/2}$ to the value of 1 according to the equation of Fig. 17.

In ray theory this variation of refractive index focuses all incoming parallel rays of any direction to a single point on the opposite surface, where an appropriate antenna element can be located. Of course, the far-field pattern is divergent and has the shape of conventional antenna beams, i.e., finite half-power beam width and side lobes, due to the finite aperture size of the sphere and the wave character of the radiation. However, there are several advantages of this lens type. Since the dielectric constant is one on the surface, proper impedance matching to the antenna element is simplified (► Impedance Matching and Baluns). The maximum value of 2 in the center and the decrease to 1 at the surface implies that the sphere can be manufactured of light-weight materials, so that even for large spheres the total weight can be kept reasonable. Nevertheless, the production of a sphere following exactly the required refractive index is difficult, and mostly the sphere is made of a number of different spherical layers with constant index using

DOI 10.1007/978-981-4560-75-7_125-1

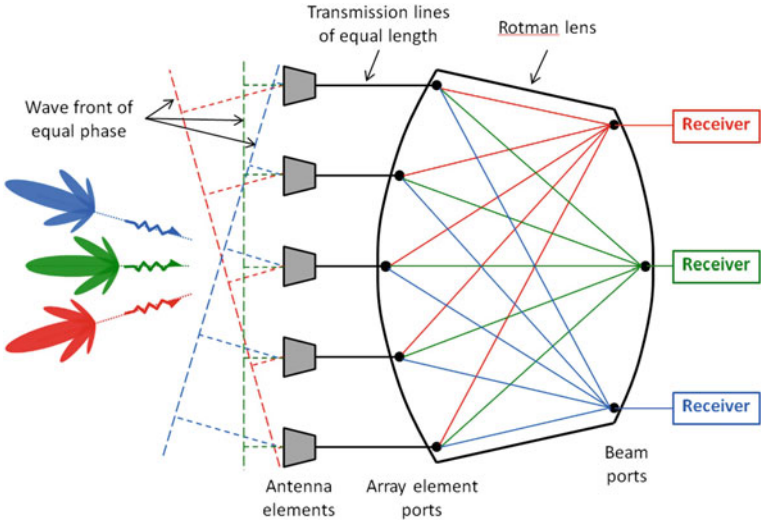


Fig. 18 Imaging principle of a Rotman lens radiometer system shown for m = 5 element ports and n = 3 beam ports. Radiation from a certain direction results in an increasing linear phase shift from the first to the last antenna element of the row. These phase shifts are compensated by the lens such that only at one beam port, all contributions are in phase. The lens has to generate that condition at each beam port, resulting in n simultaneous focused beams *for* n directions

a foam-type material, where the foam density can be used for adjusting the permittivity value of each layer (Bor et al. 2014). Note that antenna coupling of the adjacent antenna elements can lead to impedance mismatch and pattern degradation (► Suppression of Mutual Coupling between Elements in Multiple Element Antennas), and the distance to the surface of the sphere can be different from zero for proper operation. Suitable antenna element types for Luneburg lens systems are horn antennas or open waveguides, for instance. Dielectric losses of the sphere impact the temperature resolution of the radiometer system, especially for larger spheres. However, if sufficiently integrated antenna/receiver circuits are available, a rather simple scanner design for two-dimensional imaging radiometers without mechanical motion can be realized. Although basically considered as an appropriate imaging principle, radiometric imaging systems based on the Luneburg lens have not been realized quite often in the past (Shannikov et al. 2005).

Another interesting lens type for radiometric imaging systems is the Rotman lens (► Multi-Beam Antenna Arrays, ► Radio-Frequency Beam-Forming for Scanned and Multi-Beam Antenna Systems). This lens is a multi-port device consisting of the array element ports, the beam ports, and the lens region according to Fig. 18. The lens can be manufactured in various embodiments like waveguide structures or microstrip technology, for instance, where it typically consists of a port region and a parallel plate region. The first is responsible for proper impedance match (► Impedance Matching and Baluns) and the physical connection to the external circuit parts, and the latter provides the required phase shifts for all beam directions. The shape of the lens is derived from the necessary phase or delay relationships required for each antenna beam simultaneously (Tudosie 2009). The single-antenna element can be of any type useful for the specific realization and frequency range, e.g., horn antennas, patches (► Microstrip Patch Antennas), dipoles, etc.

The antenna elements should be arranged as close as necessary to avoid grating lobes for the single-antenna patterns. The number of antenna elements m has to be chosen according to the demanded field of view and angular resolution. The number of beam ports n has to be selected according to the number of required beam directions to sample the field of view according to the Nyquist criterion. However, it is obvious that a large number of ports are difficult to realize, and the practical implementation of the theoretical principle has limits. Hence, there are restrictions in shaping the beam and, depending on

DOI 10.1007/978-981-4560-75-7_125-1

frequency and realization, high losses of the lens, and the required transmission line structures can be existent, reducing the temperature resolution of the radiometer system. Multiple beams in parallel can be generated, thus scanning the field of view at an instant. Furthermore, several Rotman lenses can be stacked and connected such that two-dimensional scanning is possible.

However, only few radiometer systems based on Rotman lens scanning have been realized in the past (Hall et al. 2001). The main reason might be the complexity and size for realizing systems with a high number of simultaneous beams and the relatively high losses of lens structures especially at higher frequencies. In general, lens-type scanning antenna systems for radiometric applications have been realized rather rarely in the past, although their optical counterpart is widely used since many centuries for focusing light and related imaging systems. A main reason might be the bulky, weighty, and voluminous embodiment required for generating a reasonable number of beams simultaneously especially at lower frequencies. Furthermore, the complex design and manufacturing for suitable beam shapes and low-loss reception of radiation are still a challenge. Last but not least, the FPA-like type of scanning, in particular in two dimensions, requires a large number of single receivers being still expensive at present, as already mentioned throughout this section.

Up to now possible and already realized antenna concepts for imaging microwave radiometer systems have been discussed. The following section will address the basic quantities of radiometer antenna systems as they are important for the design and verification of such devices.

Basic Quantities of Radiometer Antennas

The previous section has illustrated a multitude of applications, realized systems, and scanning methods for one- or two-dimensional radiometric imaging, where the antenna system plays a dominant role in the obtained overall system performance. The contribution of the antenna system can be divided into two major parts, the impact of the antenna pattern and the signal losses introduced by the antenna system. The antenna pattern determines the spatial quality of an image, i.e., the spatial resolution and the shape of intensity transitions in the image. The signal losses are generated by attenuation losses of the antenna structure and material (▶ Transmission Lines), being of ohmic and/or dielectric origin, and losses by impedance mismatch of the antenna itself, the radiometer receiver(s), and the connecting transmission line (s). The losses in total impact the noise level, known as the temperature resolution or sensitivity of the imaging process, i.e., the capability of discriminating small brightness temperature differences in the final image.

As discussed in the previous sections, the imaging process can be basically discriminated between real-beam scanning and synthetic-beam scanning. Real beams are produced by real-aperture systems, i.e., mechanical scanners, focal plane arrays (including reflector- and lens-based systems), phased arrays, and frequency scanning. Synthetic beams are produced by aperture synthesis or digital beamforming. The following discussion is mainly performed for real-aperture antennas, and the peculiarities of aperture synthesis are comprehensively outlined.

Impact of the Antenna Pattern

The basic imaging relation between the brightness temperature distribution to be mapped and the scanning antenna pattern was given for spherical coordinates by Eq. 5a:

$$T_A(\Theta',\Phi') = \frac{1}{\Omega_A} \int\limits_{\substack{antenna\\pattern}} T_B(\Theta,\Phi)\ F_{Pn}(\Theta-\Theta',\Phi-\Phi')\ \sin(\Theta-\Theta')\ d\Theta\ d\Phi. \tag{5a}$$

Because of the convolution character of the antenna power pattern F_{Pn} with the original brightness temperature distribution T_B, it is obvious that the antenna pattern has a smoothing influence on T_B resulting in the blurred version, the antenna temperature T_A, as the imaging result. The amount of smoothing or blurring depends of course on the width and particular shape of the antenna pattern. The width of the antenna pattern can be represented by its half-power beam width $\Delta\Theta$ usually given for the two orthogonal directions at $\Phi = 0°$ and $\Phi = 90°$ or in a similar way by its maximum directivity D_{max}. Moreover, the width or $\Delta\Theta$ is a measure of the spatial resolution which can be achieved in the image, being defined as a measure for the discrimination capability of two adjacent point sources. Since the maximum directivity is related to the beam angle of the antenna by

$$D_{\max} = \frac{4\pi\ \ (\mathrm{sr})}{\Omega_A\ \ (\mathrm{sr})} \approx \frac{41,253\ \ \left(\mathrm{deg}^2\right)}{\Delta\Theta_{\Phi=0}\ \ \Delta\Theta_{\Phi=90}\ \ \left(\mathrm{deg}^2\right)}, \tag{6}$$

the directivity value can be understood as well as the number of point sources $N = D_{max}$, which can be resolved in the solid angle of 4π (Kraus and Marhefka 2001). In the one-dimensional case, it is the number of sources of width $\Delta\Theta$ being resolvable within the angular range of 360°. However, it should be noted that the directivity value does not directly give any statement about the shape of the antenna pattern and its side-lobe distribution.

The shape of the antenna pattern and especially the side-lobe contribution can be expressed in the beam efficiency η_A, which is given by the ratio of power transferred by the main lobe and the total transferred power:

$$\eta_A = \frac{\Omega_{ML}}{\Omega_A}, \tag{6a}$$

$$\Omega_{ML} = \int_0^{2\pi}\int_0^{\Theta_0} F_{Pn}(\Theta,\Phi)\ \ \sin\Theta\ \ d\Theta\ \ d\Phi. \tag{6b}$$

The angle Θ_0 has to be chosen adequately, e.g., as the position of the first null or the first minimum of the antenna pattern. In the hypothetical case of very low side lobes, the beam efficiency approaches the value of 1, and for strong side lobes, it is significantly less than 1. However, it has to be noted that the beam efficiency makes no direct statement about the spatial resolution behavior of the antenna pattern.

Consequently, both quantities – directivity and beam efficiency – have to be used for characterizing the spatial imaging quality of a radiometer system. For a more striking illustration of the spatial imaging quality, the analysis of the response on some canonical scenarios is very helpful. So, for a few basic brightness temperature distributions, the imaging relation (Eq. 4) can be directly evaluated.

DOI 10.1007/978-981-4560-75-7_125-1

(A) Constant brightness temperature scene:

$$T_B(\Theta, \Phi) = T_0, \tag{7a}$$

$$T_A(\Theta', \Phi') = \frac{T_0}{\Omega_A} \underbrace{\int}_{\substack{antenna \\ pattern}} F_{Pn}(\Theta - \Theta', \Phi - \Phi') \sin(\Theta - \Theta') d\Theta d\Phi = T_0. \tag{7b}$$

If the brightness temperature is constant at T_0 for a larger region around the actual field of view of the imaging process, the observed value for the antenna temperature is identical. Hence, the antenna pattern has no impact on the image quality, but of course, the image itself does not contain very much information. On the other hand such a situation is suitable for calibration purposes if the value of T_0 is known with sufficient accuracy.

(B) Brightness temperature source of very narrow width embedded in an otherwise constant background:

If the solid angle covered by the source $\delta\Omega_S$ is very small compared to the equivalent solid angle of the antenna pattern Ω_A, the antenna pattern can be assumed to be constant within $\delta\Omega_S$ and the obtained antenna temperature is given by

$$T_A(\Theta', \Phi') = T_0 + (T_1 - T_0) \frac{\delta\Omega_S}{\Omega_A} F_{Pn}(\Theta_S - \Theta', \Phi_S - \Phi'), \tag{8a}$$

$$\frac{\delta\Omega_S}{\Omega_A} << 1. \tag{8b}$$

Again, T_0 is the brightness temperature of the constant background, T_1 is the brightness temperature of the point source, and Θ_S and Φ_S indicate the angular source position. It is obvious that in this case the obtained antenna temperature for the source shows only a small deviation from the background due to the small value of the solid angle ratio. However, it can be recognized that the angular point source distribution shows with sufficient accuracy the shape of the antenna power pattern. This opportunity can be used for the power pattern measurement of large dish antennas used in radio astronomy (▶ Antennas in Radio Telescope Systems), which cannot be characterized by conventional antenna test ranges due to their size and the challenging compliance of the far-field condition. The following example shall illustrate the situation.

Consider the sun at L band ($\lambda = 20$ cm) having a surface brightness temperature of up to 100,000 K (Reul et al. 2007). Observed from the Earth, the angular diameter of the sun is about 0.5°. Using a dish antenna of $W = 5$ m diameter will have a half-power beam width of about $\Delta\Theta \approx 70° \lambda/W = 2.8°$, and the solid angle ratio is about $\delta\Omega_S/\Omega_A \approx 0.5^2/2.8^2 = 0.032$, which fulfills roughly the small-size requirement for the assumption of a point source. Hence, the observed maximum brightness temperature will be 3,200 K, which would allow with respect to the cosmic background of around 3 K a dynamic range in the order of 30 dB for a pattern measurement.

Equation 7b can be extended by adding a second point source of brightness temperature T_2 at a different location:

DOI 10.1007/978-981-4560-75-7_125-1

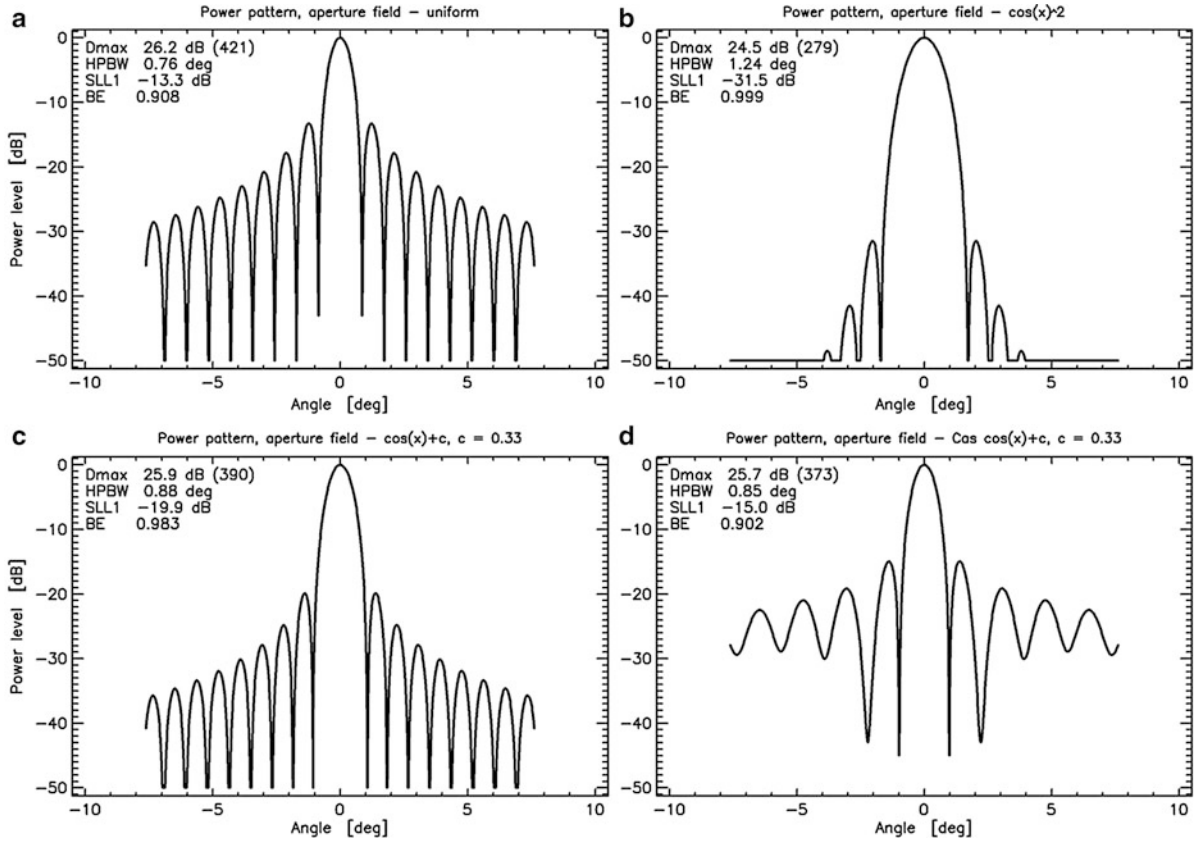


Fig. 19 Various different normalized antenna patterns produced by specific aperture fields having a maximum value of 1: (**a**) uniform, (**b**) $\cos(x)^2$, (**c**) $\cos(x) + c$, and (**d**) $\cos(x) + c$ with 5 % center blockage (Cassegrain type). The cos(x) function is zero at the aperture edges $x = \pm W/2$. Pattern values < -50 dB are not shown. Each graph indicates values for maximum directivity (Dmax), half-power beam width (HPBW), first side-lobe level (SLL1), and beam efficiency (BE). The patterns have been computed for an aperture size of W = 2 m and a frequency of 10 GHz

$$\begin{aligned} T_A(\Theta', \Phi') &= T_0 \\ &+ (T_1 - T_0)\ \frac{\delta\Omega_{S1}}{\Omega_A}\ F_{Pn}(\Theta_{S1} - \Theta', \Phi_{S1} - \Phi') \\ &+ (T_2 - T_0)\ \frac{\delta\Omega_{S2}}{\Omega_A}\ F_{Pn}(\Theta_{S2} - \Theta', \Phi_{S2} - \Phi'). \end{aligned} \tag{9}$$

The numberings 1 and 2 indicate the individual quantities for each source. If the two-point sources are identical but differently located, the observed antenna temperature is given by

$$T_A(\Theta', \Phi') = T_0 + T_\Delta\ (F_{Pn}(\Theta_{S1} - \Theta', \Phi_{S1} - \Phi') + F_{Pn}(\Theta_{S2} - \Theta', \Phi_{S2} - \Phi')), \tag{10a}$$

$$T_\Delta = (T_S - T_0)\ \frac{\delta\Omega_S}{\Omega_A}. \tag{10b}$$

The quantity T_Δ is the maximum change in brightness temperature due to the point sources with respect to the constant background. Equation 10a now can be used for analyzing graphically the impacts of the half-power beam width and the antenna pattern shape on the separation capabilities of the two-point targets, i.e., the spatial resolution, and the contrast of those with respect to the background. For that purpose a

DOI 10.1007/978-981-4560-75-7_125-1

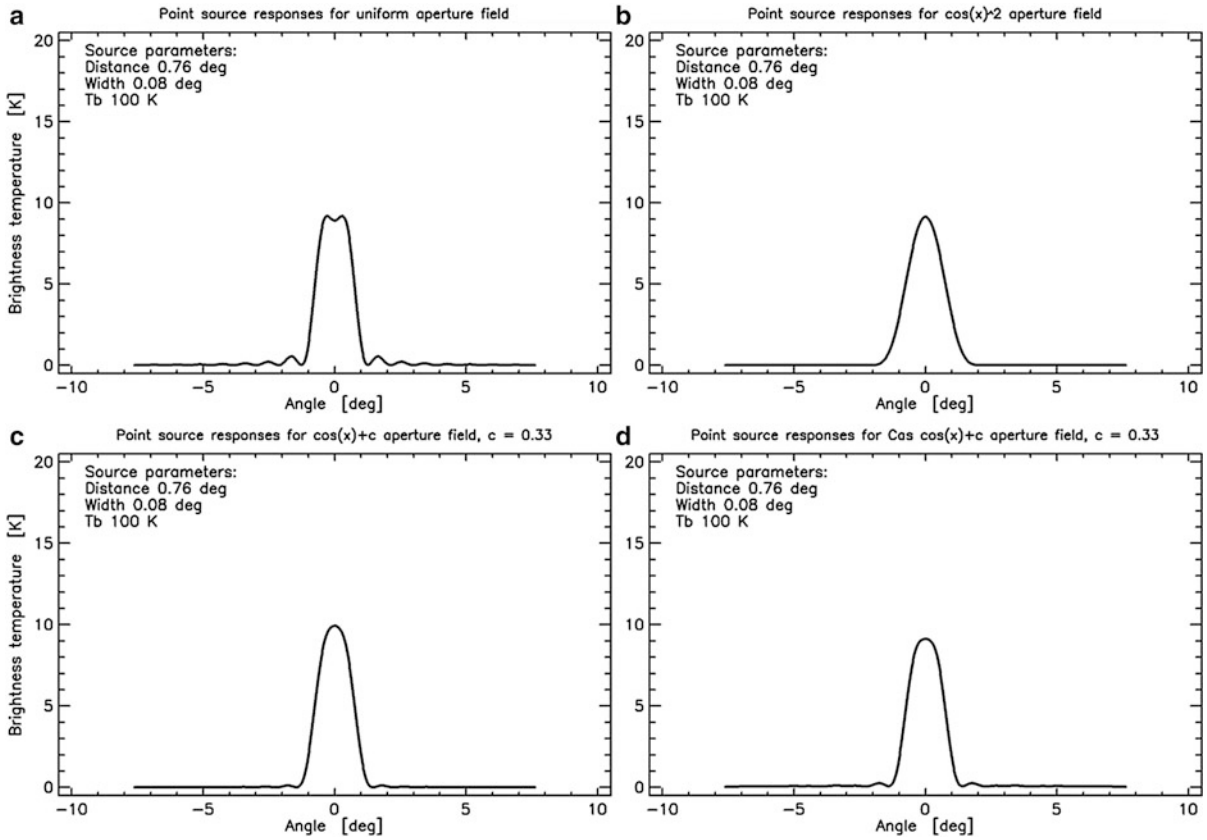


Fig. 20 Imaging result for two identical point sources separated by 0.76°. The brightness temperature level and width of the original sources are 100 K and 0.08°. The scanning antenna patterns correspond to the description of Fig. 19. No additional noise was considered for the simulation

one-dimensional analysis is appropriate and the results are representative for the two-dimensional case as well:

$$T_A(\Theta') \quad = \quad T_0 \quad + \quad T_\Delta \ (F_{Pn}(\Theta_{S1} - \Theta') \ + \ F_{Pn}(\Theta_{S2} - \Theta')). \tag{11}$$

The antenna power patterns can have quite a different shape, according to the half-power beam widths and the side-lobe levels, depending on the specific antenna type and its source field distribution in the aperture plane. For practical reasons the antenna aperture size shall be kept constant in this analysis, since most applications dictate a finite size, i.e., diameter or length and width, due to sensor volume and cost restrictions. In this case the only parameter to influence the antenna pattern is the aperture field distribution. Figure 19 summarizes field distributions and the corresponding power patterns used for the analysis. The patterns show a high variety with respect to half-power beam width and side-lobe level. Note that such aperture distributions can be adjusted in a wide range by choice of amplitude taper for array antennas or the feed pattern in case of dish antenna systems. Summaries of radiation characteristics from aperture field can be found in Olver et al. (1994) and Johnson et al. (1993). For example, the normalized power pattern for the uniform distribution of the aperture field of width W is given by

$$F_{Pn}(\Theta) = \left(\frac{\sin q}{q}\right)^2; \quad q = \frac{\pi W}{\lambda} \sin \Theta \tag{12}$$

DOI 10.1007/978-981-4560-75-7_125-1

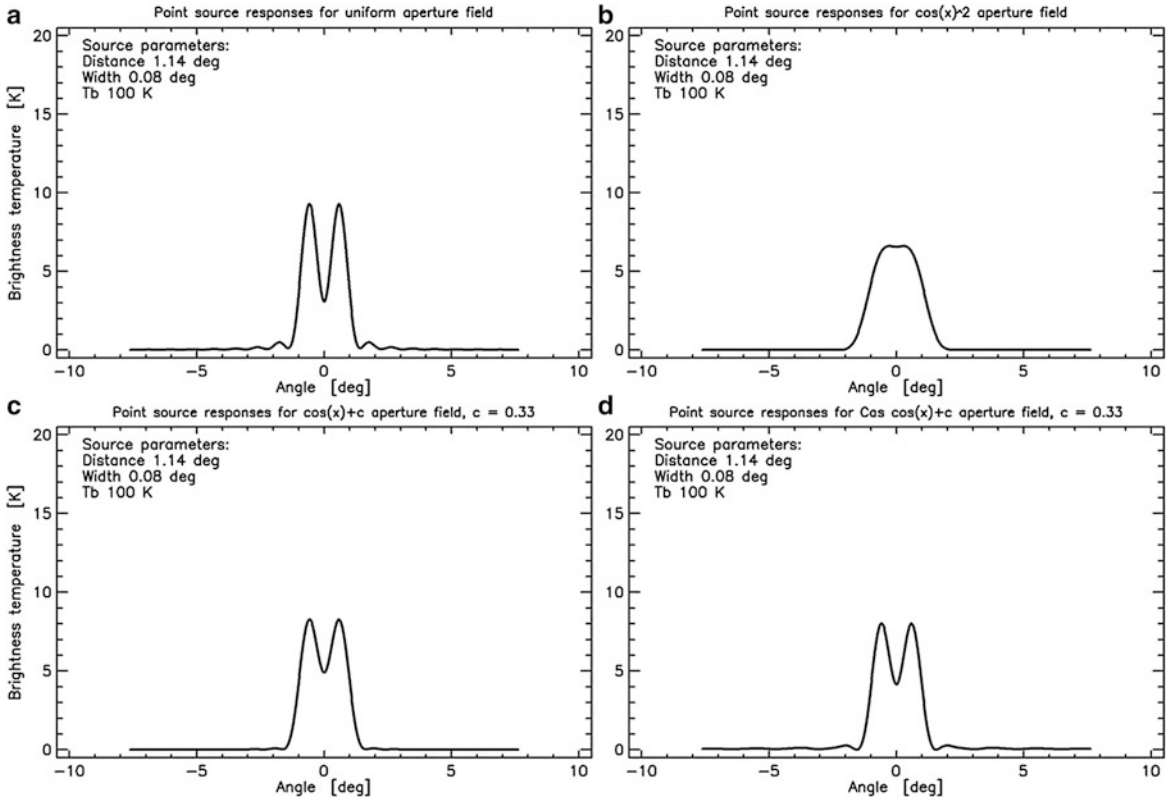


Fig. 21 Imaging result for two identical point sources separated by 1.14°. The brightness temperature level and width of the original sources are 100 K and 0.08°. The scanning antenna patterns correspond to the description of Fig. 19. No additional noise was considered for the simulation

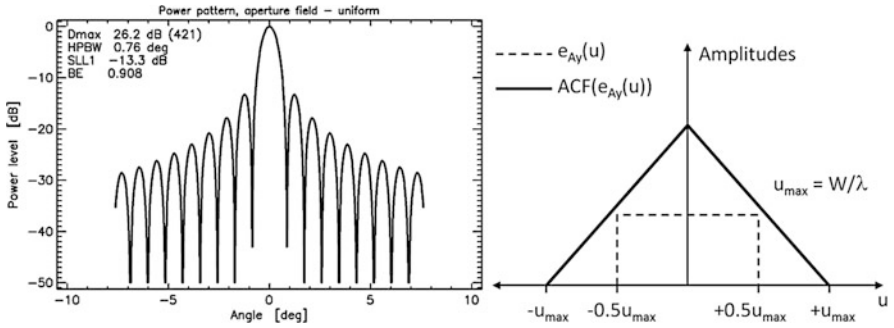


Fig. 22 *Left side*: One-dimensional antenna power pattern according to Fig. 19 and Eq. 12 for a uniform aperture field distribution $e_{Ay}(u)$, W = 2 m, and λ = 3 cm; *right side*: corresponding power spectrum in the spatial frequency domain being the autocorrelation (ACF) of $e_{Ay}(u)$. Note that here for the angle of the left Fig. 1 = sin Θ is valid according to Eq. 16a

The impact of such antenna patterns on the imaging of two identical point sources located at different separation widths is shown in Figs. 20, 21, and 22. The imaging result of Fig. 20 indicates that the two sources can be only separated by the antenna pattern produced by the uniform aperture field. However, this signature shows the highest ringing of the brightness temperature profile in the vicinity of the sources, which can lead to misinterpretation of the true values. In all other cases only one source can be identified from the profile, but always lower-ringing impacts are observed. Consequently the uniform aperture approach is useful where maximum detection capability of any target within a scene is the goal. A strongly tapered aperture field as in the case of the $\cos(x)^2$ case produces maximum smoothing of the true profile

DOI 10.1007/978-981-4560-75-7_125-1

but shows practically no corruption of the brightness temperature values in the vicinity of the transitions from the sources to the background. That behavior is useful for a requirement of high absolute accuracy of the observed profiles. Note that the separation in this case was exactly the half-power beam width of the pattern Fig. 19a corresponding to the uniform aperture field.

Now in Fig. 21, the sources were separated by 1.5 times the half-power beam width of the uniform case. Now the two sources can be separated for all situations with exception of the pattern for the $\cos(x)^2$ aperture field. Again, the separation is best for case (a) where the reconstructed source levels show again as well highest amplitudes, but again the ringing around the transition area is worst.

According to previous analysis, the choice of the antenna pattern, within restrictions on the feasibility, should be driven by the specific application, either if best detection performance or best absolute accuracy performance is desired. It should be noted that these results are valid in general although shown for a specific example.

Spatial Sampling Requirements

In most cases of radiometric imaging, aperture antennas are used for focusing an antenna beam. According to Olver et al. (1994), the co-polar field radiation pattern E_{cp} in the far field of an aperture field e_{Ay}, located in the x-y plane of a Cartesian coordinate system and directed in y direction, can be expressed with sufficient accuracy by

$$E_{cp}(\Theta,\Phi) = A_0(1 + \cos\Theta)\ F_y(\Theta,\Phi), \tag{13a}$$

$$A_0 = \frac{jk\exp(-jkr)}{4\pi r}, \quad k = \frac{2\pi}{\lambda}, \tag{13b, c}$$

$$F_y(\Theta,\Phi) = \iint\limits_{aperture} e_{Ay}(x,y)\ \exp(jk\ (x\ \sin\Theta\ \cos\Phi\ +\ y\ \sin\Theta\ \sin\Phi))\ dx\ dy. \tag{13d}$$

The quantity A_0 represents the far-field phase variation and amplitude decay with distance r, and λ is again the wavelength. If the aperture is large and thus the beam highly directive, the approximation $\cos(\Theta) \approx 1$ is furthermore valid in the beam region. Hence, the normalized far-field and power patterns F_{Fn} and F_{Pn} can be represented simply by

$$F_{Fn}(\Theta,\Phi) = \frac{F_y(\Theta,\Phi)}{\mathrm{Max}\{F_y(\Theta,\Phi)\}} = c_F\ F_y(\Theta,\Phi), \tag{14a}$$

$$F_{Pn}(\Theta,\Phi) = F_{Fn}(\Theta,\Phi)\ F_{Fn}{}^*(\Theta,\Phi) = |F_{Fn}(\Theta,\Phi)|^2, \tag{14b}$$

where c_F represents the normalization constant. Using Eqs. 13d and 14a, most typical antenna patterns for aperture antennas neglecting cross-polarized contributions can be described theoretically. Next, some useful transformations can be applied:

$$u = \frac{x}{\lambda}, \quad v = \frac{y}{\lambda}, \tag{15a, b}$$

$$l = \sin\Theta\ \cos\Phi, \quad m = \sin\Theta\ \sin\Phi, \quad \cos\Theta = \sqrt{1 - l^2 - m^2}, \tag{16a, b, c}$$

$$l^2 + m^2 \leq 1. \tag{16d}$$

Hence, Eq. 13d results in the following form:

$$F_y(l,m) = \iint\limits_{aperture} e_{Ay}(u,v)\ \exp(j2\pi\ (u\ l\ +\ v\ m))\ du\ dv. \tag{17}$$

Now, the well-known fact is obvious that the far-field pattern of an antenna is related to the aperture field by a two-dimensional Fourier transform. Since the directional cosines l and m represent the spatial coordinates of the far-field pattern, the coordinates u and v can be interpreted as the spatial frequency components of the aperture field distribution. This interpretation is identical to the case of the conventional time domain – frequency domain relationship in signal processing. It should be further noted that the aperture field method to estimate radiation patterns is a valuable tool for aperture and especially radiometer antennas, since in most cases the required aperture fields are known with sufficient accuracy. From Eqs. 14b and 17, it is also known that the autocorrelation function of the aperture field is related to the Fourier transform of the antenna power pattern (Kraus and Marhefka 2001):

$$F_{Pn}(l,m) = \mathrm{FT}\{\mathrm{ACF}(e_{Ay}(u,v))\}, \tag{18}$$

where FT denotes the Fourier transform operator and ACF the autocorrelation operator. For the one-dimensional antenna pattern of Eq. 12, as shown in Fig. 19a, the Fourier transform in the spatial frequency domain is shown in Fig. 22.

Now, Eq. 5a can be expressed in the new coordinates given by Eq. 16:

$$T_A(l',m') = \frac{1}{\Omega_A} \int\limits_{\substack{antenna\\pattern}} T_B(l,m)\ F_{Pn}(l-l',m-m')\ \frac{dl\ dm}{\sqrt{1-(l-l')^2-(m-m')^2}}. \tag{19}$$

For narrow antenna patterns (magnitude of l−l′, m−m′ ≪1, for larger values the antenna pattern becomes zero), the square root in the denominator can be approximated by the value 1. If that equation again is interpreted as a cross-correlation integral in the spatial domain, it can be expressed in the spatial frequency domain as the product of the single Fourier transforms:

$$T_A(l,m) = \frac{1}{\Omega_A}\ \mathrm{FT}\left\{t_B(u,v)\ f^*_{Pn}(u,v)\right\}. \tag{20}$$

Theoretically, the spectrum of the brightness temperature distribution $t_B(u,v)$ can have infinite extent, since in nature a transition from one to another brightness temperature level with nearly infinite slope is possible (e.g., a metal plate reflecting cold sky radiation located on a warm background). However, the known smoothing character of the antenna pattern in the spatial domain is produced by a finite extent of its spatial frequency spectrum $f_{Pn}(u,v)$, as can be seen by the one-dimensional example of the triangle function of Fig. 22 (right).

DOI 10.1007/978-981-4560-75-7_125-1

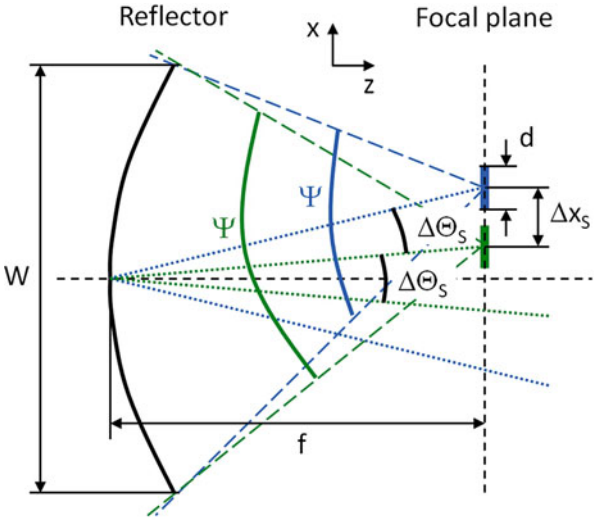


Fig. 23 One-dimensional illustration of focal plane array (FPA) relationships. The FPA elements (*green* and *blue bold lines* on the *right*) are drawn oversized and are usually small compared to dimensions W and f as the overall FPA size should be. Thus, the angular sampling interval $\Delta\Theta_S$ and the illumination angle ψ are similar for each beam generation, whose pointing direction is indicated schematically by the *dotted lines*. Only two elements are shown in asymmetric arrangement for clarity

Now, it is obvious that the imaging of a brightness temperature distribution requires discrete sampling at a finite rate in the spatial domain, either due to a finite image acquisition time or due to the given sampling pattern of a multi-beam imaging system like a focal plane array, for instance (▶ Multi-Beam Antenna Arrays, ▶ Radio-Frequency Beam-Forming for Scanned and Multi-Beam Antenna Systems). Hence, sampling by the antenna at an interval of Δl in the spatial domain produces an infinite repetition of the corresponding spatial frequency spectrum at an interval of $\Delta u_R = 1/\Delta l$, assuming a one-dimensional consideration. Consequently, in order to avoid an overlap of adjacent spectra, i.e., the generation of aliasing impacts on the imaging quality, the sample rate in spatial domain has to follow the sampling theorem:

$$\Delta u_R \geq 2\,u_{\max} \Rightarrow \Delta l \leq \frac{1}{2\,u_{\max}} = \frac{\lambda}{2\,W}. \tag{21}$$

Furthermore, since $l = \sin\Theta$ is valid, it follows that $d\Theta = dl/\cos\Theta$. Hence, the smallest value for $d\Theta$ is given for $\Theta = 0$, leading to the angular sampling rule in radians:

$$\Delta\Theta_S \leq \frac{\lambda}{2\,W}. \tag{22}$$

The same rule of course is valid for the second image dimension assuming a rectangular sampling grid. Now, considering the spatial resolution of the antenna pattern of Fig. 22a and Eq. 12 to be defined by the half-power beam width in radians,

$$\Delta\Theta_{HPBW} = 0.886\ \frac{\lambda}{W}, \tag{23}$$

it follows finally that the angular sampling has to be performed at a spacing of considerably less than the spatial resolution:

DOI 10.1007/978-981-4560-75-7_125-1

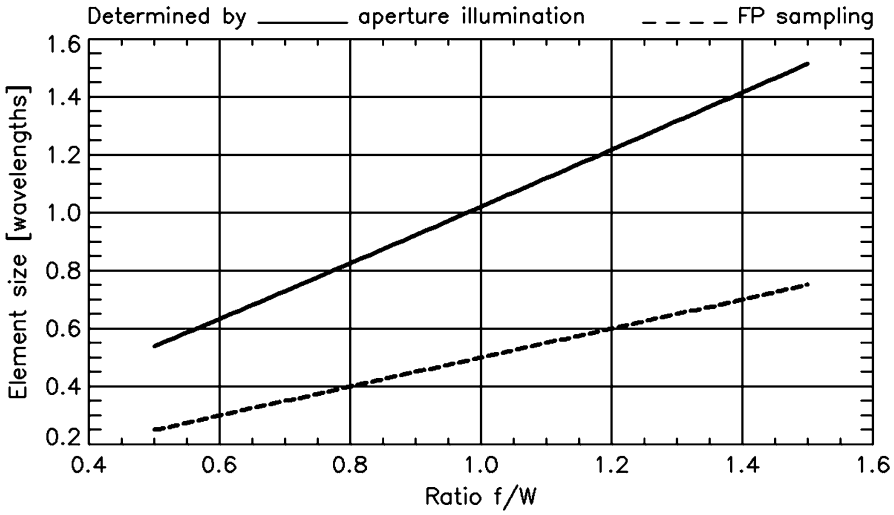


Fig. 24 Requirements on the element size of a focal plane array antenna as a function of f/W ratio of a direct-fed parabolic reflector as shown in Fig. 23. The *solid curve* represents the size as determined by a suitable reflector illumination, and the *dashed line* represents the requirement dictated by proper sampling of the focal plane

$$\Delta\Theta_S \leq 0.564 \;\; \Delta\Theta_{HPBW}. \tag{24}$$

However, it has to be noted that the sampling requirement of Eq. 22 can be relaxed to some extent for a different aperture field distribution producing a lower maximum spatial frequency. This leads on one hand to a reduction of the criterion of Eq. 22 and on the other hand to less spatial resolution (higher value) as for Eq. 23. Furthermore, in the two-dimensional case, the sampling can be performed on a hexagonal grid instead of the rectangular one, providing some higher sampling density.

Focal Plane Arrays

In a first-order approximation, the angular sampling grid as dictated by Eq. 22 is directly related to the sampling pattern in the focal plane of a reflector- or lens-based focal plane array system as given next (Johansson 1995):

$$\Delta\Theta_S \approx \frac{1}{f} \quad \Delta x_S, \text{ or } \Delta x_S \approx \frac{f}{2} \frac{\lambda}{W}, \tag{25}$$

where $\Delta\Theta_S$ is given by Eq. 22, Δx_S is the spacing of the array elements in the focal plane, and f is the focal length of the focusing unit according to Fig. 23.

Now, it has to be considered that in most applications a compact design is required, leading to values for the f/W ratio in the range ≤ 1. Thus, for the maximum element size, $d \leq \Delta x_S \leq \lambda/2$ is valid according to Eq. 25. On the other hand, the reflector has to be illuminated properly by the single element; hence,

$$\Psi \approx \frac{\lambda}{d} \tag{26}$$

is a typical choice being valid for many types of feed antennas as an estimate for the half-power beam width. The illumination angle ψ can be computed, for instance, for the shape of a parabolic reflector (Stutzman and Thiele 1998; ▶ Reflector Antennas), so that for the element size in wavelengths, the following relation is valid:

DOI 10.1007/978-981-4560-75-7_125-1

$$\frac{d}{\lambda} \approx \frac{1}{\Psi}; \quad \Psi = 4 \ \tan^{-1}\left(\frac{1}{4\ (f/W)}\right). \tag{27}$$

Now, the illumination-driven element size of Eq. 27 is plotted versus the sampling-driven element size of Eq. 25 using $d = x_S$ in Fig. 24.

This result shows that it is basically impossible to fulfill both requirements at the same time (▶ Multi-Beam Antenna Arrays, ▶ Radio-Frequency Beam-Forming for Scanned and Multi-Beam Antenna Systems). Either some undersampling of the focal plane and thus of the scene (use of larger element size and hence larger element spacing) or some more spillover contribution from the background behind the reflector (use of smaller element size producing wider beam width) or a compromise of both has to be accepted to some degree. Note that the result is valid for a staring FPA system and approximations to the basic quantities. In order to fulfill both requirements to a more satisfying level, the following measures can be undertaken:

1. Fine-tuning of parameters of a staring FPA: Since the previous derivation is based on assumptions and approximations, some fine-tuning of the single parameters can move both curves of Fig. 24 closer together. Parameters are the f/W ratio, the aperture size W (i.e., spatial resolution and spillover contribution), the feed element type in the focal plane, the shape of the reflector illumination (i.e., the aperture field distribution of the FPA elements), and the arrangement of the elements (e.g., hexagonal versus rectangular). However, it should be noted that these measures produce only limited improvement and some image distortions might remain. It should be further mentioned as well that by arranging small antenna elements very close together, problems with beam distortion, mutual coupling, and impedance mismatch can occur (▶ Suppression of Mutual Coupling between Elements in Multiple Element Antennas), which may cancel the benefits obtained by denser spatial sampling.
2. The use of other reflector shapes and multi-reflector systems: The use of a parabolic main reflector will produce higher beam distortions for elements at a large off-axis distance. Hence, the use of spherical reflectors can provide improved beam quality. In that case the FPA elements are no more to be located in a plane but rather on a spherical surface at half of the main reflector radius. Curved arrangement of the elements can also improve somewhat the performance for parabolic reflectors especially for large off-axis elements (▶ Reflector Antennas). Furthermore, the use of multi-reflector systems like Cassegrain or Gregorian arrangements can provide another degree of freedom for optimizing FPA performance. Here, as well only limited improvement can be expected, although it will be higher than for Case 1.
3. Additional mechanical motion: When a staring FPA does not sample properly a scene according to the sampling theorem, additional mechanical motion of the whole system can overcome that problem to fill the gaps. That is the way how FPA performance mostly is improved nowadays especially for systems with less constraints on imaging time but high absolute accuracy (e.g., in radio astronomy or for experimental systems) or close-to-real-time images with less accuracy requirements. In addition such an approach can increase the field of view of a radiometric imaging system considerably, since staring FPA systems have rather narrow fields of view due to the limited number of elements and degrading performance for elements at large off-axis operation. However, it should be noted that this approach increases the expense considerably especially for challenges in the FPA design and the necessary high-precision and high-dynamic mechanics.
4. FPA elements designed as phased-array or digital beamforming array: These approaches have been considered in the last decades especially for radio astronomy systems (Padman 1995; Fisher 1996; Fisher and Bradley 2000; Veidt 2006; Warnick 2013, ▶ Antennas in Radio Telescope Systems). The electronic beam-shaping capability of such a design allows applying full sampling in the focal plane

DOI 10.1007/978-981-4560-75-7_125-1

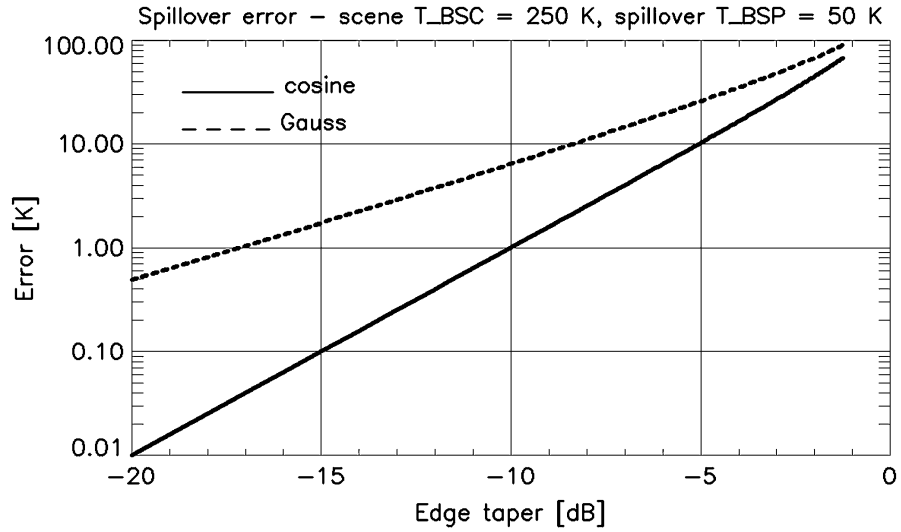


Fig. 25 Spillover error in brightness temperature determination of a scene as a function of edge taper ($p_{dB} = 10 \log p$) for a direct-fed reflector antenna. Two extreme temperature cases for the scene and the spillover region have been used for maximum error impact. The normalized power pattern of the feed element was assumed to be of shape $\cos(b\Theta)$ and $\exp(-b\Theta^2)$, b being in each case a width parameter

and thus the elimination of aliasing impacts without any mechanical motion. However, limitations on the achievable field of view remain and sequential sampling is required in the case of a phased-array approach, which again can be overcome by digital beamforming techniques due to digital generation of all antenna beams in parallel in the computer system. The latter might be the most promising technique for future FPA systems.

Spillover Impacts

Next, the error impact of spillover contributions to radiometric observations shall be briefly discussed for the one-dimensional situation, while results are representative for the two-dimensional case as well. According to Fig. 23, the antenna temperature for a feed element can be expressed as

$$T_A = \frac{T_{BSC}}{\Omega_A} \int\limits_{|\Theta| \le \Psi/2} F_{Pn}(\Theta) \quad d\Theta + \frac{T_{BSP}}{\Omega_A} \int\limits_{|\Theta| > \Psi/2} F_{Pn}(\Theta) \quad d\Theta, \tag{28}$$

where constant brightness temperature of the scene T_{BSC} and constant brightness temperature of the spillover region T_{BSP} have been assumed. Note that T_{BSC} is the brightness temperature within the focused beam of the reflector, i.e., the intended quantity to be measured. Now, the normalized antenna power pattern of the feed shall be represented by $F_{Pn}(\Theta) = \cos(b\Theta)$, which approximates a real situation quite realistically. The constant b is determined such that $\cos(b\psi/2) = p$, $0 \le p \le 1$ is valid. Furthermore, $F_{Pn}(\Theta) = 0$ for $b\Theta > \pi/2$ is assumed. Then, the integrals of Eq. 28 can be computed and the spillover error in the antenna temperature is estimated by

$$\Delta T_{spill} = T_{BSC} \quad - \quad T_A = (T_{BSC} - T_{BSP})\left(1 - \sqrt{1 - p^2}\right). \tag{29}$$

Figure 25 shows a plot of the spillover error for which $T_{BSC} = 250$ K and $T_{BSP} = 50$ K have been assumed. Those values simulate a situation where the radiometer antenna (the main beam) is pointing

toward the Earth and hence the spillover contributions are received from the sky, for instance. In addition, the situation for a Gaussian power pattern is shown which was evaluated numerically.

The spillover error is most important for measurements requiring high absolute accuracy. As can be identified, even an edge taper of −10 dB can produce a brightness temperature error of 1 K for the cosine case and up to 6 K for the Gaussian case so that at least taper values in the range of −10 to −15 dB are preferable for most Earth observation applications. In case of a bad side-lobe level distribution, the spillover impact can be even higher as shown in the examples. Since a high edge taper is accompanied by a loss in spatial resolution of the imaging system, a smaller edge taper magnitude of only a few decibels might be beneficial for security applications. However, a high spillover contribution can lower or blur the achievable brightness temperature contrast within an image. Note that the spillover impact is present for each focusing system using any single-feed or multi-feed subsystem in combination with other focusing units like reflectors (▶ Reflector Antennas) or lenses (▶ Dielectric Lens Antennas).

Sensitivity Considerations

Another important impact influencing strongly the spatial sampling of a scene is given by the requested temperature resolution, also called sensitivity, of the radiometric imaging system, since all other error sources should be kept considerably smaller than this. That instance is based on the first-order assumption that most error sources can be treated like noise and are statistically independent so that the Gaussian law of error propagation can be applied. The sensitivity describes the minimum difference in brightness temperature which can be detected by the system. It cannot be made arbitrarily good since it depends on application-driven system parameters. In a general formulation, the temperature resolution is given by Ulaby (1982):

$$\Delta T = k \frac{T_{sys}}{\sqrt{B\ \tau}} = k \frac{T_A + T_{REC}}{\sqrt{B\ \tau}}. \tag{30}$$

The finite sensitivity is caused by the noisy character of the signals to be detected and depends on the system noise temperature as the sum of the antenna temperature T_A and the average receiver noise temperature T_{REC}, and the product of the system bandwidth B and the integration time τ per spatial sample. The factor k comprises the impact of the specific receiver architecture and the imaging method to be used and can have thus a wide range of values ≥1. The value 1 is valid for a single total-power receiver integrated in a mechanical scanner sequentially sampling a scene, for instance.

Now, an imaging task usually is determined by the requirement to image a scene of certain size at a certain spatial resolution within a given time. Consequently, the integration time per spatial sample can be expressed as follows:

$$\tau = \frac{\Delta\Omega}{4\Omega_{FOV}} t_{frame}. \tag{31}$$

The angle increment ΔΩ represents the two-dimensional angular resolution cell and Ω_{FOV} the two-dimensional field of view, both expressed in square degrees. The factor 4 results from the requirement of sampling at a rate of about half the spatial resolution as discussed previously, and t_{frame} is the duration to acquire a complete image. Combining both equations results in

$$\Delta T = \frac{2kT_{sys}}{\sqrt{Bt_{frame}}} \sqrt{\frac{\Omega_{FOV}}{\Delta\Omega}}, \tag{32}$$

DOI 10.1007/978-981-4560-75-7_125-1

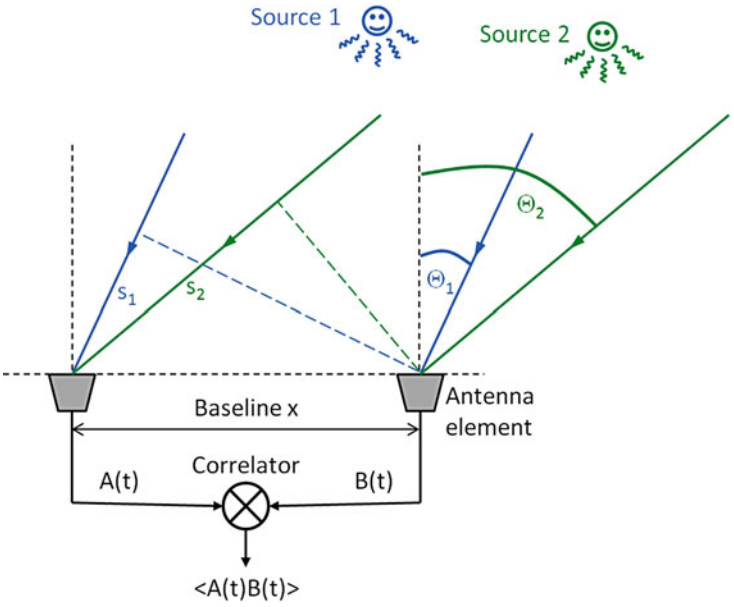


Fig. 26 Schematic principle of a two-element interferometer. Radiation of a certain direction is received by two antennas and fed to a complex correlator providing the time-averaged quadrature components of the signal multiplication. The distance of the two antenna elements is called the baseline x

which is a general relation between spatial resolution and sensitivity of the imaging process. The second term on the right side considers the quantities of the antenna system, i.e., the beam size and the range of scan angles. Equation 32 and practical numbers for the single quantities indicate that for most applications, a compromise between spatial resolution and sensitivity is mandatory. As an example consider the real-beam scanner described in Peichl et al. (2013). Here, the system temperature T_{sys} is about 800 K for an assumed antenna temperature of 300 K, the effective system bandwidth B is about 15 GHz, the image acquisition time t_{frame} is 1 s, the scanned field of view Ω_{FOV} is 360° × 30°, the angular resolution cell is about $\Delta\Omega = 0.75° \times 0.75°$, and the factor k has the value 1 for a single total-power-receiver imaging system. Thus, the corresponding temperature resolution ΔT is 1.8 K.

Aperture Synthesis Imaging

The interest in aperture synthesis imaging for radiometric applications started in the 1980s when high spatial resolution at lower frequencies became of interest (Ruf et al. 1988; Le Vine 1990). That combination requires large antenna structures which cannot be moved any more mechanically at a given speed for reasonable frame rates. As an alternative for focal plane arrays or phased arrays (► Phased Array, ► Applications of Phased Array Feeders in Reflector Antennas), the method of aperture synthesis is a vivid approach, which is used in radio astronomy since the 1950s (Thompson et al. 2001, ► Antennas in Radio Telescope Systems). The big advantage here is that the required number of single-antenna/receiver elements can be reduced considerably compared to the other electronic scanning approaches. In the last decades increased research and development for applying aperture synthesis to various applications has been undertaken (Camps 1996; Peichl et al. 1998; Harvey and Appleby 2003; Barré et al. 2008; Rautiainen et al. 2008; Martín-Neira et al. 2014).

DOI 10.1007/978-981-4560-75-7_125-1

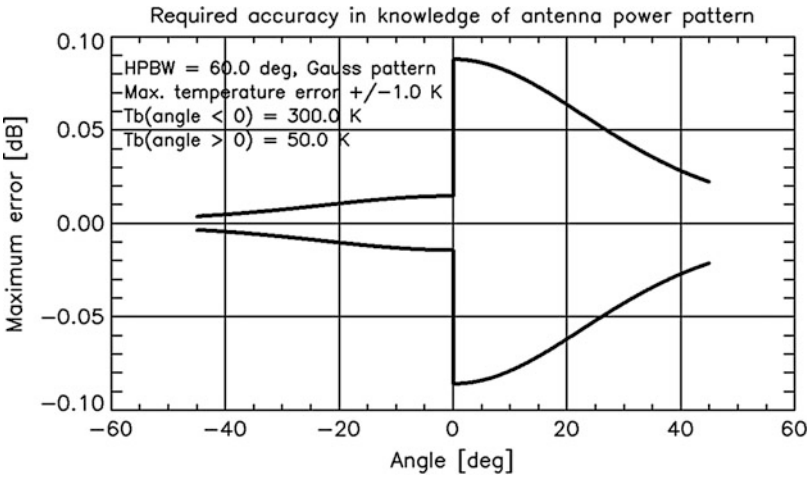


Fig. 27 Required absolute accuracy in the knowledge of the normalized antenna power pattern for two as constant assumed brightness temperature levels (at angles $< 0°$ and $> 0°$). The power pattern was assumed to be of Gaussian shape with a half-power beam width of 60°. A maximum tolerable temperature error of ±1 K was set

The Van Cittert-Zernike Theorem

Aperture synthesis imaging is based on the use of a multitude of two-element interferometers in a suitable arrangement, each one providing a different distance between the elements. The situation for a single interferometer is shown schematically in Fig. 26 for the one-dimensional case.

The interferometer correlates the two signals received by the antenna elements in phase quadrature so that the correlator output is the time average of two signal products, $<A(t)B(t) >$ and $< A(t)B(t\text{-}90°/(2\pi f)>$. Assuming that $B(t) = E_0 \sin(2\pi ft)$ and $A(t) = E_0 \sin(2\pi f(t\text{-}s/c))$, where E_0 is the signal amplitude, f the microwave frequency, and s the path difference between the two signals for a source in the far field located at an angle Θ_i with respect to boresight, the quadrature output can be combined in a complex notation:

$$cor(x) = I(x) + jQ(x) = \frac{1}{2}E_0^2\left[\cos\left(2\pi\frac{x}{\lambda}\sin\Theta_i\right) + j\sin\left(2\pi\frac{x}{\lambda}\sin\Theta_i\right)\right]. \tag{33}$$

Now, it is evident that in the case of a single source, one complex measurement allows to determine the angular position Θ_i of the source and its radiation power proportional to E_0^2, since the baseline x and the wavelength $\lambda = f/c$ are known. If there are two sources as shown in Fig. 26, one single measurement will not be sufficient, since the signals from both sources superimpose and hence no clear discrimination is possible. However, by adding another measurement at a different baseline, x will provide the lacking information. Consequently, by adding more and more sources, more and more baselines will be required. In the end a continuous distribution of sources in space requires a continuous correlation function cor (x) and the relation between the radiation power distribution $E_0^2(\Theta)$ and cor(x) is a Fourier transform, provided that the sources are located in the far field of the interferometers and that the sources are statistically independent or uncorrelated, which is valid for thermal radiation as in the case of radiometric imaging. That instance is called the Van Cittert-Zernike theorem and provides the basis for aperture synthesis imaging. For the two-dimensional case and an ideal system (i.e., identical antenna elements, ideal and identical frequency responses of the receivers, narrow bandwidth, ideal complex correlation), the theorem is given by

DOI 10.1007/978-981-4560-75-7_125-1

$$vis(u,v) = \frac{1}{\Omega_A} \iint_{\Omega_A} \frac{F_{Pn}(l,m)}{\sqrt{1-l^2-m^2}} \; T_B(l,m)\exp(-j2\pi(ul+vm))dl\,dm. \tag{34}$$

The function vis(u,v) now is called visibility function and represents a measurement of the two-dimensional spatial frequency spectrum of the brightness temperature distribution T_B(l,m) of the scene being of course proportional to the radiation power distribution, F_{pn}(l,m) is the antenna power pattern of the single-antenna elements, and l and m are the directional cosines as defined by Eq. 16. The quantity Ω_A again is the antenna element's solid angle, which in the case of aperture synthesis determines the field of view to be imaged. The coordinates u and v are the baseline components in a rectangular coordinate system and represent the metric length in wavelengths, i.e., u = x/λ and v = y/λ.

Impact of the Antenna Patterns

Now, it is evident that from a measurement of vis(u,v), the demanded brightness temperature distribution T_B(l,m) can be determined by an inverse Fourier transform as the image reconstruction procedure. However, the direct transformation result is given by the transform of the whole kernel in the integral, from which T_{Brec} as the reconstructed version of the true temperature T_B can be computed by

$$T_{Brec}(l,m) = T_B{}'(l,m) \; \frac{\sqrt{1-l^2-m^2}}{F_{Pn}(l,m)} \; \Omega_A; T_B{}'(l,m) = \mathrm{FT}\{vis(u,v)\}. \tag{35}$$

Basically, the image could be reconstructed in the whole angular area of the antenna pattern. In doing so the problem is that the power level of the antenna pattern vanishes usually more rapidly than the square root function of Eq. 35 does, and hence a division by small numbers can occur. That instance will amplify unavoidable errors in the knowledge of the antenna power pattern considerably, making results at these locations useless. Now, the uncertainty in the knowledge of the antenna pattern can be assumed to be F_{Pn}'(l,m) = F_{Pn}(l,m) + ΔF_{Pn}(l,m). So if a maximum temperature error of ± ΔT is given, the required accuracy in the knowledge of the antenna power pattern can be determined from Eq. 35 where F_{Pn}' is used instead of F_{Pn}:

$$\Delta F_{Pn}(l,m) = -F_{Pn}(l,m) \; \frac{\Delta T}{T_B(l,m) \; + \; \Delta T}. \tag{36}$$

A plot of the required accuracy for the knowledge of the antenna power pattern is shown for a Gaussian shape F_{Pn}(l) = exp(−b l^2) in Fig. 27 where again l = sinΘ is valid. It can be identified that even in the center direction, a high accuracy in the knowledge of the antenna pattern is required to keep the desired temperature error in the given limit. In the outer regions of the pattern, this requirement is even much stronger. Pattern knowledge accuracy in the order of a few hundreds of a decibel is hard to achieve by measurements or keep over the lifetime of an instrument so that the regions far out from the center are not to be used when high absolute-temperature accuracy is demanded. However, the accuracy requirement is more relaxed when a lower brightness temperature level is observed.

Based on the previous insight, it is evident that another approach to address the influence of the antenna pattern in the application of aperture synthesis has to be used. A practical way is to combine this task with the overall calibration of the imaging system. While for classical scanner systems using single receivers, the so-called hot-cold method is applicable for calibration (Ulaby 1982), where two well-known but different temperature references are measured by the receivers, a synthesis system has to use the overall scene within the field of view of the single-antenna element. The challenge now is to find large enough

DOI 10.1007/978-981-4560-75-7_125-1

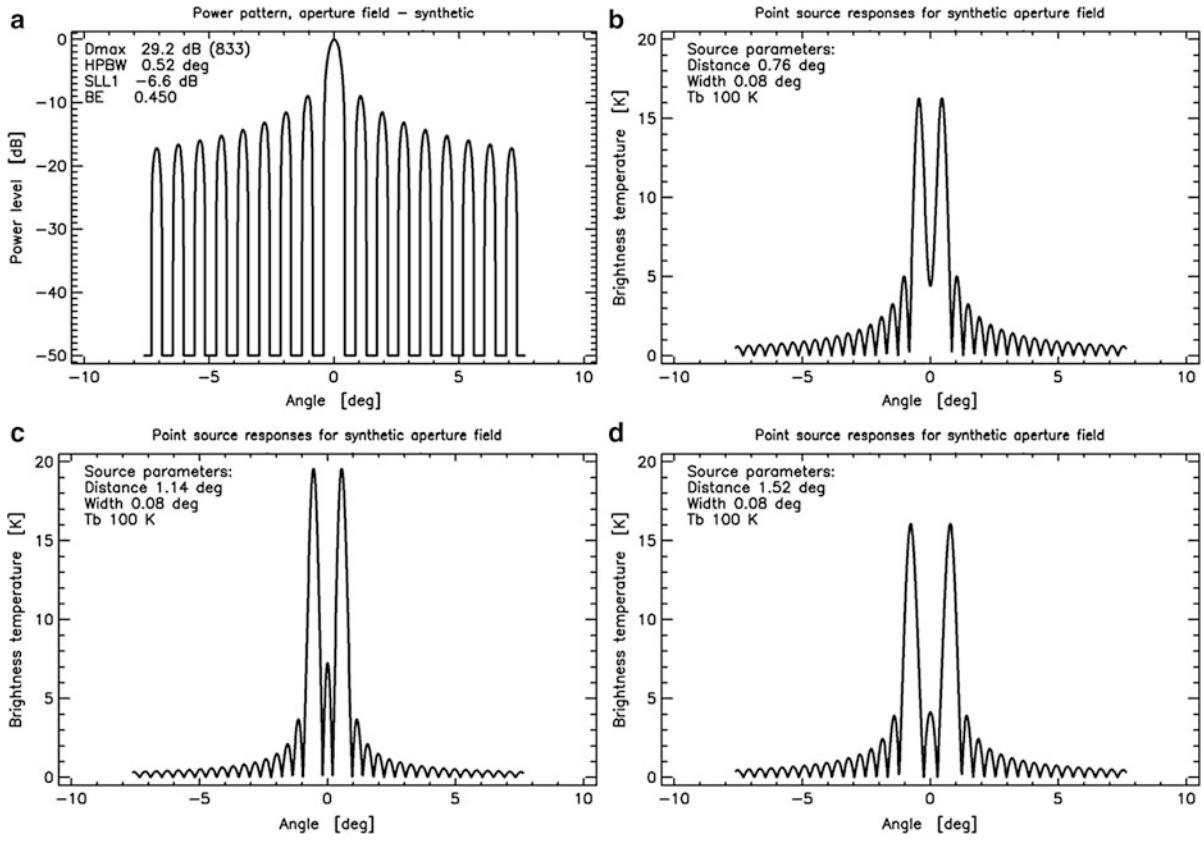


Fig. 28 Image (**a**): Normalized synthetic power pattern of an aperture synthesis system having identical parameters as the antenna types of Fig. 19. The spacing W here corresponds to the maximum baseline. Images (**b**–**c**): Point source responses of the identical situation as in Figs. 20 and 21. Image (**d**): Response for an additional separation of the two-point sources. Note that the magnitude of the brightness temperature is shown in (**b**–**d**) for better comparison (some side-lobe areas would have negative values)

angular regions in nature offering a well-known brightness temperature distribution. In the case of spaceborne radiometers, such an area is given by the galactic pole to which the instrument is pointed from time to time to perform the calibration measurement (Brown et al. 2008). Other regions on Earth are, for instance, the open ocean or regions on the Antarctic or the rain forest in South America. For ground-based systems, the sky region can be used under clear-sky conditions. However, in most of the last examples a model of the true brightness temperature distribution has to be used additionally in order to address angular dependencies of the brightness temperature. A more controlled calibration environment is given by the use of an anechoic chamber at known physical temperature, but in that case no in situ calibration during a later system operation is possible and only a characterization prior to a mission is feasible. A more detailed treatment of the impact of antenna patterns in aperture synthesis systems is discussed in Camps et al. (2006).

Sampling the U-V Plane and Spatial Resolution

Since the visibility function can only be sampled by a finite number of antennas in an aperture synthesis array, again the sampling theorem has to be considered. As described previously, the antenna power pattern determines the field of view to be imaged and hence the range of spatial coordinates in l and m direction to be considered. Assuming an ideally symmetric power pattern of maximum extensions l_{max}

DOI 10.1007/978-981-4560-75-7_125-1

and m_{max} in positive and negative coordinate direction, the sampling requirement for the u-v plane to avoid aliasing in the reconstructed image can be formulated in rectangular coordinates as follows:

$$\Delta u \leq \frac{1}{2\ l_{max}}; \quad -l_{max} \leq l \leq +l_{max}$$
$$\Delta v \leq \frac{1}{2\ m_{max}}; \quad -m_{max} \leq m \leq +m_{max} \tag{37a, b}$$

Note that Δu and Δv again are given in wavelengths. Considering now the full hemisphere for which the l-m representation is valid, i.e., $0° \leq \Theta \leq \pi/2$ and $0° \leq \Phi \leq 2\pi$, the values $l_{max} = 1$ and $m_{max} = 1$ hold and $\Delta u = \Delta v = 0.5$ results. Thus, the smallest spacing of the single-antenna elements has to be half a wavelength for imaging a full hemisphere. If, for instance, $\Theta_{max} = 30°$, then $l_{max} = m_{max} = 0.5$ and $\Delta u = \Delta v = 1$ is required. It should be noted that the single-element antenna pattern is not completely zero for values larger than l_{max} and m_{max}, and hence some aliasing may remain in the final image depending on the pattern power level at those ranges.

Since furthermore the dimensions of the synthesis array cannot extend to infinity, the spatial frequency spectrum has upper limits u_{max} and v_{max}. In accordance with the treatise of section Spatial Sampling Requirements, that instance results in a finite spatial resolution of the imaging process. If now an ideal point source at location l_0, m_0 is measured according to Eq. 35, the visibility function takes the form $vis(u,v) = a_0 \exp(-j(ul_0 + vm_0))$, where a_0 is a constant including all other prefactors. Inserting that visibility function in the inverse transform of Eq. 35, the result for the normalized reconstructed point source response is given by

$$F_{Pn,\,syn}(l,m) = \frac{\sin\left(2\pi\ u_{max}\ (l-l_0)\right)}{2\pi\ u_{max}\ (l-l_0)} \quad \frac{\sin\left(2\pi\ v_{max}\ (m-m_0)\right)}{2\pi\ v_{max}\ (m-m_0)}. \tag{38}$$

Note the notation $F_{Pn,syn}$ for the point source response, being similar to the one of the antenna power patterns of the spatially scanning radiometer systems. The reason for that is that, as previously mentioned, the antenna pattern can be determined by scanning a point source, and hence Eq. 38 is simultaneously interpreted as the synthesized antenna beam of the aperture synthesis array. The synthetic character can be recognized as well by the fact that it can have negative values according to the known sin(x)/x behavior, but it represents a power pattern. Furthermore, Eq. 38 represents a type of natural response, since no weighting was applied to the ideal visibility function prior to reconstruction. Note that Eq. 38 is the result of a continuous function in the u-v plane according to Eq. 34, which can be closely approximated by adequate sampling.

Now, by applying Eq. 11 to the synthesized beam of Eq. 38 again for the one-dimensional case, the imaging result for two adjacent point sources at different distances can be computed again for the previous examples. The synthetic beam and the result for three point source distances are shown in Fig. 28. Note the narrow beam width of 0.52° compared to the real-beam systems shown in Fig. 19 and the very high side-lobe level at −6.6 dB. Since the synthetic beam has negative side lobes, the computation of the directivity is a high number due to the cancelation of positive and negative side lobes in evaluating the integral of Eqs. 5b and 6. However, by considering the directivity as the number of point sources which can be resolved within 360°, this approach is useful. For computing the beam efficiency, the antenna solid angle in this case was computed by using the magnitude of the synthetic power pattern in Eq. 5b. This approach more realistically considers the degrading impact of the side lobes in the imaging quality by providing a rather low value for the beam efficiency.

DOI 10.1007/978-981-4560-75-7_125-1

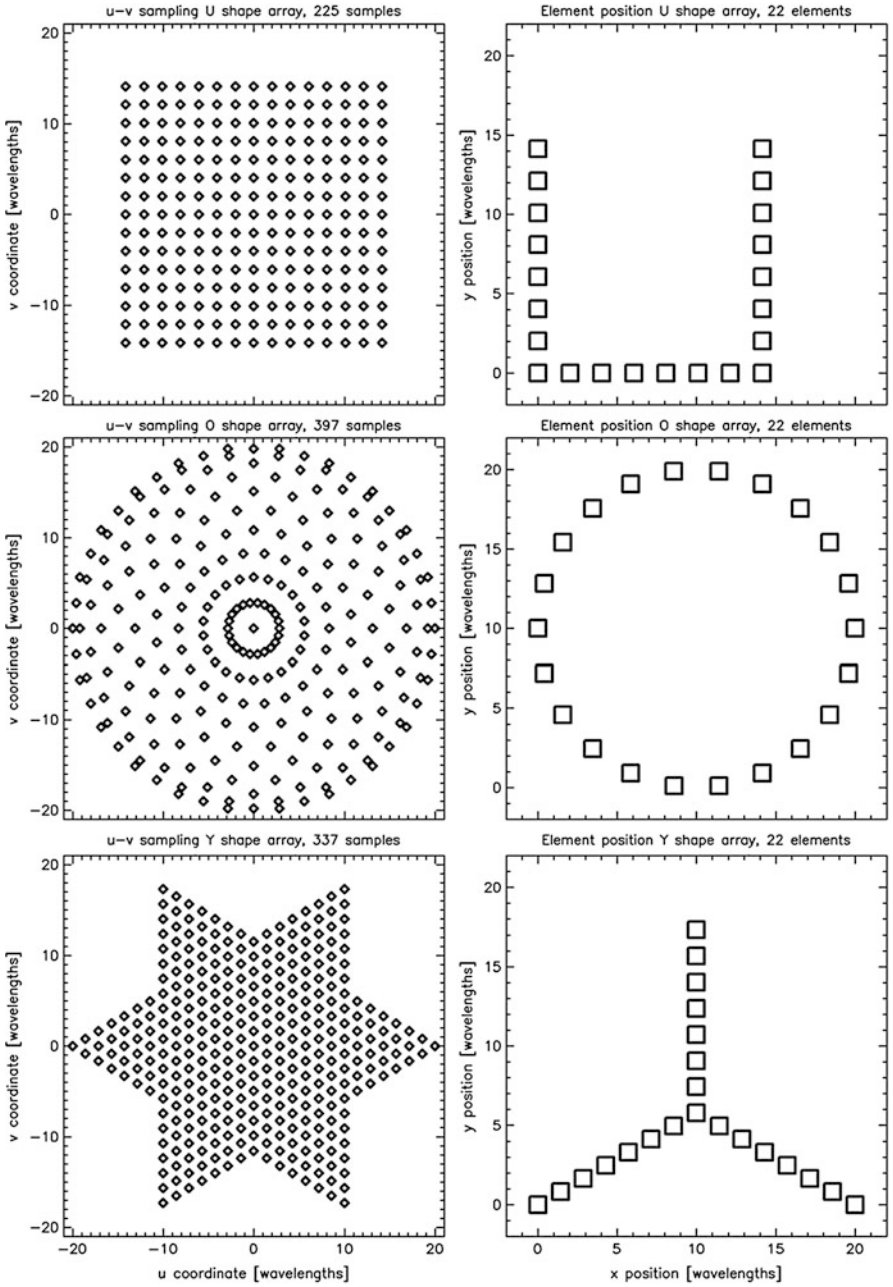


Fig. 29 Array configurations (*right*) for aperture synthesis (U, O, Y shape) and corresponding sampling patterns in the u-v plane (*left*). The maximum baseline for all configurations was set to 20 wavelengths and the number of antenna/receiver elements is always 22

The images (b)–(d) clearly show the result that for a synthetic imaging system of the identical size, the highest separation quality of the two sources can be achieved, thus providing superior capabilities for object detection in security applications, for instance. On the other hand, the high side-lobe contributions can degrade the absolute accuracy considerably. This can be observed by the fact that for the case of image (c), the maximum point source levels are higher than for situations (b) and (d) although this source separation is in between those two examples.

Now, it should be noted that the single-visibility samples can be weighted prior to image reconstruction, in order to manipulate the shape of the synthesized beam, by modifying Eq. 35:

DOI 10.1007/978-981-4560-75-7_125-1

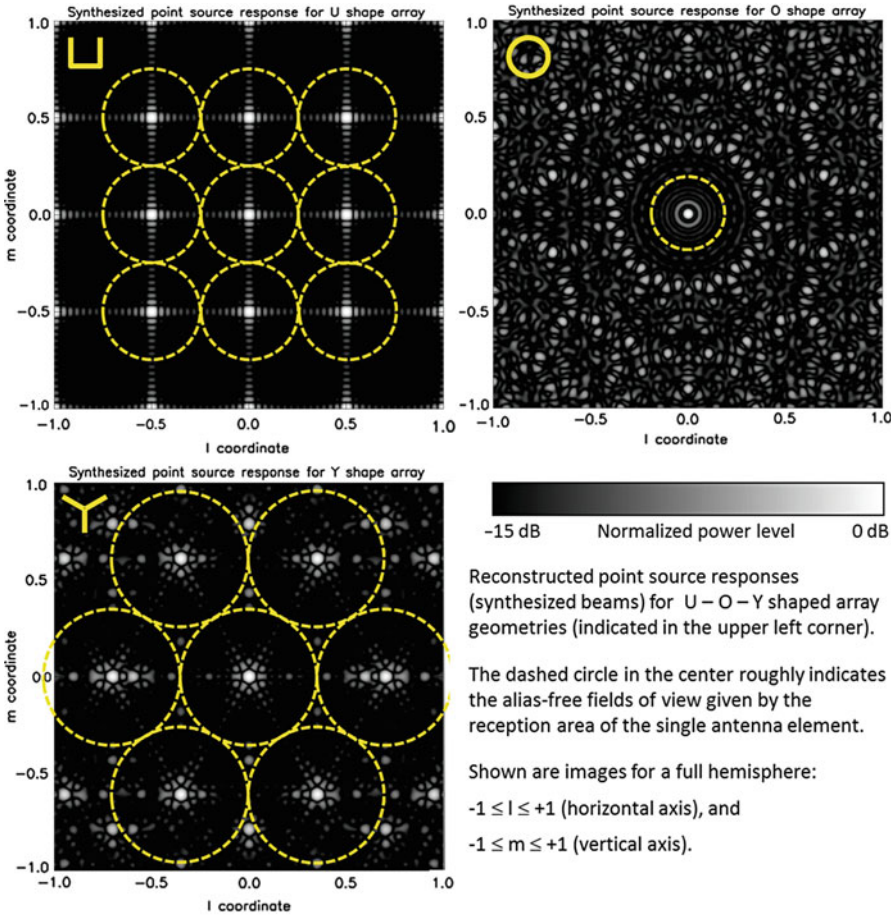


Fig. 30 Reconstructed point source responses for the aperture synthesis arrays of Fig. 29

$$T_B''(l,m) = \mathrm{FT}\{vis(u,v) \quad wsf(u,v)\}. \tag{39}$$

The function wsf(u,v) now can be understood as a twofold operator, one for weighting of the visibility amplitudes at locations (u, v), and another one for the shape of the sampling grid in the u-v plane. The weighting functionality follows exactly the well-known windowing operation of standard signal theory (Harris 1978), and the sampling functionality is briefly discussed next.

Array Design in Aperture Synthesis

Besides high spatial resolution, real-time capability, and large fields of view, the dominant power of aperture synthesis is given by the capability of using sparse arrays to synthesize a full aperture. Since each single element contributes to several baselines, an array can be designed to have low redundancy so that each baseline is only once existent, providing a unique visibility sample. On regular grids for the single elements, some redundancy will remain, but for the addition of some statistical shift in each position, all baselines will be at least slightly different. So, the goal of designing an array besides other constraints will be the maximum performance in synthesized beam shape by using the lowest number of single-antenna/receiver elements.

Now, Eq. 39 can be transformed to the spatial domain, resulting in a convolution of the brightness temperature $T_B'(l,m)$ and the weighting and sampling function WSF(l,m):

DOI 10.1007/978-981-4560-75-7_125-1

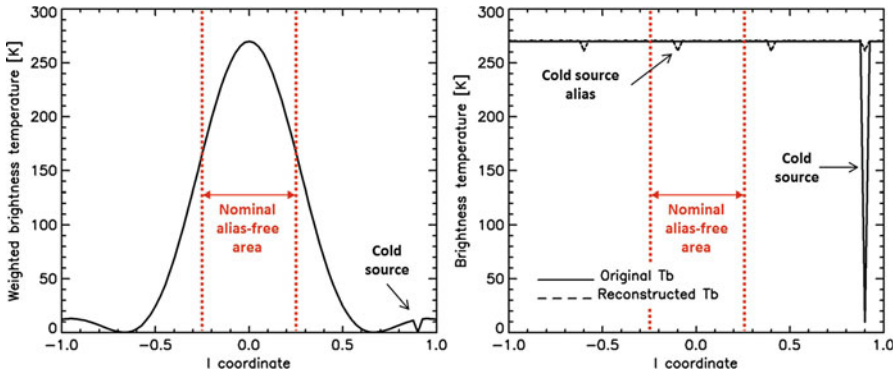


Fig. 31 *Left*: Simulated brightness temperature distribution of a scene weighted by an antenna power pattern as given by Eq. 12. The scene itself consists of a constant background at 270 K and a cold source of 10 K at l = 0.9; *right*: original and reconstructed brightness temperature distributions. For the synthesis array, an element size of W = 1.5λ, a constant distance of Δu = 2λ, and a maximum baseline of $u_{max} = 20\lambda$ have been assumed. The shape of the antenna pattern was corrected for in the reconstruction

$$T_B''(l,m) = \mathrm{FT}\{vis(u,v)\} * \mathrm{FT}\{wsf(u,v)\} = T_B'(l,m) * WSF(l,m). \tag{40}$$

The function WSF(l,m) now is the synthesized beam which the synthesis instrument can provide at best, assuming in other respects an ideal system as done before. If the sampling theorem is fulfilled and the weighting is constant for each u-v sample, the function WSF(l,m) becomes identical to the synthesized beam of Eq. 38. Now, the interesting question is which element locations, together with adequate weighting, deliver what type of synthesized beam. In Fig. 29 some simple array articulations and their corresponding sampling pattern in the u-v plane are shown.

For performance comparison, two important design parameters were set identically, the maximum baseline length of the array, i.e., the displacement between the two most distant elements, and the total number of elements. Three array geometries are considered, U, O, and Y shapes, indicating the basic array structure. As well the distance between adjacent elements is set constant along the basic structures. All three arrays generate very different sampling patterns in the u-v plane. The U shape produces a rectangular pattern and the lowest number of independent baselines, since due to the array geometry, a lot of identical baselines are produced, proving no additional spatial frequency information. The O geometry generates the largest synthesized area and no regular sampling pattern and the medium number of independent samples. The density of samples as well varies depending on the location in the u-v plane. The Y shape produces a hexagonal sampling grid, the highest sampling density in average, and the highest number of independent samples. The size of the synthesized area is in between the O and U cases.

For such arrays the synthesized beam or the point source response is computed for the whole l-m plane, i.e., $l^2 + m^2 \leq 1$, and it is shown in Fig. 30. The impact of the sampling pattern can be clearly identified by the different appearance of fully pronounced spatial repetitions of the point source responses, i.e., the aliases. The rectangular sampling lattice produces the aliases and the side lobes along the cardinal lines in horizontal and vertical direction. At other locations the side-lobe levels are smaller than for the two other cases.

The dashed circles are produced by the boundaries of the symmetrically assumed reception area of an antenna, just avoiding aliasing in the central region of the l-m plane, thus being the alias-free region of the imaging process. It can be observed that the Y array generates a replica of the point source on a hexagonal lattice and the largest alias-free field of view. Accordingly, the U-shape array produces a replica on a rectangular grid and a smaller alias-free field of view. A very narrow alias-free field of view is generated by the O array, since the average sampling density is the lowest one. Due to the irregular sampling grid, no

DOI 10.1007/978-981-4560-75-7_125-1

distinct replica of the point source response is visible, but a high number of lower-level sources and side lobes appear distributed over the whole l-m plane. The O array under the given conditions produces the highest spatial resolution (narrowest synthesized beam), since the highest number of large baselines is available.

It should be further noted that the responses are only shown down to −15 dB and that the high side-lobe levels may be not suitable for radiometric measurements. However, as mentioned before, standard windowing can be applied to reduce side lobes at the cost of spatial resolution. In no case of course, the aliases can be suppressed by such means. As a conclusion the array design has large impact on the imaging quality of an aperture synthesis instrument. Depending on the application, the largest field of view and the highest spatial resolution together with the acceptable side-lobe level have to be balanced against each other for the most suitable design. Several other geometries than the shown ones have been investigated and applied in the past (Thompson et al. 2001; Hall 2006; Jirousek et al. 2010; Dong and Li 2010; Schuetz et al. 2013).

The sampling requirements according to Eq. 37a,b require in theory that from outside the limits given by l_{max} and m_{max}, no radiation is received. Since the visible scene is determined by the single-element antenna pattern, this requirement cannot be fulfilled in practice and a certain nonzero decay of the pattern outside that boundary (i.e., outside the centric circles in Fig. 30) has to be accepted. Mostly, aperture antennas are used for synthesis arrays so that a first-order approximation of such an antenna pattern can be given by Eq. 12, considering a one-dimensional analysis for the moment. If now the visible scene limit is set to the first null of the antenna pattern and by using Eq. 37a,b, it follows

$$l_{\max} = \sin \Theta_{\max} = \frac{\lambda}{W}, \qquad \Delta u = \frac{\Delta x}{\lambda} \leq \frac{1}{2\ l_{\max}} \text{ or } \Delta x \leq \frac{W}{2}. \tag{41a, b}$$

Hence, the metric distance Δx between the spatial frequency samples should be at least half the aperture size W of the antenna element, which cannot be fulfilled everywhere in the u-v plane. Physically, only $\Delta x \geq W$ can be realized as minimum spacing. Possible aliasing impacts on synthesis imaging shall be illustrated exemplarily in the one-dimensional simulation of Fig. 31.

The original brightness temperature distribution weighted by the antenna power pattern, as it is observed by the single element of the synthesis array, is shown on the left. The nominal alias-free area as given by Eq. 37a,b is indicated at $l = \pm 0.25$. A cold source is located significantly outside the alias-free area but it is folded back in the alias-free range in the reconstructed image. This is caused by the fact that the antenna pattern is not zero outside the alias-free range due to its finite aperture size of $W = 1.5\lambda$ and that the element spacing of the synthesis array was chosen on a regular grid at a distance of $\Delta u = 2\lambda$.

As a conclusion from the previous analysis, the following recommendations can be formulated:

- The use of antenna elements whose pattern sufficiently vanishes outside the demanded alias-free field of view to negligible values. This requirement can hardly be fulfilled. Only antenna types whose physical width or diameter can be made sufficiently small in parallel to producing high directivity and low side-lobe levels are valid. Useful antenna types are based on dielectrics like microstrip-based antennas or dielectric rods. However, it should be noted that such antennas in a close arrangement can produce strong mutual coupling (▶ Suppression of Mutual Coupling between Elements in Multiple Element Antennas), which can influence the antenna pattern and the power match in an unacceptable way (reconstruction errors, loss of sensitivity).

- No use of rectangular sampling grids since possible aliases superimpose constructively in the nominal alias-free region. Improved behavior is achieved by a hexagonal sampling grid in the two-dimensional case due to the higher sample density or by non-regular sampling grids, since in that case possible aliases distribute across the whole nominal alias-free area and thus do not impact massively a certain region.

Sensitivity in Aperture Synthesis

In contrast to real-beam imaging systems, an aperture synthesis radiometer generates an image by contributions from all individual antenna/receiver elements to each pixel. Hence, the same is valid for the noise content of the image. Furthermore, a synthesis array artificially produces an effective antenna aperture according to its focused beam, but the physical radiation collecting area is much smaller or, in other words, the number of collected microwave photons from the scene for a certain duration is less. Obviously, that instance has impact on the final temperature resolution of the reconstructed image. The corresponding equation for an aperture synthesis instrument can be approximated by (Le Vine 1990)

$$\Delta T = \frac{T_{sys}}{\sqrt{2\,B\,t_{frame}}}\sqrt{\frac{2}{n\ (n-1)}}\frac{\Omega_{FOV}}{\Delta\Omega} \approx \frac{T_{sys}}{\sqrt{B\ t_{frame}}}\frac{1}{n}\frac{\Omega_{FOV}}{\Delta\Omega}. \tag{42}$$

Here, identical quantities have been assumed for each antenna/receiver, i.e., the system noise temperature T_{sys}, the system bandwidth B, and the antenna element solid angle Ω_{FOV} defining the scene size. The quantity $\Delta\Omega$ represents the solid angle of the synthesized beam, and the number n is the number of antenna/receiver elements given the case that $n \gg 1$ is large and that only nonredundant baselines are accounted for. Note that zero redundancy of a synthesis array is only possible for a small number of elements and that redundancy counts for large n and regular element arrangements have single cluster points instead of a uniform distribution. Hence, the redundancy is often not used for improving the sensitivity in the image in order to avoid entering other errors due to element imbalances. Furthermore, Eq. 42 assumes that no errors are produced by aliasing impacts as described previously. As an example consider the SMOS system (Barré et al. 2008; Font et al. 2010) for $T_{sys} = 450$ K, B = 19 MHz, $t_{frame} = 1.2$ s, n = 69, $\Omega_{FOV} \approx 70° \times 70°$, and $\Delta\Omega = 1.5° \times 1.5°$, resulting in a temperature resolution of $\Delta T \approx 3$ K, which is well in accordance with predicted and measured values.

A comparison of Eq. 32 for the real-beam case and Eq. 42 for the synthetic-beam case results in

$$\Delta T_{syn} = \frac{1}{2\,k\,n}\quad\sqrt{\frac{\Omega_{FOV}}{\Delta\Omega}}\quad\Delta T_{real}. \tag{43}$$

Thus, the sparse array character of aperture synthesis takes effect such that the synthetic-beam sensitivity ΔT_{syn} is degraded by the square root of the ratio of scene solid angle and beam solid angle compared to the real-beam sensitivity ΔT_{real}. That disadvantage is compensated to some extent by two times the number of elements used for the synthetic array. As a rough tendency it should be noted that extreme thinning of the synthesis array and large fields of view at very high spatial resolution will degrade sensitivity, unless some significant redundancy in form of increasing the number of elements (without increasing additionally spatial resolution) is accepted.

DOI 10.1007/978-981-4560-75-7_125-1

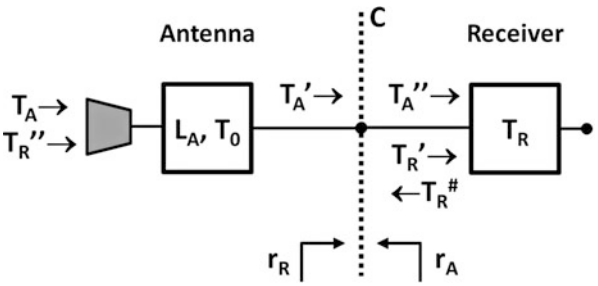


Fig. 32 Simple model for noise computation of the antenna/receiver interaction considering antenna losses due to ohmic or dielectric attenuation and reflection losses due to impedance mismatch (▶ Impedance Matching and Baluns). L_A and T_0 represent the attenuation factor and physical temperature of the antenna, T_R is the receiver noise temperature, and r_A and r_R are the complex voltage reflection coefficients at reference plane C as to be measured toward the antenna and toward the receiver. *Arrows* indicate flow direction of the temperatures or noise power, respectively

Losses of the Antenna System

Sufficient sensitivity of a radiometric imaging system, i.e., sufficient temperature resolution in the final image product, requires a low-noise figure or noise temperature of the system itself, according to Eqs. 30, 32, and 42. The noise contribution of the imaging system has basically two sources: (a) dissipation losses and (b) reflection losses (▶ Transmission Lines). The fundamental rule for computing the overall noise behavior of a receiving system is given by the cascade formula for noisy two-ports (Pozar 2005):

$$T_{cas} = T_{e1} + \frac{T_{e2}}{G_1} + \frac{T_{e3}}{G_1 G_2} + \ldots \tag{44}$$

Here, T_{cas} is the total or equivalent noise temperature of a serial connection of noisy two-ports to be located at the input of the first two-port, assuming in this case all two-ports to be noise-free. The quantities T_{ei} and G_i represent the equivalent noise temperatures and power gains of the single two-ports. It is evident that for the lowest noise temperature T_{cas}, the first components dominate and should have a low-noise temperature and high gain. Note that the term gain can mean either amplification, i.e., $G > 1$, or attenuation, i.e., $0 \leq G \leq 1$. In the latter case it is common to use the reciprocal of G, i.e., $L = 1/G$ with $1 \leq L \leq \infty$, and call it attenuation factor or loss factor.

Since the antenna system is the first component of any radiometric sensing system, its losses are of special interest. Usually, the antenna is a passive microwave structure built to physically collect free-space waves and guide those to a following transmission line and the receiver input. Thus, the basic antenna is an attenuator and its equivalent noise temperature can be described by (Ulaby 1982)

$$T_{att} = (L_{att} - 1)T_0, \tag{45}$$

where L_{att} is the power attenuation factor and T_0 is the physical temperature of the antenna or attenuator, respectively. Furthermore, both, the antenna and the receiver, have individual complex input impedances leading to complex individual power reflection coefficients (▶ Impedance Matching and Baluns), where the latter of course usually should be as low as possible.

Impact of Dissipation and Reflection Loss

In order to estimate both loss types for a radiometer antenna – dissipation and reflection losses –, the equivalent block diagram of Fig. 32 can be used.

The dissipation loss is considered by the single-primed antenna temperature T_A' according to

DOI 10.1007/978-981-4560-75-7_125-1

Table 1 Computed power transfer efficiency p for various values of power reflection coefficients. The worst case has been assumed for which the phase difference of the product $r_A \times r_R$ in the denominator of Eq. 47b is 180°

$\lvert r_A\rvert^2$ [dB] $\lvert r_R\rvert^2$ [dB]	−10	−20	−∞
−10	0.67	0.84	0.90
−20	0.84	0.96	0.99
−∞	0.90	0.99	1.00

$$T_A' = \frac{1}{L_A}\ T_A + \left(1 - \frac{1}{L_A}\right) T_0. \tag{46}$$

Consequently, for minimum dissipation loss $L_A = 1$, the true antenna temperature is transferred to the reference plane C, i.e., $T_A' = T_A$, and for very large losses $L_A \to \infty$, the true antenna temperature is completely blocked, i.e., $T_A'' = T_0$. Usually, the loss factor will be close to 1 for practical antennas but not identical to 1.

The reflection loss can be determined by the double-primed antenna temperature T_A'' following (Otoshi 1968)

$$T_A'' = p\ T_A' + (1-p) T_R^{\#},$$
$$p = \frac{\left(1 - |r_A|^2\right)\left(1 - |r_R|^2\right)}{|1 - r_A r_R|^2}. \tag{47a, b}$$

The quantity p represents the power transfer efficiency determined by the mismatch between the antenna and the receiver reflection coefficients. For ideal match, i.e., $r_A = r_R^*$, the efficiency becomes maximum, i.e., $p = 1$ and $T_A'' = T_A'$. If one or both power reflection factors are maximum, i.e., $|r_A|^2 = 1$ or $|r_R|^2 = 1$, the efficiency becomes minimum, i.e., $p = 0$ and $T_A'' = T_R^{\#}$. Obviously, the practical cases are somewhere in between and the efficiency should become close to one for minimum reflection losses.

The quantity $T_R^{\#}$ represents the noise temperature generated by the receiver and transmitted toward the reference plane C. Two special cases of receiver design have to be discriminated for its value. In the first

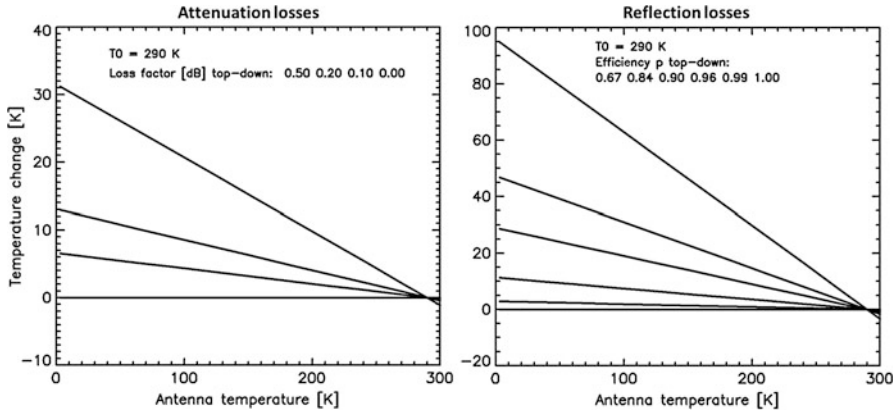


Fig. 33 Change in the antenna temperature T_A'' at the input of the receiver according to Fig. 32. On the left side only dissipation losses have been considered, i.e., $p = 1$. On the right side only reflection losses were simulated, i.e., $L_A = 1$. The values of the efficiency p correspond to the reflection coefficients of Table 1

Table 2 Computed values for receiver noise temperatures at different locations according to Fig. 32 and ratios of temperature resolution including antenna losses ΔT_A over temperature resolution without antenna losses ΔT. ΔT and ΔT_A are computed according to Eq. 30 for k = 1. Parameters are as follows: $L_A = 1.047$ (0.2 dB), $p = 0.84$, and $T_0 = 290$ K

L_I [dB]	T_R [K]	T_R' [K]	T_R'' [K]	$\Delta T_A/\Delta T$ $T_A = 300$ K	$\Delta T_A/\Delta T$ $T_A = 3$ K
No isolator	30	30	57	1.08	1.82
No isolator	300	300	447	1.25	1.49
No isolator	1,000	1,000	1,460	1.35	1.46
1.0	30	113	129	1.04	1.14
1.0	300	453	483	1.04	1.07
1.0	1,000	1,334	1,400	1.04	1.05

case, the first component of the receiver directly at the input is an isolator with a low forward loss factor L_I assumed to have as well a physical temperature T_0. Isolators are often used to improve impedance match (▶ Impedance Matching and Baluns) and to block local oscillator leakage from a possible mixer toward the antenna. Since the isolator has a high attenuation in its backward direction toward the antenna, its equivalent noise temperature is T_0, and hence $T_R^{\#} = T_0$ is valid. In order to avoid additional losses and for very broadband receivers, the isolator sometimes is omitted and then $T_R^{\#} = T_R$ is valid.

Combining Eqs. 46 and 47a,b now allows considering the full impact of all loss types together:

$$T_A'' = p\ \left(\frac{1}{L_A}\ T_A + \left(1 - \frac{1}{L_A}\right) T_0\right) + (1-p)\ T_R^{\#}, \tag{48}$$

$$T_R^{\#} = \begin{cases} T_0 & \text{with isolator,} \\ T_R & \text{without isolator.} \end{cases} \tag{49}$$

In the following the single impact of both loss types shall be illustrated graphically. For the attenuation factor of the antenna, practical values of 0.5, 0.2, 0.1, and 0.0 dB have been considered. The power reflection coefficients were selected according to Table 1.

Figure 33 shows the change of the antenna temperature T_A'' at the input of the receiver in dependence of the true antenna temperature at the antenna free-space terminal and the two different sources of loss. An isolator at physical temperature T_0 on the input of the receiver has been considered. In both loss cases a significant increase of the modified antenna temperature can be observed, especially for low levels of the true antenna temperature. That instance is important for calibration issues when the antenna is not part of the calibration loop, which is the case when internal loads are used and fed to the receiver by switches. In that case, a very precise knowledge of the attenuation and reflection behavior of the antenna/receiver combination is mandatory. If calibration is performed using external reference sources to be observed by the antenna, the losses are included in the calibration process and do not distort correct brightness temperature measurements. However, it is evident that antenna dissipation in the order of 0.1 dB or less and power reflection coefficients of at least −20 dB are recommended.

The other important impact of antenna losses is the increase of the equivalent receiver noise temperature T_R'' at the antenna input terminal according to Fig. 32, leading on the one hand to a decrease in sensitivity or temperature resolution and on the other hand to a reduction of the receiver's dynamic range. Since the total noise temperature directly at the receiver input is $T_A'' + T_R'$, that quantity can be transformed to the antenna input, resulting in $T_A + T_R''$, where T_R'' can now be computed as follows:

DOI 10.1007/978-981-4560-75-7_125-1

Table 3 Conductivity values for different metals

Metal type	Silver (AG)	Copper (CU)	Gold (AU)	Aluminum (AL)	Brass (BR)	Stainless steel (SS)
σ [S/m] $\times 10^7$	6.17	5.81	4.10	3.82	1.50	0.11

$$T_R'' = (L_A - 1)\ T_0 + L_A\left(\frac{1}{p} - 1\right) T_R^{\#} + \frac{L_A}{p} T_R', \tag{50}$$

$$T_R' = \begin{cases} T_R & \text{without isolator,} \\ (L_I - 1)\ T_0 + L_I\ T_R & \text{with isolator.} \end{cases} \tag{51}$$

In case of an isolator at the receiver input, the quantity L_I is again the power loss factor and its physical temperature is assumed to be T_0. In the following Table 2, a few values for practical numbers are given to illustrate the impact.

It can be recognized that the isolator lowers the impact of changing receiver noise temperatures on the degradation of temperature resolution, since for nearly all six cases in the lower-right edge of Table 2, the ratio is quite similar. However, it has to be noted that the original receiver noise temperature T_R is always increased by the isolator according to Eq. 51, being an attenuator as the first input component, and thus the sensitivity ΔT is generally degraded. If no isolator is used, especially the impedance mismatch losses (▶ Impedance Matching and Baluns) can produce a stronger dependence of the sensitivity on the actual antenna temperature. As a rule of thumb, the antenna losses can produce significant degradation of temperature resolution except very low levels $L_A < 0.1$ dB, and high power transfer efficiency $p > 95$ % (see Table 1 for reflection coefficients) can be achieved.

Magnitude of Antenna Dissipation Loss

Generally, it is hard to determine precisely the power dissipation, i.e., ohmic and dielectric losses, of an antenna, especially for the desired low-loss conditions. In fact this quantity is included in the ratio of antenna directivity and antenna gain, but mostly only the gain is measured and the directivity is computed numerically from electromagnetic models. So, the dissipation loss cannot be estimated with sufficient precision. Since radiometric measurements are very sensitive to antenna losses, as pointed out before, they are suitable to perform this task. It should be remembered that even a dissipation loss of 0.01 dB can produce an increase of the antenna temperature of about 0.7 K according to Eq. 46 for $T_A = 10$ K and $T_0 = 300$ K, which can be measured by a properly calibrated radiometer with sufficient accuracy.

The range of antenna systems used for radiometric measurements is quite large as indicated already in section Overview on Radiometer Antennas. There are pure feed-horn-reflector systems in fully mechanical scanners (▶ Reflector Antennas), folded optical systems using quarter-wave plates for focal plane arrays, sparse arrays of patch (▶ Microstrip Patch Antennas) or horn antennas for aperture synthesis, slotted waveguides for frequency scanning (▶ Waveguide Slot Antennas and Arrays, ▶ Tapered Slot Antennas, ▶ Travelling-wave Antennas (Including Leaky Wave and Surface Wave Antennas)), or lens-based approaches (▶ Dielectric Lens Antennas) for multi-beam generation just to mention few of them. Depending on the technology, the antenna dissipation losses can be higher or lower depending on the media (metals and dielectrics) and the path lengths along which the waves have to interact with those. Lowest losses should be expected for pure metallic interaction like for parabolic reflectors and horn feeds, for which a theoretical estimate shall be given next.

For antennas being small compared to the wavelength like short dipoles or narrow loops, the attenuation factor can be computed by

DOI 10.1007/978-981-4560-75-7_125-1

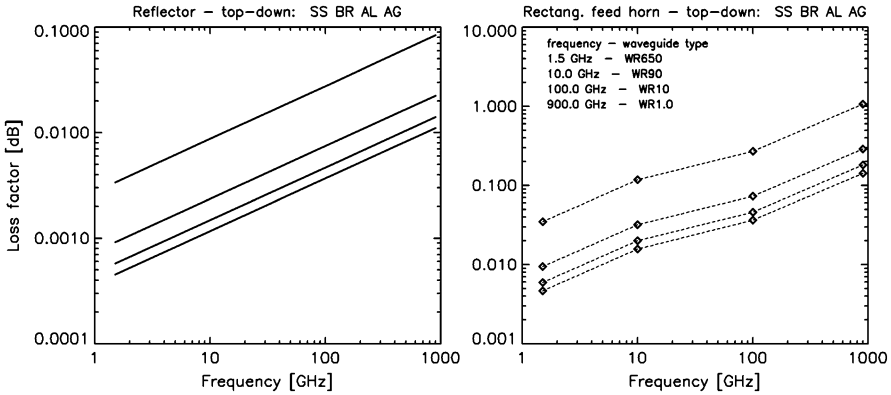


Fig. 34 Computed power loss factors of the reflector part (*left*) and the feed horn (*right*) of a reflector antenna for various metal types as a function of frequency. The metals are indicated as given in Table 3. For the horns standard rectangular waveguides have been chosen according to the frequency, and a length of 5λ has been assumed

$$L_A = \frac{R_{rad} + R_{dis}}{R_{rad}}, \tag{52}$$

where R_{rad} is the radiation resistance accounting for the really radiated/received power conversion and R_{dis} is the dissipation resistance representing power conversion by ohmic and dielectric losses. Note that the sum of both resistances represents the real part of the complex antenna impedance (▶ Impedance Matching and Baluns). For electrically small antennas, analytic expressions for both quantities exist (Stutzman and Thiele 1998), but unfortunately those antennas are rather not used for radiometric applications due to their relatively high dissipation loss and badly shaped antenna pattern. For very low-loss reflector antennas (▶ Reflector Antennas), an experimental approach is described in Skou (1997) using radiometric sky measurements. According to this investigation, the power reflectivity of a well-conducting smooth surface is given by

$$r = 1 - \frac{4R_S}{Z_0}, \tag{53a}$$

$$R_S = \sqrt{\frac{\pi f \mu}{\sigma}}, \qquad Z_0 = \sqrt{\frac{\mu}{\varepsilon}}. \tag{53b, c}$$

The quantity R_S is the surface resistivity of the reflector material and Z_0 is the intrinsic impedance of the medium in which the wave propagates. In case of air or vacuum and a non-ferromagnetic reflector material, the free-space permeability $\mu = \mu_0 = 4\pi \times 10^{-7}$ H/m and the free-space permittivity $\varepsilon = \varepsilon_0 = 8.854 \times 10^{-12}$ F/m can be used and $Z_0 \approx 377\ \Omega$ is valid. The quantity σ is the conductivity of the reflector material and some values for different metals are given in Table 3.

Using Eq. 1 and $t = 0$, and the relation $e = 1 - 1/L_{A,refl} = 1 - r$, the power attenuation factor for a well-conducting reflector now can be computed:

$$L_{A,refl} = \left(1 - \frac{4R_S}{Z_0}\right)^{-1}. \tag{54}$$

Since the reflector antenna furthermore consists of the reflecting surface and a feed antenna, the loss factor of the latter one has to be estimated as well. For that purpose the impact of the feed is approximated by the

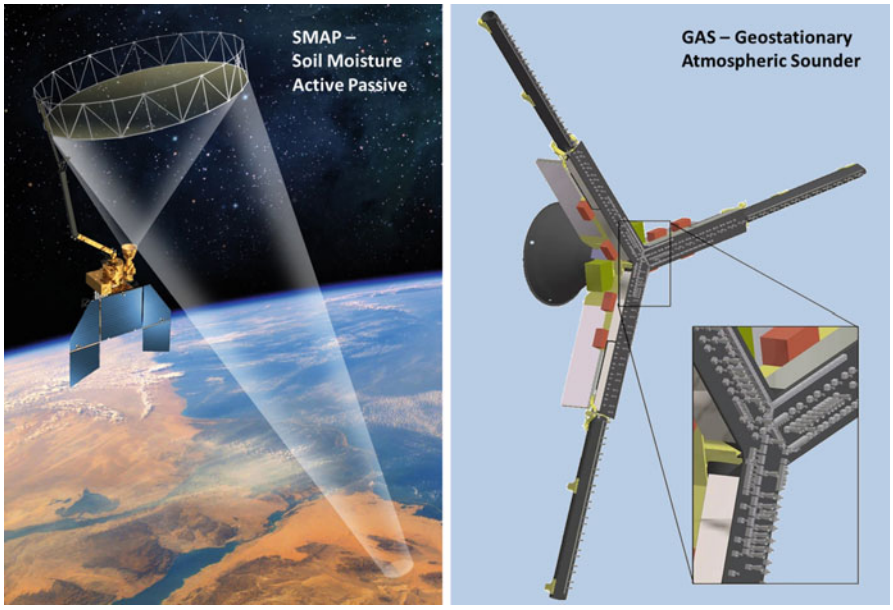


Fig. 35 *Left*: artist view of the SMAP conical scanner system in space (Courtesy of NASA/JPL-Caltech, USA); *right*: drawing of the GAS Y-shaped antenna array. Each frequency band is realized as a separate Y array scaled for identical spatial resolution (Courtesy of ESA-ESTEC, Netherlands)

attenuation of a piece of rectangular waveguide (▶ Transmission Lines), which is representative if the flare angle of such a horn antenna opens smoothly over a certain length Δz. In that case the wall currents in the flare region are similar to those of the waveguide, and the attenuation constant α and the loss factor $L_{A,\text{feed}}$ for the dominant mode TE_{10} are estimated by (Orfanidis 2008)

$$\alpha = \frac{R_S}{Z_0}\left(\frac{1}{b} + \frac{2}{a}\left(\frac{\lambda}{2a}\right)^2\right)\left(1 - \left(\frac{\lambda}{2a}\right)^2\right)^{-1/2}, \tag{55}$$

$$L_{A,\,\text{feed}} = \exp(2\alpha\,\Delta z). \tag{56}$$

The quantities a and b are the width and height of the rectangular waveguide and λ is the free-space wavelength. Thus, the total power loss factor L_A for the reflector antenna is finally estimated to be

$$L_A = L_{A,\,refl}\;\; L_{A,\,\text{feed}}. \tag{57}$$

Figure 34 shows computed results of L_A for the reflector material and the pseudo feed horn as described before. It has to be noted that the total loss of both antennas is given by Eq. 57 and thus by the sum of the decibel values from the figures. As can be observed, the total loss factor is dominated by the feed horn and can reach values of 0.1 dB at high millimeter-wave or even more at submillimeter-wave frequencies (▶ Millimeter-Wave Antennas and Arrays, ▶ Mm-Wave Sub-mm-wave Antenna Measurement, ▶ Photomixer Antennas at Terahertz, ▶ Terahertz Antennas and Measurement). Hence, even the most lossless antenna structure, as the purely metallic reflector antenna is, has to be taken into account when high absolute accuracy in brightness temperature measurements is required. Fortunately, the feed horn has most impacts, and it can be included much easier in external calibration loops as the whole reflector due to size constraints. However, it should be noted that the computation of the reflector losses assumes that the surface is sufficiently smooth with respect to the wavelength and clean and not oxidized, respectively. In

DOI 10.1007/978-981-4560-75-7_125-1

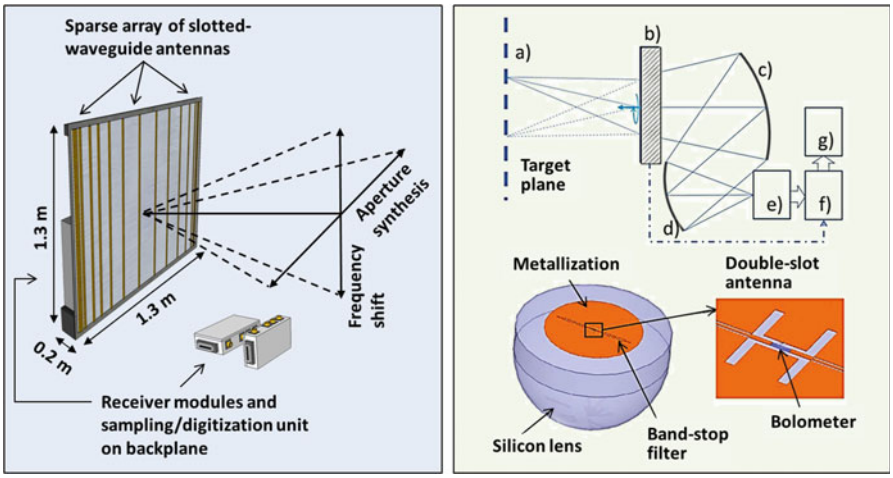


Fig. 36 *Left*: schematic of a hybrid fully electronic scanner at Ka band using aperture synthesis for horizontal and frequency scanning for vertical beamsteering; *right*: focal plane array scanner operating at three submillimeter-wave bands around 250, 450, and 720 GHz. The upper diagram shows the system architecture consisting of (**b**) rotating scan unit for conical beam generation, (**c**) primary focusing mirror (54 cm size), (**d**) secondary mirror, (**e**) closed-cycle cryostat incorporating 1,868-element focal plane array, (**f**) data acquisition, and (**g**) processor. The lower images show one channel of the focal plane array consisting of the integration of a silicon lens, a double-slot antenna, and a bolometer as the radiation detector (Courtesy of Asqella Oy, Finland, ▶ On-Chip Antennas, ▶ Photomixer Antennas at Terahertz, ▶ Terahertz Antennas and Measurement)

reality, the estimated values can be even much higher due to certain surface roughness being typically in the order of up to few micrometers.

Summary

The previous sections have demonstrated that a multitude of antenna types and antenna systems are used and can be used for radiometric applications in microwave remote sensing. The main candidates are mechanical scanners using typically moving reflector antennas (▶ Reflector Antennas), focal plane arrays using folded microwave optics and horn arrays to produce multiple real antenna beams (▶ Multi-Beam Antenna Arrays, ▶ Radio-Frequency Beam-Forming for Scanned and Multi-Beam Antenna Systems), and aperture synthesis where sparse arrays form synthetically multiple antenna beams based on sampling the spatial frequency spectrum. Promising as well is digital beamforming where the signals of the single channels of an array are sampled and digitized directly and the beam generation is as well done synthetically later in a digital processor. Since this type of sampling an aperture allows highest flexibility in signal processing, a multitude of beam-shaping approaches like in aperture synthesis or phased-array imaging is possible. The present bottleneck here is the amount of data to be stored and processed which can be tremendous for large antenna systems and high bandwidths. Lens-based imaging systems producing multiple real beams are rather rarely used but may become more attractive when highly integrated receiver modules (systems on chip) are available from mass production at affordable cost (▶ On-Chip Antennas).

Independent of the imaging principle, care must be applied to the antenna system design, since it determines the spatial resolution, the field of view or scene size, the amount of aliasing impacts, and the imaging time. In addition, since the antenna is the first component of the radiometer system, it can have a large influence on the sensitivity by dissipation and reflection losses in the case of finite conductivity and/or dielectric attenuation and impedance mismatch between antennas and receivers (▶ Impedance

Matching and Baluns). Although such quantities in the value range of tenths or hundredths of decibels for low-loss antenna structures are not a problem for radar systems or communication links, the error in brightness temperature estimation can be too high since the signals to be measured are similar in power to the radiometer's own noise floor. Hence, the antenna should be incorporated in the calibration loop, if possible, but this is very often not feasible due to its size. In that case precise measurement on component level and sophisticated modeling is required to account for these impacts mathematically.

In Earth observation from space, next-generation systems are still using mechanical scanner technology as it is intended for the National Aeronautic and Space Administration - Jet Propulsion Laboratory (NASA-JPL) SMAP mission to be launched early in 2015 (Entekhabi et al. 2010). Soil Moisture Active Passive (SMAP) is using a 6-m mesh reflector (► Reflector Antennas, ► Space Antennas including Terahertz Antennas) in a conical scan to determine soil moisture over land passively and actively at L-band frequencies, achieving a 40-km ground resolution from a 685-km orbit. An artist view of the system in space is shown in Fig. 35 (left). NASA and ESA both performed studies (Geostationary Synthetic Thinned Aperture Radiometer (GeoSTAR) and Geostationary Atmospheric Sounder (GAS)) on using aperture synthesis technology for high-frequency Earth observation from a geostationary orbit, i.e., at about 36,000 km distance (Tanner 2007; Christensen et al. 2007, ► Space Antennas including Terahertz Antennas, ► Terahertz Antennas and Measurement). Using bands around 50–56, 118, 183, and 380 GHz for the monitoring of atmospheric temperature, humidity profiles, and rain mapping, the Y-shaped arrays have an arm length of maximum 4 m, achieving a ground resolution of about 30 km for the lowest frequency. A schematic of the GAS antenna array is shown in Fig. 35 (right). While the GeoSTAR system is intended to use a stationary Y-shaped array requiring a very high number of single-antenna elements for proper u-v-plane sampling, the GAS system is using a highly thinned around boresight rotating Y array in order to fill up sufficiently the u-v plane. For both systems, the realization is not known yet and a possible launch could be toward 2020.

In a more distant future, it could be possible and attractive for spaceborne Earth observation to use multiple satellites on different orbits for synthesizing very large antenna apertures. Each single satellite can carry an autonomous aperture synthesis array in order to perform like SMOS, but it can be synchronized in phase with the other satellites to build a large aperture when the satellites have a suitable constellation in orbit. Thus, selected areas on the ground could be imaged at very high resolution, provided the satellite constellation can be adjusted accordingly. Such an approach is similar to very-long-baseline interferometry (VLBI) from radio astronomy (► Antennas in Radio Telescope Systems), where array sizes of thousands of kilometers are achieved by connecting coherently several single radio telescopes across the world (Felli and Spencer 1989; Thompson et al. 2001).

Another potential present and future application area for radiometric imaging is given by the security and safety market, where each type of new information source of physically different nature is of big interest. Here, very high spatial resolution and real-time capability are often the main drivers for system design, being especially useful for enhanced vision in surveillance, reconnaissance, or status control in general (e.g., assisted viewing for pilotage, detection of concealed objects, indication of potential dangers, etc.). In order to avoid too large antenna structures, very often millimeter waves or submillimeter waves have to be used (► Millimeter-Wave Antennas and Arrays, ► Mm-Wave Sub-mm-wave Antenna Measurement) if sufficient penetration to media of interest is guaranteed (e.g., atmospheric obstacles, dielectric covers, etc.). The real-time requirement demands for systems of high frame rates, being only feasible by electronically scanning concepts or hybrids of those, and some optional mechanical scanning support.

Figure 36 (left) shows a schematic of an approach combining aperture synthesis and frequency scanning in order to reduce considerably the number of required channels (Schreiber 2014). Such a system is intended for large-area surveillance at longer distances of up to few hundred meters. For a first

DOI 10.1007/978-981-4560-75-7_125-1

system design, it is estimated that by using about 35 slotted waveguide antennas (► Waveguide Slot Antennas and Arrays, ► Tapered Slot Antennas) and receivers, a field of view of 120° × 30° (aperture synthesis × frequency scan direction) at a resolution of 0.4° can be scanned in Ka band (&30-40 GHz). By using adequate high-speed samplers and digitizers, a frame rate of up to 10 f/s is expected for 2 K sensitivity.

Figure 36 (right) shows an approach for three-band imaging at millimeter and submillimeter waves based on a focal plane array consisting of cryogenically cooled integrated antenna-bolometer structures (Luukanen et al. 2013, ► Millimeter-Wave Antennas and Arrays, ► Mm-Wave Sub-mm-Wave Antenna Measurement, ► Photomixer Antennas at Terahertz, ► Terahertz Antennas and Measurement). The system is intended for concealed weapon detection on persons at a close distance of a few meters. The vertical arrangement of the single channels allows in combination with a rotating beam tilting unit to scan an area of about 22° × 11° at angular resolutions of 0.17°–0.46° depending on the frequency band. Sensitivity between 0.2 and 0.8 K shall be feasible at a frame rate of 10 f/s. High losses of the overall antenna structure and the lack of low-noise amplification are compensated by cryogenic cooling of the receivers at 7 K and bandwidths of hundred to several hundred GHz (► Ultra-Wideband Antennas).

Previous examples of near-future approaches have indicated the way forward by using highly integrated receiver technology in combination with suitable antenna array structures as the most promising solutions for many applications. Here, the design of the antenna system plays a central role in the overall system development in order to achieve high sensitivity and high spatial resolution at high frame rates and sufficiently large fields of view.

Cross-References

- ► Antennas in Radio Telescope Systems
- ► Applications of Phased Array Feeders in Reflector Antennas
- ► Beam-scanning Leaky Wave Antennas
- ► Dielectric Lens Antennas
- ► Impedance Matching and Baluns
- ► Microstrip Patch Antennas
- ► Millimeter-Wave Antennas and Arrays
- ► Mm-Wave Sub-mm-wave Antenna Measurement
- ► Multi-Beam Antenna Arrays
- ► On-Chip Antennas
- ► Phased Arrays
- ► Photomixer Antennas at Terahertz
- ► Radio Frequency Beam-Forming for Scanned and MultiBeam Antenna Systems
- ► Radio Wave Propagation Models
- ► Reflector Antennas
- ► Space Antennas including Terahertz Antennas
- ► Suppression of Mutual Coupling between Elements in Multiple Element Antennas
- ► Taped Slot Antennas
- ► Terahertz Antennas and Measurement
- ► Transmission Lines
- ► Travelling-wave Antennas (Including Leaky Wave and Surface Wave Antennas)
- ► Ultra-wideband Antennas
- ► Waveguide Slot Antennas and Arrays

DOI 10.1007/978-981-4560-75-7_125-1

► Waveguide Slot Array Antennas

References

Appleby R, Wallace HB (2007) Standoff detection of weapons and contraband in the 100 GHz to 1 THz region. IEEE Trans Ant Propagat 55(11):2944–2956

Appleby R, Anderton RN, Jack JW (2004) The design of a real-time 94 GHz passive millimeter-wave imager for helicopter operations. Proceedings of SPIE security & defense symposium – passive MMW imaging technology VII, Orlando, vol 5619

Barré HMJP, Duesmann B, Kerr YH (2008) SMOS: the mission and the system. IEEE Trans Geosci Remote Sens 46(3):587–593

Bor J, Lafond O, Merlet H, Le bars P, Himdi M (2014) Foam based Luneburg lens antenna at 60 GHz. Progr Electromagn Res Lett 44:1–7

Bosch-Lluis X (2011) On the design of microwave radiometers with digital beamforming and polarization synthesis for earth observation. PhD dissertation, Universitat Polytècnica de Catalunya, Barcelona

Brown MA, Torres F, Corbella I, Colliander A (2008) SMOS calibration. IEEE Trans Geosci Remote Sens 46(3):646–658

Camps AJ (1996) Application of Interferometric radiometry to earth observation. PhD dissertation, Universitat Polytècnica de Catalunya, Barcelona

Camps A, Skou N, Torres F, Corbella I, Duffo N, Vall-llossera M (2006) Considerations about antenna pattern measurements of 2-D aperture synthesis radiometers. IEEE Geosci Remote Sensing Lett 3(2):259–261

Christensen J, Carlström A, Ekström A, de Maagt P, Colliander A, Emrich A, Embretsén J (2007) GAS: the geostationary atmospheric sounder. In: International geoscience and remote sensing symposium, IGARSS, Barcelona, Spain, 2007

Clark S, Lovberg J, Kolinko V, Martin C, McCoy J (2004) Pupil plane array based millimeter-wave imaging radiometer, final technical report, Trex Enterprises Corporation

Dong J, Li Q (2010) Antenna array design in aperture synthesis radiometers. In: Mini I (ed) Microwave and millimeter wave technologies modern UWB antennas and equipment. InTech 2010. Available from http://www.intechopen.com/books/microwave-and-millimeter-wave-technologiesmodern uwb-antennas-and-equipment/antenna-array-design-in-aperture-synthesis-radiometers

ECMWF. http://old.ecmwf.int/products/forecasts/d/charts/monitoring/satellite/smos/

Entekhabi D et al (2010) The soil moisture active passive (SMAP) mission. Proc IEEE 98(5):704–716

ESA Bulletin 137 (2009) European Space Agency, February 2009

Felli M, Spencer RE (1989) Very long baseline interferometry. Kluwer Academic Publishers, Dordrecht, The Netherlands

Fisher JR (1996) Phased-array feeds for low-noise reflector antennas. National radio astronomy observatory, Green Bank, West Virginia, Electronics division internal report no. 307

Fisher JR, Bradley RF (2000) Full-sampling focal plane arrays. In: Imaging at radio through submillimeter wavelengths, ASP conference proceedings, vol 217. Astronomical Society of the Pacific

Font J, Camps A, Borges A, Martin-Neira M, Boutin J, Reul N, Kerr YH, Hahne A, Mecklenburg S (2010) SMOS: the challenging sea surface salinity measurement from space. Proc IEEE 98(5):649–665

Hall PJ (2006) The square kilometre array: an engineering perspective. Springer Science & Business Media, Dordrecht

Hall L, Hansen H, Abbott D (2001) A microstrip-based Rotman lens for mm-wave sensing operations. In: Proceedings of SPIE passive millimeter wave imaging technology V, Orlando, vol 4373, pp 40–48.
Harris FJ (1978) On the use of windows for harmonic analysis with the discrete fourier transform. Proc IEEE 66(1):51–83
Harvey AR, Appleby R (2003) Passive mm-wave imaging from UAVs using aperture synthesis. Aeronaut J 107(1068):87–97
Heinz E, May T, Born D, Ziegler G, Peiselt K, Vyacheslav Z, Krause T, Krüger A, Schulz M, Bauer F, Meyer HG (2014) Progress in passive submillimeter-wave video imaging. In: Proceedings of SPIE defense security + sensing symposium – passive and active MMW imaging XVII, Baltimore, vol 9078
Imaoka K, Kachi M, Kasahara M, Ito N, Nakagawa K, Oki T (2010) Instrument performance and calibration of AMSR-E and AMSR2. International archives of the photogrammetry, remote sensing and spatial information science, vol XXXVIII, part 8, Kyoto, 2010
Janssen MA (1993) Atmospheric remote sensing by microwave radiometry. Wiley, New York
JAXA. http://sharaku.eorc.jaxa.jp/AMSR/gcom/gcom.html
Jirousek M, Peichl M, Suess H (2010) A microwave imaging spectrometer for security applications. In: Proceedings of SPIE defence security + sensing symposium – passive millimeter-wave imaging technology XIII, Orlando, vol 7670, 2010
Johansson JF (1995) Fundamental limits for focal-plane array efficiency. In: Multi-feed systems for radio telescopes, ASP conference series, vol 75
Johnson RC, Jasik H, Crawford HB (1993) Antenna engineering handbook. McGraw-Hill, New York
Kim W-G, Moon N-W, Chang Y-S, Lee M-K, Jung S-W, Choi J-H, Jung J-M, Kim Y-H (2011) System design of focal plane array based millimeter-wave imaging radiometer for concealed weapon detection. In: Proceedings of IGARSS, IEEE, Vancouver, 2011
Kraus J, Marhefka R (2001) Antennas for all applications. McGraw-Hill, New York
Le Vine DM (1990) The sensitivity of synthetic aperture radiometers for remote sensing applications from space. Radio Sci 25(4):441–453
Luukanen A, Kiuru T, Leivi MM, Rautiainen A, Varis J (2013) Passive three-color sub-millimeter-wave video camera. In: Proceedings of SPIE defence security + sensing symposium – passive and active MMW imaging XVI, Baltimore, vol 8715, 2013
Martín-Neira M, LeVine DM, Kerr Y, Skou N, Peichl M, Camps A, Corbella I, Hallikainen M, Font J, Wu J, Mecklenburg S, Drusch M (2014) Microwave interferometric radiometry in remote sensing: an invited historical review. Radio Sci 49. doi:10.1002/2013RS005230
Mätzler C (2006) Thermal microwave radiation: applications for remote sensing. The Institution of Engineering and Technology, London
National Academy of Sciences (2007) Assessment of millimeter-wave and terahertz technology for detection and identification of concealed explosives and weapons. Committee on Assessment of Security Technologies for Transportation, National Research Council, National Academies Press, Washington, DC
Nielsen PH, Pontoppidan K, Heeboell J, Le Stradic B (1989) Design, manufacture and test of a pushbroom radiometer. In: 6th international conference on antennas and propagation, ICAP, Coventry, 89, vol 1, 1989
Olver AD, Clarricoats PJB, Kishk AA, Shafai L (1994) Microwave horns and feeds, IEE electromagnetic waves series. Institution of Engineering and Technology, London
Orfanidis SJ (2008) Electromagnetic waves and antennas. Rutgers University, Piscataway. http://www.ece.rutgers.edu/&orfanidi/ewa/
Otoshi TY (1968) The effect of mismatched components on microwave noise-temperature calibrations. IEEE Trans Microwave Theor Techn 16(9):675–686

Padman R (1995) Optical fundamentals for array feeds. In: Multi-feed systems for radio telescopes, ASP conference series, vol 75

Peichl M, Süß H, Süß M, Kern S (1998) Microwave imaging of the brightness temperature distribution of extended areas in the near and far field using two-dimensional aperture synthesis with high spatial resolution. Radio Sci 33(3):781–801

Peichl M, Dill S, Jirousek M, Süß H (2007) Microwave radiometry – imaging technology and applications. In: Review lecture, ITG conference on wave propagation in communication, microwave systems and navigation, WFMN 2007, Chemnitz

Peichl M, Dill S, Rudolf D (2013) SUMIRAD – a low-cost fast millimeter-wave radiometric imaging system. In: Proceedings of SPIE defence security + sensing symposium – passive and active MMW imaging XVI, vol 8715

Pozar DM (2005) Microwave engineering, 3rd edn. Wiley, Hoboken

Rautiainen K, Kainulainen J, Auer T, Pihlflyckt J, Kettunen J, Hallikainen MT (2008) Helsinki University of technology L-band airborne synthetic aperture radiometer. IEEE Trans Geosci Remote Sens 46(3):717–726

Reul N, Tenerelli J, Chapron B, Waldteufel P (2007) Modeling sun glitter at L-band for sea surface salinity remote sensing with SMOS. IEEE Trans Geosci Remote Sens 45(7):2073–2087

Richter J, Nötel D, Klöppel F, Huck J, Essen H, Schmidt L-P (2006) A multi -channel radiometer with focal plane array antenna for W-band passive millimeter-wave imaging. In: IEEE international MTT symposium, San Francisco, 2006

Ruf CS, Swift CT, Tanner AB, Le Vine DM (1988) Interferometric synthetic aperture microwave radiometry for the remote sensing of the Earth. IEEE Trans Geosci Remote Sens 26(5):597–611

Schreiber E (2014) Fully-electronic passive microwave imaging system using beam steering by frequency shift and aperture synthesis. In: DLR Forschungsbericht 2014–06, DLR Bibliotheks- und Informationswesen, Cologne, 2014 (in German)

Schreiber E, Peichl M (2013) The design of a broadband slotted waveguide antenna for electronic beam steering applications in MW radiometry. In: Proceedings of European microwave conference EuMC, Nuremberg, 2013

Schuetz CA, Martin RD, Biswas I, Mirotznik MS, Shi S, Schneider GJ, Murakowski J, Prather DW (2007) Sparse aperture millimeter-wave imaging using optical detection and correlation techniques. In: Proceedings of SPIE defence security + sensing symposium – passive and active MMW imaging X, Orlando, vol 6548, 2007

Schuetz CA, Martin R, Dillon T, Yao P, Mackrides D, Harrity C, Zablocki A, Shreve K, Bonnett J, Curt P, Prather DW (2013) Realization of a video-rate distributed aperture millimeter-wave imaging system using optical upconversion. In: Proceedings of SPIE defence security + sensing symposium – passive and active MMW Imaging XVI, vol 8715, Baltimore, 2013

Shannikov DV, Kuzmin SV, Voronin VA, Rybakov YV, Shchukin GG, Khaikin VB (2005) Microwave radiometric and antenna systems for mapping of atmosphere and underlying surface, International Symposium on Remote Sensing of Environment, ISRSE. http://www.isprs.org/publications/related/ISRSE/html/papers/935.pdf

Skou N (1997) Measurement of small antenna reflector losses for radiometer calibration budget. IEEE Trans Geosci Remote Sens 35(4):967–971

Skou N, Le Vine D (2006) Microwave radiometer systems – design and analysis, 2nd edn. Norwood, Artech House

Sönmez MK, Trew RJ, Hearn CP (1992) Front-end topologies for phased array radiometry. In: 22nd European microwave conference, vol 2, 5–9 Helsinki, Sept 1992

DOI 10.1007/978-981-4560-75-7_125-1

Srinivasan K, Limaye A, Laymon C, Meyer P (2010) Phased array radiometer calibration using a radiated noise source. In: IEEE geoscience and remote sensing symposium, 26–30, Honolulu, July 2010

Stutzman WL, Thiele GA (1998) Antenna theory and design. Wiley, New York

Tanner AB (2007) Initial results of the geostationary synthetic thinned array radiometer (GeoSTAR) demonstrator instrument. IEEE Trans Geosci Remote Sens 45(7):1947–1957

Thompson AR, Moran JM, Swenson GW (2001) Interferometry and synthesis in radio astronomy, 2nd edn. Wiley, New York

Tudosie G (2009) Analysis and design of compact passive distribution networks for microwave applications. Dissertation ETH No. 18340, Swiss Federal Institute of Technology (ETH) Zurich

Ulaby FT, Moore RK, Fung AK (1981) Microwave remote sensing, active and passive. In: Microwave remote sensing fundamentals and radiometry, vol I. Addison-Wesley, Reading

Veidt B (2006) Focal-plane array architectures: horn clusters vs. phased-array techniques, SKA Memo 71. www.skatelescope.org/pages/page_memos.htm

Warnick KF (2013) Phased-array antennas for radio astronomy. In: Radio science meeting (USNC-URSI NRSM), US National Committee of URSI National, 2013

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_129-1

Transmission Lines

Cam Nguyen*
Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA

Abstract

This chapter presents the theory of transmission lines, which are essential for the design of radiofrequency (RF) circuits associated with antennas and transmission-line-based antennas. It includes the transmission-line equations and important transmission-line parameters such as characteristic impedance, propagation constant, phase velocity, effective relative dielectric constant, dispersion, loss, distortion, impedance, reflection coefficient, etc. It also presents synthetic transmission lines, a class of transmission lines that can be implemented using only lumped elements, which could be useful for RF circuits needed for on-chip antenna interface. Moreover, the chapter covers the most commonly used printed-circuit transmission lines for RF systems, which also include those that can be realized in multilayer structures that are attractive for printed-circuit antennas and associated RF circuits.

Keywords

Transmission lines; TEM transmission line; Quasi-TEM transmission line; TEM mode; Quasi-TEM mode; Coplanar waveguide; CPW; Microstrip line; Coplanar strips; Slot line; Strip line; Multilayer transmission lines; Antenna; Microwave; Millimeter wave; RF

Introduction

Transmission lines are important elements for radiofrequency (RF) systems, which operate at different frequencies in the electromagnetic (EM) spectrum up to terahertz (THz). Transmission lines, both theory and design, are needed not only for specific RF circuits required for antennas to function properly, such as antenna-feeding or antenna-interface circuitry, but also for various antennas such as transmission-line-based traveling antennas – for example, slot-line antennas. The design of RF circuits needed for antennas, and hence antennas eventually, would not be complete, or even optimum, without considering transmission lines. Especially, at extremely high frequencies, such as those in the upper end of the millimeter-wave range, and for on-chip implementations, such as direct integration of RFICs (RF-integrated circuits) and antennas, transmission lines are inevitable. Even when the use of on-chip transmission lines to provide interface with on-chip antennas is not appealing due to their relatively large size as compared to on-chip lumped elements, RF designers should make use of the theory of transmission lines to properly implement on-chip lumped elements to simulate transmission lines for antenna interface to achieve certain features that lumped elements cannot achieve otherwise. Therefore, it is crucial that RF engineers have sufficient knowledge in transmission lines and consider their possible use in antennas. The basics of transmission lines can be found in various textbooks in electromagnetics (e.g., Paul et al. 1998). This chapter will address the fundamentals of transmission lines including transmission-line equations and important transmission-line parameters such as characteristic impedance, propagation constant, phase velocity,

*Email: cam@ece.tamu.edu

DOI 10.1007/978-981-4560-75-7_129-1

effective relative dielectric constant, dispersion, loss, distortion, impedance, reflection coefficient, etc. It also discusses synthetic transmission lines, which enable the design of transmission lines using lumped elements useful for on-chip antenna interface and commonly used printed-circuit transmission lines including those in multilayer structures that are essential for antennas and associated RF circuits.

Transmission-Line Equations

Transmission lines, like other elements in RF circuits, possess certain electrical properties which are needed for circuit design. The characteristics of any transmission line are governed by a set of equations commonly known as transmission-line equations. To facilitate the analysis of transmission lines under different excitations and operating conditions, these equations are classified into two kinds: general transmission-line equations, which are valid for excitations of any time variation, and sinusoidal steady-state or time-harmonic transmission-line equations, which are applicable only when the excitation is a sinusoidal waveform and the transmission line is under a steady-state operation. The transmission-line parameters such as characteristic impedance, however, are the properties of a transmission line itself regardless of the excitation, provided that the magnitude of the excitation is within a certain limit.

General Transmission-Line Equations

A transmission line exists only when it has at least two conductors. Consequently, a (two-conductor) transmission line is normally represented as two parallel lines as shown in Fig. 1a. The transmission line is assumed to be uniform; i.e., its geometry, including the conductor dimensions, spacing between them, and dielectric media around the conductors remain constant along the transmission-line length or its longitudinal (or axial) direction. From electromagnetic (EM) theory, it is known that a signal, or wave, propagating along the line is due to the interchange of the electric and magnetic energy. As a signal traverses the line, it carries with it the electric and magnetic fields. It is also known that as a signal propagates along the line, it suffers losses due to imperfect conductors and dielectrics constituting the line. Using circuit theory, a transmission line can then be characterized electrically as a combination of inductors, representing the magnetic energy; capacitors, representing the electric energy; resistors, representing the loss due to conductors; and other resistors representing the loss due to dielectrics. In view of this and the fact that circuit theory is only valid at frequencies where dimensions are very small compared to the wavelengths at those frequencies, the transmission line is divided into many infinitesimal elements, each having a length of *dz* much shorter than the operating wavelength. Each tiny element can

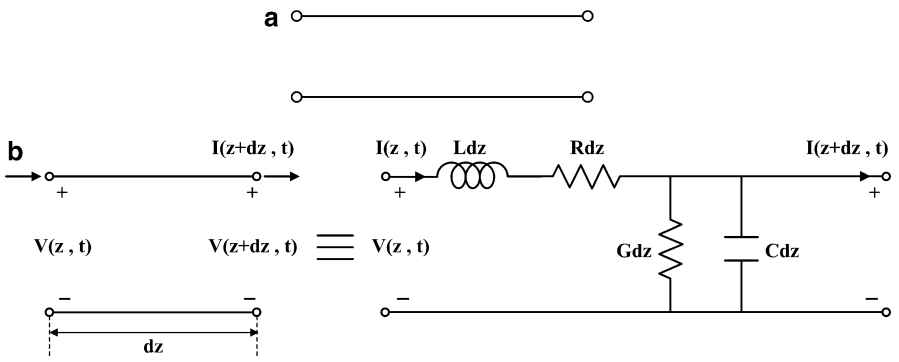


Fig. 1 (**a**) A transmission line represented by two parallel conductors and (**b**) an equivalent circuit of an infinitesimal transmission line; $V(z,t)$, $V(z+dz,t)$, and $I(z,t)$, $I(z+dz, t)$ are the instantaneous voltages and currents at the input (z) and output ($z+dz$), respectively

DOI 10.1007/978-981-4560-75-7_129-1

now be represented by an equivalent circuit consisting of a series inductance *Ldz*, a series resistance *Rdz*, a shunt capacitance *Cdz*, and a shunt conductance *Gdz*, as shown in Fig. 1b. *L*, *R*, *C*, and *G* are the series inductance (*H*/*m*), series resistance (Ω/*m*), shunt capacitance (*F*/*m*), and shunt conductance (*S*/*m*) per unit length of the transmission line, respectively. *L* and *C* account for the respective magnetic and electric energy of the signal stored in the transmission line. *R* and *G* account for the respective losses due to the imperfect conductors and dielectrics used in the transmission line. The loss due to the dielectrics is sometimes called leakage loss. This loss is associated with a shunt current existed across the dielectrics due to their finite conductivity, which represents the leakage current. This leakage current is substantial for high-loss substrates, such as silicon used for RFICs, and has significant effects in RF circuits realized on these substrates, not only because of the resulting dielectric loss in the substrates but also because of the coupling that current causes among different circuit elements. Each transmission line is characterized by its own *R*, *L*, *G*, and *C*, or its series impedance $Z = R + j\omega L$ and shunt admittance $Y = G + j\omega C$ per unit length.

A (finite-length) transmission line can be considered as consisting of many minuscule sections connected in cascade or, electrically, a continuous distribution of R, L, G, and C along the line. As such, a transmission line can be modeled as a cascade of many equivalent circuits, each corresponding to an infinitesimal section as shown in Fig. 1b. This distributed-circuit model for a transmission line is valid only when we assume no mutual impedance exists between adjacent sections making up the transmission line. This implies that there is no mutual coupling between these adjacent elements. Consequently, it can be assumed that there are no longitudinal fields E_z and H_z for the wave traversing the line, assuming *z* is the direction of propagation. The electric and magnetic fields of the wave thus have only *x* and *y* components; that is, they lie in the planes orthogonally transverse to the propagation direction. Such a wave is called transverse electromagnetic (TEM) wave or mode. In practice, however, there exists very small longitudinal fields along transmission lines, and the corresponding wave is referred to as quasi-TEM wave. All equations derived in this section are thus valid only for the TEM or quasi-TEM mode. Other waves having either E_z or H_z, or both, can also exist on transmission lines and are called higher-order waves or modes. These modes are significant only near the feed points to the transmission line, around a discontinuity on the line, or at very high frequencies. For most uniform transmission lines used in practice, TEM of quasi-TEM mode is the principal mode and is referred to as the transmission line's dominant mode.

Applying Kirchhoff's voltage and current laws to the circuit in Fig. 1b gives, after some manipulations,

$$\frac{\partial V(z,t)}{\partial z} + L\frac{\partial I(z,t)}{\partial t} + RI(z,t) = 0 \tag{1}$$

$$\frac{\partial I(z,t)}{\partial z} + C\frac{\partial V(z,t)}{\partial t} + GV(z,t) = 0 \tag{2}$$

$$\frac{\partial^2 V(z,t)}{\partial z^2} - LC\frac{\partial^2 V(z,t)}{\partial t^2} - (RC + LG)\frac{\partial V(z,t)}{\partial t} - RGV(z,t) = 0 \tag{3}$$

$$\frac{\partial^2 I(z,t)}{\partial z^2} - LC\frac{\partial^2 I(z,t)}{\partial t^2} - (RC + LG)\frac{\partial I(z,t)}{\partial t} - RGI(z,t) = 0 \tag{4}$$

For lossless transmission lines, Eqs. 1, 2, 3, and 4 become, after letting $R = G = 0$,

DOI 10.1007/978-981-4560-75-7_129-1

$$\frac{\partial V(z,t)}{\partial z} + L\frac{\partial I(z,t)}{\partial t} = 0 \tag{5}$$

$$\frac{\partial I(z,t)}{\partial z} + C\frac{\partial V(z,t)}{\partial t} = 0 \tag{6}$$

$$\frac{\partial^2 V(z,t)}{\partial z^2} - LC\frac{\partial^2 V(z,t)}{\partial t^2} = 0 \tag{7}$$

$$\frac{\partial^2 I(z,t)}{\partial z^2} - LC\frac{\partial^2 I(z,t)}{\partial t^2} = 0 \tag{8}$$

Equations 1–2 and 5–6 are classified as the telegraphist's equations, and Eqs. 3–4 and 7–8 are called the transmission-line equations. Equations 3 and 4 can also be viewed as the wave equations for lossless transmission lines – similar to the wave equations involving electric and magnetic fields derived from Maxwell's equations. All these equations are general and hold for voltages and currents with arbitrary time dependence and for any transmission line. They become simpler when the applied voltage or current is the sinusoidal time-varying function.

Practical transmission lines are lossy, and the voltage and current waves or signals will be attenuated as they propagate along the line.

Sinusoidal Steady-State Transmission-Line Equations

Another form of transmission-line equations is now derived by assuming that the applied voltage varies sinusoidally with time and the transmission line is operated under a steady-state condition (i.e., no transients). Using the phasor representations, the instantaneous voltage and current can be written, with reference to cosωt, as

$$V(z,t) = \mathrm{Re}\left[V(z)e^{j\omega t}\right] \tag{9}$$

$$I(z,t) = \mathrm{Re}\left[I(z)e^{j\omega t}\right] \tag{10}$$

where $V(z)$ and I(z) represent the phasor voltage and current, respectively, which are functions of location (z) only, and Re (.) denotes the real part. Substituting Eqs. 9–10 into 1–4 yields

$$\frac{dV(z)}{dz} + (R + j\omega L)I(z) = 0 \tag{11}$$

$$\frac{dI(z)}{dz} + (G + j\omega C)V(z) = 0 \tag{12}$$

$$\frac{d^2V(z)}{dz^2} - \gamma^2 V(z) = 0 \tag{13}$$

$$\frac{d^2I(z)}{dz^2} - \gamma^2 I(z) = 0 \tag{14}$$

DOI 10.1007/978-981-4560-75-7_129-1

where

$$\gamma = \alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)} \tag{15}$$

As will be seen later, γ dictates the propagation of voltage and current waves along the line and hence is called the propagation constant of the transmission line. Similarly, α (in Neper/m or Np/m) and β (in radian/m) determine the attenuation and phase per unit length of the transmission line and are called the attenuation and phase constant, respectively. α in Np/m can be converted into α in decibel/m by $\alpha[\mathrm{dB/m}] = 8.686\alpha[\mathrm{Np/m}]$.

Equations 11, 12, 13, and 14 can also be obtained by replacing the partial derivative with respect to t, $\partial/\partial t$, with $j\omega$ in Eqs. 1, 2, 3, and 4, recognizing the equivalence between them for sinusoidal signals. Equations 11 and 12 are called the sinusoidal steady-state or time-harmonic telegraphist's equations, and Eqs. 13 and 14 are known as the sinusoidal steady-state or time-harmonic transmission-line equations or the transmission-line equations in the frequency domain. The transmission-line Eqs. 13 and 14 are especially used to analyze transmission lines operating under sinusoidal steady state, which is encountered in continuous-wave (CW) operations of electrical circuits. It should also be noted that all the transmission-line equations are applicable to any transmission lines. For lossless transmission lines, these equations become, after setting R=G=0 in Eqs. 11, 12, 13, and 14,

$$\frac{dI(z)}{dz} + j\omega CV(z) = 0 \tag{16}$$

$$\frac{dV(z)}{dz} + j\omega LI(z) = 0 \tag{17}$$

$$\frac{d^2V(z)}{dz^2} + \beta^2 V(z) = 0 \tag{18}$$

$$\frac{d^2I(z)}{dz^2} + \beta^2 I(z) = 0 \tag{19}$$

Equations 13 and 14 have the following well-known solutions:

$$\begin{aligned} V(z) &= V_o^+ e^{-\gamma z} + V_o^- e^{\gamma z} = V_o^+ e^{-\alpha z} e^{-j\beta z} + V_o^- e^{\alpha z} e^{j\beta z} \\ &= V^+(z) + V^-(z) \end{aligned} \tag{20}$$

$$\begin{aligned} I(z) &= I_o^+ e^{-\gamma z} + I_o^- e^{\gamma z} = I_o^+ e^{-\alpha z} e^{-j\beta z} + I_o^- e^{\alpha z} e^{j\beta z} \\ &= I^+(z) + I^-(z) \end{aligned} \tag{21}$$

respectively. $V^+(z)$ and $I^+(z)$ are commonly referred to as the forward-traveling voltage and current waves (in the $+z$ direction), and $V^-(z)$ and $I^-(z)$ are called the backward-traveling waves (in the $-z$ direction). The total instantaneous voltage and current at any location on the line and at any time can be written, using Eqs. 9, 10, 15, 20, and 21, as

DOI 10.1007/978-981-4560-75-7_129-1

$$V(z,t) = V_o^+ e^{-\alpha z} \cos\left[\omega\left(t - \frac{\beta}{\omega}z\right)\right] + V_o^- e^{\alpha z} \cos\left[\omega\left(t + \frac{\beta}{\omega}z\right)\right] \tag{22}$$

$$I(z,t) = I_o^+ e^{-\alpha z} \cos\left[\omega\left(t - \frac{\beta}{\omega}z\right)\right] + I_o^- e^{\alpha z} \cos\left[\omega\left(t + \frac{\beta}{\omega}z\right)\right] \tag{23}$$

The first and second terms in the right-hand side of Eqs. 22 and 23 correspond to the forward-traveling waves $V^+(t-z/v)$, $I^+(t-z/v)$ and backward-traveling waves $V^-(t+z/v)$, $I^-(t+z/v)$, respectively. These waves travel at a velocity of

$$v = \frac{\omega}{\beta} \tag{24}$$

Equations 22 and 23 indicate that as the forward or backward voltage and current wave travel along the line at any instant and frequency, their amplitudes at a particular location z are reduced exponentially in accordance with $-\alpha|z|$ and their phases undergo a phase change of βz. α and β determine the attenuation and phase per unit length of the transmission line and are therefore called the attenuation and phase constant, respectively. $\gamma = \alpha + j\beta$ dictates the propagation of voltage and current waves along the line and hence is called the propagation constant of the transmission line.

The instantaneous (sinusoidal) voltage $V(z,t)$ versus time at different locations along the line can be sketched to demonstrate that the voltage travels along the line as time progresses. Note that in contrary to the circuit theory, in which voltage and current are assumed stationary and constant along the conductor of a transmission line, they are indeed (propagating) waves and are a function of the transmission line's characteristics.

Transmission-Line Parameters

It is recalled that components in electronic circuits are physical elements and electrical parameters are used to describe them so that they can be used in circuit analysis and design. For example, an inductor is described by an electrically equivalent inductance, a resistor is represented by a resistance, etc. Similarly, a transmission line is a physical structure, and it is also described by certain electrical parameters. The most useful parameters of a transmission line, at least for circuit design and analysis, are characteristic impedance, propagation constant or attenuation and phase constants, and velocity. The phase constant and velocity are directly related, and, many times, the parameter "effective relative dielectric constant" is also used in lieu of these. Transmission-line parameters can be determined from the per-unit-length inductance, resistance, capacitance, and conductance of the transmission line described in section "Per-Unit-Length Parameters R, L, C, and G." They are independent of the time variation of the applied voltage or current and, in general, the magnitude of the applied signal. As such, expressions for the transmission-line parameters can be derived considering a transmission line operating under a sinusoidal steady state. This would facilitate the solution process. To aid the use of these expressions, three cases are considered: lossy (or general), lossless, and low-loss transmission lines.

General Transmission Lines

For any transmission line operated under a sinusoidal steady state, the phasor current can be written using Eq. 11 as

DOI 10.1007/978-981-4560-75-7_129-1

$$I(z) = -\frac{1}{R + j\omega L}\frac{dV(z)}{dz} \tag{24}$$

Taking the derivative of the voltage in Eq. 20 and substituting it into Eq. 24 results in

$$I(z) = \frac{\gamma}{R + j\omega L}\left(V_o^+ e^{-\gamma z} - V_o^- e^{\gamma z}\right) \tag{25}$$

Comparing Eqs. 21 and 25 and making use of Eq. 15 gives

$$Z_o \triangleq \frac{V_o^+}{I_o^+} = -\frac{V_o^-}{I_o^-} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \tag{26}$$

which is defined as the characteristic impedance of the transmission line (in Ohms). Its reciprocal

$$Y_o \triangleq \frac{I_o^+}{V_o^+} = -\frac{I_o^-}{V_o^-} = \sqrt{\frac{G + j\omega C}{R + j\omega L}} \tag{27}$$

is called the characteristic admittance of the line (in *mhos*). Note that the characteristic impedance and admittance, as expected, are generally complex as can be seen in Eqs. 26 and 27. It is important to note that these parameters are simply mathematically defined quantities and, as expected through their expressions in Eqs. 26 and 27, they are not and cannot be used as the normal impedance and admittance encountered in the electrical circuit theory. The propagation constant $\gamma = \alpha + j\beta$ of the transmission line is given in Eq. 15, where α and β, as described in section "Sinusoidal Steady-State Transmission-Line Equations," are the attenuation and phase constants of the line, respectively. They are obtained by taking the respective real and imaginary parts of the propagation constant. It should be noted that the attenuation constant is contributed by the conductor attenuation constant (due to imperfect conductors) and the dielectric attenuation constant (due to imperfect dielectrics), which will be examined in section "Dielectric and Conductor Losses in Transmission Lines." As the per-unit-length resistance R, inductance L, conductance G, and capacitance C of a transmission line are different for different transmission lines, the characteristic impedance Z_o and propagation constant γ are also different. In practice, Z_o and γ (or α and β) are normally used to characterize a transmission line instead of R, L, G, and C.

A closed-form expression for the attenuation constant can also be derived based on the time-average transmitted power and power loss per unit length on the line. To that end, an infinitely long transmission line is considered, and the phasor voltage and current at any location are written, from Eqs. 20 to 21, as

$$V(z) = V_o^+ e^{-\alpha z} e^{-j\beta z} \tag{28}$$

$$I(z) = I_o^+ e^{-\alpha z} e^{-j\beta z} \tag{29}$$

The time-average power or, simply, average power transmitted at any point on the line is given by

$$P_T(z) = \frac{1}{2}\mathrm{Re}[V(z)I^*(z)] \tag{30}$$

DOI 10.1007/978-981-4560-75-7_129-1

which becomes, after using Eqs. 28 and 29 and replacing $I_o^+ = V_o^+/Z_o$ with $Z_o = R_o + jX_o$,

$$P_T(z) = \frac{1}{2}\mathrm{Re}\left[\frac{\left(V_o^+\right)^2}{R_o - jX_o}e^{-2\alpha z}\right] = \frac{\left(V_o^+\right)^2 R_o}{2|Z_o|^2}e^{-2\alpha z} \tag{31}$$

Let P_{T0} being the power flow at $z = 0$ and P_{T0}–ΔP_T being the transmitting power at $z = \Delta z$. The average power loss per unit length is obtained as the difference in these powers over the distance, which is essentially the rate of decrease of the average transmitted power as the wave travels along the line and can, therefore, be derived from Eq. 31 as

$$P_L(z) = \lim_{\Delta z \to 0}\frac{(P_{T0} - \Delta P_T) - P_{T0}}{\Delta z} = -\frac{dP_T(z)}{dz} = 2\alpha P_T(z) \tag{32}$$

from which

$$\alpha = \frac{P_L(z)}{2P_T(z)} \tag{33}$$

Note that z is arbitrary and so, for uniform transmission lines whose geometry along the direction of propagation is constant, the power loss per unit length itself is independent of z. This power loss is due to the currents flowing through the resistance R and the conductance G and, hence, can be determined from Fig. 1b as

$$P_L(z) = \frac{1}{2}\left[R|I(z)|^2 + G|V(z)|^2\right] \tag{34}$$

which becomes, after utilizing Eqs. 28 and 29 and $I_o^+ = V_o^+/Z_o$,

$$P_L(z)^L = \frac{\left(V_o^+\right)^2}{2|Z_o|^2}\left[R + G|Z_o|^2\right]e^{-2\alpha z} \tag{35}$$

The attenuation constant can now be obtained from Eqs. 31, 33, and 35 as

$$\alpha = \frac{1}{2R_o}\left[R + G|Z_o|^2\right] \tag{36}$$

For transmission lines fabricated using good dielectrics and conductors, which are normally used in practice, the rate of change of the attenuation constant with frequency is, in general, relatively small. At low frequencies, the attenuation constant is very small and may be neglected. However, the attenuation of a transmission line cannot be ignored when the line is either long, used at high frequencies, or carries a large amount of power creating heat due to ½ (I^2R+V^2G) that must be dissipated. It is noted that the attenuation calculated from Eq. 36 neglects the attenuation due to radiation. An additional resistance representing the loss due to radiation can be defined and included in the per-unit-length resistance R, and hence Z_o, to account for the loss due to radiation. However, radiation loss is usually very small in practical

DOI 10.1007/978-981-4560-75-7_129-1

transmission lines and is, therefore, normally neglected. For some RF circuits that implement transmissions directly on a high-loss substrate such as silicon or without proper shield from that substrate, the attenuation constant, however, cannot be neglected due to substantial loss caused by the substrate. As this attenuation constant is large, it not only reduces the signal amplitude significantly but also causes large (substrate) coupling among circuit elements, degrading the circuit performance and, if not suitably taken care of, may cause the circuit not functioning properly under certain conditions.

The velocity can be derived by letting the phase $\omega t - \beta z$ of the forward-traveling wave in Eq. 22 equal to a constant and differentiating it with respect to time, as

$$v = \frac{dz}{dt} = \frac{\omega}{\beta} \tag{37}$$

This velocity corresponds to a constant phase of the wave and so is normally called the phase velocity, or simply velocity, of the wave propagating on the line. It indicates how fast points of equal phase on the wave move along the transmission line. As all points move at the same velocity, the phase velocity describes the rate at which the wave propagates along the transmission line and, hence, can be considered as the velocity of propagation. The distance, over which the phase is changed by 2π radians at any instant, is defined as the wavelength and is thus derived as

$$\lambda = \frac{2\pi}{\beta} = \frac{v}{f} \tag{38}$$

where f denotes the frequency.

Lossless Transmission Lines

Although all practical transmission lines have loss, lossless transmission lines are sometimes used for easy and fast circuit analysis and design. This assumption is also often used to illustrate some concepts in transmission lines without lost of generality. A lossless transmission line is defined as one having perfect conductors and dielectrics. This results in no resistance and conductance per unit length; i.e., $R=G=0$. Setting $R=G=0$ in Eqs. 26 and 15, we obtain

$$Z_o = \sqrt{\frac{L}{C}} \tag{39}$$

$$\gamma = j\beta = j\omega\sqrt{LC} \tag{40}$$

The velocity is obtained from Eqs. 37 and 40 as

$$v = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}} \tag{41}$$

Low-Loss Transmission Lines

For most practical transmission lines, the per-unit-length resistance R and conductance G are very small as compared to the per-unit-length reactance ωL and susceptance ωC, respectively. Under these conditions, a transmission line is considered as low-loss transmission line and approximate closed-form equations can

be derived for the characteristic impedance and attenuation and phase constants. Equation 26 can be rewritten as

$$Z = \sqrt{\frac{L}{C}}\left(1+\frac{R}{j\omega L}\right)^{1/2}\left(1+\frac{G}{j\omega C}\right)^{-1/2} \tag{42}$$

Applying the binomial series

$$(1+x)^a = 1 + ax + \frac{a(a-1)}{2!}x^2 + \cdots + \frac{a(a-1)\cdots(a-n+1)}{n!}x^n + \cdots \tag{43}$$

where x and a are real numbers, and retaining up to second-order terms, we obtain

$$Z \cong \sqrt{\frac{L}{C}}\left(1+\frac{R}{j2\omega L}\right)\left(1-\frac{G}{j2\omega C}\right) + \frac{1}{8}\left(\frac{R}{\omega L}\right)^2 - \frac{3}{8}\left(\frac{G}{\omega C}\right)^2 \tag{44}$$

Expanding Eq. 44 and making use of the fact that $R << \omega L$ and $G << \omega C$ yields

$$Z_o \cong \sqrt{\frac{L}{C}}\left[\left(1+\frac{1}{8}\left(\frac{R}{\omega L}\right)^2 - \frac{3}{8}\left(\frac{G}{\omega C}\right)^2 + \frac{RG}{4\omega^2 LC}\right) + j\left(\frac{G}{2\omega C} - \frac{R}{2\omega L}\right)\right] \tag{45}$$

Similarly, we can rewrite Eq. 15 and apply the binomial expression (43) and the conditions $R << \omega L$ and $G << \omega C$ to get

$$\gamma \cong \frac{1}{2}\left(R\sqrt{\frac{C}{L}} + G\sqrt{\frac{L}{C}}\right) + j\omega\sqrt{LC}\left(1+\frac{1}{8}\left(\frac{R}{\omega L}\right)^2 + \frac{1}{8}\left(\frac{G}{\omega C}\right)^2 - \frac{RG}{4\omega^2 LC}\right) \tag{46}$$

from which, we can obtain

$$\alpha \cong \frac{1}{2}\left(R\sqrt{\frac{C}{L}} + G\sqrt{\frac{L}{C}}\right) \tag{47}$$

$$\beta \cong \omega\sqrt{LC}\left[1+\frac{1}{8\omega^2}\left(\frac{R}{L} - \frac{G}{C}\right)^2\right] \tag{48}$$

For most engineering design purposes, it is sufficient to retain only the first-order terms in the binomial expansions. Under this approximation, Eqs. 45, 47, and 48 reduce to

$$Z_o \cong \sqrt{\frac{L}{C}}\left[1 + j\left(\frac{G}{2\omega C} - \frac{R}{2\omega L}\right)\right] \tag{49}$$

DOI 10.1007/978-981-4560-75-7_129-1

$$\alpha \cong \frac{1}{2}\left(R\sqrt{\frac{C}{L}} + G\sqrt{\frac{L}{C}}\right) \tag{50}$$

$$\beta \cong \omega\sqrt{LC} \tag{51}$$

The phase velocity is then obtained from Eq. 37 as

$$v \cong \frac{1}{\sqrt{LC}} \tag{52}$$

It is apparent that the phase constant and velocity of a low-loss transmission line are approximately equal to those of a lossless line.

It should be mentioned here that, for transmission lines implemented directly on (or near) a high-loss substrate such as silicon or without proper shield from that substrate, the transmission lines cannot be considered low-loss and the above-derived equations for low-loss transmission lines may not be applicable.

Per-Unit-Length Parameters R, L, C, and G

In the previous sections, it has been learned that a transmission line and all of its parameters are characterized by its series inductance L, series resistance R, shunt capacitance C, and shunt conductance G per unit length. Expressions for these per-unit-length parameters are now derived.

General Formulation

An infinitely long transmission line consisting of two conductors C_1 and C_2, as shown in Fig. 2, is considered. The inductance per unit length is contributed by two parts, one from the conductors as if they were perfect and another from the skin effect resulted from the imperfect conductors. The earlier inductance portion is called the external inductance and the latter is referred to as the internal inductance. The external inductance L_e, capacitance C, and conductance G per unit length are given, under static conditions, as

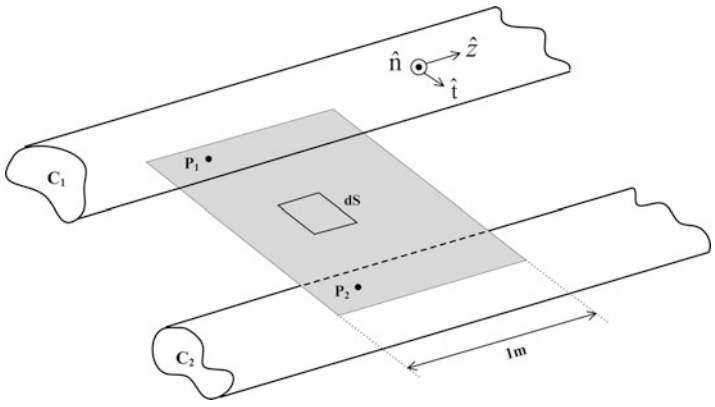


Fig. 2 A transmission line consisting of two conductors C_1 and C_2. $\hat{n}$ is the unit vector normal to the conducting surface; $\hat{t}$ is the unit vector tangential to the surface; and $\hat{z}$ is the unit vector along the longitudinal direction

DOI 10.1007/978-981-4560-75-7_129-1

$$L_e = \frac{\psi}{I_o} \tag{53}$$

$$C = \frac{Q}{V_o} \tag{54}$$

$$G = \frac{I_G}{V_o} \tag{55}$$

where ψ is the total magnetic flux per unit length linking the current, I_o is the total current on the conductor C_1, V_o is the voltage difference between the two conductors, Q is the total charge per unit length on the conductor C_1, and I_G represents the current flowing across the conductance G, which is the total shunt conduction current per unit length. Note that L_e, C, and G are not a function of frequency, assuming frequency-independent permittivity, permeability, and conductivity for the dielectric material.

The total charge per unit length on the conductor C_1 can be determined, assuming perfect conductor, as

$$Q = \oint_{C_1} \rho_s d\ell = \varepsilon \oint_{C_1} \vec{E} \cdot \hat{n}\, d\ell \tag{56}$$

where $d\ell$ is the differential length and the charge density per unit length $\rho_s = D_n = \varepsilon \vec{E} \cdot \hat{n}$ from the boundary condition, with D_n representing the normal component of the electric flux density, $\vec{E}$ being the electric field, $\hat{n}$ standing for the unit vector perpendicular to the conductor surface, and ε being the permittivity of the medium surrounding the conductors. Note that the integral is carried along a contour enclosing the conductor C_1. The voltage between the two conductors is obtained by taking the integral of the electric field along a path between any points on the surfaces of the conductors C_1 and C_2 as

$$V_o = \int_{C_1}^{C_2} \vec{E} \cdot d\,\vec{\ell} \tag{57}$$

where $d\ell$ is the differential-length vector. The capacitance per unit length can be obtained from Eqs. 54, 56, and 57 as

$$C = \frac{Q}{V_o} = \frac{\varepsilon \oint_{C_1} \vec{E} \cdot \hat{n}\, d\ell}{\int_{C_1}^{C_2} \vec{E} \cdot d\,\vec{\ell}} \tag{58}$$

Applying Ampere's law of Maxwell's equations, the total current flowing on the conductor C_1 is obtained by taking the line integral of the tangential magnetic field H_t around a contour enclosing the conductor as

$$I_o = \oint_{C_1} H_t d\ell \tag{59}$$

At the perfectly conducting surface C_1, the electric ($\vec{E}$) and magnetic ($\vec{H}$) fields have only normal (n) and tangential (t) components, respectively, and, for TEM propagating wave, they are related by

DOI 10.1007/978-981-4560-75-7_129-1

$$H_t = \frac{E_n}{\eta} = \frac{\vec{E} \cdot \hat{n}}{\eta} \tag{60}$$

where $\eta = \sqrt{\mu/\varepsilon}$ is the intrinsic impedance of the dielectric medium of the transmission line. Substituting H_t from Eqs. 60 into 59 and using Eqs. 56 yields

$$I_o = \frac{1}{\eta} \oint_{C_1} \vec{E} \cdot \hat{n} d\ell = \frac{Q}{\sqrt{\mu\varepsilon}} \tag{61}$$

The characteristic impedance of the transmission line can be derived, making use of Eqs. 58 and 61, as

$$Z_o = \frac{V_o}{I_o} = \frac{\sqrt{\mu\varepsilon}}{C} \tag{62}$$

from which,

$$C = \frac{\sqrt{\varepsilon\mu}}{Z_o} \tag{63}$$

The total magnetic flux linkage per unit length is obtained as an integration of the magnetic flux density over a surface area extended from the conductor C_1 to conductor C_2 that has a longitudinal length of 1 m between the two conductors, as shown in Fig. 2,

$$\psi = \int_S \vec{B} \cdot d\vec{S} = \int_{P_1}^{P_2} \mu H d\ell = \frac{\mu}{\eta} \int_{P_1}^{P_2} - \vec{E} \cdot d\vec{l} = \frac{\mu}{\eta} V_o \tag{64}$$

where P_1 and P_2 are arbitrary points on the surfaces of the conductors C_1 and C_2, respectively, and P_1P_2 is perpendicular to the flux lines. The external inductance per unit length is, from Eqs. 53 and 64,

$$L_e = \frac{\psi}{I_o} = \frac{\mu}{\eta} \frac{V_o}{I_o} = \sqrt{\varepsilon\mu} \frac{V_o}{I_o} \tag{65}$$

or, upon using Eq. 62,

$$L_e = Z_o \sqrt{\varepsilon\mu} \tag{66}$$

From Eqs. 63 and 66, we can also obtain

$$Z_o = \sqrt{\frac{L_e}{C}} \tag{67}$$

which is the same as that obtained in Eq. 39, with $L_e = L$ for lossless transmission lines.

Equations 63 and 66 also lead to

$$v = \frac{1}{\sqrt{L_e C}} = \frac{1}{\sqrt{\mu\varepsilon}} \tag{68}$$

DOI 10.1007/978-981-4560-75-7_129-1

as given in Eq. 41. It is noted again that the velocity for a lossless transmission line immersed in a medium characterized by ε and μ is exactly the same as that for a uniform plane wave propagating in the same medium due to their same TEM mode.

The conductance per unit length given by Eq. 55 can be rewritten as

$$G = \frac{I_G}{I_D}\frac{I_D}{V_o} \tag{69}$$

where I_D is the (total) displacement current in the transmission line and is given as

$$I_D = \int_S \vec{D}\cdot d\vec{S} = j\omega\varepsilon'\int_S \vec{E}\cdot d\vec{S} \tag{70}$$

where S represents a surface enclosing one conductor (e.g., C_1) and ε' is the real part of the dielectric's complex permittivity $\hat{\varepsilon} = \varepsilon' - j\varepsilon''$ with ε'' being the imaginary part of the dielectric's complex permittivity accounting for the loss in the dielectric. Note that $\varepsilon' = \varepsilon$ where ε is the permittivity of the medium used earlier. This current flows through the shunt capacitance per unit length C and is, thus, also given as

$$I_D = j\omega C V_o \tag{71}$$

The total shunt conduction current per unit length can be obtained as

$$I_G = \sigma\int_S \vec{E}\cdot d\vec{S} = \omega\varepsilon''\int_S \vec{E}\cdot d\vec{S} \tag{72}$$

upon replacing the dielectric's conductivity σ with $\omega\varepsilon''$. The conductance per unit length is then obtained from Eqs. 69 to 72 as

$$G = \frac{\omega\varepsilon''}{\varepsilon'}C \tag{73}$$

Equation 73 with $\varepsilon'' = \frac{\sigma}{\omega}$ shows that

$$\frac{G}{C} = \frac{\omega\varepsilon''}{\varepsilon'} = \frac{\sigma}{\varepsilon'} = \omega\tan\delta \tag{74}$$

upon using the loss tangent of the dielectric $\tan\delta = \sigma/\omega\varepsilon'$, which is a useful relationship between the shunt conductance G and capacitance C per unit length of a transmission line, allowing one parameter to be determined from the other for a given transmission-line medium.

The resistance per unit length R, as recalled, is used to represent the power loss due to imperfect conductors and can be determined from the following known relation:

$$P_L = \frac{1}{2}RI_o^2 \tag{75}$$

where P_L is the total power loss per unit length along the conductors. P_L is the sum of the individual power losses on the two conductors and is given as

DOI 10.1007/978-981-4560-75-7_129-1

$$P_L = \frac{R_S}{2}\oint_{C_1+C_2} \vec{J}_S \cdot \vec{J}_S^{*} dl = \frac{R_S}{2}\oint_{C_1+C_2} \left(\hat{n}\times \vec{H}\right)\cdot\left(\hat{n}\times \vec{H}^{*}\right) dl = \frac{R_S}{2}\oint_{C_1+C_2} \left|\vec{H}\right|^2 dl \tag{76}$$

upon using

$$\left(\hat{n}\times \vec{H}\right)\cdot\left(\hat{n}\times \vec{H}^{*}\right) = \hat{n}\cdot\vec{H}\times\left(\hat{n}\times \vec{H}^{*}\right) = \hat{n}\cdot\left[\left(\vec{H}\cdot\vec{H}^{*}\right)\hat{n} - \left(\vec{H}\cdot\hat{n}\right)\vec{H}^{*}\right] = \left|\vec{H}\right|^2$$

which makes use of the fact that $\hat{n}\cdot\vec{H} = 0$ at a perfect conductor. $\vec{J}_S$ is the surface current density along the z axis; $R_S = 1/\sigma_c\delta_s$ is the (frequency-dependent) surface resistance of the conductor with the skin depth $\delta_s = \sqrt{2/\omega\mu_c\sigma_c}$, σ_c being the conductor's skin depth and conductivity, respectively, and μ_c being the permeability of the conductor (typically equal to μ_o). The magnetic field intensity $\vec{H}$ is taken as that for perfect conductors. Note that the integration is carried around both conductors to obtain the total loss. The resistance per unit length is obtained from Eqs. 59 and 75–76 as

$$R = R_s\frac{\oint_{C_1+C_2}\left|\vec{H}\right|^2 dl}{\left(\oint_{C_1}\left|\vec{H}\right| dl\right)^2} \tag{77}$$

after replacing H_t with $\left|\vec{H}\right|$, which is valid since the magnetic field has only tangential component at the surface of a perfect conductor.

Equation 77 implies that R depends on the frequency as expected due to the frequency-dependent skin effect of the conductors. When the conductors are imperfect, there will be a penetration of the magnetic field into the conductors, causing currents to flow internally within the conductors. This results in an additional inductance for the transmission line and hence increasing the inductance per unit length of the line. This extra inductance (L_i) is called "internal inductance" or "skin-effect inductance." The magnetic energy stored in this inductance is given by

$$W_{L_i} = \frac{1}{4}L_i I_o^2 \tag{78}$$

As the conductors have a finite conductivity, the transmission line exhibits a surface (or skin effect) impedance on the conductors as

$$Z_S = R_S + j\omega L_S = \frac{1}{\sigma_c\delta_S}(1+j) \tag{79}$$

where R_s and L_s are the surface resistance (or resistivity) and inductance of the conductor, respectively. The magnetic energy stored in the surface inductance L_s is

$$\begin{aligned} W_{L_S} &= \frac{1}{4}L_S\oint_{C_1+C_2}\left|\vec{J}_S\right|^2 dl \\ &= \frac{1}{4}L_S\oint_{C_1+C_2}\left|\vec{H}\right|^2 dl \end{aligned} \tag{80}$$

DOI 10.1007/978-981-4560-75-7_129-1

making use of $\vec{J}_s = \hat{n} \times \vec{H}$ on a perfect conductor. Replacing the integral with that from Eq. 76, and using Eq. 75 and the fact that $R_s = \omega L_s$ from Eq. 80, we get

$$W_{L_s} = \frac{RI_o^2}{4\omega} \tag{81}$$

which must be equal to the magnetic energy stored in the internal inductance. This leads to

$$\omega L_i = R \tag{82}$$

The total series inductance per unit length can now be obtained as

$$L = L_e + \frac{R}{\omega} \tag{83}$$

and is frequency dependent. For low-loss transmission lines, $R << \omega L_e$, and hence this additional inductance due to the conductor's finite conductivity is very small and normally neglected. Note that, from Eqs. 77 and 82, R increases as $\sqrt{\omega}$, whereas L_i decreases as $\sqrt{\omega}$. However, the reactance ωL_i is proportional to $\sqrt{\omega}$ and hence cannot be neglected at high frequencies.

As have been seen, the per-unit-length resistance R and internal inductance L_i can be evaluated from Eqs. 77 and 82, respectively. In practice, transmission lines with good conductors are used. For these transmission lines, R and L_i are typically small. Therefore, the transmission line's parameters like characteristic impedance can be accurately determined, and even approximations are used for R and L_i. For good conductors, we can derive approximate closed-form equations for R and L_i as follows.

A plane conductor as shown in Fig. 3 is considered. Its resistance is given, assuming the current is uniformly distributed over the cross section of the conductor, as

$$R_{dc} = \frac{\ell}{\sigma_c A} = \frac{\ell}{\sigma_c h W} \tag{84}$$

where A is the cross-sectional area of the conductor and σ_c is the conductivity, whose value remains almost the same from dc to infrared frequencies. At dc, the skin depth is infinity and the current would reside completely within the conductor, thus distributing uniformly over the entire cross section. Equation 84 is therefore strictly valid at dc and approximate at low frequencies where the skin depth is large. At high frequencies, the skin depth is small, and the current approximately resides within a few skin depths from

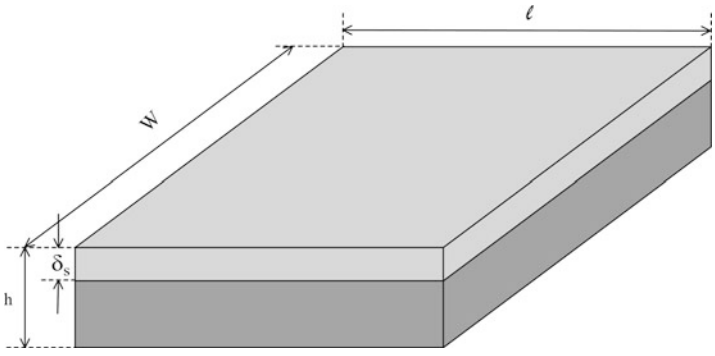


Fig. 3 A plane conductor

DOI 10.1007/978-981-4560-75-7_129-1

the surface, causing nonuniform current distribution over the conductor's entire cross section. However, within one skin depth δ_s of the surface, the current does not vary much and we can approximate the current distribution as uniform. As a result, the resistance per unit length of the conductor at high frequency, corresponding to a cross-sectional area of $\delta_s W$, can be approximated as

$$R_f = \frac{1}{\sigma_c \delta_s W} \tag{85}$$

It is noted herein that the current does not vanish at a depth equal to the skin depth; instead, it takes several skin depths to reduce to a negligible value (e.g., $4.6\delta_s$ to reach 1 % of its initial value) which should be used in determining R_f (for conductors or very lossy substrates.) These distances (e.g., $4.6\delta_s$), however, are not considered here since the resistance is not uniformly distributed over several skin depths. The use of one skin depth results in possible inaccuracy for R_f.

For good conductors, $\delta_s = 1/\alpha \simeq \sqrt{2/\omega\sigma_c\mu_o}$, where α is the attenuation constant, and so

$$R_f = \frac{1}{W}\sqrt{\frac{\omega\mu_o}{2\sigma_c}} \tag{86}$$

for good conductors. Note the relative permeability μ_r for (nonmagnetic) conductors is approximately equal to one. It is then observed that, for a given conductor with conductivity σ_c, as the frequency is increased, the skin depth reduces leading to increased resistance and attenuation constant and hence loss. This phenomenon is indeed expected for increasing frequency. R_f can be rewritten using the surface resistance R_S and δ_s as

$$R_f = \frac{R_s}{W} \tag{87}$$

which is expected since the conductor width can be considered as consisting of W unit widths connected in parallel. The transmission line's per-unit-length resistance can now be estimated as

$$R = \frac{R_{s1}}{W_1} + \frac{R_{s2}}{W_2} \tag{88}$$

where W_1, W_2 and R_{s1}, R_{s2} are the widths and surface resistances of the transmission line's two conductors, respectively, assuming plane conductors.

The internal inductance per unit length L_i of the transmission line is obtained from Eqs. 82 and 88 as

$$L_i = \frac{1}{\omega}\left(\frac{R_{s1}}{W_1} + \frac{R_{s2}}{W_2}\right) \tag{89}$$

As an example, a coaxial transmission line, whose cross section is shown in Fig. 4a, is now considered. Considering the unit-length inner conductor shown in Fig. 4b and assuming the current is concentrated within a skin depth of the surface and the frequency is sufficiently high so that the skin depth is relatively small as compared to the conductor's radius, we can neglect the conductor's curvature and approximate the inner conductor per unit length as a plane conductor having width equal to the circumference $2\pi a$ and thickness equal to the skin depth δ_s, as shown in Fig. 4c. Similarly, the outer conductor of radius b per unit length can also be approximately represented by a plane conductor of width $2\pi b$ and thickness δ_s. The

DOI 10.1007/978-981-4560-75-7_129-1

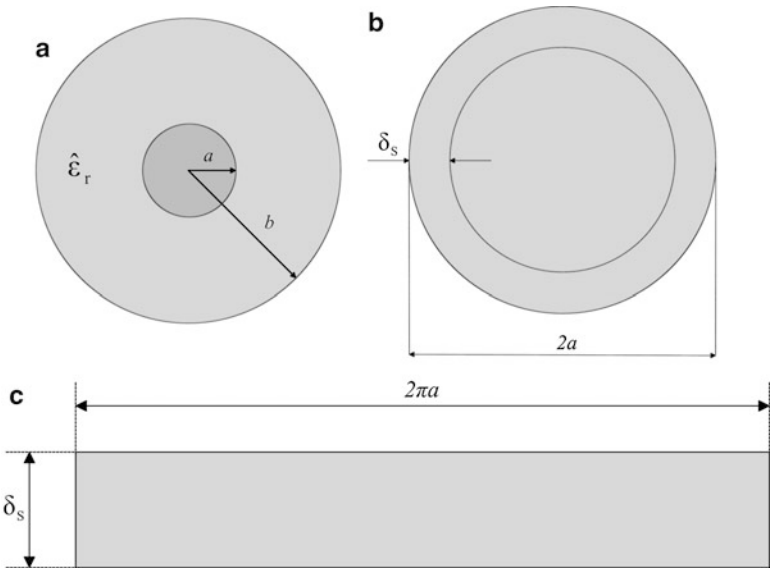


Fig. 4 Cross section of a coaxial line (**a**) and the unit-length inner conductor (**b**) with its approximate plane conductor (**c**)

per-unit-length resistance R and internal inductance L_i can be obtained from Eqs. 88 and 89, respectively, as

$$R = \frac{R_s}{2\pi a} + \frac{R_s}{2\pi b} = \frac{(a+b)R_s}{2\pi ab} \tag{90}$$

$$L_i = \frac{(a+b)R_s}{2\pi\omega ab} \tag{91}$$

assuming same material for the inner and outer conductors. As will be seen later, these results are exactly the same as those in Eqs. 101 and 102 obtained from a more rigorous analysis. It should be noted that for some planar transmission lines, the width of the second conductor (normally, ground plane), along which the current concentrates, may be much wider than that of the first conductor; in that case, a width smaller than the actual width should be used in the calculations. Otherwise, the portions of R and L_i corresponding to that conductor would be zero. For instance, for microstrip lines, the width of the ground plane used in calculations can be approximated as about five times of the width of the strip. Figure 5 shows sketches of the current distributions within the conductors of coaxial and microstrip lines.

Formulation for Simple Transmission Lines

For transmission lines with simple geometry, such as coaxial line, closed-form expressions for R, L, G, and C can be easily derived. For most practical printed-circuit transmission lines, however, these expressions are very difficult – if not impossible – to derive, and numerical methods, for example, spectral-domain method, are normally employed to obtain numerical results for the parameters.

As an illustration of the procedure to determine the transmission line's per-unit-length parameters for simple transmission lines, a (lossy) coaxial transmission line as shown in Fig. 4a is considered. The relative dielectric constant of the medium, assumed to be nonmagnetic, between the two conductors is characterized by $\hat{\varepsilon} = \varepsilon' - \varepsilon''$ and the conductors' conductivity is σ_c. First, the electric and magnetic fields for the coaxial line are determined. To facilitate this formulation, it is assumed that the conductors are perfect ($\sigma \to \infty$) so that the transmission line supports TEM mode regardless whether the medium is

DOI 10.1007/978-981-4560-75-7_129-1

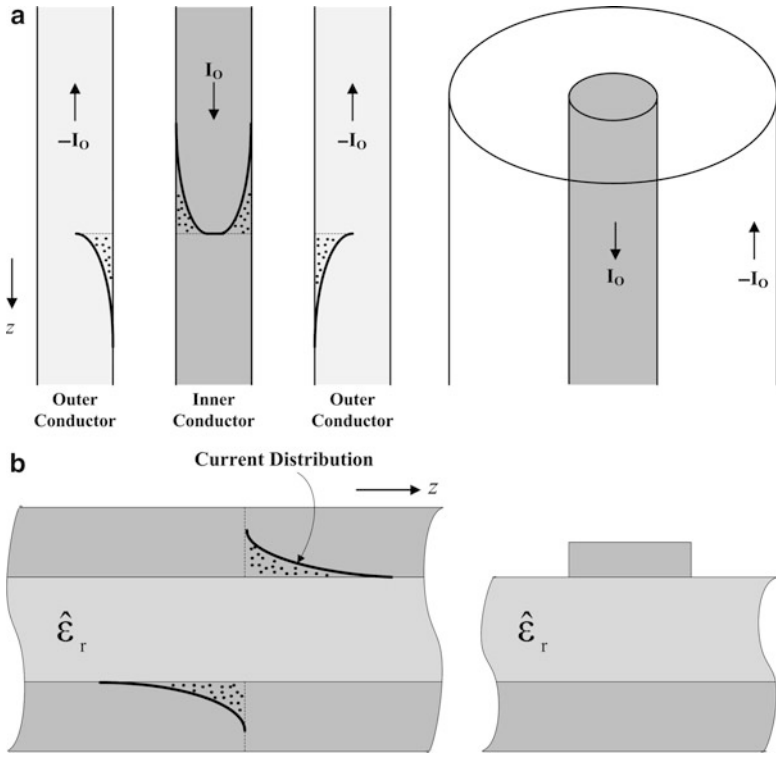


Fig. 5 Sketch of the current distribution in coaxial line (**a**) and microstrip line (**b**)

lossless or lossy. It is noted that TEM mode only exists on transmission lines with perfect conductors immersed in a single dielectric (i.e., the medium surrounding the conductors is homogeneous). For practical transmission lines including coaxial line, the conductors are typically good, and, consequently, the fields can be assumed to be almost the same as that of the TEM mode with perfect conductors. Therefore, the fields for a lossless transmission line are normally employed to evaluate the parameters of these (low-loss) practical transmission lines. For TEM mode, the fields lie on a transverse plane and, consequently, the electric field for the coaxial line has only radial component. Applying Gauss' law of Maxwell's equations over a cylindrical surface enclosing the inner conductor, the total charge per unit length on the inner conductor is obtained as

$$Q = \oint_S \vec{D} \cdot d\vec{S} = \varepsilon \int_0^{2\pi} E_r \hat{a}_r \cdot \hat{a}_r r d\varphi = 2\pi\varepsilon r E_r \tag{92}$$

The voltage between the two conductors is

$$V_o = -\int_b^a \vec{E} \cdot d\vec{l} = \int_a^b E_r dr = \frac{Q}{2\pi\varepsilon} \int_a^b \frac{dr}{r} = \frac{Q}{2\pi\varepsilon} \ln \frac{b}{a} \tag{93}$$

upon substituting E_r from Eq. 92 and taking the integral. The electric field is then obtained from Eqs. 92 to 93 as

$$E = \frac{V_o}{r \ln(b/a)} a_r \tag{94}$$

The magnetic field is given by

$$\vec{H} = \frac{\hat{a}_z \times \vec{E}}{\eta} = \frac{V_o}{r\eta \ln(b/a)} \hat{a}_\varphi \tag{95}$$

The total current is obtained from Eq. 59 as

$$I_o = \oint_{C_1} \vec{H} \cdot d\hat{l} = \int_0^{2\pi} \vec{H} \cdot \hat{a}_\varphi a d\varphi = \frac{2\pi V_o}{\eta \ln(b/a)} \tag{96}$$

Note that the contour C_1 corresponds to $r = a$. The external series inductance per unit length L_e can now be obtained by substituting ψ from Eq. 85 and I_o from Eqs. 96 to 53 as

$$\begin{aligned} L_e &= \frac{\psi}{I_o} \\ &= \frac{\mu}{2\pi} \ln\left(\frac{b}{a}\right) \end{aligned} \tag{97}$$

The shunt capacitance per unit length C is given from Eq. 54 as

$$\begin{aligned} C &= \frac{Q}{V_o} \\ &= \frac{2\pi\varepsilon}{\ln(b/a)} \end{aligned} \tag{98}$$

upon using Eq. 93.

The shunt conductance per unit length G is obtained by substituting C from Eqs. 98 to 73 as

$$G = \frac{2\pi\omega\varepsilon''}{\ln(b/a)} \tag{99}$$

The power loss per unit length on the conductors is obtained from Eq. 76, making use of Eq. 95, as

$$\begin{aligned} P_L &= \frac{R_s}{2} \oint_{C_1 + C_2} \left|\vec{H}\right|^2 dl \\ &= \frac{R_s}{2} \left\{ \frac{V_o^2}{a^2 \eta^2 [\ln(b/a)]^2} \int_0^{2\pi} a d\varphi + \frac{V_o^2}{b^2 \eta^2 [\ln(b/a)]^2} \int_0^{2\pi} b d\varphi \right\} \\ &= \frac{R_s}{2} \frac{2\pi V_o^2}{\eta^2 [\ln(b/a)]^2} \left(\frac{1}{a} + \frac{1}{b}\right) \\ &= \frac{a+b}{ab} \frac{\pi R_s V_o^2}{[\eta \ln(b/a)]^2} \end{aligned} \tag{100}$$

Substituting P_L in Eq. 100 into Eq. 75 and solving for R, leads to, upon using Eq. 96,

DOI 10.1007/978-981-4560-75-7_129-1

$$R = \frac{(a+b)R_s}{2\pi ab} \tag{101}$$

The internal inductance per unit length is obtained from Eq. 82 as

$$L_i = \frac{(a+b)R_s}{2\pi\omega ab} \tag{102}$$

Note that R and L_i in Eqs. 101 and 102, respectively, are the same as those derived in Eq. 90 and 91. The total series inductance per unit length can be obtained from Eqs. 97 and 102 as

$$L = \frac{\mu}{2\pi}\ln\frac{b}{a} + \frac{(a+b)R_s}{2\pi\omega ab} \tag{103}$$

The foregoing analysis can be employed to determine the parameters per unit length for other transmission lines with simple geometry.

Dielectric and Conductor Losses in Transmission Lines

It has been seen in sections "Transmission-Line Parameters" and "Per-Unit-Length Parameters R, L, C, and G" that the loss in transmission lines can be determined exactly from Eqs. 15 or 36 for (lossy) transmission lines and approximately from Eq. 47 for low-loss lines. This loss is attributed by three different kinds: dielectric, conduction, and radiation loss. It is a common practice to compute these three constituent losses, and the results enable us to get insight to the loss phenomena and helps identify the magnitude of each loss contribution. This information is useful for circuit design – for instance, it can help choose a proper transmission line or possibly optimize a transmission-line structure including its substrates and dimensions to reduce a certain loss.

Radiation loss of uniform transmission lines used in practice is normally very small and usually neglected, especially at low RF regions. As a result, the attenuation constant of a transmission line is commonly described as

$$\alpha = \alpha_d + \alpha_c \tag{104}$$

where α_d and α_c represent the dielectric and conductor attenuation constant, respectively, and depend on frequency. Most transmission lines have $\alpha_c >> \alpha_d$, except for highly lossy dielectrics, such as silicon, in which the dielectric loss is the dominant contribution. For homogeneous transmission lines (e.g., strip line), α_d is independent of the geometry of the line, whereas for inhomogeneous transmission lines (e.g., microstrip line), α_d is in general a function of the line's geometry. On the other hand, α_c is always a function of the geometry. It can be written, from Eq. 33,

$$\alpha = \alpha_d + \alpha_c = \frac{P_{Ld} + P_{Lc}}{2P_T} \tag{105}$$

where

$$\alpha_d = \frac{P_{Ld}}{2P_T} \tag{106}$$

$$\alpha_c = \frac{P_{Lc}}{2P_T} \tag{107}$$

with P_{Ld} and P_{Lc} representing the power losses per unit length due to imperfect dielectric and conductors, respectively. Practical transmission lines normally have small losses, and so their electric and magnetic fields are only slightly changed or perturbed from the lossless case. Therefore, a perturbation method can be employed, in which the fields for the lossless transmission line are used to approximate those of the corresponding lossy transmission line, to determine α_d and α_c. The perturbation method is convenient for analyzing low-loss transmission lines since the actual fields of these lines are not needed.

Dielectric Attenuation Constant

From the Poynting theorem, the power loss per unit length of a transmission line due to imperfect dielectric can be written as

$$P_{Ld}(z) = \frac{\sigma}{2}\int_S \vec{E}\cdot\vec{E}^* dS = \frac{\sigma}{2}\int_S \left|\vec{E}\right|^2 dS = \frac{\omega\varepsilon''}{2}\int_S \left|\vec{E}\right|^2 dS \tag{108}$$

where $\sigma = \omega\varepsilon''$ is the conductivity of the dielectric and $\vec{E}$ is the electric field intensity. Note that because of the dielectric's finite conductivity, there is current $\vec{J} = \sigma\vec{E}$ existing between the two conductors. The average power flow along the line is given as

$$P_T = \frac{1}{2}\mathrm{Re}\int_S \vec{E}\times\vec{H}^*\cdot d\vec{S} \tag{109}$$

Replacing $\vec{H}$ with

$$\frac{\vec{H} = \hat{a}_z \times \vec{E}}{\eta} \tag{110}$$

where $\eta = \sqrt{\mu/\varepsilon}$ is the intrinsic impedance of the dielectric medium and applying the vector identity for three arbitrary vectors $\vec{A}$, $\vec{B}$, and $\vec{C}$,

$$\vec{A}\times\vec{B}\times\vec{C} = \left(\vec{A}\cdot\vec{C}\right)\vec{B} - \vec{C}\left(\vec{A}\cdot\vec{B}\right) \tag{111}$$

gives

$$P_T(z) = \frac{1}{2\eta}\int_S \vec{E}\cdot\vec{E}^* dS = \frac{1}{2\eta}\int_S \left|\vec{E}\right|^2 dS \tag{112}$$

which can also be derived as

DOI 10.1007/978-981-4560-75-7_129-1

$$P_T(z) = \frac{\eta}{2}\int_S \vec{H}\cdot\vec{H}^* dS = \frac{\eta}{2}\int_S \left|\vec{H}\right|^2 dS \tag{113}$$

Here, η is assumed real, which is a good approximation for low-loss materials. The dielectric attenuation constant can be derived from Eqs. 106, 108, and 112 as

$$\alpha_d = \frac{P_{Ld}}{2P_T} = \frac{\sigma\eta}{2} = \frac{\omega\varepsilon''\eta}{2} \tag{114}$$

For a low-loss dielectric with (complex) relative dielectric constant $\hat{\varepsilon}_r = \varepsilon_r' - j\varepsilon'', \varepsilon_r'' << \varepsilon_r'$ and its intrinsic impedance

$$\eta = \sqrt{\frac{\mu}{\hat{\varepsilon}}} = \frac{\eta_o}{\sqrt{\varepsilon_r' - j\varepsilon_r''}} \simeq \frac{\eta_o}{\sqrt{\varepsilon_r'}} \tag{115}$$

where $\eta_o = 120\pi\Omega$ is the intrinsic impedance of air. Note that $\varepsilon_r' = \varepsilon_r$ of the dielectric. The dielectric attenuation constant for a low-loss TEM transmission line can now be approximated as

$$\alpha_d \cong \frac{k_o}{2}\frac{\varepsilon_r''}{\sqrt{\varepsilon_r}} \tag{116}$$

where $k_o = \omega\sqrt{\varepsilon_o\mu_o}$ is the wave number in air and ε_r is used in place of ε_r'.

Conductor Attenuation Constant

It is well known that as a signal propagates along a perfect conductor, its energy cannot penetrate into the conductor. However, if the conductor is nonperfect, some of the energy would enter the conductor, yet attenuated in accordance to the attenuation function $e^{-2\alpha_c r}$, where r is the distance from the conductor surface in the direction perpendicular to it. The density of power flow into a conductor is equal to the power dissipated within the conductor itself. Therefore, the power loss per unit length due to an imperfect conductor is given as

$$\begin{aligned} P_{Lc}(z) &= \frac{1}{2}\mathrm{Re}\int_{C_1+C_2}\left(\vec{E}\times\vec{H}^*\right)\cdot\hat{a}_r dS = \frac{1}{2}\mathrm{Re}\int_{C_1+C_2}\left(\hat{a}_r\times\vec{E}\right)\cdot\vec{H}^* = \frac{1}{2}\mathrm{Re}\int_{C_1+C_2} Z_S\left|\vec{H}\right|^2 dS \\ &= \frac{1}{2}R_S\int_{C_1+C_2}\left|\vec{H}\right|^2 dS = \frac{1}{2}R_S\int_{C_1+C_2}\left|\vec{H}\right|^2 d\ell \end{aligned} \tag{117}$$

where C_1 and C_2 represent the two conductors of the transmission line, η_c is the intrinsic impedance of the conductor, Z_S is the surface impedance, and R_S is its real part or the surface resistance (or resistivity) of the conductor as given in Eq. 79. $\vec{H}$ is the magnetic field assuming no loss; that is, its magnitude is equal to the magnitude of the magnetic field at the conductor surface. The conductor attenuation constant can be derived from Eqs. 107, 113, and 117 as

DOI 10.1007/978-981-4560-75-7_129-1

$$\alpha_c = \frac{P_{Lc}}{2P_T} = \frac{R_S \int_{C_1+C_2} \vec{H} \cdot \vec{H}^* dl}{2\eta_c \int_S \vec{H} \cdot \vec{H}^* dS} \tag{118}$$

It can also be derived as

$$\alpha_c = \frac{R_S \int_{C_1+C_2} \vec{E} \cdot \vec{E}^* dl}{2\eta \int_S \vec{E} \cdot \vec{E}^* dS} \tag{119}$$

It is particularly noted that, as a signal propagates along transmission lines, the electric and magnetic fields and current of the signal penetrate into the surfaces of the transmission lines' conductors and reside within a few skin depths of the conductors. The thickness of the conductors in transmission lines should therefore be at least larger than the skin depth at the operating frequencies to minimize the loss due to the conductors. If the conductor thickness is comparable to the skin depth, a significant conduction loss would occur.

As an example for determining the attenuation constant of transmission lines, a coaxial transmission line whose inner and outer radii are a and b, respectively, as shown in Fig. 4a is considered. The dielectric is assumed to be lossy and characterized by $\hat{\varepsilon}_r = \varepsilon_r' - j\varepsilon_r''$ where $\varepsilon_r' = \varepsilon_r$. The conductors are imperfect with a finite conductivity σ_c. Using perturbation method, it is assumed that the fields of this transmission line are given as those of the lossless coaxial line:

$$\vec{E} = \hat{a}_r \frac{V_o}{\ln\left(\frac{b}{a}\right)} \frac{1}{r} e^{-j\beta z} \tag{120}$$

$$\vec{H} = \hat{a}_r \frac{V_o}{\eta \ln\left(\frac{b}{a}\right)} \frac{1}{r} e^{-j\beta z} \tag{121}$$

where V_o is the potential of the inner conductor, r is the radius between a and b, and $\beta = k = k_o\sqrt{\varepsilon_r}$. The potential of the outer conductor is assumed to be zero. The power flow along the coaxial line is obtained as

$$P_T = \frac{1}{2}\mathrm{Re}\int_S \vec{E} \times \vec{H}^* \cdot d\vec{S} = \frac{1}{2}\int_0^{2\pi}\int_a^b \vec{E} \times \vec{H}^* \cdot \hat{a}_z r dr d\varphi = \frac{\pi V_o^2}{\eta \ln(b/a)} \tag{122}$$

The power loss per unit length due to dielectric loss is

DOI 10.1007/978-981-4560-75-7_129-1

$$\begin{aligned} P_{Ld} &= \frac{\omega\varepsilon''}{2}\int_S \left|\vec{E}\right|^2 dS \\ &= \frac{\omega\varepsilon''}{2}\int_0^{2\pi}\int_a^b \left|\vec{E}\right|^2 r dr d\varphi \\ &= \frac{\pi\omega\varepsilon'' V_o^2}{\ln(b/a)} \end{aligned} \tag{123}$$

The dielectric attenuation constant can be derived from Eqs. 106, 122, to 123 as

$$\alpha_d = \frac{k_o\varepsilon_r''}{2\sqrt{\varepsilon_r}} \tag{124}$$

which is the same as Eq. 116 as expected for TEM transmission lines.

The power loss per unit length due to conductor loss is

$$\begin{aligned} P_{Lc} &= \frac{R_S}{2}\int_{C_1+C_2} \left|\vec{H}\right|^2 dl \\ &= \frac{R_S}{2}\left[\int_0^{2\pi}\left|\vec{H}(r=a)\right|^2 a d\varphi + \int_0^{2\pi}\left|\vec{H}(r=b)\right|^2 b d\varphi\right] \\ &= \frac{R_S\pi V_o^2}{\eta^2\ln^2(b/a)}\left(\frac{1}{a}+\frac{1}{b}\right) \end{aligned} \tag{125}$$

The conductor attenuation constant is obtained from Eqs. 107, 122, and 125 as

$$\alpha_c = \frac{R_S}{2\eta\ln(b/a)}\left(\frac{1}{a}+\frac{1}{b}\right) \tag{126}$$

which depends on the transmission line's geometry.

The (total) attenuation constant can now be determined using Eqs. 104, 124, and 126 as

$$\alpha = \frac{k_o\varepsilon_r''}{2\sqrt{\varepsilon_r}} + \frac{R_S}{2\eta\ln(b/a)}\left(\frac{1}{a}+\frac{1}{b}\right) \tag{127}$$

Dispersion and Distortion in Transmission Lines

Dispersion

When the velocity of wave or signal (normally TEM or quasi-TEM) propagating along a transmission line is dependent upon frequency, the transmission line is said to have dispersion. There is no dispersion in an (ideal) lossless transmission line because the velocity, as given by Eq. 41, is independent of frequency.

DOI 10.1007/978-981-4560-75-7_129-1

Any practical transmission line, however, is lossy and its phase constant, in general, does not vary linearly with frequency. This results in a velocity depending on frequency, thus causing dispersion. As an example, the phase constant for low-loss transmission lines, as derived in Eq. 48, is in general a nonlinear function of frequency, and the resultant velocity, from Eq. 37, is hence dependent upon frequency. For good transmission lines, the rate of change of the velocity versus frequency is, in general, relatively small until the frequency reaches an extremely high value. Also, the velocity changes very little at low frequencies. Consequently, narrow-band signals and signals whose spectra consist of very low frequencies suffer relatively small dispersion within the operating bandwidth.

Dielectric substrates used in practical transmission lines usually have relative dielectric constants almost constant versus frequency, particularly in the low RF region. Homogeneous transmission line, such as strip line, has velocity depending exclusively on the relative dielectric constant, as given by $v = c/\sqrt{\varepsilon_r}$, hence possessing very little dispersion. Most transmission lines used in practice, however, are inhomogeneous, having more than one dielectric substrate – for example, microstrip line. An inhomogeneous transmission line has velocity depending on the effective relative dielectric constant, as shown in Eq. 217. Assuming the relative dielectric constant is independent of frequency, calculations show that in general the effective relative dielectric constant increases with frequency, and this change can be substantial, especially for high dielectric-constant substrates. Consequently, the dispersion for inhomogeneous transmission lines can be large. Due to the dispersion, signals of different frequencies travel at different velocities. Since the lengths of transmission lines used in RF circuits are typically viewed in terms of wavelength which varies as $\lambda = v/f$ (even they are given as physical dimension), which are affected by the transmission line's dispersion, the dispersion in transmission lines needs to be considered in the design and analysis of wideband RF circuits. Another effect of the dispersion is the distortion it imposes on a propagating signal whose spectrum contains multiple frequencies occurring simultaneously, which is critical for high-frequency and high-speed circuits.

Figure 6 shows the frequency-dependent effective relative dielectric constant for microstrip lines with SiO_2 and silicon as the substrate that can be used in RFICs. In practice, the relative dielectric constant slightly decreases as the frequency is increased. However, this rate of reduction is relatively small as compared to the rate of increase of the effective relative dielectric constant versus frequency, making it a secondary effect as compared to the effective relative dielectric constant. Some transmission lines have effective relative dielectric constant staying almost constant from dc up to a certain cutoff frequency and then increasing as the operating frequency exceeding that cutoff frequency. These transmission lines can hence have very small dispersion up to the cutoff frequency.

Distortion

There are various kinds of waveforms that may be used in RF circuits. However, for the discussion purposes, three kinds of waveforms are considered here: continuous sinusoidal waveform, periodic waveform, and non-periodic waveform. The sinusoid contains only single frequency and hence retains its original wave shape through linear networks; therefore, signal distortion is not much of an issue except a reduction in amplitude. It is noted that, even for wideband circuits, each signal still only has one frequency. The non-periodic and periodic signals, such as a single sinusoidal pulse or train of pulses, respectively, however, contain multiple constituent signals operating at different frequencies simultaneously. As these composite signals propagate along a transmission line, their amplitude and phase change versus frequency due to the frequency-dependent loss and dispersion of the transmission line, respectively, causing distortion in the signal waveform. This distortion can be significant for some transmission lines and signals with wide bandwidths, especially at very high frequencies, as it can alter the signal substantially as the signal propagates along a transmission line. Distortion, therefore, needs to be considered in transmission-line design as well as in circuit analysis and design. An example is pulse signals propagating in ultra-wideband (UWB) RF circuits or digital circuits, where distortion can be

DOI 10.1007/978-981-4560-75-7_129-1

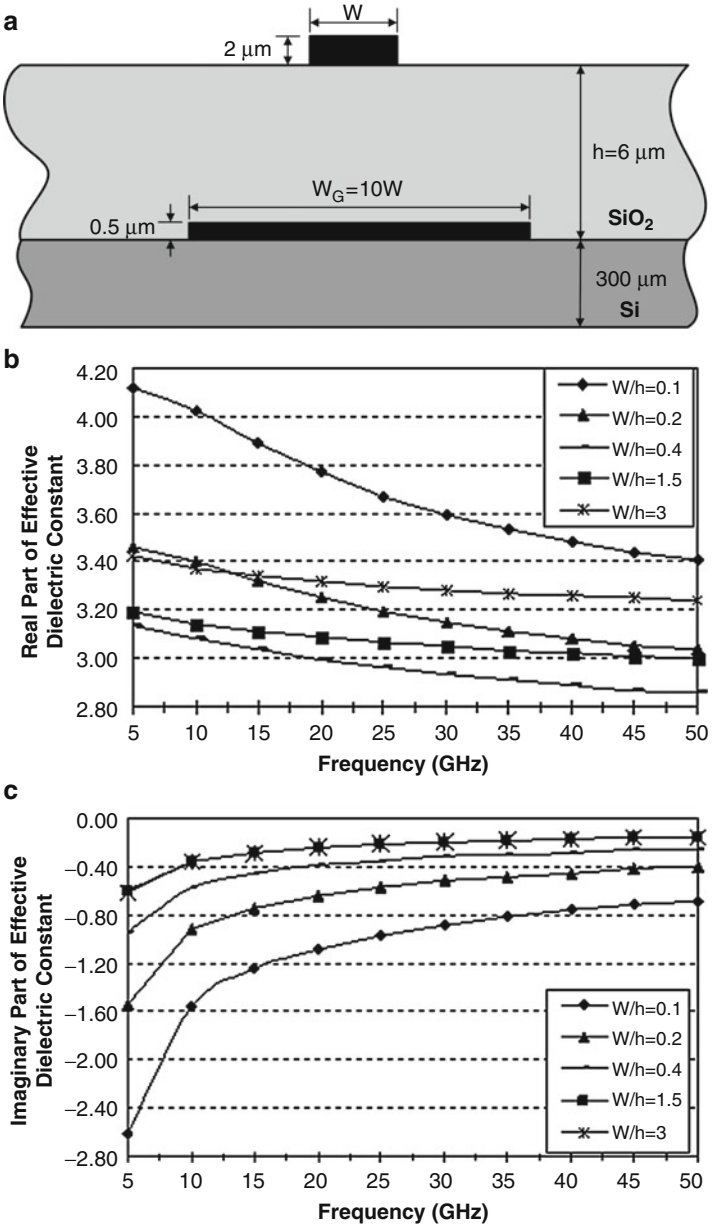


Fig. 6 Microstrip line (**a**) and its real (**b**) and imaginary (**c**) part of effective relative dielectric constant versus frequency for SiO_2 with loss tangent tan $\delta = 0.0002$ and $\varepsilon_r = 4.1$ and silicon with resistivity ρ=5Ω−cm and $\varepsilon_r = 12.5$ (**b**). The conductor is copper having conductivity $\sigma = 5.8 \times 10^7$mho/m

significant within the operating frequency range. It is noted that UWB frequencies are from 3.1 to 10.6 GHz and UWB circuits for both communications and radar can be based on pulse signals operating within this frequency range.

Periodic signals such as a train of pulses can be expressed using Fourier series as

$$\begin{aligned} V(t) &= a_0 + \sum_{n=1}^{\infty} [a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t)] \\ &= \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t} \end{aligned} \tag{128}$$

where ω_0 is the fundamental frequency and a_0, a_n, b_n, and c_n are the Fourier coefficients. The signal, as can be seen, consists of multiple sinusoidal signals at discrete frequencies $\omega = n\omega_0$. Non-periodic signals such as a single pulse may also be represented using the exponential Fourier series given in Eq. 128, in which the period is extended to infinity, or using Fourier integral as

$$V(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} V(\omega)e^{j\omega t}d\omega \tag{129}$$

where $V(\omega)$ is the Fourier transform of $V(t)$, which shows that a non-periodic signal also contains multiple sinusoidal signals of different frequencies in a continuous spectrum.

An infinitely long transmission line is now considered, and it is assumed that a signal propagating along the line has a spectrum containing multiple frequencies $f_1, f_2, \ldots, f_N$, that is, the signal contains multiple signals operating at these frequencies concurrently. Since the propagation constant of the transmission line is function of frequency, the amplitude and phase constants of these constituent signals are different. The (phasor) voltage of the composite signal at any location on the line at any given time can be determined as the summation of its constituent signals, using Eq. 28, as

$$V(z) = \sum_{n=1}^{N} V_n(z) = \sum_{n=1}^{N} V_n(0)e^{-\alpha_n z}e^{-j\beta_n z} \tag{130}$$

where $V_n(z)$ and $V_n(0)$ are the voltages of the constituent signal at frequency f_n at z and $z = 0$, respectively, and α_n and β_n are the respective attenuation and phase constant at frequency f_n. Since the constituent signals arrive at z with different amplitudes and phases, they produce a distorted signal when added together. The phase distortion is caused by the dispersion of the transmission line, which typically causes the waveform to spread out and change in shape, while the amplitude distortion is due to the transmission line's loss which results in amplitude reduction. Additional distortion is caused by the reflection and cross-coupling occurred in the transmission line. Figure 7 illustrates the distortion of a signal propagating in a (practical) lossy and dispersive transmission line.

Distortionless Transmission Lines

The idea of distortionless was first proposed by Oliver Heaviside in 1887. It is postulated that the per-unit-length parameters R, L, G, and C of a transmission line are related by the following condition:

$$\frac{R}{L} = \frac{G}{C} \tag{131}$$

The propagation constant of such a transmission line can be written using Eq. 15 and the condition in Eq. 131 as

$$\begin{aligned}\gamma &= \alpha + j\beta = \sqrt{(R+j\omega L)(G+j\omega C)} \\ &= \sqrt{(R+j\omega L)\left(\frac{RC}{L}+j\omega C\right)} = \sqrt{\frac{L}{C}}(R+j\omega L)\end{aligned} \tag{132}$$

from which,

DOI 10.1007/978-981-4560-75-7_129-1

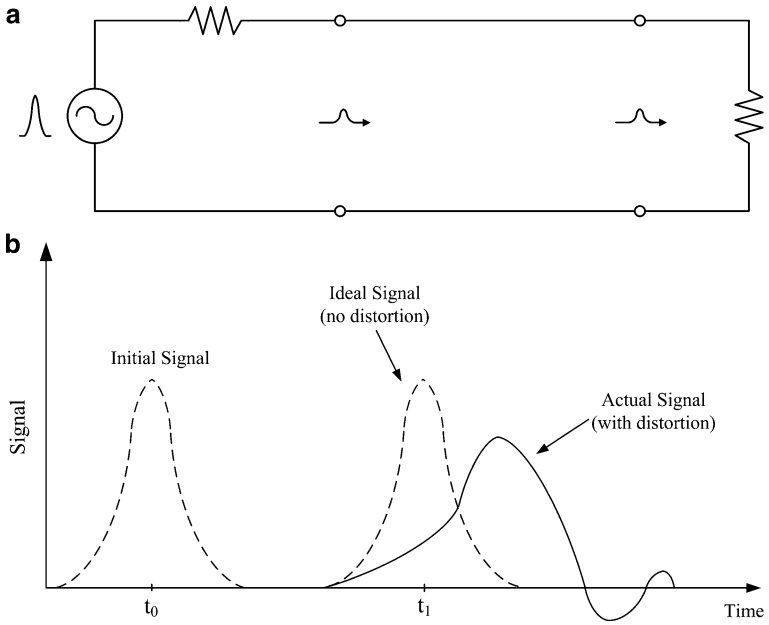


Fig. 7 Pulse signal propagating in a transmission line (**a**) and its distortion at t_1 with respect to the initial signal at t_o

$$\alpha = R\sqrt{\frac{C}{L}} \tag{133}$$

and

$$\beta = \omega\sqrt{LC} \tag{134}$$

are obtained. The characteristic impedance and phase velocity are obtained from Eqs. 26 and 37, making use of Eqs. 131 and 134, respectively, as

$$Z_o = \sqrt{\frac{R+j\omega L}{G+j\omega C}} = \sqrt{\frac{R+j\omega L}{\frac{RC}{L}+j\omega C}} = \sqrt{\frac{L}{C}} \tag{135}$$

and

$$v = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}} \tag{136}$$

which are exactly the same as those for lossless transmission lines. Equations 133 and 136 show that the transmission line has constant loss and velocity versus frequency, implying that the waveform of signals traversing the line retains its shape (with reduced amplitude) and hence suffers no (phase) distortion. It can therefore be concluded that any transmission line that satisfies the condition stated in Eq. 131 is distortionless. Since R and G affect the loss of transmission lines, a distortionless transmission should only be formed by adjusting L and/or C to meet the distortionless condition – for instance, placing additional series inductors periodically along the conductors of a transmission line to increase the inductance per unit length, assuming the added inductance is distributed uniformly over the segment

DOI 10.1007/978-981-4560-75-7_129-1

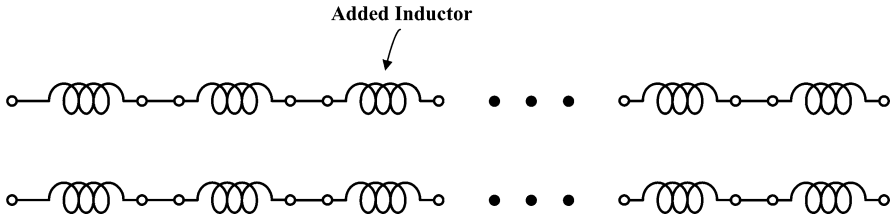


Fig. 8 Distortionless transmission formed by adding series inductors periodically

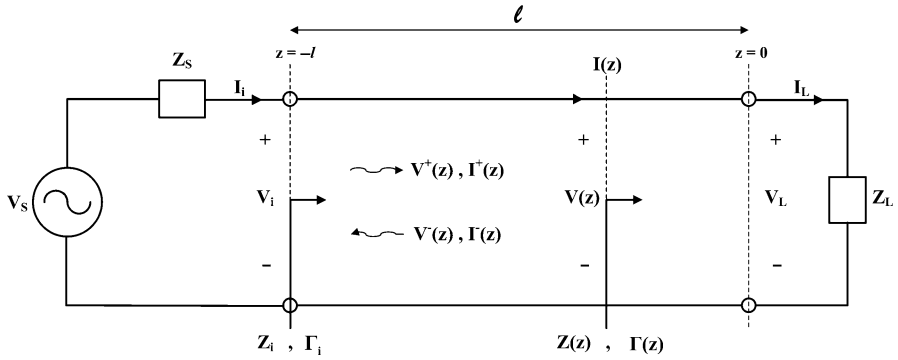


Fig. 9 A finitely long transmission line

where the inductor is added, as shown in Fig. 8. As R, L, G, and C represent a transmission line whose length is very small as compared to wavelength, the smaller the interval (period) that the additional inductors and/or capacitors are placed, the better representation for the transmission line and, hence, the better distortionless transmission that can be realized. Using smaller intervals also makes the assumption of uniform distribution for the added inductors and/or capacitors more valid.

Impedance, Reflection Coefficients, and Standing-Wave Ratios

Impedance

A transmission line of finite length, terminated with a load $Z_L \neq Z_o$, as shown in Fig. 9, is considered. It is also assumed that the transmission line is driven by a steady-state sinusoidal voltage source to facilitate the derivation of equations for the impedance, reflection coefficients, and standing-wave ratios (SWR). This assumption is valid since, like the transmission-line parameters discussed previously, these parameters are also independent of the driving source's time variation. It is assumed that the origin of the coordinate ($z = 0$) is located at the load Z_L and the transmission-line input is at $z = -\ell$. The voltage at any location, obtained as the solution of transmission-line equation, consists of the forward and backward voltage waves as described earlier. The forward voltage wave $V^+(z) = V_o^+ e^{-\gamma z}$ is also called the incident voltage wave, whereas the backward voltage wave $V^-(z) = V_o^- e^{\gamma z}$ can be called the reflected voltage wave. This reflected voltage wave is due to the reflection of the incident voltage wave along the finitely long transmission line.

The voltage V_L and current I_L at the load ($z = 0$) can be derived from Eqs. 20 to 21 as

$$V_L = V_o^+ + V_o^- \tag{137}$$

and

$$I_L = I_o^+ + I_o^- \tag{138}$$

Equation 138 can be rewritten, making use of Eq. 26, as

$$I_L = \frac{V_o^+}{Z_o} - \frac{V_o^-}{Z_o} \tag{139}$$

Dividing Eq. 137 by Z_o and adding the resulting equation to Eq. 139 gives

$$V_o^+ = \frac{1}{2}(V_L + I_L Z_o) \tag{140}$$

or

$$V_o^+ = \frac{1}{2} I_L (Z_L + Z_o) \tag{141}$$

after substituting $V_L = Z_L I_L$. Similarly, dividing Eq. 137 by Z_o, subtracting the resulting equation from Eq. 140, and using $V_L = Z_L I_L$ yield

$$V_o^- = \frac{1}{2} I_L (Z_L - Z_o) \tag{142}$$

Now substituting V_o^+ and V_o^- from Eqs. 141 and 142, respectively, into Eq. 20 gives

$$V(z) = \frac{I_L}{2}[(Z_L + Z_o)e^{-\gamma z} + (Z_L - Z_o)e^{\gamma z}] \tag{143}$$

from which

$$V^+(z) = \frac{I_L}{2}(Z_L + Z_o)e^{-\gamma z} \tag{144}$$

and

$$V^-(z) = \frac{I_L}{2}(Z_L - Z_o)e^{\gamma z} \tag{145}$$

are obtained. Similarly, the current at any point along the transmission line can be derived from Eq. 21, making use of Eqs. 26, 141, and 142, as

$$I(z) = \frac{I_L}{2Z_o}[(Z_L + Z_o)e^{-\gamma z} - (Z_L - Z_o)e^{\gamma z}] \tag{146}$$

which is formed by

$$I^{+}(z) = \frac{I_L}{2Z_o}(Z_L + Z_o)e^{-\gamma z} \tag{147}$$

and

$$I^{-}(z) = -\frac{I_L}{2Z_o}(Z_L - Z_o)e^{\gamma z} \tag{148}$$

The impedance at any location z looking toward the load is obtained from Eqs. 143 and 146 as

$$Z(z) = \frac{V(z)}{I(z)} = Z_o \frac{(Z_L + Z_o)e^{-\gamma z} + (Z_L - Z_o)e^{\gamma z}}{(Z_L + Z_o)e^{-\gamma z} - (Z_L - Z_o)e^{\gamma z}}$$

$$= Z_o \frac{Z_L(e^{\gamma z} + e^{-\gamma z}) - Z_o(e^{\gamma z} - e^{-\gamma z})}{Z_L(e^{-\gamma z} - e^{\gamma z}) + Z_o(e^{\gamma z} + e^{-\gamma z})} \tag{149}$$

Using the trigonometry identities

$$\cos h\,\gamma z = \frac{e^{\gamma z} + e^{-\gamma z}}{2} \tag{150}$$

and

$$\sin h\,\gamma z = \frac{e^{\gamma z} - e^{-\gamma z}}{2} \tag{151}$$

Eq. 149 becomes

$$Z(z) = Z_o \frac{Z_L \cos h\,\gamma z - Z_o \sin h\,\gamma z}{-Z_L \sin h\,\gamma z + Z_o \cos h\,\gamma z} \tag{152}$$

Dividing Eq. 152 by cosh γz gives

$$Z(z) = Z_o \frac{Z_L - Z_o \tan h\,\gamma z}{Z_o - Z_L \tan h\,\gamma z} \tag{153}$$

Taking a reciprocal of Eq. 153, the admittance at any location z looking toward the load can be obtained as

$$Y(z) = Y_o \frac{Y_L - Y_o \tan h\,\gamma z}{Y_o - Y_L \tan h\,\gamma z} \tag{154}$$

where $Y_o = 1/Z_o$ and $Y_L = 1/Z_L$ are the characteristic admittance of the transmission line and the load admittance, respectively. Note that the impedance and admittance calculations based on Eqs. 153 and 154, respectively, only take "negative" values for z. Of particular interest are the impedance and admittance at the input of the transmission line ($z = -\ell$) looking into the load

$$Z_i = Z_o \frac{Z_L + Z_o \tan h\,\gamma l}{Z_o + Z_L \tan h\,\gamma l} \tag{155}$$

DOI 10.1007/978-981-4560-75-7_129-1

and

$$Y_i = Y_o \frac{Y_L + Y_o \tan h\,\gamma l}{Y_o + Y_L \tan h\,\gamma l} \tag{156}$$

Reflection Coefficients

The ratio of the reflected voltage wave to the incident voltage wave is defined as the voltage reflection coefficient and is obtained from Eqs. 144 to 145 as

$$\Gamma(z) \equiv \frac{V^-(z)}{V^+(z)} = \frac{Z_L - Z_o}{Z_L + Z_o} e^{2\gamma z} \tag{157}$$

Equation 149 can be rewritten, upon using Eq. 157, as

$$Z(z) = Z_o \frac{1 + \Gamma(z)}{1 - \Gamma(z)} \tag{158}$$

from which it can be derived that

$$\Gamma(z) = \frac{Z(z) - Z_o}{Z(z) + Z_o} \tag{159}$$

At the load ($z = 0$), the reflection coefficient, namely, the load reflection coefficient, is obtained as

$$\Gamma_L = |\Gamma_L| e^{j\phi_L} = \frac{V_o^-}{V_o^+} = \frac{Z_L - Z_o}{Z_L + Z_o} \tag{160}$$

where ϕ_L is its phase. Combining Eqs. 157 and 160 gives

$$\Gamma(z) = \Gamma_L e^{2\gamma z} \tag{161}$$

Similarly, the current at any point along the transmission line can be derived from Eq. 21, making use of Eqs. 26, 140, and 141, as

$$I(z) = \frac{I_L}{2Z_o}[(Z_L + Z_o)e^{-\gamma z} - (Z_L - Z_o)e^{\gamma z}] \tag{162}$$

from which the current reflection coefficient $\Gamma_I(z)$ can be obtained as the ratio between the reflected and incident current waves

$$\Gamma_I(z) \equiv \frac{I^-(z)}{I^+(z)} = -\frac{Z_L - Z_o}{Z_L + Z_o} e^{2\gamma z} \tag{163}$$

Comparing Eqs. 157 and 163 leads to

DOI 10.1007/978-981-4560-75-7_129-1

$$\Gamma_I(z) = -\Gamma(z) \tag{164}$$

It can be seen, from Eqs. 157 and 163, that the magnitudes of the reflection coefficients in a lossless transmission line are always constant; only their phases vary as a function of location. In practice, the voltage reflection coefficient is commonly used to indicate the reflection coefficient. Note that the reflection coefficient is dimensionless and its magnitude varies between 0 and 1. The magnitude of reflection coefficient is also often expressed in decibel (dB) as $10 \log |\Gamma(z)|^2$, which is known as the return loss.

Standing-Wave Ratio

The voltage at any location along a transmission line is obtained from Eq. 20, upon using Eq. 160, as

$$\begin{aligned} V(z) &= V_o^+ e^{-\alpha z} e^{-j\beta z} + V_o^- e^{\alpha z} e^{j\beta z} \\ &= V_o^+ e^{-\alpha z} e^{-j\beta z} \left(1 + |\Gamma_L| e^{2\alpha z} e^{j(2\beta z + \phi_L)}\right) \end{aligned} \tag{165}$$

Equation 165 indicates that the amplitude and phase of the incident voltage wave reduces as the load is approached, as illustrated in Fig. 10, whereas those of the reflected voltage wave increases. So it is expected that there are locations along the transmission line, at which the amplitudes of the incident and reflected waves are equal, and there are also locations at which the phases of these opposite waves are equal or 180° out of phase. At the equal-phase locations, the two waves will add constructively while they will subtract at the out-of-phase locations.

The magnitude of the voltage at any point on a transmission line can be obtained from Eq. 165 as

$$|V(z)| = \left|V_o^+ e^{-\alpha z}\right| \left|1 + |\Gamma_L| e^{2\alpha z} e^{j(2\beta z + \phi_L)}\right| \tag{166}$$

Replacing $e^{j(2\beta z + \phi_L)}$ with $\cos(2\beta z + \phi_L) + j \sin(2\beta z + \phi_L)$ and taking the magnitude of the resulting equation gives

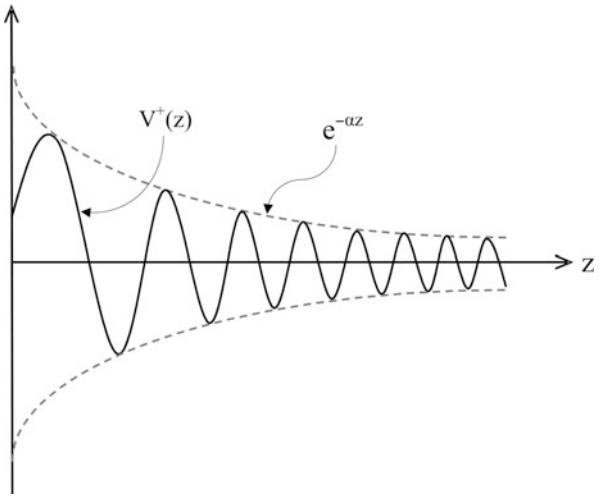


Fig. 10 Illustration of incident wave $V^+(z)$ along a transmission line

DOI 10.1007/978-981-4560-75-7_129-1

$$\begin{aligned}|V(z)| &= \left|V_o^+ e^{-\alpha z}\right| \left\{\left[1+|\Gamma_L| e^{2\alpha z}\cos\left(2\beta z+\phi_L\right)\right]^2 + \left|\Gamma_L^2\right| e^{4\alpha z}\sin^2\left(2\beta z+\phi_L\right)\right\}^{1/2} \\ &= \left|V_o^+ e^{-\alpha z}\right| \left\{\left(1+|\Gamma_L| e^{2\alpha z}\right)^2 - 2|\Gamma_L| e^{2\alpha z}\left[1-\cos\left(2\beta z+\phi_L\right)\right]\right\}^{1/2}\end{aligned} \tag{167}$$

which becomes, after using the trigonometry $\cos 2x = 1 - 2\sin^2 x$,

$$|V(z)| = \left|V_o^+ e^{-\alpha z}\right| \left\{\left(1+|\Gamma_L| e^{2\alpha z}\right)^2 - 4|\Gamma_L| e^{2\alpha z}\sin^2\left(\beta z+\phi_L/2\right)\right\}^{1/2} \tag{168}$$

Examining Eq. 168 reveals that it has maximum values of

$$|V(z)|_{\max} = \left|V_o^+ e^{-\alpha z}\right| \left(1+\left|\Gamma_L e^{2\alpha z}\right|\right) \tag{169}$$

corresponding to

$$z = \frac{n\pi - \phi_L/2}{\beta}, \quad n = 0, 1, 2, \cdots \tag{170}$$

and minimum values of

$$|V(z)|_{\min} = \left|V_o^+ e^{-\alpha z}\right| \left(1-\left|\Gamma_L e^{2\alpha z}\right|\right) \tag{171}$$

corresponding to

$$z = \frac{(n+1/2)\pi - \phi_L/2}{\beta}, \quad n = 0, 1, 2, \cdots \tag{172}$$

These results show that the magnitude of the voltage along a transmission line fluctuates as a function of location between maximum and minimum values at periodic positions. The maximum and minimum voltages are due to the addition between the incident and reflected voltage waves at locations where they have same phase and the subtraction at locations where they are 180° out of phase, respectively. The magnitude also reduces as the signal propagates along the transmission line due to loss in the line, which also results in different values for maximum (minimum) at different locations. This is in contrast with the circuit theory, which indicates that the voltage along a conductor (which is part of a two-conductor transmission line) is always constant. The spacing between two successive maxima or minima is a half-wavelength as can be determined from Eqs. 170 and 172, respectively. Equations 170 and 172 also indicate that the distance between a maximum and its nearest minimum is a quarter-wavelength.

The resultant voltage waveform formed by the incident and reflected signals, as given in Eqs. 165 and 168, is referred to as the voltage standing-wave pattern. It is noted that, at the load, the incident voltage amplitude is smallest with respect to those away from the load, whereas the amplitude of the reflected voltage is largest as compared to others at different locations. These incident and reflected voltage amplitudes at the load are thus not equal (unless $Z_L = 0$ or ∞), thereby do not create a pure standing wave at the load except when $Z_L = 0$ or ∞. However, at certain location z_o along the line, the amplitudes of the incident and reflected waves are equal as discussed earlier, and hence, standing waves would be formed in the vicinity of this location, as illustrated in Fig. 11. Far away from the location z_o in the +z and −z directions, the incident and reflected waves die out, respectively, resulting in the standing waves gradually vanishing as the waves move away from the location. The standing wave is therefore a function

DOI 10.1007/978-981-4560-75-7_129-1

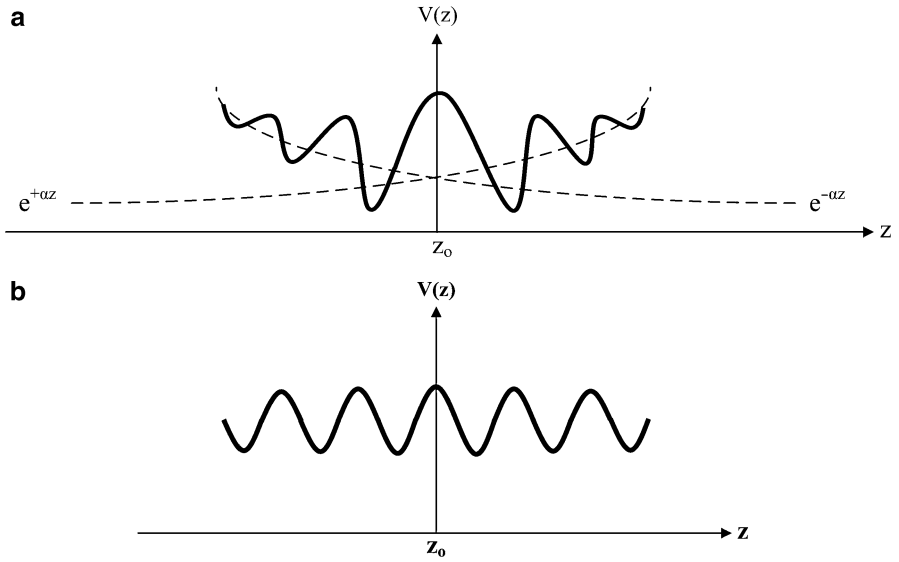


Fig. 11 Standing-wave pattern in high-loss (**a**) and low-loss (**b**) transmission line terminated with load $Z_L \neq Z_o$. The standing-wave pattern for lossless line is similar to that in (**b**)

of location along the line; its magnitude is large near the location z_o and reduces at locations away from z_o. It should be noted that, in general, a pure standing wave does not exist along a transmission line except when $\left|V_o^+ e^{-\alpha z}\right| = \left|V_o^- e^{\alpha z}\right|$, e.g., in a short-circuited lossless transmission line. Indeed, the wave can be considered as having a standing wave and a traveling wave simultaneously at any location and at any time along a transmission line. In practical transmission lines, the loss is normally small, causing an almost constant-magnitude standing-wave pattern along the line as seen in Fig. 11b, which is similar to the standing-wave pattern for a lossless transmission line. True standing wave is only possible in transmission lines terminated with a short, open, or reactance as seen later.

The ratio between $|V(z)|_{\max}$ and $|V(z)|_{\min}$ is defined as the voltage standing-wave ratio (VSWR):

$$VSWR = \frac{1 + \left|\Gamma_L e^{2\alpha z}\right|}{1 - \left|\Gamma_L e^{2\alpha z}\right|} \tag{173}$$

which is, like reflection coefficient, dimensionless and varies between 1 and infinity.

Following the same procedure, it can also be shown that the current along a transmission line oscillates between its maximum and minimum values as the current propagates, forming a current wave consisting of both standing and traveling waves. The maximum and minimum values for the current, however, occur at locations corresponding to the minimum and maximum voltages, respectively. These maximum or minimum locations are also successively spaced a half-wavelength apart. The current standing-wave ratio is equal to VSWR. In practice, VSWR is commonly used instead of the current standing-wave ratio.

Perfect Match and Total Reflection

Perfect Match

When the load impedance is equal to the transmission-line characteristic impedance, it can be obtained respectively from Eqs. 160 and 173 that

$$\Gamma_L = 0 \tag{174}$$

and

$$VSWR = 1 \tag{175}$$

which imply that the transmission line has no reflection and is perfectly matched at the load. Under this condition, all incident voltage and current waves, or incident power, will be absorbed by the load. Similarly, when the source impedance is equal to the transmission-line characteristic impedance, the transmission line is perfectly matched to the source, and no reflection occurs at the source location.

As can be seen in Eq. 161, the magnitude of the reflection coefficient reduces exponentially as the wave moves away from the load. In the limit when the distance is large, this magnitude reduces to zero. At that location, VSWR is also equal to 1. This phenomenon is also encountered in a short transmission line when the loss is high. So when a transmission line has a large loss or long length, it has no reflection at locations far away from the load.

Total Reflection

Total reflection in a transmission line occurs when the reflection coefficient or VSWR equals to 1 or infinity, respectively, as can be seen in the following cases:

Open-Circuited Transmission Line

An open-circuited transmission line is formed when it is terminated by an open ($Z_L \simeq \infty$). The reflection coefficient at the load and VSWR can be obtained from Eqs. 160 and 173 as

$$\Gamma_L = 1 \tag{176}$$

$$VSWR \simeq \infty \tag{177}$$

Equation 176 shows that the reflected and incident voltages are equal, signifying that the incident voltage and power are completely reflected back with equal phase. The transmission line has total reflection. It is noted that VSWR is only equal to infinity when the transmission line has no loss ($\alpha = 0$) or at the load ($z = 0$), as can be determined from Eq. 202. The voltage on the line can be expressed using Eqs. 165 and 176 as

$$\begin{aligned} V(z) &= V_o^+ e^{-\gamma z}(1 + \Gamma_L e^{2\gamma z}) \\ &= V_o^+ (e^{-\gamma z} + e^{\gamma z}) \\ &= 2V_o^+ \cos h(\gamma z) \end{aligned} \tag{178}$$

which clearly represents a pure standing wave. Traveling wave no longer exists along an open-circuited transmission line.

Short-Circuited Transmission Line

A short-circuited transmission line is formed when it is short-circuited at the load ($Z_L = 0$). The reflection coefficient at the load and VSWR are obtained from Eqs. 160 and 173 as

$$\Gamma_L = -1 \tag{179}$$

$$VSWR \simeq \infty \tag{180}$$

The reflected and incident voltages have equal magnitude but opposite phase, indicating that all incident voltage and power are reflected back with the phase reversed. There is total reflection along the line. It is noted that VSWR is only equal to infinity when the transmission line has no loss ($\alpha = 0$) or at the load ($z = 0$), as can be determined from Eq. 173. The total voltage along the line is found from Eqs. 165 and 179 to be

$$V(z) = -2V_o^+ \sin h(\gamma z) \tag{181}$$

representing a pure standing wave. Traveling wave no longer exists along a short-circuited transmission line.

Transmission Line Terminated with Reactance

When a transmission line is terminated with a capacitor or inductor, the magnitude of the reflection coefficient at the load is

$$|\Gamma| = 1 \tag{182}$$

The reflected and incident voltages have equal magnitude, and all incident voltage and power are reflected back.

Lossless Transmission Lines

All equations derived in sections "Impedance," "Reflection Coefficients," "Standing-Wave Ratio," and "Perfect Match and Total Reflection" are applicable for lossy as well as lossless transmission lines. For lossless transmission lines, the attenuation constant $\alpha = 0$ and the propagation constant $\gamma = j\beta$, where β is the phase constant. Making use of these and recognizing that $tan(j\beta l) = jtan(\beta l)$, the following equations can be derived from Eqs. 153 to 156:

$$Z(z) = Z_o \frac{Z_L - jZ_o \tan \beta z}{Z_o - jZ_L \tan \beta z} \tag{183}$$

$$Y(z) = Y_o \frac{Y_L - jY_o \tan \beta z}{Y_o - jY_L \tan \beta z} \tag{184}$$

$$Z_i = Z_o \frac{Z_L + jZ_o \tan \beta l}{Z_o + jZ_L \tan \beta l} \tag{185}$$

$$Y_i = Y_o \frac{Y_L + jY_o \tan \beta l}{Y_o + jY_L \tan \beta l} \tag{186}$$

$\theta = \beta l$ is called the electrical length of the line and depends on frequency. Electrical length actually represents the phase of the transmission line and is frequently used in high-frequency circuit design. Note that the input impedance Z_i and input admittance Y_i are periodic functions with respect to βl and therefore repeat themselves every half-wavelength along the transmission line. From Eqs. 157, 161, and 163, it can be derived that

DOI 10.1007/978-981-4560-75-7_129-1

$$\Gamma(z) = \frac{Z_L - Z_o}{Z_L + Z_o} e^{2j\beta z} \tag{187}$$

$$\Gamma(z) = \Gamma_L e^{2j\beta z} \tag{188}$$

$$\Gamma_I(z) = -\frac{Z_L - Z_o}{Z_L + Z_o} e^{2j\beta z} \tag{189}$$

from which

$$|\Gamma(z)| = |\Gamma_I(z)| = |\Gamma_L| = \left|\frac{Z_L - Z_o}{Z_L + Z_o}\right| \tag{190}$$

From Eq. 173, it can be written that

$$VSWR = \frac{1 + |\Gamma_L|}{1 - |\Gamma_L|} \tag{191}$$

and hence,

$$|\Gamma(z)| = |\Gamma_L| = \frac{VSWR - 1}{VSWR + 1} \tag{192}$$

These (lossless) equations are commonly used in practice. It is noted that the VSWR and magnitude of the reflection coefficient in a lossless transmission line are always constant. This is expected as the magnitudes of the incident and reflected voltage waves remain constant as they propagate along a lossless transmission line. The voltage standing-wave pattern on a lossless transmission line is similar to that for a low-loss transmission line as shown in Fig. 11b.

The input impedance of an open-circuited lossless transmission line can be derived from Eq. 185, by letting $Z_L = \infty$, as

$$Z_{io} = Z_o \frac{Z_L + jZ_o \tan\beta l}{Z_o + jZ_L \tan\beta l} = -jZ_o \cot an\beta l \tag{193}$$

The input admittance is

$$Y_{io} = jY_o \tan\beta l \tag{194}$$

Similarly, the input impedance and admittance of a short-circuited lossless transmission line can be derived from Eq. 185 by letting $Z_L = 0$ as

$$Z_{is} = jZ_o \tan\beta l \tag{195}$$

$$Y_{is} = -jY_o \cot\beta l \tag{196}$$

Examining Eqs. 193, 194, 195, and 196 reveals that open- and short-circuited lossless lines behave as an inductor or a capacitor depending on the operating frequency and the transmission line's physical length. At a given frequency, any inductor or capacitor can be theoretically realized by choosing a proper length

DOI 10.1007/978-981-4560-75-7_129-1

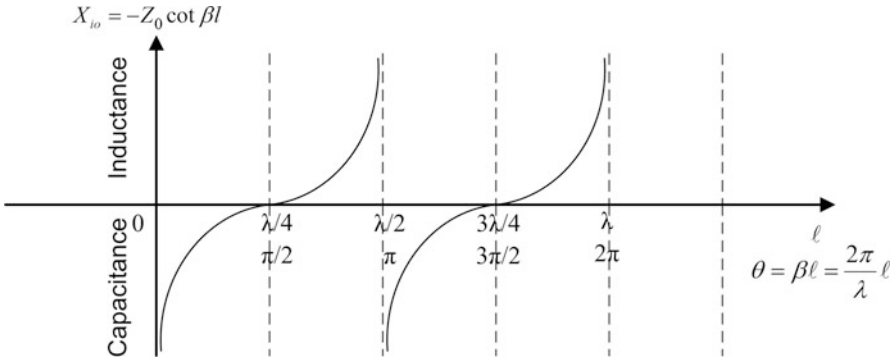


Fig. 12 Impedance behavior of an open-circuited lossless transmission lines versus lengths

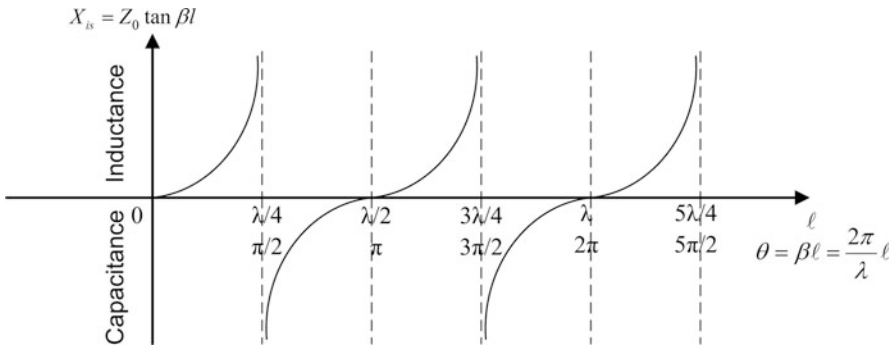


Fig. 13 Impedance behavior of a short-circuited lossless transmission lines versus lengths

for these transmission lines. Figures 12 and 13 illustrate the inductive and capacitive behaviors of open- and short-circuited lossless transmission lines, respectively, versus physical and electrical lengths.

When the transmission line is very short compared to wavelength, $\beta l = 2\pi l/\lambda << 1$. $\tan \beta l$ can then be approximated as βl and it can be obtained from Eq. 193:

$$Z_{io} \cong -j\frac{Z_o}{\beta l} \tag{197}$$

Replacing $Z_o = \sqrt{L/C}$ from Eq. 39 and $\beta = \omega\sqrt{LC}$ from Eq. 40 leads to

$$Z_{io} \cong \frac{1}{j\omega(Cl)} \tag{198}$$

which shows that a very short open-circuited lossless transmission line behaves as a capacitor having a capacitance value of Cl. Note that C is the capacitance per unit length of the line and a transmission line (whether open-circuited or not), whose length is not very short compared to wavelength, does not behave as a capacitor of Cl. Similarly, the following equation for a short-circuited lossless transmission line can be derived when its length is very short compared to wavelength:

$$Z_{is} \cong j\omega(Ll) \tag{199}$$

which indicates that a very short short-circuited line behaves as an inductor with an inductance of Ll. Note

DOI 10.1007/978-981-4560-75-7_129-1

that L is the inductance per unit length of the line and a transmission line (whether short-circuited or not), whose length is not very short compared to wavelength, does not behave as an inductor of *Ll*.

Two special cases worth mentioned are the open- and short-circuited transmission lines whose lengths are multiples of quarter-wavelength. Substituting $l = n\lambda/4$, where n is a positive integer, into Eqs. 222 and 224, we obtain

$$Z_{io} = 0 \qquad \text{for open-circuited lossless line} \tag{200}$$

and

$$Z_{is} = \infty \qquad \text{for short-circuited lossless line} \tag{201}$$

These results indicate that open- and short-circuited transmission lines of multiple quarter wavelengths behave as short and open circuits at their inputs, respectively. Quarter-wavelength open- and short-circuited transmission lines ($l = \lambda/4$) are often used in bias networks for high-frequency narrow-band circuits.

Synthetic Transmission Lines

As discussed in section "Transmission-Line Equations," a transmission line can be approximately represented as an equivalent circuit consisting of many identical sub-circuits, each sub-circuit corresponding to an infinitesimal section of the transmission line, as shown in Fig. 14. For lossless transmission lines, R = G = 0 as seen in Fig. 14b, and a signal would traverse without loss, implying that uniform lossless transmission lines can be operated over extremely wide bandwidths up to the limit of the TEM or quasi-TEM cutoff frequency (theoretically infinite bandwidth.)

Examination of Fig. 14b reveals that synthetic (or artificial) transmission lines may be formed by cascading multiple identical LC sections or cells. It is noted that R and G seen in Fig. 14a represent losses due to the conductors and substrate, respectively, which are undesirable. Therefore, they are not used in creating synthetic transmission lines. Nevertheless, R and G are implicitly included in synthetic transmission lines due to the losses associated with the constituent inductors and capacitors, respectively. These losses essentially limit the operating bandwidth of synthetic transmission lines. It is expected that the larger the number of sections (N) for a synthetic transmission, the closer the resemblance of an actual transmission line. Synthetic transmission lines can exploit the properties of transmission lines to achieve certain functions without using actual transmission lines or eliminate the need of configuring actual

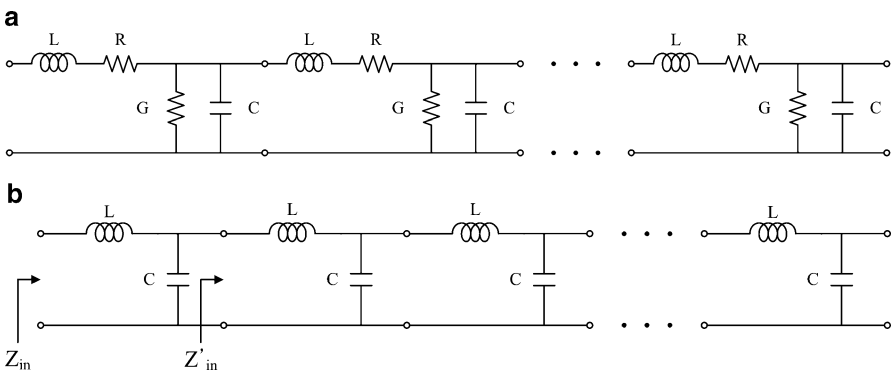


Fig. 14 Approximate equivalent circuit of lossy (**a**) and lossless (**b**) transmission lines

DOI 10.1007/978-981-4560-75-7_129-1

transmission lines in particular fashions (e.g., microstrip line), which might be difficult and/or inconvenient to realize under certain cases. For instance, for silicon-based system-on-chip implementations with embedded antennas, synthetic transmission lines could be used in place of real transmission lines in antenna-feeding circuitry such as Wilkinson power divider to reduce the die size, which is crucial for on-chip applications.

A lossless synthetic transmission line with infinite number of sections ($N \to \infty$) as represented in Fig. 14b is now considered, and the inductors and capacitors are assumed ideal. Its characteristic impedance and phase velocity can be treated the same as those for lossless transmission lines. That is, they can be obtained from Eqs. 39 and 41 as

$$Z_o = \sqrt{\frac{L}{C}} \tag{202}$$

and

$$v_p = \frac{1}{\sqrt{LC}} \tag{203}$$

respectively. Note that herein L and C are the inductance and capacitance of each section making up the synthetic transmission line. The time delay per section of a synthetic transmission line is considered as equivalent to the time delay per unit length of a corresponding actual transmission line and, hence, can be obtained as

$$t_d = \sqrt{LC} \tag{204}$$

It can be seen that a synthetic transmission line provides flexibility in realizing a particular characteristic impedance and phase velocity by properly choosing L and C. It can also achieve certain time delay (small or large) with suitable L, C, and N. Using actual transmission lines may require a long physical length in order to obtain a large time delay.

When the constituent inductors and capacitors have small loss, the corresponding synthetic transmission line can be considered as having low loss, and, similar to an actual low-loss transmission line, its characteristic impedance, attenuation constant, phase constant, and velocity can be obtained from Eqs. 49 to 52 as

$$Z_o \cong \sqrt{\frac{L}{C}}\left[1 + j\left(\frac{G}{2\omega C} - \frac{R}{2\omega L}\right)\right] \tag{205}$$

$$\alpha \cong \frac{1}{2}\left(R\sqrt{\frac{C}{L}} + G\sqrt{\frac{L}{C}}\right) \tag{206}$$

$$\beta \cong \omega\sqrt{LC} \tag{207}$$

$$v \cong \frac{1}{\sqrt{LC}} \tag{208}$$

DOI 10.1007/978-981-4560-75-7_129-1

where R and G are the resistance and conductance representing loss in the inductor and capacitor of each cell, respectively.

The lossless synthetic transmission line as shown in Fig. 14b resembles a low-pass filter, and therefore, it is associated with a cutoff frequency ω_c which defines passband and stopband below and above ω_c, respectively. As N is of infinity, it is justified to assume that (N-1) also approaches infinity. As such, the input impedances Z_{in} and Z'_{in} would be equal. As indicated in Fig. 14b, Z_{in} can be written as a combination of series L and parallel (C, Z'_{in}):

$$Z_{\text{in}} = j\omega L + \frac{1}{\frac{1}{Z_{\text{in}}} + j\omega C} \tag{209}$$

where Z'_{in} in the right-high side has been replaced with Z_{in}. Expanding Eq. 209 gives

$$Z_{\text{in}}^2 - j\omega L Z_{\text{in}} - \frac{L}{C} = 0 \tag{210}$$

which can be solved for Z_{in}:

$$Z_{\text{in}} = \frac{1}{2} j\omega L \pm \frac{1}{2} j \sqrt{\frac{L}{C}} \sqrt{\omega^2 LC - 4} \tag{211}$$

Equation 211 shows that when $\omega^2 LC - 4 < 0$, corresponding to $\omega < 2/\sqrt{LC}$, Z_{in} is complex, and signals with $\omega < 2/\sqrt{LC}$ can propagate along the transmission line, hence signifying a passband. In the limit of ω approaching 0, Z_{in} approaches the characteristic impedance $Z_o = \sqrt{L/C}$. On the other hand, when $\omega^2 LC - 4 > 0$, $\omega > 2/\sqrt{LC}$ and correspondingly Z_{in} becomes imaginary. Signals whose (radian) frequencies are above $2/\sqrt{LC}$ thus cannot propagate on the transmission line, thereby indicating a stopband corresponding to a cutoff frequency of

$$\omega_c = \frac{2}{\sqrt{LC}} \tag{212}$$

For ideal (lossless) synthetic transmission line, the cutoff frequency becomes, making use of Eq. 202,

$$\omega_c = \frac{2}{CZ_o} \tag{213}$$

The cutoff frequency effectively sets the operating bandwidth of synthetic transmission lines from *dc* to ω_c.

A synthetic transmission line should therefore be designed so that its highest operating frequency is below $\omega_c = 2/\sqrt{LC}$. In practice, the number of sections N making up a synthetic transmission line is always finite, so the actual cutoff frequency is smaller (in fact, could be much smaller) than the theoretical value calculated from Eq. 212 which corresponds to infinite N. Therefore, the highest operating frequency may need to be much smaller than ω_c for proper operation. Typically, the cutoff frequency is chosen to be at least three times of the highest operating frequency. Actual simulation of a designed synthetic transmission line needs to be performed to make sure its cutoff frequency is well above the intended operating frequency range. It is particularly noted that a synthetic transmission line is not an exact

DOI 10.1007/978-981-4560-75-7_129-1

transmission line; it only approximates a transmission line over a certain bandwidth, which may be wide or narrow depending on the number of sections used and the values of L and C.

In some applications, such as those requiring signals at multiple frequencies to be transmitted concurrently, such as (time-domain) pulse signals including ultra-wideband (UWB) signals that contain simultaneous components at frequencies within 3.1–10.6 GHz, frequency-independent velocity or linear phase response versus frequency in the passband is needed to avoid signal distortion. The lossless and low-loss synthetic transmission lines have exact and approximate frequency-independent velocities as seen in Eqs. 203 and 208, respectively, and correspondingly producing no and minimum signal distortion theoretically.

A synthetic transmission line can be designed by choosing L and C to achieve desired characteristic impedance per Eq. 202 and desired cutoff frequency (or bandwidth) per Eq. 212. For wide bandwidth, LC is kept sufficiently small to produce a high cutoff frequency. Effectively, there is a trade-off between the operating bandwidth and characteristic impedance (and hence matching for circuits implementing the synthetic transmission line). In order to achieve a certain time delay, the number of required sections can be determined from

$$N = \frac{T_d}{t_d} = \frac{T_d}{\sqrt{LC}} \tag{214}$$

where T_d is the desired time delay.

It is noted that synthetic transmission lines, as suggested in their equivalent circuit, can also be realized using solid-state devices, such as Schottky diodes or MOSFETs, operated in the linear (small-signal) regime to implement both inductors and capacitors, or a combination between lumped elements and solid-state devices. Using solid-state devices would make smaller synthetic transmission lines, which are attractive for low-cost RF applications, yet at possible degradation of circuit performance such as increased noise and reduced linearity, which are not desirable, especially in the receiver front end. Particularly, for large-signal circuits, such as power amplifiers, solid-state devices should not be used due to their nonlinear behavior under large signals, unless they are intended to be used as nonlinear elements in the designed synthetic transmission line.

The main function of a synthetic transmission line is to maintain as close as possible its desired characteristic impedance within a specified tolerance over a bandwidth as wide as possible. Additionally, under some operating conditions, it is also desired to maintain linear phase response as much as possible to avoid signal distortion due to phase nonlinearity. In practice, inductors, capacitors, and semiconductor devices used in synthetic transmission lines have their own parasitics, generally consisting of resistors, inductors, and capacitors, which may severely degrade the transmission-line performance, especially in high RF regions. For instance, the parasitics cause increase in loss and more distortion in the transmission phase, leading to reduced operating bandwidth. Therefore, care must be exercised when designing a synthetic transmission line; all parasitics of its constituent elements must be considered in simulations to make sure that the introduced attenuation and distortion will not ruin the integrity of the signal passing through the transmission line within the operating frequency range. In order to lessen these effects, the number of sections needs to be limited considering trade-off with the operating bandwidth.

TEM and Quasi-TEM Transmission-Line Parameters

Transmission lines with perfect conductors embedded in a homogeneous medium, such as strip line, support pure transverse electromagnetic (TEM) mode or wave. However, transmission lines with more

DOI 10.1007/978-981-4560-75-7_129-1

than two dielectrics (or embedded in an inhomogeneous medium), such as microstrip line, can only support quasi-TEM mode which resembles TEM mode. The TEM or quasi-TEM mode is the dominant mode in transmission lines and has no cutoff frequency, implying that transmission lines can support signals from DC up to very high frequencies at which higher-order modes begin to propagate. The most important parameters of a transmission line for circuit design purposes, assuming the line is not very lossy, are perhaps the characteristic impedance and effective dielectric constant for the quasi-TEM mode. For transmission lines with high loss, the attenuation constant is also an important parameter.

As the operating frequency increases, an infinite number of modes including quasi-TEM and higher-order modes may exist in transmission lines, particularly near discontinuities. Higher-order modes have cutoff frequencies which limit the operating frequency ranges and are undesirable. These modes have either longitudinal electric or magnetic field or both. The modes having longitudinal electric field is called transverse magnetic (TM) modes, and those having longitudinal magnetic field is referred to as transverse electric (TE) modes. The modes having both longitudinal electric and magnetic fields are known as hybrid modes. When a higher-order mode coexists with the quasi-TEM mode above a certain frequency, it will disrupt the circuit performance since it will take away some of the energy of the quasi-TEM mode and, particularly, will generate spurious responses in active circuits. The share of energy that sustains the higher-order mode is inevitable due to the conservation of power and the fact that the total energy propagating in a transmission line is finite.

There are two approaches in analyzing a transmission line: static or quasi-static and dynamic or full-wave approaches. The static or quasi-static approach only produces transmission-line parameters for the TEM or quasi-TEM mode that are theoretically valid only at DC. On the other hand, the dynamic approach not only can produce the frequency-dependent transmission-line parameters for the TEM or quasi-TEM mode but also those for the higher-order modes, whose parameters are functions of frequency.

As stated earlier, transmission line parameters obtained by the static or quasi-static approach are strictly valid at DC. In practice, however, these results can also be used at higher frequencies. The question most engineers would ask is how high of frequency they can use static results. There is no clear-cut answer for this. Some tends to believe that as long as the frequency is different from zero, static results cannot be used. On the other hand, most practical engineers perhaps use static results up to high frequencies such as around 18 GHz. In fact, a number of millimeter-wave circuits operating up to *W*-band (75–110 GHz) developed in the 1980s has been successfully designed using only static results. Nevertheless, at high frequencies, especially those in the millimeter-wave region, a dynamic approach should be employed for more accurate determination of the parameters of transmission lines. The main trade-off between using static versus dynamic results is that the former is simpler to calculate but less accurate.

Static or Quasi-Static Analysis

Static or quasi-static analysis produces transmission-line parameters that are frequency independent. Simple equations can now be derived for the static or quasi-static characteristic impedance and effective relative dielectric constant of transmission lines. Without loss of generality, a lossless microstrip line, as shown in Fig. 15a, is considered as a representative transmission line.

The characteristic impedance of this transmission line is given as

$$Z_o = \sqrt{\frac{L}{C}} \tag{215}$$

where L and C are the inductance and capacitance per unit length of the transmission line, respectively. Let L_a and C_a represent the inductance and capacitance per unit length of the transmission line when the

DOI 10.1007/978-981-4560-75-7_129-1

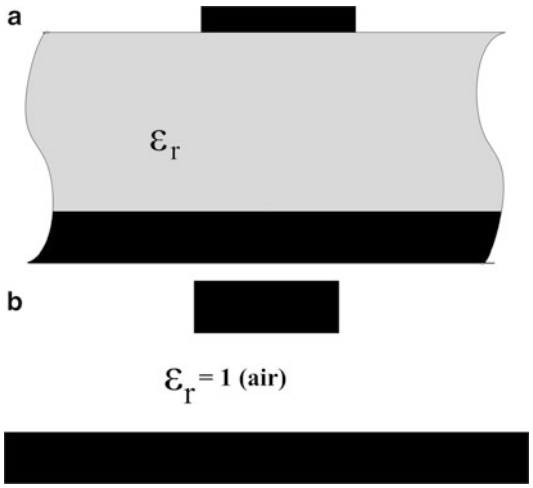


Fig. 15 Microstrip line with (**a**) and without (**b**) substrate

dielectric is replaced by air (i.e., $\varepsilon_r = 1$) as shown in Fig. 15b. Equation 215 can then be rewritten, using the fact that the inductance per unit length does not depend upon the surrounding dielectrics, as

$$Z_o = \sqrt{\frac{L_a}{C_a}\frac{C_a}{C}} = \frac{1}{c\sqrt{CC_a}} \tag{216}$$

where $c = 3 \times 10^8$ m/s is the speed of light in air. The phase velocity of the quasi-TEM wave propagating along the transmission line is given as

$$v_p = \frac{c}{\sqrt{\varepsilon_{\text{reff}}}} \tag{217}$$

where $\varepsilon_{\text{reff}}$ is defined as the effective relative dielectric constant. Squaring both sides of Eq. 217 and making use of $L = L_a$, we obtain

$$\varepsilon_{\text{reff}} = \frac{C}{C_a} \tag{218}$$

The corresponding wavelength is given as

$$\lambda = \frac{\lambda_o}{\sqrt{\varepsilon_{\text{reff}}}} \tag{219}$$

where $\lambda_o = c/f$ is the free-space wavelength. It is now apparent that the static or quasi-static characteristic impedance and effective relative dielectric constant of any transmission line can be determined solely in terms of the transmission line's capacitances per unit length with and without dielectrics.

Dynamic Analysis

Dynamic analysis produces frequency-dependent characteristic impedance and propagation constant and, hence, attenuation and effective dielectric constant of transmission lines for TEM or quasi-TEM as well as hybrid modes. This approach begins by solving the wave equations derived from Maxwell's equations for

DOI 10.1007/978-981-4560-75-7_129-1

the electric and magnetic fields subject to appropriate boundary conditions. The process also yields an eigenvalue equation whose eigenvalue is the propagation constant $\gamma = \alpha + j\beta$.

The phase constant β can be used to calculate the effective dielectric constant as

$$\varepsilon_{\text{eff}} = \left(\frac{\beta}{\text{k}_o}\right)^2 \tag{220}$$

where $k_o = \omega\sqrt{\varepsilon_o \mu_o}$ is the free-space wave number. It should be noted that the phase constant must follow the condition

$$\omega\sqrt{\varepsilon_{\text{o}}\mu_o} \leq \beta \leq \omega\sqrt{\varepsilon_{\text{o}}\mu_o\varepsilon_r} \tag{221}$$

where ε_r has the highest value of relative dielectric constants used for the transmission line.

Using the electric and magnetic fields, the (dynamic) characteristic impedance can be calculated as a function of frequency. It is well known that the characteristic impedance for the TEM or quasi-TEM mode is unique since the mode's voltage and current are uniquely defined. For a hybrid mode, however, the characteristic impedance is not unique. There exist various definitions for the characteristic impedance, and the common ones are based on the voltage and current, power and current, and power and voltage as given respectively below:

$$Z_o^{VI} = \frac{V_o}{I_o} \tag{222}$$

$$Z_o^{PI} = \frac{2P_{\text{avg}}}{|I_o|^2} \tag{223}$$

$$Z_o^{PV} = \frac{|V_o|^2}{2P_{\text{avg}}} \tag{224}$$

where V_o is the voltage across the conductors, I_{o} is the longitudinal current, and P_{avg} represents the time-average power transmitted across the transmission line's cross section. These three characteristic impedances are related by

$$Z_o^{VI} = \sqrt{Z_o^{PI} Z_o^{PV}} \tag{225}$$

It should be noted that different definitions of the characteristic impedance give different numerical results, except at DC. For TEM or quasi-TEM mode, however, all the three definitions give identical results. These definitions are also identical if $P_{\text{avg}} = \frac{1}{2} VI^*$. Generally, this condition is, however, not valid for hybrid modes in a transmission line. The definitions based on the power and current and power and voltage are deduced from a two-conductor transmission line. The choice of a particular definition is not very clear. A possible choice is to use a definition according to a specific use of the transmission line in a circuit. A common definition for transmission lines containing slots, such as slot line and coplanar waveguide (CPW), is based on the power and voltage. For instance, the characteristic impedance of CPW, shown in Fig. 18, can be obtained from Eq. 224 with the voltage and power given as

DOI 10.1007/978-981-4560-75-7_129-1

$$V_o = \int_a^b E_x(x,h)\ dx \tag{226}$$

$$P_{\text{avg}} = \frac{1}{2} R_e \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(E_x H_y^* - E_y H_x^* \right)\ dxdy \tag{227}$$

The characteristic impedance of the slot line, shown in Fig. 22, can also be defined by Eq. 224, where the power is again given by Eq. 227 and the voltage is obtained as

$$V_o = \int_{-W/2}^{W/2} E_x(x,h)\ dx \tag{228}$$

For other transmission lines without slots, the definitions based on the current and power and voltage and current are normally employed. For the case of microstrip line, shown in Fig. 17, the current, voltage, and power are given as

$$I_o = \int_{-W/2}^{W/2} J_z(x,h)\ dx \tag{229}$$

$$V_o = -\int_o^h E_y dy \tag{230}$$

$$P_{\text{avg}} = \frac{1}{2} R_e \int_o^{\infty} \int_{-\infty}^{\infty} \left(E_x H_y^* - E_y H_x^* \right)\ dxdy \tag{231}$$

where J_z is the current density on the strip along the z direction.

Printed-Circuit Transmission Lines

Printed-circuit transmission lines, typically configured as planar and uniplanar structures, are important in RF circuits. Printed-circuit transmission lines have evolved from the advances of electronic integrated circuits (ICs), and in turn, they have helped pushing the progress of the IC technology, such as making the ICs more compact, more versatile, having better interconnections and improved performance, etc. Not only that printed-circuit transmission lines fulfill their most fundamental objective of delivering signals; they can also be exploited to create various RF components, such as wideband hybrid junctions, by appropriately combining them. Various printed-circuit transmission lines have also been developed. The most commonly used structures are microstrip line, coplanar waveguide (CPW), coplanar strip (CPS), strip line, and slot line; among them, the microstrip line and strip line form the first planar version of the well-known coaxial transmission line. Figure 16 describes these printed-circuit transmission lines.

This section presents a brief discussion of these transmission lines along with the closed-form formulas for computing their characteristic impedances, effective dielectric constants, and losses. In all these equations, infinitely wide ground planes are assumed. However, they are also valid for narrow ground planes provided that the field distributions are not significantly altered. Although these parameters can be accurately determined using various dynamic methods, the closed-form expressions allow convenient and fast computer-aided design and analysis of RF circuits. It is especially noted that all the printed-circuit

DOI 10.1007/978-981-4560-75-7_129-1

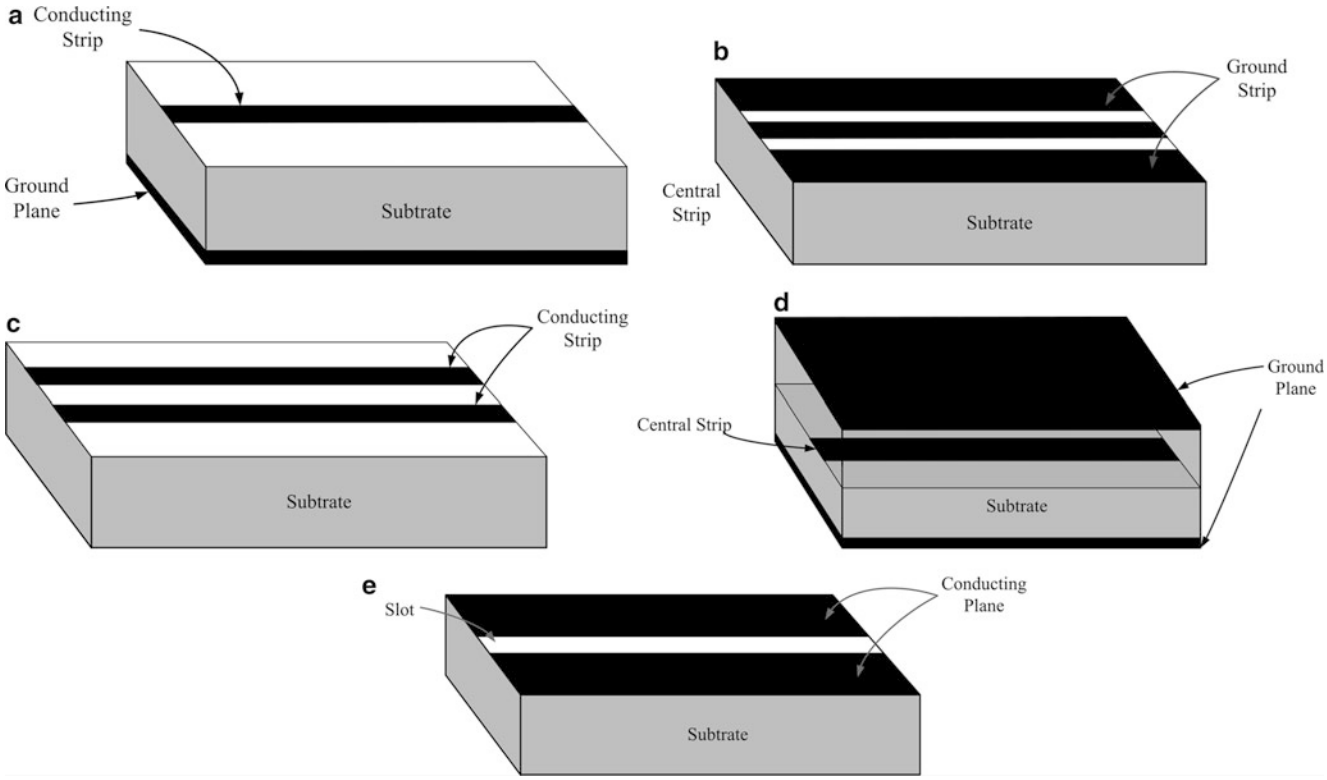


Fig. 16 Commonly used printed-circuit transmission lines: (**a**) microstrip line, (**b**) CPW, (**c**) CPS, (**d**) strip line, and (**e**) slot line

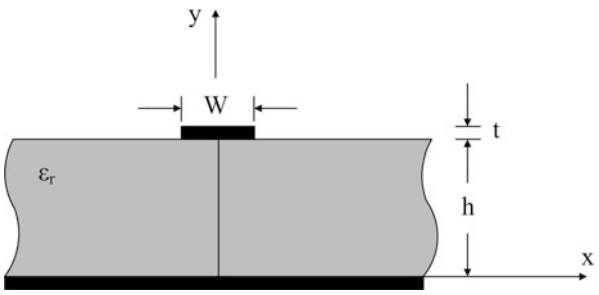


Fig. 17 Cross section of microstrip line

transmission lines are presented with single dielectric, single metal layer, and large ground plane (assumed to be infinitely wide) as conventionally employed. However, multiple dielectrics and metals and/or narrow ground planes can be used to optimize the performance as well as the size of the transmission lines. This kind of implementation is particularly suitable for RF hybrid integrated circuits designed on multilayer substrates such as low-temperature co-fired ceramic (LTCC) and liquid crystal polymer (LCP) materials or RFICs on silicon, which has inherent multiple dielectric and metal layers. Such implementations are, in fact, needed since reducing size is one of the main concerns in RF circuit design.

Microstrip Line

The microstrip line, as shown in Fig. 17, was first proposed in 1952 (Grieg and Englemann 1952), and since then it has become a widely used transmission line due to its planar nature, ease of fabrication using

DOI 10.1007/978-981-4560-75-7_129-1

photolithographic processes, easy integration with solid-state devices, good heat sinking, and good mechanical support.

The closed-form expressions for the characteristic impedance and effective dielectric constant of the microstrip line, assuming zero strip thickness ($t = 0$), are given as (Edwards 1984)

$$Z_o = \begin{cases} \dfrac{119.9}{\sqrt{2(\varepsilon_r+1)}} \left\{ \ln\left[4\dfrac{h}{W} + \sqrt{16\left(\dfrac{h}{W}\right)^2 + 2}\right] - \dfrac{1}{2}\left(\dfrac{\varepsilon_r - 1}{\varepsilon_r - 1}\right)\left(\ln\dfrac{\pi}{2} + \dfrac{1}{\varepsilon_r}\ln\dfrac{4}{\pi}\right)\right\}, & \dfrac{W}{h} < 3.3 \\ \dfrac{119.9\pi}{2\sqrt{\varepsilon_r}} \left\{ \dfrac{W}{2h} + \dfrac{\ln 4}{\pi} + \dfrac{\ln\left(\frac{e\pi^2}{16}\right)}{2\pi}\left(\dfrac{\varepsilon_r - 1}{\varepsilon_r^2}\right) + \dfrac{\varepsilon_r + 1}{2\pi\varepsilon_r}\left[\ln\dfrac{e\pi}{2} + \ln\left(\dfrac{W}{2h} + 0.94\right)\right]\right\}^{-1}, & \dfrac{W}{h} > 3.3 \end{cases} \tag{232}$$

and

$$\varepsilon_{\text{reff}} = \begin{cases} \dfrac{\varepsilon_r + 1}{2} + \dfrac{\varepsilon_r - 1}{2}\left(1 + 10\dfrac{h}{W}\right)^{-0.555}, & Z_o < (63 - 2\varepsilon_r)\Omega \\ \dfrac{\varepsilon_r + 1}{2}\left[1 - \dfrac{1}{2K}\left(\dfrac{\varepsilon_r - 1}{\varepsilon_r + 1}\right)\left(\ln\dfrac{\pi}{2} + \dfrac{1}{\varepsilon_r}\ln\dfrac{4}{\pi}\right)\right]^{-2}, & Z_o > (63 - 2\varepsilon_r)\Omega \end{cases} \tag{233}$$

where $e = 2.7182818$ is the Euler constant and

$$K = \frac{Z_o\sqrt{2(\varepsilon_r + 1)}}{119.9} + \frac{1}{2}\left(\frac{\varepsilon_r - 1}{\varepsilon_r + 1}\right)\left(\ln\frac{\pi}{2} + \frac{1}{\varepsilon_r}\ln\frac{4}{\pi}\right) \tag{234}$$

The accuracy of the characteristic impedance and effective dielectric constant calculated from Eqs. 232 to 233 is within ±1 and ±0.25 %, respectively.

The normalized strip width, W/h, can also be determined from the characteristic impedance and relative dielectric constant as

$$\frac{W}{h} = \begin{cases} \dfrac{2}{\pi}[D - 1 - \ln(2D - 1)] + \dfrac{\varepsilon_r - 1}{\pi\varepsilon_r}\left[\ln(D - 1) + 0.293 - \dfrac{0.517}{\varepsilon_r}\right], & Z_o < (44 - 2\varepsilon_r)\Omega \\ \left(\dfrac{e^K}{8} - \dfrac{1}{4e^K}\right)^{-1}, & Z_o > (44 - 2\varepsilon_r)\Omega \end{cases} \tag{235}$$

where

$$D = \frac{59.95\pi^2}{Z_o\sqrt{\varepsilon_r}} \tag{236}$$

Using Eq. 235, W/h can be determined within ±1 % of accuracy.

Practical microstrip line has finite strip thickness t, which effectively increases the strip width. Equations 232 and 233 can be modified to take into account t for more accurate results as (Gupta et al. 1979a)

DOI 10.1007/978-981-4560-75-7_129-1

$$Z_o(t) = \begin{cases} \dfrac{60}{\sqrt{\varepsilon_{\text{reff}}(t)}} \ln\left[\dfrac{8h}{W_e} + 0.25\dfrac{W_e}{h}\right], & \dfrac{W}{h} \le 1 \\ \dfrac{120\pi}{\sqrt{\varepsilon_{\text{reff}}(t)}}\left[\dfrac{W_e}{h} + 1.393 + 0.667\ln\left(\dfrac{W_e}{h} + 1.444\right)\right]^{-1}, & \dfrac{W}{h} \ge 1 \end{cases} \tag{237}$$

and

$$\varepsilon_{\text{reff}}(t) = \varepsilon_{\text{reff}} - C \tag{238}$$

where

$$\frac{W_e}{h} = \begin{cases} \dfrac{W}{h} + \dfrac{1.25}{\pi}\dfrac{t}{h}\left(1 + \ln\dfrac{4\pi W}{t}\right), & \dfrac{W}{h} \le \dfrac{1}{2\pi} \\ \dfrac{W}{h} + \dfrac{1.25}{\pi}\dfrac{t}{h}\left(1 + \ln\dfrac{2h}{t}\right), & \dfrac{W}{h} \ge \dfrac{1}{2\pi} \end{cases} \tag{239}$$

$$C = \frac{\varepsilon_r - 1}{4.6}\frac{t/h}{\sqrt{W/h}} \tag{240}$$

and $\varepsilon_{\text{reff}}$ is the effective relative dielectric constant with $t = 0$ obtained from Eq. 233.

The frequency-dependent effective dielectric constant and characteristic impedance can be found from Hammerstad and Jensen (1980) as

$$\varepsilon_{\text{reff}}(f) = \varepsilon_r - \frac{\varepsilon_r - \varepsilon_{\text{reff}}(0)}{1 + G\left(\frac{f}{f_p}\right)^2} \tag{241}$$

and

$$Z_o(f) = Z_o(0)\frac{\varepsilon_{\text{reff}}(f) - 1}{\varepsilon_{\text{reff}}(0) - 1}\sqrt{\frac{\varepsilon_{\text{reff}}(0)}{\varepsilon_{\text{reff}}(f)}} \tag{242}$$

where

$$\mathrm{f_p} = \frac{\mathrm{Z_o}(0)}{2\mu_{\mathrm{o}}\mathrm{h}} \tag{243}$$

$$G = \frac{\pi^2}{12}\frac{\varepsilon_r - 1}{\varepsilon_{\text{reff}}(0)}\sqrt{\frac{Z_o(0)}{60}} \tag{244}$$

$\varepsilon_{\text{eff}}(0)$ and $Z_o(0)$ are the quasi-static effective dielectric constant and characteristic impedance, which can be determined using Eqs. 233 or 238 and 232 or 237, respectively, and $\mu_o = 4\pi \times 10^{-7}$ H/m is the permeability of free space.

DOI 10.1007/978-981-4560-75-7_129-1

As for any transmission line, the loss in microstrip line is due to imperfect conductors and dielectrics and is characterized by the attenuation constant $\alpha = \alpha_{c+}\alpha_d$, where α_c and α_d represent the conductor and dielectric attenuation constants, respectively. α_c (in dB/cm) can be determined from Pucel et al. (1968a, b)

$$\alpha_c = \begin{cases} \dfrac{8.68R_s}{2\pi Z_o h}\left[1-\left(\dfrac{W_e}{4h}\right)^2\right]\left\{1+\dfrac{h}{W_e}+\dfrac{h}{\pi W_e}\left[\ln\left(\dfrac{4\pi W}{t}\right)+\dfrac{t}{W}\right]\right\}, & 0<\dfrac{W}{h}\le 1/2\pi \\ \dfrac{8.68R_s}{2\pi Z_o h}\left[1-\left(\dfrac{W_e}{4h}\right)^2\right]\left\{1+\dfrac{h}{W_e}+\dfrac{h}{\pi W_e}\left[\ln\left(\dfrac{2h}{t}\right)-\dfrac{t}{h}\right]\right\}, & 1/2\pi<\dfrac{W}{h}\le 2 \\ \dfrac{8.68R_s}{Z_o h}\left\{1+\dfrac{h}{W_e}+\dfrac{h}{\pi W_e}\left[\ln\left(\dfrac{2h}{t}\right)-\dfrac{t}{h}\right]\right\}\left[\dfrac{W_e}{h}+\dfrac{\dfrac{W_e}{\pi h}}{\dfrac{W_e}{2h}+0.94}\right]\left[\dfrac{W_e}{h}+\dfrac{2}{\pi}\ln\left(\dfrac{W_e}{2h}+0.94\right)\right]^{-2}, & \dfrac{W}{h}\ge 2 \end{cases} \tag{245}$$

where $R_s = \sqrt{\omega\mu_d/2\sigma}$ is the surface resistivity of the conductor, with σ being the conductivity. W_e is the effective strip width, taking into account the strip's finite metallization thickness, given in Eq. 268. α_d can be found by (Welch and Pratt 1966)

$$\alpha_d = \frac{27.3\varepsilon_r(\varepsilon_{reff}-1)\tan\delta}{\sqrt{\varepsilon_{reff}}(\varepsilon_r-1)\lambda_0} \qquad (dB/m) \tag{246}$$

where tanδ is the loss tangent of the dielectric and λ_o is the free-space wavelength. For the microstrip line or any other printed-circuit transmission lines employing low-loss dielectrics, such as SiO_2 used in RFICs, the dielectric loss is normally much lower than the conductor loss. However, when these transmission lines are deposited directly on high-loss substrates such as silicon in RFICs without proper shielding from the high-loss substrates, the dielectric loss becomes dominant.

Coplanar Waveguide

Figure 18 shows cross sections of the conventional and conductor-backed coplanar waveguide (CPW). The (conventional) CPW was first proposed in 1969 (Wen 1969). Since then, it has rapidly gained widespread use due to its many attractive features, such as elimination of via-holes in connecting circuit elements to ground, easy integration with solid-state devices, ease in realizing compact-balanced circuits, and reduction of cross talk between lines. The CPW consists of a central conducting strip (signal line) and two ground conducting strips (ground lines). The two ground lines must be tied together so that they act as a single conductor in order to form a two-conductor transmission line together with the signal line. Otherwise, it can cause propagation problems at high frequencies due to the existence of additional (quasi-TEM) modes of a three-conductor transmission line. In practice, they are connected together via air-bridges. In the simplest equivalence, an air-bridge is modeled as a network consisting of series inductor, resistor, and shunt capacitor, hence resulting in different potentials at the ground lines. For the conductor-backed CPW, via-holes are typically used to connect the back conductor to the ground lines to keep them at equal potential and prevent or minimize the propagation of the other transmission-line modes and parallel-plate waveguide modes within the operating frequency range. However, since a via-hole would electrically consist of at least a resistor in series with an inductor, it inhibits the back conductor from being at the same potential with the ground strips. The unequal potentials are more pronounced at very high frequencies, resulting in potential problems for the CPW to maintain as a two-conductor

DOI 10.1007/978-981-4560-75-7_129-1

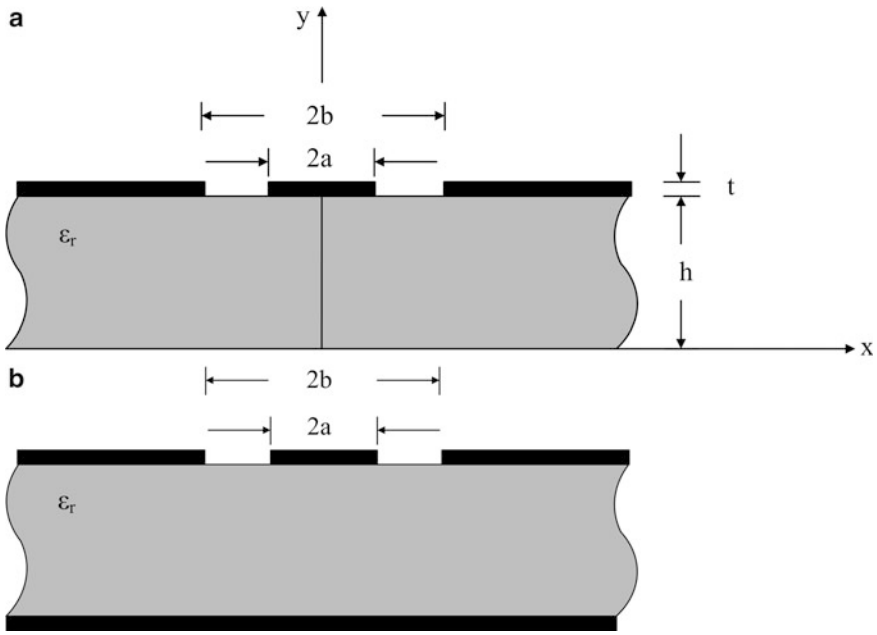


Fig. 18 Cross section of conventional (**a**) and conductor-backed (**b**) CPW

transmission line. The conductor-backed CPW particularly increases the mechanical strength as well as facilitate heat sinking and packaging needed for practical applications. It is operated at narrower bandwidth as compared to the conventional CPW due to the possibility of parallel-plate waveguide mode propagation.

Closed-form expressions for the effective relative dielectric constant and characteristic impedance of CPWs with zero strip thickness and infinitely wide ground strips were derived using conformal-mapping methods (Ghione and Naldi 1983, 1984). For the conventional CPW, they are given according to (Ghione and Naldi 1984)

$$\varepsilon_{\text{eff}} = 1 + \frac{\varepsilon_r - 1}{2} \frac{K(k')}{K(k)} \frac{K(k_1)}{K\left(k_1'\right)} \tag{247}$$

and

$$Z_o = \frac{30\pi}{\sqrt{\varepsilon_{\text{eff}}}} \frac{K(k')}{K(k)} \tag{248}$$

where

$$k = \frac{a}{b} \tag{249}$$

$$k' = \sqrt{1 - k^2} \tag{250}$$

DOI 10.1007/978-981-4560-75-7_129-1

$$k_1 = \frac{\sin h\left(\frac{\pi a}{2h}\right)}{\sin h\left(\frac{\pi b}{2h}\right)} \tag{251}$$

$$k_1' = \sqrt{1\text{-}k_1^2} \tag{252}$$

K represents the complete integral of the first kind, whose values can be determined from an integral or tabulated tables. The $K(k)/K(k')$ ratio can also be approximately obtained by (Hilberg 1965)

$$\frac{K(k)}{K(k')} = \begin{cases} \dfrac{\pi}{\ln\left(2\dfrac{1+\sqrt{k'}}{1-\sqrt{k'}}\right)}, & 0 \le k \le 0.707 \\ \frac{1}{\pi}\ln\left(2\frac{1+\sqrt{k}}{1-\sqrt{k}}\right), & 0.707 \le k \le 1 \end{cases} \tag{253}$$

For the conductor-backed CPW, the effective dielectric constant and characteristic impedance can be evaluated from (Ghione and Naldi 1983)

$$\varepsilon_{\text{eff}} = \frac{1 + \varepsilon_r \dfrac{K(k')}{K(k)} \dfrac{K(k_1)}{K\left(k_1'\right)}}{1 + \dfrac{K(k')}{K(k)} \dfrac{K(k_1)}{K\left(k_1'\right)}} \tag{254}$$

and

$$Z_o = \frac{60\pi}{\sqrt{\varepsilon_{\text{eff}}}} \frac{1}{\dfrac{K(k)}{K(k')} + \dfrac{K(k_1)}{K\left(k_1'\right)}} \tag{255}$$

where

$$k = \frac{a}{b} \tag{256}$$

$$k' = \sqrt{1 - k^2} \tag{257}$$

$$k_1 = \frac{\tan h\left(\frac{\pi a}{2h}\right)}{\tan h\left(\frac{\pi b}{2h}\right)} \tag{258}$$

$$k_1' = \sqrt{1 - k_1^2} \tag{259}$$

DOI 10.1007/978-981-4560-75-7_129-1

When the thickness t of the central strip and ground strips is considered, the widths of the strip and gaps effectively increase and decrease, respectively. The effective dielectric constant and characteristic impedance of the (conventional) CPW can be found, taking into account this effect, using the following formulas (Gupta et al. 1979b):

$$\varepsilon_{\text{reff}}(t) = \varepsilon_{\text{reff}} - \frac{0.7(\varepsilon_{\text{reff}} - 1)\dfrac{t}{b-a}}{\dfrac{K(k)}{K(k')} + 0.7\dfrac{t}{b-a}} \tag{260}$$

where $\varepsilon_{\text{reff}}$ is the effective relative dielectric constant when $t = 0$ given in Eq. 247 and

$$Z_o = \frac{30\pi}{\sqrt{\varepsilon_{\text{reff}}(t)}} \frac{K(k_e')}{K(k_e)} \tag{261}$$

where

$$k_e = \frac{S_e}{S_e + 2W_e} \tag{262}$$

$$k_{\text{e}}' = \sqrt{1 - k_e^2} \tag{263}$$

$$S_e = 2a + \Delta \tag{264}$$

$$W_e = b - a - \Delta \tag{265}$$

$$\Delta = \frac{1.25t}{\pi}\left[1 + \ln\left(\frac{8\pi a}{t}\right)\right] \tag{266}$$

The conductor attenuation constant for the conventional CPW is given as (Gupta et al. 1981)

$$\alpha_c = \frac{4.88 \times 10^{-4}}{\pi} R_s \varepsilon_{\text{reff}} Z_0 P \frac{b+a}{(b-a)^2} \left\{ \frac{\dfrac{1.25t}{\pi}\ln\left(\dfrac{8\pi a}{t}\right) + 1 + \dfrac{1.25t}{2\pi a}}{\left[2 + \dfrac{2a}{b-a} - \dfrac{1.25t}{\pi(b-a)}\left(1 + \ln\dfrac{8\pi a}{t}\right)\right]^2} \right\} (dB/m) \tag{267}$$

where

$$P = \begin{cases} \dfrac{k}{(1-k')(k')^{3/2}}\left[\dfrac{K(k)}{K(k')}\right]^2, & 0 \le k \le 0.707 \\ \dfrac{1}{(1-k)\sqrt{k}}, & 0.707 \le k \le 1 \end{cases} \tag{268}$$

The expression for the dielectric attenuation constant is the same as that for the microstrip line, which is given in Eq. 246.

DOI 10.1007/978-981-4560-75-7_129-1

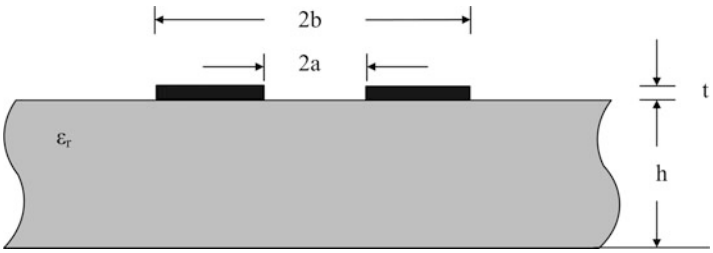


Fig. 19 Cross section of coplanar strips

Coplanar Strips

Coplanar strips (CPS), as shown in Fig. 19, employ two parallel strips on the same side of a dielectric substrate. The CPS structure is also useful for RF circuits, especially balanced circuits due to its inherent balance nature. The CPS structure allows easy connections for series and shunt solid-state devices. The effective dielectric constant and characteristic impedance of the CPS for zero strip thickness can be evaluated using the following closed-form equations derived from a conformal-mapping method (Ghione and Naldi 1984):

$$\varepsilon_{\text{eff}} = 1 + \frac{\varepsilon_r - 1}{2} \frac{K(k')}{K(k)} \frac{K(k_1)}{K(k_1')} \tag{269}$$

and

$$Z_o = \frac{120\pi}{\sqrt{\varepsilon_{\text{eff}}}} \frac{K(k)}{K(k')} \tag{270}$$

where

$$k = \frac{a}{b} \tag{271}$$

$$k' = \sqrt{1 - k^2} \tag{272}$$

$$k_1 = \frac{\sin h\left(\frac{\pi a}{2h}\right)}{\sin h\left(\frac{\pi b}{2h}\right)} \tag{273}$$

$$k_1' = \sqrt{1 - k_1^2} \tag{274}$$

When the strip thickness is considered, the effective dielectric constant and characteristic impedance can be determined using (Gupta et al. 1979b)

$$\varepsilon_{\text{reff}}(t) = \varepsilon_{\text{eff}} - \frac{0.7(\varepsilon_{\text{reff}} - 1)\frac{t}{a}}{\frac{K(k')}{K(k)} + 0.7\frac{t}{a}} \tag{275}$$

DOI 10.1007/978-981-4560-75-7_129-1

and

$$Z_o(t) = \frac{120\pi}{\sqrt{\varepsilon_{\text{reff}}(t)}} \frac{K(k_e)}{K(k'_e)} \tag{276}$$

where

$$k_e = \frac{S_e}{S_e + 2W_e} \tag{277}$$

$$k'_e = \sqrt{1 - k_e^2} \tag{278}$$

$$S_e = S - \Delta \tag{279}$$

$$W_e = W + \Delta \tag{280}$$

$$\Delta = \frac{1.25\text{t}}{\pi}\left[1 + \ln\left(4\pi\frac{b-a}{t}\right)\right] \tag{281}$$

The attenuation due to the conductor loss can be found from Gupta et al. (1981)

$$\alpha_c = \frac{4.34}{\pi}\frac{R_s}{Z_o}P\frac{a+b}{a^2}\left\{\frac{\frac{1.25}{\pi}\ln\left(4\pi\frac{b-a}{t}\right) + 1 + \frac{1.25}{\pi}\frac{t}{b-a}}{\left\{1 + \frac{b-a}{a} + \frac{1.25}{2\pi}\frac{t}{a}\left[1 + \ln\left(4\pi\frac{b-a}{t}\right)\right]\right\}^2}\right\} \tag{282}$$

where P is given in (4.297). The dielectric attenuation constant can be determined from Eq. 268.

Strip Line

Strip line, as shown in Fig. 20, is essentially a printed-circuit version of the coaxial transmission line. Its dominant mode of propagation is pure TEM, assuming perfect conductors. Because of possible excitement of unwanted parallel-plate modes at high frequencies due to the two parallel ground planes, the strip line is in general more suitable for use in low RF regimes. Its characteristic impedance for zero-thickness strip can be determined using the following conformal-mapping formulas (Cohn 1954):

$$Z_o = \frac{30\pi}{\sqrt{\varepsilon_r}} \frac{K(k')}{K(k)} \tag{283}$$

where

$$k = \tan h\left(\frac{\pi W}{4a}\right) \tag{284}$$

$$k' = \sqrt{1 - k^2} \tag{285}$$

DOI 10.1007/978-981-4560-75-7_129-1

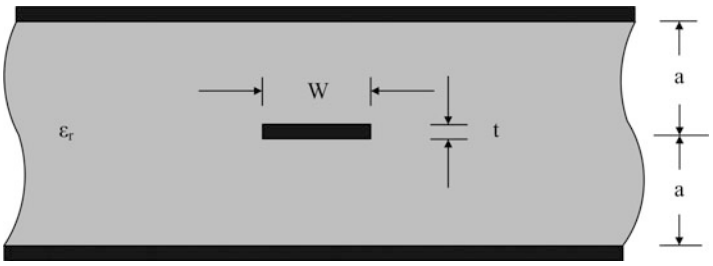


Fig. 20 Cross section of strip line

For finite strip thickness, the characteristic impedance can be found from Wheeler (1978)

$$Z_o = \frac{30}{\sqrt{\varepsilon_r}} \ln\left\{1 + \frac{4}{\pi}\frac{2a-t}{W_e}\left[\frac{8}{\pi}\frac{2a-t}{W_e} + \sqrt{\left(\frac{8}{\pi}\frac{2a-t}{W_e}\right)^2 + 6.27}\right]\right\} \tag{286}$$

where

$$\frac{W_e}{2a-t} = \frac{W}{2a-t} + \frac{\Delta W}{2a-t} \tag{287}$$

$$\frac{\Delta W}{2a-t} = \frac{x}{\pi(1-x)}\left\{1 - \frac{1}{2}\ln\left[\left(\frac{x}{2-x}\right)^2 + \left(\frac{0.0796x}{\frac{W}{2a}+1.1x}\right)^m\right]\right\} \tag{288}$$

$$m = \frac{2}{1 + \frac{2x}{3(1-x)}} \tag{289}$$

$$x = \frac{t}{2a} \tag{290}$$

The strip width can also be determined from the characteristic impedance and relative dielectric constant. For a zero-thickness strip, the expression for the strip width can be derived from Eqs. 283 to 285 as

$$\frac{W}{a} = \frac{4}{\pi} \tan h^{-1}\sqrt{p} \tag{291}$$

where

$$p = \begin{cases} \sqrt{1 - \left[\frac{e^{\pi q}-2}{e^{\pi q}+2}\right]^4}, & q \geq 1 \\ \left[\frac{\frac{\pi}{e^q}-2}{\frac{\pi}{e^q}+2}\right]^2, & 0 \leq q \leq 1 \end{cases} \tag{292}$$

$$q = \frac{Z_o\sqrt{\varepsilon_r}}{30\pi} \tag{293}$$

DOI 10.1007/978-981-4560-75-7_129-1

Using Eqs. 286, 287, 288, 289, and 290, we can drive the following expression for the strip width, taking into account the strip thickness

$$\frac{W}{2a-t} = \frac{W_e}{2a-t} - \frac{\Delta W}{2a-t} \tag{294}$$

where

$$\frac{W_e}{2a-t} = \frac{8}{\pi A} \tag{295}$$

$$\frac{\Delta W}{2a-t} = \frac{x}{\pi(1-x)}\left\{1 - \frac{1}{2}\ln\left[\left(\frac{x}{2-x}\right)^2 + \left(\frac{0.0796x}{\frac{W_e}{2a}+1.1x}\right)^m\right]\right\} \tag{296}$$

$$A = \frac{2B}{C} \tag{297}$$

$$B = \exp\left(\frac{Z_o\sqrt{\varepsilon_r}}{30}\right) - 1 \tag{298}$$

$$C = \sqrt{4B + 6.27} \tag{299}$$

$$m = \frac{2}{1 + \frac{2x}{3(1-x)}} \tag{300}$$

$$x = \frac{t}{2a} \tag{301}$$

The conductor attenuation constant (in dB/m) can be determined from

$$\alpha_c = \begin{cases} \dfrac{23.4x10^{-3}R_s\varepsilon_r Z_o A}{30\pi(2a-t)}, & Z_o < \dfrac{120}{\sqrt{\varepsilon_r}} \\ \dfrac{1.4R_s B}{2Z_o a}, & Z_o > \dfrac{120}{\sqrt{\varepsilon_r}} \end{cases} \tag{302}$$

where

$$A = 1 + \frac{2W}{2a-t} + \frac{1}{\pi}\frac{2a+t}{2a-t}\ln\left(\frac{4a-t}{t}\right) \tag{303}$$

$$B = 1 + \frac{2a}{0.5W + 0.7t}\left(0.5 + \frac{0.414t}{W} + \frac{1}{2\pi}\ln\frac{4\pi W}{t}\right) \tag{304}$$

The dielectric attenuation constant is given by

DOI 10.1007/978-981-4560-75-7_129-1

$$\alpha_d = \frac{27.3\sqrt{\varepsilon_r}\tan\delta}{\lambda_o} \qquad (dB/m) \tag{305}$$

Slot Line

The slot line, whose cross section is shown in Fig. 21, is also useful for RF circuits. Its balance nature is especially attractive for circuits requiring balance topology. It is noted that, in contrast with conventional transmission lines, the two conducting strips forming the slot are kept at equal potential (typically grounded) and the slot line does not support a TEM or quasi-TEM propagation mode. Modes propagating on the slot line are quasi-TE modes, which resemble the TE type. The dominant mode is quasi-TE_{10}, similar to the TE_{10} of the rectangular waveguides. The quasi-TE_{10} mode of the slot line, however, has no cutoff frequency. Due to the propagating mode being quasi-TE_{10}, when the slot line is used together with other transmission lines supporting TEM or quasi-TEM mode in circuits, suitable transitions between them need to be used to allow proper transformation of the fields of the slot line into those of the other transmission lines. In order to be used as a good transmission line, the slot line should be fabricated using a substrate with a high dielectric constant to minimize radiation.

Closed-form expressions for the characteristic impedance, Z_o, based on voltage and power and the wavelength, λ_g, for the slot line on high dielectric constant substrates ($9.7 \le \varepsilon_r \le 20$) were obtained (Garg and Gupta 1976) by curve fitting to the numerical results based on Cohn (1969). They are given as

$$\lambda_g/\lambda_o = \begin{cases} 0.923 - 0.195\ln\varepsilon_r + 0.2\dfrac{W}{h} - \left(0.126\dfrac{W}{h} + 0.02\right)\ln\left(\dfrac{h}{\lambda_o}\times 10^2\right), & 0.02 \le \dfrac{W}{h} \le 0.2 \\ 0.987 - 0.21\ln\varepsilon_r + \dfrac{W}{h}(0.111 - 0.0022\varepsilon_r) - \left(0.053 + 0.041\dfrac{W}{h} - 0.0014\varepsilon_r\right)\ln\left(\dfrac{h}{\lambda_o}\times 10^2\right) & \\ \qquad , \ 0.2 \le \dfrac{W}{h} \le 1 \end{cases} \tag{306}$$

$$Z_o = \begin{cases} 72.62 - 15.283\ln\varepsilon_r + 50\dfrac{\left(\dfrac{W}{h} - 0.02\right)\left(\dfrac{W}{h} - 0.1\right)}{W/h} + \ln\left(\dfrac{W}{h}\times 10^2\right)[19.23 - 3.693\ln\varepsilon_r] \\ -\left[0.139\ln\varepsilon_r - 0.11 + \dfrac{W}{h}(0.465\ln\varepsilon_r + 1.44)\right]\left(11.4 - 2.636\ln\varepsilon_r - \dfrac{h}{\lambda_o}\times 10^2\right)^2, \ 0.02 \le \dfrac{W}{h} \le 0.2 \\ 113.19 - 23.257\ln\varepsilon_r + 1.25\dfrac{W}{h}(114.59 - 22.531\ln\varepsilon_r) + 20\left(\dfrac{W}{h} - 0.2\right)\left(1 - \dfrac{W}{h}\right) \\ -\left[0.15 + 0.1\ln\varepsilon_r + \dfrac{W}{h}(-0.79 + 0.899\ln\varepsilon_r)\right] \\ \times\left[10.25 - 2.171\ln\in_r + \dfrac{W}{h}(2.1 - 0.617\ln\in_r) - h/\lambda_o x 10^2\right]^2, \ 0.2 \le \dfrac{W}{h} \le 1 \end{cases} \tag{307}$$

These equations were obtained assuming infinitesimally thin conductors and are accurate to within 2 % for the following ranges:

DOI 10.1007/978-981-4560-75-7_129-1

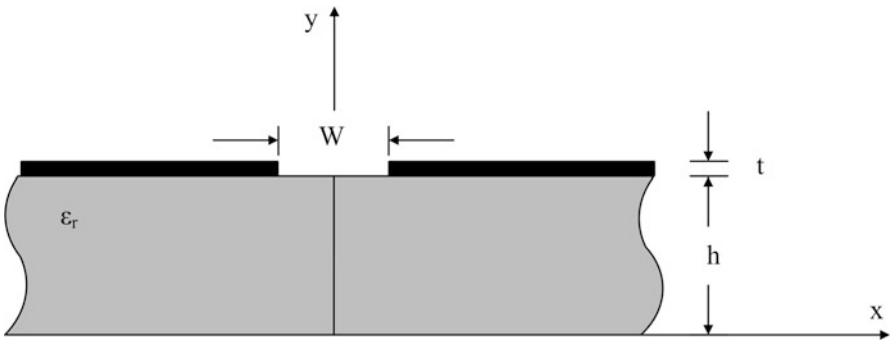


Fig. 21 Cross section of slot line

$$9.7 \leq \varepsilon_r \leq 20 \tag{308}$$

$$0.01 \leq \frac{h}{\lambda_o} \leq \left(\frac{h}{\lambda_o}\right)_c \tag{309}$$

where $(h/\lambda_o)_c$ is the cutoff value for the TE_{10} surface-wave mode on the slot line and is given as

$$\left(\frac{h}{\lambda_o}\right)_c = 0.25\sqrt{\varepsilon_r - 1} \tag{310}$$

Closed-form expressions for the characteristic impedance and wavelength for low dielectric constant substrates were also derived by curve fitting to results of the spectral-domain method (Janaswamy and Schaubert 1986).

Field Distributions

The behavior of a transmission line and the signal propagating on it, governed by the power $\vec{P} = 1/2\mathrm{Re}\left(\vec{E} \times \vec{H}^*\right)$, are dictated by the transmission line's electric and magnetic fields, which affect circuit performance at all frequencies, particularly more pronounced at high frequencies. Therefore, in RF circuit design, it is important and useful to understand the field distributions of printed-circuit transmission lines used in RF circuits. As signals propagate along a transmission line, they carry with them the electric and magnetic fields, and the signals may couple from the transmission line to other circuit elements including transmission lines. The coupling between a transmission line and other transmission lines or circuit elements is caused by the springing of the transmission line's electric and/or magnetic fields onto other lines or circuit elements. There are two kinds of coupling between transmission lines (or between a transmission line and other circuit elements): one is the electric coupling caused by the electric fields between the transmission lines and another is the magnetic coupling caused by the magnetic fields. The field distribution of a transmission line thus helps RF designers to visualize and comprehend the effects of coupling from one transmission line to another, effectively enabling them to optimize circuit layout while considering possible effects on circuit performance. This is especially important in RFICs where circuit elements are placed so close to each other to minimize circuit size. Another important utilization of the field distribution is for the matching between the fields of different kinds of transmission lines – for example, between microstrip line and CPW – when they are connected together. From the circuit theory point of view, two interconnected transmission lines only need to have equal characteristic impedances. This, however, is not sufficient considering EM effects; these equal-characteristic-

DOI 10.1007/978-981-4560-75-7_129-1

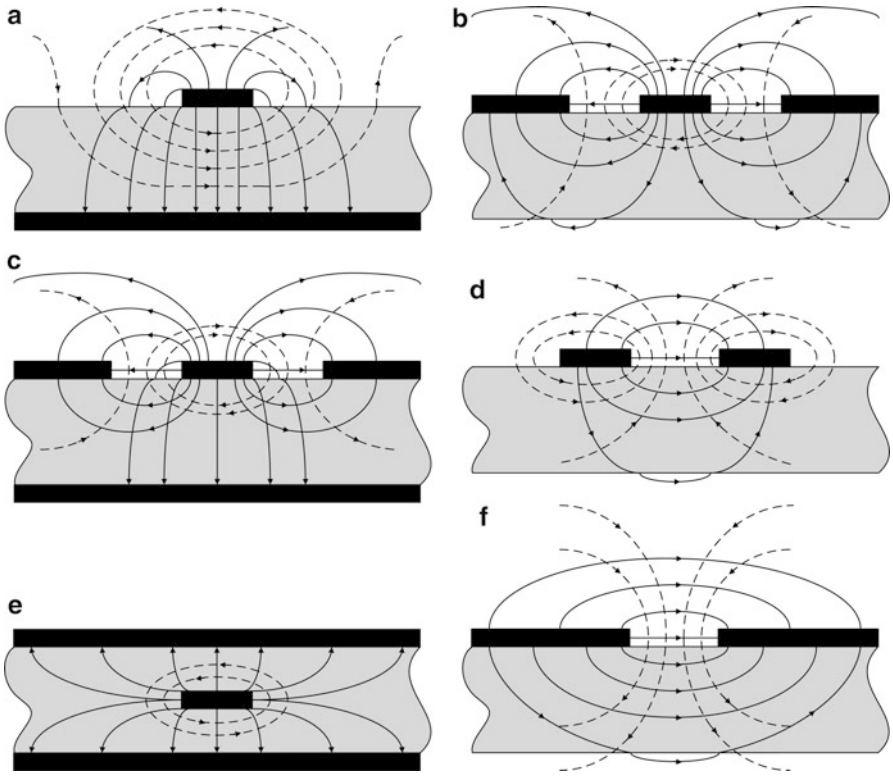


Fig. 22 Field distributions of microstrip line (**a**), CPW (**b**), conductor-backed CPW (**c**), CPS (**d**), strip line (**e**), and slot line (**f**). _____ Electric field - - - - - - - - Magnetic field

impedance transmission lines should also have their electric and magnetic fields matched to each other to minimize possible reflections due to unmatched fields at the interconnecting junction. In general, a transition between these lines should be used to allow the fields of one line to be transformed into those of the other. Figure 22 illustrates the field distributions of the printed-circuit transmission lines shown in Fig. 16. From these field distributions, we can see that the width of the ground strip, whether or not in the same plane of the central strip, does not have to be large; only sufficient width is needed to contain all or majority of the electric fields emanating from the central strip. This field confinement is particularly critical in RF circuits fabricated on high-loss substrates such as silicon-based RFICs, as the electric fields not terminated by the ground plane may enter the (high-loss) substrate, causing significant dielectric loss.

Transmission Lines in Multilayer RF Circuits

Multilayer substrates are essential for RF circuits. Figure 23 shows the cross section of a general multilayer structure that RF circuits could be realized on, which consists of multiple metal layers embedded within multilayer dielectric. The numbers of metal and dielectric layers and their thicknesses vary depending on manufacturing processes and/or intended applications. The metal layer with the thickest metallization produces the lowest conductor loss among the metal layers and is thus preferred for transmission lines. Due to the availability of multiple conductors and dielectrics as illustrated in Fig. 23, together with the use of air-bridges and via-holes, various two-conductor and multiconductor printed-circuit transmission lines can be formed for RF circuits. Such transmission lines facilitate the realization of complicated circuits and, along with the ability of achieving both horizontal and vertical integration through multilayer, ultimately allow very compact, high-density circuit integration. Moreover,

DOI 10.1007/978-981-4560-75-7_129-1

thin dielectric layers can be employed to further reduce circuit size by achieving narrow line widths. The two fundamentals that should be considered in configuring a transmission line for RF circuits are that the characteristic impedance range and the loss should be as wide and low as possible, respectively. Figures 24, 25, 26, 27, and 28 show some possible printed-circuit transmission lines with visual of the electric field lines that can be implemented for multilayer RF circuits. These transmission lines exploit the advantages of multilayer structures as well as the couplings between the constituent strips and slots to improve performance. Using a back conductor with and without slot, they can have wider characteristic-impedance ranges than those of the conventional microstrip line and CPW. In addition, they have several other desirable features such as flexibility and ability to realize complicated, highly dense circuits through appropriate arrangements of the structural elements. The analysis of these transmission lines can be accurately analyzed using full-wave dynamic approaches. Various commercially available EM simulation programs can also be used for accurate calculations of the transmission lines' parameters.

Microstrip Line

Figure 24 shows two examples of microstrip lines in multilayer RF circuits. It is noted that in Figs. 24, 25, 26, 27, and 28, the upper dielectric could consist of multiple dielectric layers having the same relative dielectric constant with different thicknesses. This is especially encountered in silicon-based CMOS/BICMOS processes used for RFICs, which employ multilayer SiO_2 dielectrics of different thicknesses (top layers) above a silicon substrate (bottom layer). The slot in the bottom conductors in Fig. 24b facilitates adjustment of the characteristic impedance by varying its location and/or size. The bottom conductors are kept at approximately equal potential using air-bridges. The top conductor can be located in any metal layer, typically the thickest metal layer for the lowest conductor loss, and the bottom conductor in another metal layer depending on desired characteristics such as characteristic impedance, loss, effective relative dielectric constant, dispersion, etc. These microstrip lines allow easy interface with other components and/or instruments for measurement purposes. The bottom conductor in Fig. 24a and the slot in Fig. 24b may be symmetrical or asymmetrical with respect to the top conductor. These bottom conductors should be sufficiently wide, and the slot should be properly located with respect to the top conductor to prevent all or majority of the electric fields coming from the top conductor from penetrating into the beneath dielectrics in order to eliminate or minimize the effects on the transmission line from other circuits and elements including transmission lines realized on the beneath dielectrics. Possible effects include noise from other circuits propagating into the microstrip line and the coupling between the microstrip line and other nearby transmission lines and circuit elements through the beneath dielectrics. The highest characteristic impedance of the microstrip line shown in Fig. 24a is limited due to the large capacitance per unit length resulting from the bottom conductor. Its characteristic impedance, effective

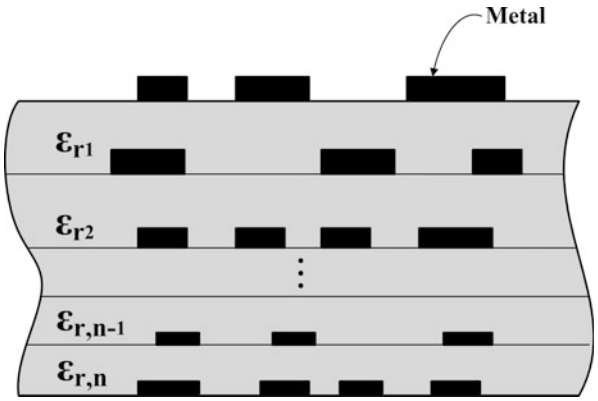


Fig. 23 General multilayer substrate structure for RF circuits

DOI 10.1007/978-981-4560-75-7_129-1

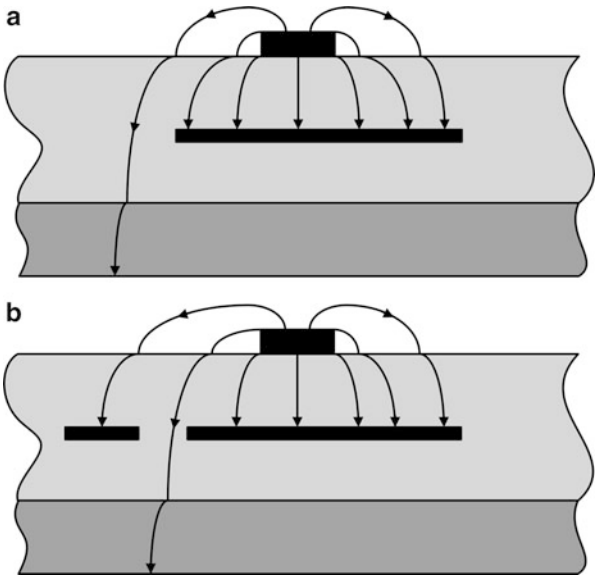


Fig. 24 Representative microstrip lines in multilayer RF circuits

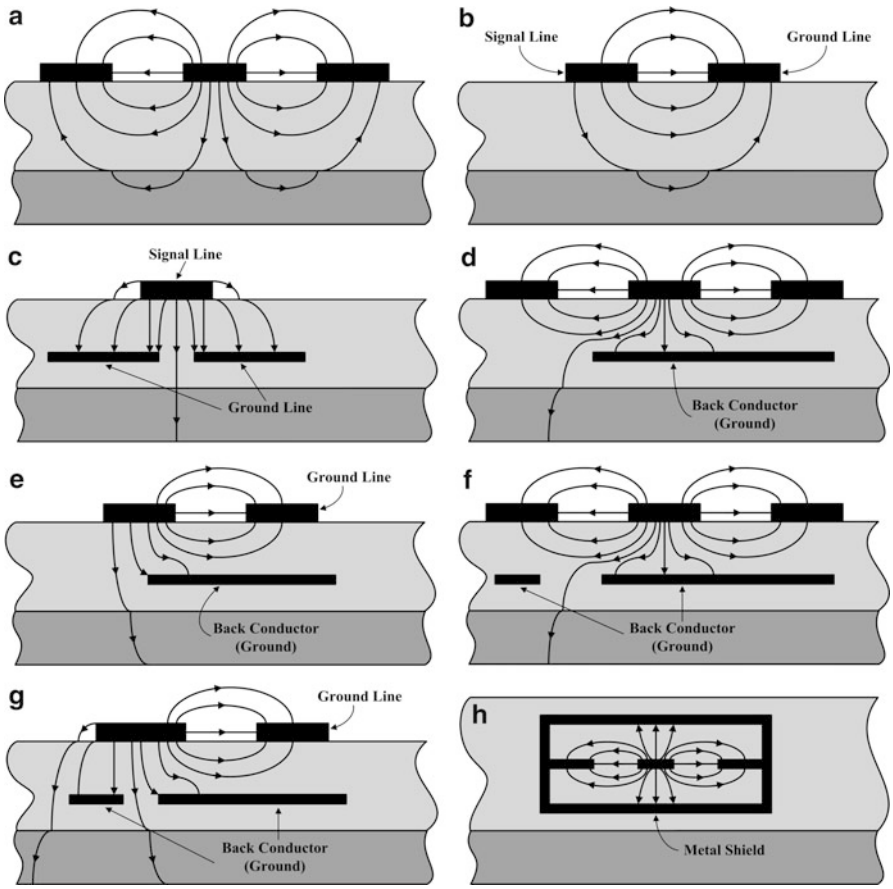


Fig. 25 Possible CPWs in multilayer RF circuits: (**a**) CPW, (**b**) single-ground CPW, (**c**) noncoplanar CPW, (**d**) conductor-backed CPW, (**e**) single-ground conductor-backed CPW, (**f**) slot-conductor-backed CPW, (**g**) single-ground slot-conductor-backed CPW, and (**h**) shielded CPW. Back conductors and ground lines are kept at equal potential through via-hole interconnects. Back conductors separated by slot are interconnected through air-bridges to maintain same potential

DOI 10.1007/978-981-4560-75-7_129-1

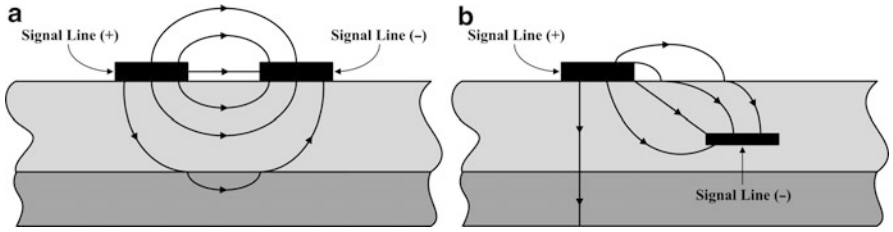


Fig. 26 Possible CPS in multilayer RF circuits: (**a**) CPS and (**b**) noncoplanar CPS

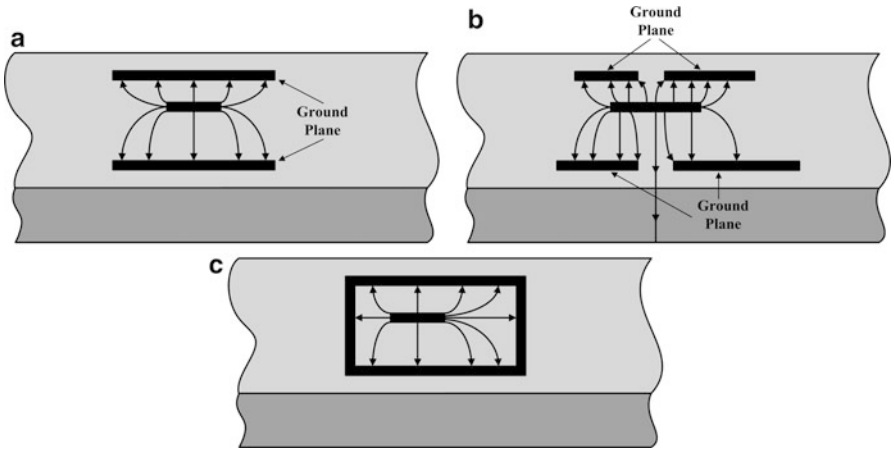


Fig. 27 Possible strip line in multilayer RF circuits: (**a**) strip line and (**b**) slitting-ground strip line

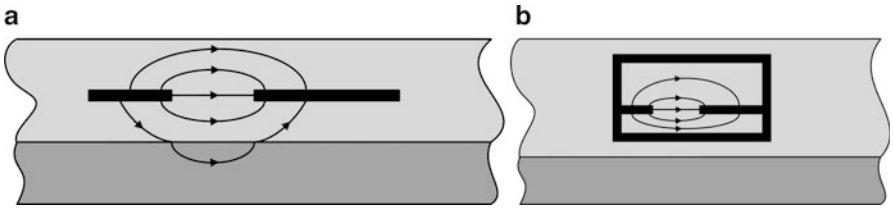


Fig. 28 Slot lines in multilayer RF circuits

relative dielectric constant, and loss can be approximately determined using the closed-form expressions presented in section "Microstrip Line." The microstrip line shown in Fig. 24b is particularly attractive since its characteristic impedance can be varied over a large range by simply changing the location and size of the slot between the bottom conductors. This unique tuning feature can be utilized to realize a larger impedance range than that of the microstrip line shown in Fig. 24a. When the slot is far away from the top conductor, the characteristic impedance and effective relative dielectric constant are constant since the slot has negligible influence on the transmission line's characteristics. However, as the slot approaches the top conductor, both the characteristic impedance and effective dielectric constant change because more of the electric field lines penetrate the dielectrics below. These changes are more drastic for larger slot and also when the slot is close to the top conductor due to increased interaction between the dielectrics below and the top conductor via the slot. The maximum change in the characteristic impedance and effective relative dielectric constant occurs when the slot is centered underneath the top conductor, where maximum field lines enter the underneath dielectric. It is noted that achieving large characteristic impedance, and hence wide impedance range, is desirable. However, if this results in more electric fields penetrating the dielectrics below, then it would cause larger dielectric loss if that dielectric has high loss

DOI 10.1007/978-981-4560-75-7_129-1

and, hence, detrimental circuit effects. Therefore, care needs to be exercised when optimizing the characteristic-impedance range using the slot.

Coplanar Waveguide

Figure 25 shows several CPWs that can be implemented in multilayer RF circuits. Any metal layers can be used; however, like the microstrip line or any other transmission lines, the thickest metal should be employed for the signal line for low conduction loss. In general, the CPW has wider characteristic impedance range than the microstrip line due to the possibility of obtaining higher characteristic impedances through using large gaps between the signal and ground lines. The CPW, in general, has larger size than the microstrip line. However, narrow ground lines can be used to reduce the size while not causing significant effects to the line's characteristic impedance and effective relative dielectric constant since the charge and current are distributed primarily along the edges of the strip and the inner edges of all the ground lines. A single ground line can also be used to reduce the size further. Unequal ground lines can be employed to accommodate circuit layout. As for the microstrip line, lower loss can be achieved by using thicker dielectric corresponding to wider signal line. All the ground lines and back conductors need to be tied together through via-holes and air-bridges to maintain at approximately equal potential. A back conductor, when connected to the ground lines, also help increase the isolation between the signal line and adjacent transmission lines since the electric fields springing away from the signal line tend to terminate in the back conductor via the ground lines. In RF circuits with very small size constraint, which causes significant interactions between elements, CPW with enhanced isolation is indeed desirable. On the other hand, when the back conductor is not interconnected with the ground lines, possibly strong coupling from the signal line to adjacent lines can result as the ground lines may serve as a bridge between them. As will be discussed later, the use of a back conductor with slot allows the characteristic impedance to be varied over a large range by simply changing the location and size of the slot. This unique tuning feature, in conjunction with the signal line and upper slots, can be utilized to realize much larger impedance ranges than those of the conventional microstrip line and CPW.

Figure 25a shows the conventional CPW, Fig. 25b shows a modified CPW with single ground line, and Fig. 25c depicts a CPW with the ground lines located in another layer. As can be seen from their electric field distributions, the CPWs in Fig. 25a, b are partially shielded from the beneath dielectrics, which results in better isolation from these dielectrics and hence other circuits and elements including transmission lines on these dielectrics. The degree of isolation from the dielectrics below depends on how far the signal line of the CPWs is located from those dielectrics. Complete isolation may be possible if the CPWs are located far from it and the gaps between the signal and ground lines are sufficiently small to cause the electric fields to focus more within the gap areas. It is noted that the single ground-line CPW, while physically similar to the coplanar strips (CPS), is electrically different since its ground line is held at zero potential, while in the CPS, this line is kept at an opposite potential of the other strip. The CPW in Fig. 25c is similar to the microstrip line shown in Fig. 24b and, as expected from its electric field lines, can be fully shielded from the beneath dielectrics and corresponding circuits even located near it if the gap between the ground lines is narrow, especially at higher frequencies of the RF range since the gap becomes negligible with respect to the operating wavelengths.

Figure 25d, e show the CPWs with a back conductor. Depending on the width and/or location of the back conductor, complete or partial isolation from the dielectrics below can be obtained. For these CPWs, while the characteristic impedance and effective relative dielectric constant stay constant when the edge of the back conductor closer to the slots is far away from the signal line, they decrease as the edge advances near the signal line.

The CPWs in Fig. 25f,g implement a slitting back conductor whose size and/or location of the slot and conductor control the degree of shielding from the dielectrics below. It is noted that the signals see the gap

DOI 10.1007/978-981-4560-75-7_129-1

in the back conductor in terms of wavelength, and so, at very high frequencies, the gap may appear opaque to the signals and hence the transmission line is completely shielded from the beneath dielectrics. In addition to the upper slot, the lower slot formed by the back conductor can also be used to change the characteristic impedance and effective relative dielectric constant. These parameters increase as the width of the lower slot is increased. The change is significantly pronounced when the width is small while virtually unnoticed for large widths. This phenomenon is expected as the transmission line's capacitance per unit length reduces when the lower slot width is increased due to more electric fields penetrating into the bottom layers. Furthermore, the range of the lower slot width that affects the behavior is smaller for smaller upper-slot width, as expected, due to strong interaction between the signal line and the coplanar ground strips when the upper slots are small. Similar to the microstrip line with slot in the bottom conductor as shown in Fig. 24b, when the slot in the back conductor is far away from the signal line, the characteristic impedance and effective relative dielectric constant are constant. However, as the slot approaches the signal line, both the characteristic impedance and effective dielectric constant increase. These changes are more drastic for larger lower slot and also when the slot is close to the signal line. The maximum characteristic impedance and effective relative dielectric constant occur when the slot is centered underneath the signal line.

Figure 25h shows a CPW enclosed by conducting walls which completely shield the transmission line from the dielectrics below and surrounding elements. The vertical walls are formed by lattices of periodic metallic via-holes placed very close to each other. Narrow metal frames on the metal layers between the top and bottom walls are also used to connect all the via-holes together to further confine the fields within the metal channel and enhance the electrical isolation between the signal line and the exterior of the channel. The metal shield prevents fields from entering and leaving the CPW from its left, right, upper, and lower sides, except through the signal line. If the size of the via-holes and distance between them can be made extremely small with respect to the operating wavelength, the via-hole lattices, together with the metal frames, can create vertical walls resembling well solid metallic walls electrically and hence can be used to replace them. Similar conducting walls can also be used for other CPWs with back conductors to provide complete shield to the dielectrics below and other surrounding elements.

Coplanar Strips

Figure 26a, b shows the conventional coplanar strips (CPS) and another version with strips located on two different metal layers, respectively, both consisting of positive (+) and negative (−) signal lines, which could be used for multilayer RF circuits. The top-most metal layer or those nearest to it are preferred for the strips. The widths of the strips are equal in typical implementations. However, unequal widths can also be used to optimize performance. Similar to the conventional CPW, the CPS is partially shielded from the dielectrics below. The coupling between the CPS and the beneath dielectrics and circuits realized on these dielectrics depends on the distance of the strips from the beneath dielectrics and the gap between the coplanar strips or the offset between the noncoplanar strips. Under certain conditions, the coupling can be significant, leading to substantial undesired coupling effects. In general, the dielectric loss depends slightly on the width of the strips, while the conductor loss is a strong function of the strip width. The CPS has larger characteristic impedance than the microstrip line. Compared to the coplanar strips, the noncoplanar strips generally have stronger electric field between the strips which results in larger capacitance per unit length and, hence, lower characteristic impedance. Noncoplanar strips, however, can have less coupling to the beneath dielectrics, and hence less effects from circuits on these dielectrics, due to possible shielding from the lower strip. The coupling is significantly reduced when the lower strip is moved closer to the upper strip. In fact, the transmission line can be completely shielded from the dielectrics below if the lower strip is directly underneath the upper strip and has sufficiently larger size than the upper strip.

DOI 10.1007/978-981-4560-75-7_129-1

Strip Line

Figure 27 shows some possible strip-line configurations for multilayer RF circuits, in which the center conductor is sandwiched between two ground planes. The upper and lower ground planes do not have to be identical, and the dielectrics above and below the central strip may have different thickness in order to optimize the strip line's characteristics. The strip line in Fig. 27b particularly has slit in either or both of the ground planes to allow high characteristic impedances to be obtained for certain applications. Figure 27c depicts a shielded strip line in which vertical walls are used together with horizontal walls to completely enclose the strip line. These vertical walls can be formed using arrays of via-holes similar to those used for the shielded CPW shown in Fig. 27h. The strip line has larger size and less convenience in layout as compared to the microstrip line and CPW and, in general, operates at lower frequencies due to possible propagation of the higher-order parallel-plate waveguide modes. However, it is inherently isolated from the surrounding dielectrics and circuits on these dielectrics and, with proper use of thin dielectrics and sufficient number of via-holes between the two ground planes, can operate at high frequencies into the millimeter-wave range. Properly configured strip line is thus valuable for multilayer RF circuits.

Slot Line

Figure 28a, b show a slot line and shielded slot line, respectively. The shield is realized with horizontal metal strips and vertical metal walls formed by lattices of via-holes. It confines the fields and isolates the slot line from other structures external to the shield including the silicon substrate, other transmission lines, and circuit elements.

Summary

Transmission lines are important and inevitable elements in antennas, not only for antennas themselves but also for associated RF circuits needed for antennas to perform in RF systems. Proper configuration and design of transmission lines, which rely primarily on the thorough understanding of transmission line theory, can help produce antennas with associated RF circuits not only optimum in performance but also optimum in size. This chapter addresses the fundamentals of transmission lines including transmission-line equations and important transmission-line parameters. It also presents synthetic transmission lines, which could be very useful for on-chip antenna interface, and printed-circuit transmission lines essential for antennas and associated RF circuits.

Cross-References

- ▶ Impedance Matching and Baluns
- ▶ Microstrip Patch Antennas
- ▶ On-Chip Antennas
- ▶ Radial-Line Slot Antennas
- ▶ Substrate-integrated Waveguide Antennas
- ▶ Taped Slot Antennas
- ▶ Transmission-Line Based Metamaterials in Antenna Engineering

DOI 10.1007/978-981-4560-75-7_129-1

References

Cohn SB (1954) Problems in strip transmission lines. IRE Trans Microw Theory Tech 2:52–55
Cohn SB (1969) Slot line on a dielectric substrate. IEEE Trans Microw Theory Tech 17:768–778
Edwards TC (1984) Foundations for microstrip circuit design. Wiley, New York, pp 44–45
Garg R, Gupta KC (1976) Expressions for wavelength and impedance of slotline. IEEE Trans Microw Theory Tech 24:532
Ghione G, Naldi C (1983) Arameters of coplanar waveguides with lower ground plane. Electron Lett 19:734–735
Ghione G, Naldi C (1984) Analytical formulas for coplanar lines in hybrid and monolithic MICs. Electron Lett 20:179–181
Grieg DD, Englemann HF (1952) Microstrip – a new transmission technique for the kilomegacycle range. Proc IRE 40:1644–1650
Gupta KC, Garg R, Bahl IJ (1979a) Microstrip lines and slotlines. Artech House, Dedham, pp 89–90
Gupta KC, Garg R, Bahl IJ (1979b) Microstrip lines and slotlines. Artech House, Dedham, pp 277–280
Gupta KC, Garg R, Chadha R (1981) Computer-aided design of microwave circuits. Artech House, Dedham, p 72
Hammerstad EO, Jensen O (1980) Accurate models for microstrip computer-aided design. 1980 IEEE. MTT-S digest, pp 407–409
Hilberg W (1965) From approximation to exact relations for characteristic impedances. IEEE Trans Microw Theory Tech 13:29–38
Janaswamy R, Schaubert DH (1986) Characteristic impedance of a wide slotline on low-permittivity substrates. IEEE Trans Microw Theory Tech 34:900–902
Paul CR, Whites KW, Nasar SA (1998) Introduction to electromagnetic fields, 3rd edn. McGraw-Hill, New York, Chap. 7
Pucel RA, Masse DJ, Hartwig CP (1968a) Correction to losses in microstrip. IEEE Trans Microw Theory Tech 16:1064
Pucel RA, Masse DJ, Hartwig CP (1968b) Losses in microstrip. IEEE Trans Microw Theory Tech 16:342–350
Welch JD, Pratt HJ (1966) Losses in microstrip transmission systems for integrated microwave circuits. NEREM Rec 8:100–101
Wen CP (1969) Coplanar waveguide: a surface strip transmission line suitable for non-reciprocal gyromagnetic device application. IEEE Trans Microw Theory Tech 17:1087–1090
Wheeler HA (1978) Transmission line properties of a stripline between parallel planes. IEEE Trans Microw Theory Tech 26:866–876

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_130-1

Gap Waveguides

Ashraf Uz Zaman* and Per-Simon Kildal
Antenna Systems Division, Signals and System Department, Chalmers University of Technology, Göteborg, Sweden

Abstract

The coming years will show new applications of wireless communications at higher frequencies (30 GHz and above). Modern wireless technologies like massive MIMO and gigabit transmission will become a reality. The industrial winners will be the companies that can provide the hardware at the lowest cost. This requires new waveguide and mmWave packaging technologies that are more cost-effective than normal rectangular waveguide technology and are more power efficient (lower losses) than PCB-based microstrip and coplanar waveguides. The gap waveguide has this potential.

The present chapter gives the historical background of gap waveguide technology until its invention in 2008 and how it has evolved since then to include many different kinds of gap waveguide types. Several useful waveguide components have been developed for integration in complete RF front ends. Passive gap waveguide parts and components like filters, couplers, and transitions have been realized very successfully, and active microwave electronics have been packaged. The chapter contains also an overview of the gap waveguide antennas that have been developed during the last years.

Keywords

PMC; PEC; AMC; Parallel-plate mode; Stopband; Gap waveguide; MMIC packaging; Wideband slot array; Corporate feed network; High-Q resonator; Narrowband filter; Transitions

Introduction

The heavy congestion at the existing radio frequency spectrum allocated for the today's wireless communications motivates and accelerates the research work at mmWave bands or even higher-frequency range where more spectrum is available for various applications. On the other hand, there are plenty of technological factors and mechanical challenges in designing high-frequency microwave RF front ends. These factors are cost pressure, smaller size requirement, increased system density, packaging and cross-talk suppression, and lower power loss dissipation. Usually, RF front ends are composed of both active circuits (ICs) and passive devices; the latter include antennas and filters which are not integrated on the semiconductor substrate. The technological progress in the area of active components in RF systems has propelled with the advancement of monolithic microwave integrated circuit (MMIC) technology, resulting in the fact that the active components of RF systems occupy only a small segment of the module space area. On the other hand, the evolution of passive components have not advanced in a similar manner due to the fabrication tolerance issues associated with geometrical scaling governed by operating frequency. On top of the mechanical tolerance issues at high frequencies, circuit features and components can have dimensions that are an appreciable fraction of wavelength. So, there are some additional complex packaging issues to deal with such as cavity modes, radiation from minor bends or

*Email: zaman@chalmers.se

DOI 10.1007/978-981-4560-75-7_130-1

discontinuities and coupling, and feedback due to surface waves. At lower frequencies, the conventional way of doing the circuit package is to use metal wall cavity and vias in the substrate to isolate each critical circuit component within a microwave front end. The size of the cavity is maintained to be lower than the half guided wavelength $\lambda_g/2$ to avoid cavity modes within the module. Also, lossy absorbers are used to dampen the high-Q cavity modes. However, the cavity size reduction becomes impractical at mmWave frequencies, i.e., the half guided wavelength $\lambda_g/2$ for a GaAs substrate at 80 GHz is only 0.52 mm. Thus, at high frequency, there is a need for rethinking in RF design techniques in terms of improved electrical performance, manufacturing flexibility, packaging, and affordability. In this regard, having upcoming high-frequency applications in mind, some new technology development seems to be important.

In section "Reconsidering Conventional Microwave Technology" of this chapter, some key factors are reconsidered regarding the usage of conventional microwave technologies such as planar microstrip line or the metal waveguide for building and manufacturing high-frequency microwave module or systems. This section puts forward the existing performance gap between the planar transmission lines such as microstrip or CPW and the nonplanar metal waveguides in terms of cost, loss performance, manufacturing flexibility, etc. and motivates the research on gap waveguide technology.

After that in section "Operating Principle of Gap Waveguides and Loss Analysis," the newly proposed and developed gap waveguide technology is presented as a promising solution for high-frequency microwave problems. The basic operating principle of the proposed gap waveguide technology is also discussed. This section shows how periodic structures can be used to form a waveguide between two parallel plates and how electromagnetic wave propagation can be controlled in such waveguides by achieving cutoff condition for global parallel-plate modes over a specific bandwidth. An overview of different variants of available gap waveguide technology is presented in this section. This section also presents a comparison of losses among different gap waveguide geometries.

Section "Design and Optimization of the Parallel-Plate Stopband" deals with parallel-plate stopband design using a periodic structure. The design of the stopband is quite important as it dictates the performance of the gap waveguide components. Typically, a unit cell of the periodic structure is analyzed with eigenmode solver of commercial software to determine the parallel-plate stopband, and this type of study consists of parametric sweeps of all parameters associated with the periodic structure.

Section "Some Analytical Approaches Used for Analyzing the Gap Waveguide" contains an overview of two analytical approaches used so far in the prediction of propagation performance of groove gap waveguide and ridge gap waveguide. One of the approaches is the spectral domain approach in which Green's functions of the considered structure are determined. On the other hand, in the second approach, the modal solution is found by dividing the field problem in different regions and connecting the regions by mode matching. This section also shows the resemblance between the ridge/groove gap waveguide and the conventional hollow ridge/rectangular waveguide, respectively, by using numerical analysis. The dispersion diagrams and characteristic impedances are compared and found to be quite similar within the stopband of the parallel-plate modes. Thus, most of the initial designs of the gap waveguide components can be done using the already known waveguide architectures and topologies instead of deriving rigorous analytical expressions for gap waveguides separately. This also opens up the opportunity to use available numerical ports in commercial full-wave simulators for designing gap waveguide components.

Section "Low-Loss Antennas Based on Gap Waveguide Technology" presents different types of antennas designed by using gap waveguide technology. Due to the low-loss properties of gap waveguide, high-efficiency and high-gain planar antennas can be designed with this technology. Single-layer wide-band slot arrays, double-layer PCB-based slot arrays, horn arrays fed by microstrip gap waveguide, and series-fed groove gap waveguide narrow-wall slot arrays have been designed at different ranges of frequency band.

DOI 10.1007/978-981-4560-75-7_130-1

Section "Transitions Between Gap Waveguide and Standard Transmission Line" covers several transitions between gap waveguide and other standard transmission lines such as microstrip and rectangular waveguide, and these transitions are very important constituent related to gap waveguide components. Most often, other transmission lines have standardized cross sections (e.g., rectangular waveguide) or characteristic impedance (e.g., microstrip line). In order to reduce the reflections from the interface between gap waveguide and standard transmission lines, impedance transformers or matching sections are required. This section contains designs of microstrip to ridge gap waveguide transitions, ridge gap waveguide to standard rectangular waveguide transitions, microstrip-ridge gap waveguide to standard rectangular waveguide, etc.

Section "Gap Waveguide Narrowband High-Q Filters" demonstrates a mechanically flexible narrowband diplexer filter design which is a very critical component in a full-duplex communication wireless link. There is a need for developing new types of high-Q filters which are more robust in terms of mechanical assembly and can be integrated with other modules of passive components such as antennas. The groove gap waveguide-based high-Q filters can be of interest in this case. These high-Q filters can be built in an open parallel-plate structure surrounded by periodic metal pins without any side walls. This opens up the possibility of placing the diplexer filter in an integrated module together with RF electronics and even with the antennas.

Section "Gap Waveguide Packaging Solution for Microwave Modules" shows the gap waveguide packaging solution which can be used in RF modules. Most of the microwave systems have to operate outdoor, under various weather conditions and therefore have to be shielded, i.e., *packaged* in a proper way to provide protection against mechanical stress and environmental condition. Also, to comply with the smaller size requirements and compactness for next-generation microwave modules, large amount of electronic components are to be placed into a confined area. For such high-density microwave modules, RF packaging is more and more important in terms of isolation and interference suppression. So, there is a need for new packaging technology to be used at high frequency, and gap waveguide-based PMC packaging can be a real good solution to overcome the packaging-related problems. In this section, the application of gap waveguide packaging solution for both passive and active microwave circuits is mentioned.

Reconsidering Conventional Microwave Technology

Conventional rectangular waveguides, planar transmission lines such as CPW, or microstrip lines are well-characterized transmission medium which are used in a variety of complex RF component and circuit designs until today. However, there are some factors to be considered while using these conventional technologies at mmWave frequencies or above, particularly for antenna applications.

Typically, standard waveguide components are quite low loss and are manufactured in split blocks. Connection between the split blocks can be achieved by screwing, diffusion bonding, or deep-brazing techniques. These techniques are costly, complex, and often not scalable to higher frequencies. As the frequency of operation of waveguide approaches mmWave frequencies, the physical dimensions of the components decrease, and a "watchmaker" level of precision is required for manufacturing and assembling the split blocks by conventional methods. To remove the tiny gaps between the two split blocks, very good electrical contact as well as good alignment of the two blocks must be achieved. Also, a high-quality surface finishing over the whole metal contact area is required for good mechanical assembly. Contacting surfaces must be protected from corrosion and oxidation over the entire life cycle of the product. These strict mechanical requirements lead to a very high-precision metal machining technique, which increases the cost of manufacturing and causes much delay in the production chain. This effectively precludes mass

DOI 10.1007/978-981-4560-75-7_130-1

production of components for commercial purposes. In the context of high-efficiency waveguide slot array antennas, this electrical contact issue is a very important factor for good electrical performance. In planar slot array applications, where multiple waveguides are needed with narrow vertical walls for separating each waveguide from the neighboring waveguides, it is very difficult to achieve electrical contact between the vertical walls of the waveguide and the radiating slot layer (Kimura et al. 2001; Kirino and Ogawa 2012). This leads to the fact that waveguide slot arrays have not found many commercial applications till today. They have mainly been used in few military applications.

Apart from the manufacturing costs and assembly complications, some other limitations of waveguide slot arrays have been reported in literature (see "Low-Loss Antennas Based on Gap Waveguide Technology"). Series-fed single-layer waveguide slot arrays are simple but have narrow bandwidth due to the long-line effect (SeHyun et al. 2006). In a single-layer structure, it is normally not possible to feed each radiating element in parallel (full corporate feed) because of the space limitations associated with keeping the element spacing smaller than one wavelength (λ_o) to avoid grating lobe (Fujii et al. 2008; Tsugawa et al. 1997). Therefore, there is need for a complex multilayer feed network. Such a double-layer corporate feed network in rectangular waveguide technology is described in Miura et al. (2011).

On the other hand, microstrip and coplanar lines are the most representative planar transmission lines. These are robust, low-cost solutions, and very suitable for integrating active microwave components on circuit boards. Especially, the CPW is widely used in monolithic microwave integrated circuits (MMICs) capable of a very high metal pattern resolution. But, the transmission properties of both microstrip and CPW lines greatly depend on the substrate parameters. Both lines suffer from high insertion loss at mmWave frequency spectrum due to the presence of lossy dielectric material. Published studies in Tsuji et al. (1995), Mesa et al. (2000), and McKinzie and Alexopoulos (1992) show that significant power leakage exists on various printed circuit transmission lines, often related to surface waves in the dielectric substrate causing serious cross-talk and interference problems (see "Gap Waveguide Packaging Solution for Microwave Modules"). Particularly in the case of top-covered microstrip line, this leakage begins at a much lower frequency than expected and becomes a matter of serious concern for power loss as well as cross-talk issue (Mesa et al. 2000). Similarly, conventional packaging of CPW modifies the CPW into a conductor-backed CPW (CBCPW), and this thereby generates power leakage in the form of coupling to parasitic parallel-plate modes. In the context of low-cost planar antennas, the microstrip antenna feed networks suffer from high ohmic and dielectric losses at high frequency (Borji et al. 2009; Pozar 1983). Spurious radiations and leakage in the form of surface waves may become major concerns in microstrip antennas and are difficult to handle (Levine et al. 1989). Spurious radiations from the discontinuities in the feeding network can have a dramatic influence on the radiation patterns of the PCB-based planar antennas. All these lead to substantial reduction in antenna efficiency and thereby gain. One way to reduce the effect of unwanted radiation is to design multilayer printed antennas in order to separate the radiating layer from the feeding line layer. However, high-performance multilayer PCBs bring added challenges to both the fabricator and designer, and most of the advantages of single-layer PCB technology do not hold anymore in that case.

So, there exists a big performance gap between the planar transmission lines such as microstrip or CPW and the nonplanar metal waveguides in terms of loss, manufacturing flexibility, and production cost. One of the main high-frequency microwave research challenges to date is to narrow down the abovementioned performance gap. Researchers aim at finding a solution which is as low cost and flexible as microstrip and also as low loss as metal waveguide. It is expected that high electrical performance and high-density integration techniques, combined with a low-cost fabrication process, should be able to offer widespread solutions for future high-frequency microwave or mmWave commercial applications. Therefore, lot of research has been done in microwave community to meet these requirements. Researchers have come out with technologies such as substrate integrated waveguide (SIW) (Deslandes and Ke 2001;

DOI 10.1007/978-981-4560-75-7_130-1

Feng and Ke 2005; Ke 2006), low-loss thin-film microstrip lines (Six et al. 2005; Nishikawa et al. 2001), and LTCC (Shafique and Robertson 2011; Tze-Min et al. 2007). Each of these proposed techniques has their own merits and demerits. In line with the research activities mentioned above, a new transmission line technology known as gap waveguide was proposed in 2009. This new technology has almost no dielectric losses and is mechanically more flexible than metal waveguide in terms of mechanical or modular assembly. The prevailing idea throughout this chapter is to describe the operating principle of this new low-loss transmission line technology and to illustrate how the critical RF passive components can be designed using this technology. A completely new packaging method for high-frequency microwave modules based on gap waveguide technology will be also presented in this chapter.

Operating Principle of Gap Waveguides and Loss Analysis

Gap waveguide technology appeared as an extension of the research about hard and soft surfaces (Kildal 1990). The soft and hard surface concept is a generic representation of the corrugated surface used in corrugated horn antennas. The soft surface has the ability to stop waves of any polarization propagating along the surface and has similar characteristics as the electromagnetic bandgap (EBG) surface (Sievenpiper et al. 1999). The hard surface can on the other hand enhance propagation along its surface (along the corrugations). This has been already used to reduce forward scattering (blockage) by cylindrical objects in Kildal et al. (1996), almost ten years before such blockage reduction became a popular research topic known under the name cloaking (Schurig et al. 2006). The blockage reduction using hard surfaces has been demonstrated to have 20 % bandwidth for dual-polarized waves, whereas most other cloaking concepts have not been proven to be useable in practice (Kildal et al. 2007). The soft surfaces have mainly been used for horn antennas. Hard horns with large-aperture efficiency have been proposed very early for application in cluster-fed reflectors for multiple beams (Kildal and Lier 1988; Lier and Kildal 1988). The hard horn work included advanced mode-matching modeling and realization of hardware (Skobelev and Kildal 2005; Sotoudeh et al. 2006). The overview article (Lier 2010) describes another trace of work resulting in very unique lightweight and wideband metasurface horns with anisotropic boundary conditions between the soft and hard cases. There is also a third path focusing more on miniaturized hard horns (Kehn et al. 2004, 2006). However, these hard waveguides were extremely narrowband. Still, they are important forerunners of the wideband ridge gap waveguide (Kildal and Kehn 2010). In early 1997, researchers found existence of local surface waves propagating along each groove of a hard surface (Sipus et al. 1997). In 2009, the existence of local waves following even the ridges of a hard corrugated surface has been discovered (Valero-Nogueira et al. 2009). The latter discovery has been further developed as gap waveguide technology by keeping only one ridge and replacing the rest of them with metal pins (Kildal et al. 2009). Ideal gap waveguide uses the basic cutoff of a PEC-PMC parallel-plate waveguide configuration to control the electromagnetic wave propagation between the two parallel plates. As long as the air gap separation between the PEC and PMC plates is less than $\lambda/4$, no wave can propagate between the plates. This can be readily seen by studying a plane wave in the air gap and trying to satisfy boundary conditions at both the PMC and PEC plates. Now, if the PMC surface incorporates a metal strip, EM waves will be able to follow the strip. This is shown in Fig. 1.

PMCs are not available in nature. Thus, in practical applications, the PMC condition is emulated by an artificial magnetic conductor (AMC) in the form of periodic textured structures such as metal pins (Silveirinha et al. 2008) or mushroom structures (Sievenpiper et al. 1999). Thus, in a real gap waveguide structure, the bottom periodic textured has high enough surface impedance to create a stopband over which no parallel-plate modes can propagate. However, the textured AMC surface must also incorporate guiding structures in the form of ridges, grooves, or strips to form the complete waveguide. As a result of

DOI 10.1007/978-981-4560-75-7_130-1

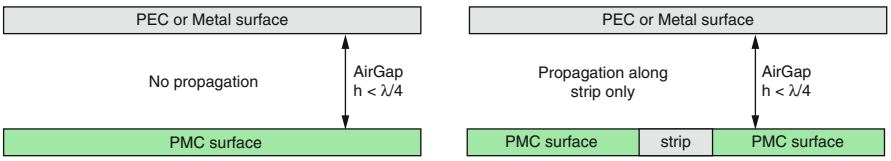


Fig. 1 Cross section of ideal gap waveguide (With permission from Zaman et al. 2014)

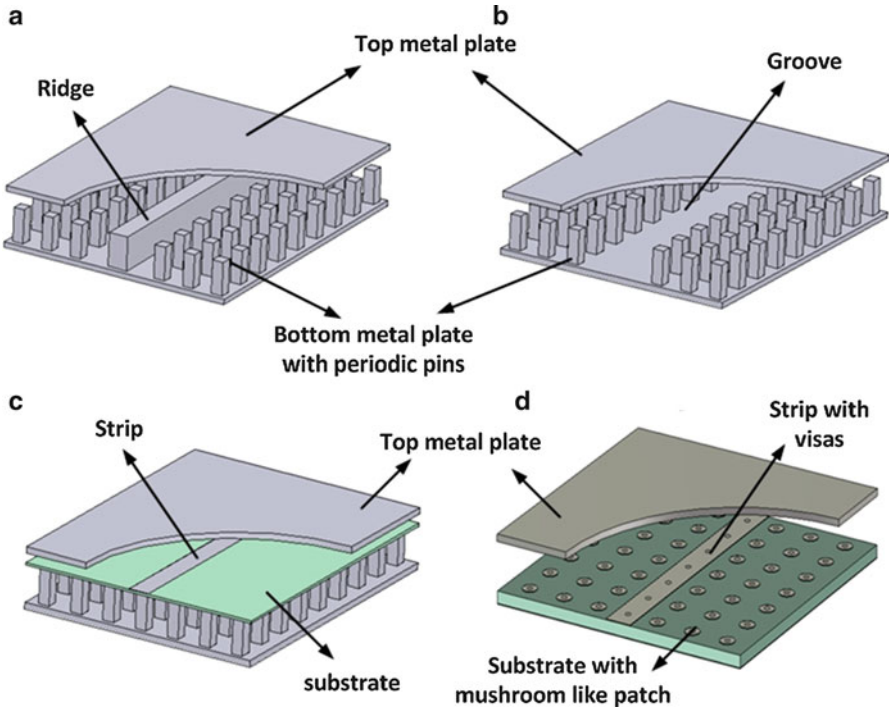


Fig. 2 Different gap waveguide geometries: (**a**) ridge gap waveguide, (**b**) groove gap waveguide, (**c**) inverted microstrip gap waveguide, (**d**) microstrip-ridge gap waveguide (With permission from Zaman et al. 2013, 2014)

the stopband (see "Design and Optimization of the Parallel-Plate Stopband"), the AMC surface creates virtual lateral walls on both sides of the guiding section and thereby prevents the lateral field leakage. Thus, the electromagnetic waves can propagate along these guiding ridges, grooves, or strips without leaking away in other directions. These guiding structures thereby define different gap waveguides or transmission lines, referred to as ridge, groove, and microstrip gap waveguides (Kildal 2009; Kildal et al. 2011; Zaman et al. 2010b; Rajo-Iglesias and Kildal 2010; Valero-Nogueira et al. 2011; Pucci et al. 2012; Raza et al. 2014; Gahete Arias et al. 2013). The different configurations of gap waveguides are shown in Fig. 2.

The main advantage of the gap waveguide structure is that it can be realized without any requirement of metal contact between the upper metal surface and the lower surface, and thereby allowing low-cost manufacturing of low-loss waveguide components at mmWave frequency bands or even higher. The relaxed mechanical requirements pave the way for low or moderate precision machining, lower fabrication time requirements, and may even make possible the usage of some low-cost fabrication techniques such as injection molding and plastic hot embossing. The operating modes in different gap waveguide geometries are also different. The desired mode in ridge gap waveguide, inverted microstrip gap waveguide, and microstrip-ridge gap waveguide is similar in nature, and all of these allow a

DOI 10.1007/978-981-4560-75-7_130-1

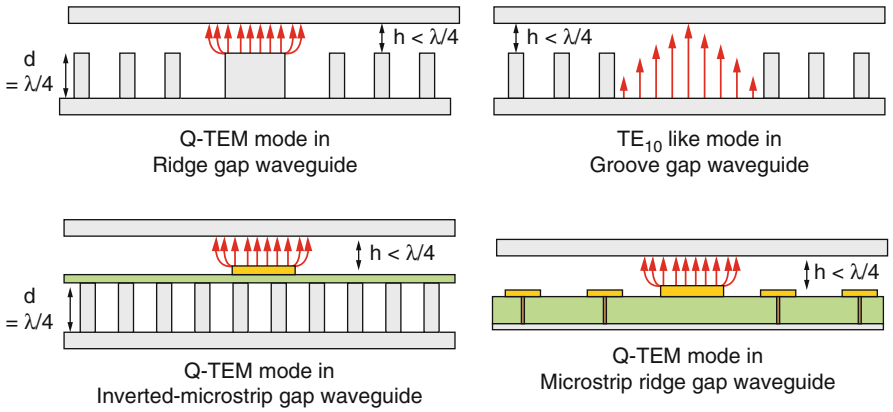


Fig. 3 Desired modes of propagation in different gap waveguide geometries

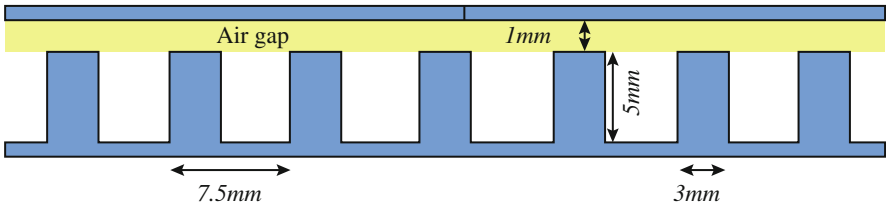


Fig. 4 Detailed dimensions of the periodic metal pin

quasi-TEM mode within the stopband. On the other hand, the desired mode in the groove gap waveguide is very similar to that of the TE_{10} mode of the rectangular waveguide (Fig. 3).

Necessity for Parallel-Plate Stopband and the Dispersion Characteristics of Gap Waveguide

As mentioned in the previous section, the main performance of the gap waveguide is determined by its ability to create parallel-plate stopband for wave propagation in the undesired direction. This stopband is usually achieved by a periodic texture located around the guiding metal ridge, groove, or the strip. The periodic structure functions as a high-impedance surface when placed closely (with an air gap smaller than λ/4) to a metal plate and is often referred to as an AMC surface. The most important thing in designing the stopband is to obtain the lower and upper cutoff frequency of this stopband. Usually, the cutoff study is performed as a function of the geometrical parameters of the periodic structure to be used. This kind of study is well described in Rajo-Iglesias and Kildal (2011). In the present section, the details of the stopband and dispersion characteristics of different gap waveguides are described. At this point, a textured surface made of square pin is considered for emulating the AMC surface over a frequency range of 10–25 GHz. Fig. 4 shows the details of the pin surface. The computed dispersion diagram for this corresponding structure is shown in Fig. 5a and 5b. A large stopband is created within 10–25-GHz frequency band where no EM waves can propagate.

Once the stopband is obtained, a guiding ridge or a groove may be incorporated in the periodic pin structure. Then, there is a mode propagation which follows the light line and is considered as the desired quasi-TEM mode in case of ridge gap waveguide. In the case of groove gap waveguide, a mode similar to TE_{10} mode starts to propagate inside the stopband, and this mode is dispersive in nature. These two scenarios of ridge gap waveguide and groove gap waveguide are shown in Fig. 5c and 5d, respectively.

DOI 10.1007/978-981-4560-75-7_130-1

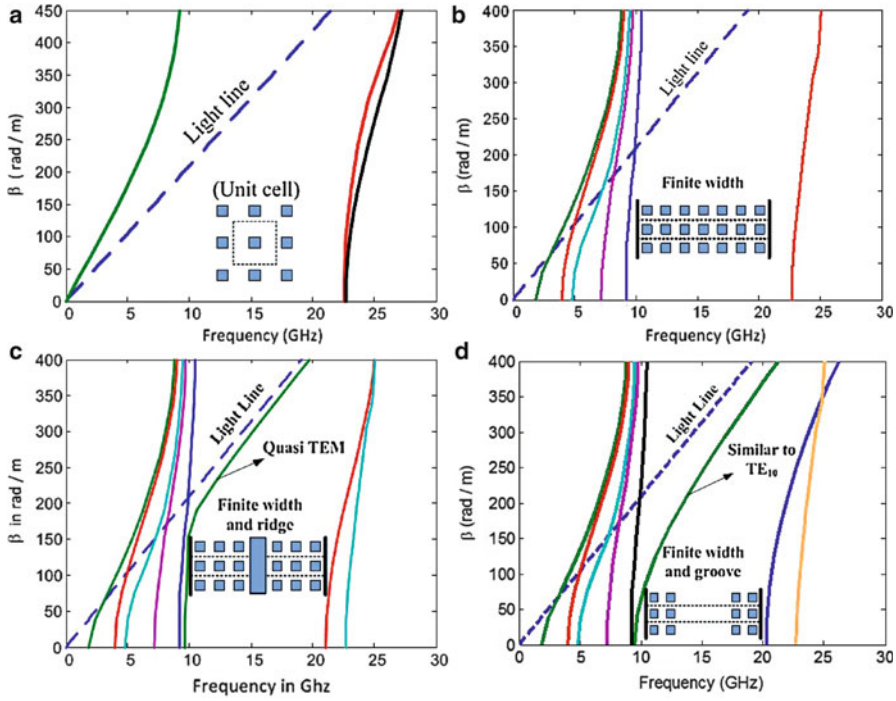


Fig. 5 Dispersion diagrams: (**a**) unit cell of a periodic pin, (**b**) periodic pin structure with finite rows of pins, (**c**) ridge gap waveguide, (**d**) groove gap waveguide (With permission from Zaman et al. 2013)

The dispersion diagram for an inverted microstrip gap waveguide and the microstrip-ridge gap waveguide would look like similar to that of ridge gap waveguide except that the stopband will shrink a bit due to the presence of dielectric material.

Field Attenuation in the Pin Region and Measured Attenuation

One very important aspect for such oversized parallel-plate gap waveguide structures is the level of field attenuation in the periodic structure which has been designed to act as high-impedance surface or AMC surface. At this point, the field decay or attenuation is studied for periodic metal pins also known as bed of nails. The modal field distribution, computed in the transverse plane inside the ridge gap waveguide structure, is shown in Fig. 6. As shown in Fig. 6, field distribution is nearly constant over the ridge, and then it falls almost at a rate of 18–20 dB per row of pins most rapidly near the upper end of the frequency band. This attenuation has been measured also by placing two ridge gap waveguide sections and separating them by three rows of pins. The measured isolation is shown in Fig. 7. The measured results correspond well with the computed attenuation level. The measured isolation between two lines placed side by side has been found to be better than 60 dB over the entire Ku-band.

Losses in Different Gap Waveguide Geometries

In order to verify the performance of the low-loss gap waveguide technology, it is important to be able to characterize its loss in an accurate way. One way to do that is to measure the attenuation of a very long line. However, at low frequencies, it will require impractically long transmission lines in order to measure the low losses accurately. As an alternative, losses can be better characterized in terms of Q-factors of resonators. Thus, several groove gap and ridge gap resonators were built and measured at Ku-band and compared with standard rectangular waveguide resonators (Pucci et al. 2013). But at higher frequency, longer ($\sim 30\lambda$) gap waveguide geometries were manufactured and measured at V-band, 50–70-GHz frequency. Tables 1 and 2 present the losses at two different frequency ranges. It is to be mentioned

DOI 10.1007/978-981-4560-75-7_130-1

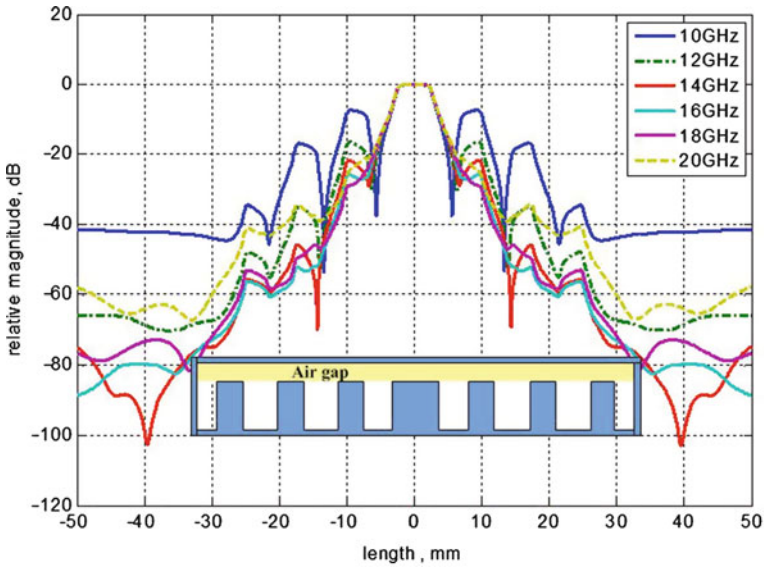


Fig. 6 Field distribution in transverse plane within the gap waveguide structure

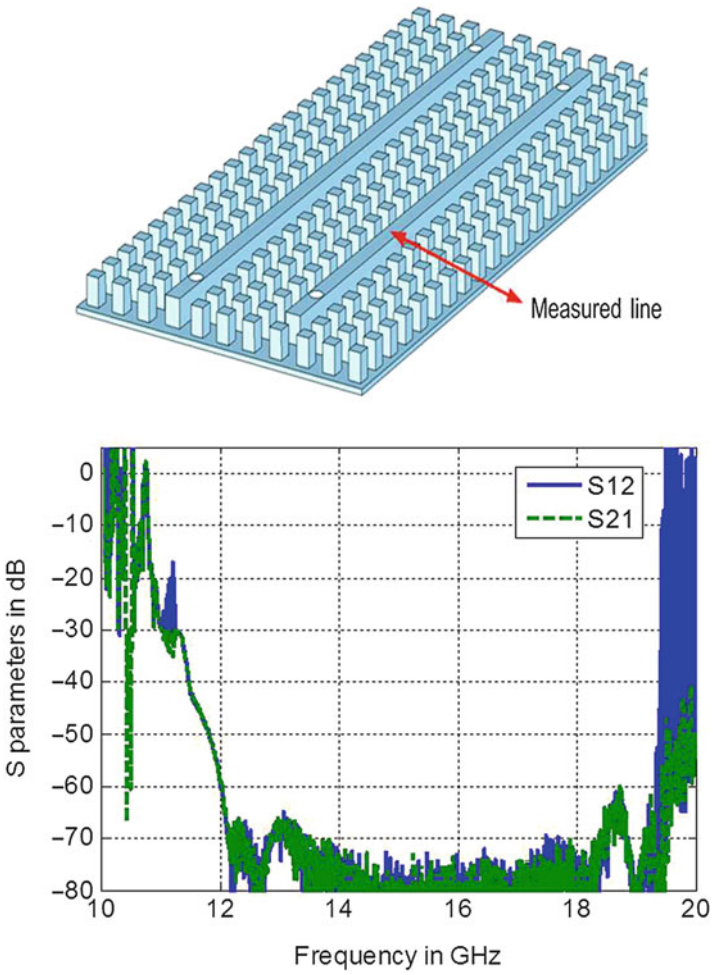


Fig. 7 Measured isolation between two lines within the gap waveguide structure (With permission from Zaman et al. 2014)

DOI 10.1007/978-981-4560-75-7_130-1

Table 1 Comparisons of simulated and measured results for gap waveguide resonators at Ku-band

	Simulations			Measured	
	Q value (eigenmode)	Q value (S_{21})	Frequency (GHz)	Q value	Frequency (GHz)
Short-circuit resonator, ridge gap waveguide	4510	4741	13.27	2255	13.308
Open-circuit resonator, ridge gap waveguide	4537	4603	13.2	4130	13.18
Short-circuit resonator, groove gap waveguide	6534	6136	13.5	5200	13.47
Open-circuit resonator, groove gap waveguide	6265	6108	13.46	5883	13.44
Rectangular waveguide resonator WR-62	8499	8462	13.543	5400	13.544

Table 2 Comparisons of insertion loss results for 30λ-long gap waveguide structures at V-band

	Simulated maximum loss (dB/cm)	Measured minimum loss (dB/cm)	Measured maximum loss (dB/cm)
Rectangular waveguide (extruded)	0.0134	0.022	0.042
Rectangular waveguide (E-plane split blocks)	0.01355	0.024	0.046
Groove gap waveguide	0.019	0.026	0.045
Ridge gap waveguide	0.0493	0.053	0.073
Microstrip-ridge gap waveguide	0.0805	0.1753	0.22
Inverted microstrip gap waveguide	(0.093)	(0.21)	(0.27)
Normal microstrip line (for Rogers 3003, 0.125-mm thick)	(0.37)	(0.63)	(0.77)

that, at lower frequency, the study was limited to only ridge gap waveguide and groove gap waveguide. But at 60 GHz, inverted microstrip gap waveguide and microstrip-ridge gap waveguide were also included in the loss study. Both these tables show that the groove gap waveguide has the lowest losses among all the gap waveguide structures. The losses of groove gap waveguide are 10–30 % more than extruded rectangular waveguide without split blocks. But the losses of groove gap waveguide become same (in the same order) as that of rectangular waveguides if the rectangular waveguide is manufactured in two parts as split blocks. Thereafter comes the ridge gap waveguide. The ridge gap waveguide is approximately 30–55 % more lossy than rectangular waveguide. The microstrip versions of the gap waveguide are more lossy than the groove and ridge gap waveguide. But they are about three to five times less lossy than a typical microstrip line depending upon the substrate used in microstrip circuit. Some of these manufactured gap waveguide lines are shown below in Fig. 8.

Design and Optimization of the Parallel-Plate Stopband

As mentioned by Kildal (2009; Kildal et al. 2011), the main performance of the gap waveguide is determined by its ability to create parallel-plate stopband which allows the EM wave to remain confined within a region or even propagate toward a desired direction of the parallel-plate waveguide instead of

DOI 10.1007/978-981-4560-75-7_130-1

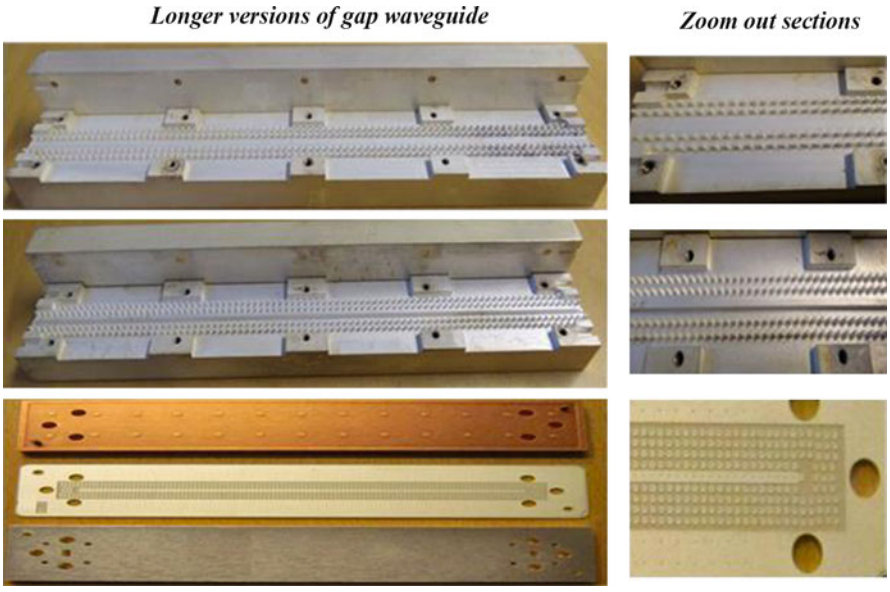


Fig. 8 The longer gap waveguide prototypes at V-band

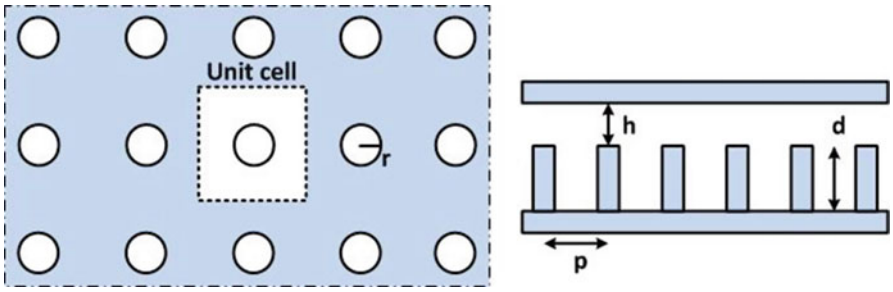


Fig. 9 Geometry of periodic metal pin (With permission from Zaman et al. 2014)

spreading out in all other directions. This stopband is achieved usually by using a metal texture in the form of a periodic structure along with a top metal plate. The periodic structure emulates a high-impedance boundary often referred to as an artificial magnetic conductor (AMC), and this high-impedance condition is achieved only over a specific bandwidth. This section explores the possibilities of creating parallel-plate cutoffs, i.e., the stopband in parallel-plate waveguides, by using different periodic structures. This is done by studying the detailed geometry of a specific periodic structure to be used as AMC and linking the geometric study with the realized bandgap inside a parallel-plate waveguide, in a similar way as people would perform the bandgap study for surface waves in open surfaces. All the analyses presented in this section are based on numerical results obtained from commercial EM solver.

Parallel-Plate Cutoff Realized by Bed of Nails

The first periodic structure to be studied in this section is the periodic metal pin structure also known as "fakir's bed of nails." This is shown in Fig. 9. For convenience, the stopband in this case is defined by the difference between the start frequency (when the parallel-mode propagation is stopped) and the end frequency (when the parallel-plate modes propagate again). The parameters that are important for designing the stopband are the height of air gap h, the height of the pin d, the period of the pin p, and the radius of the pin r.

DOI 10.1007/978-981-4560-75-7_130-1

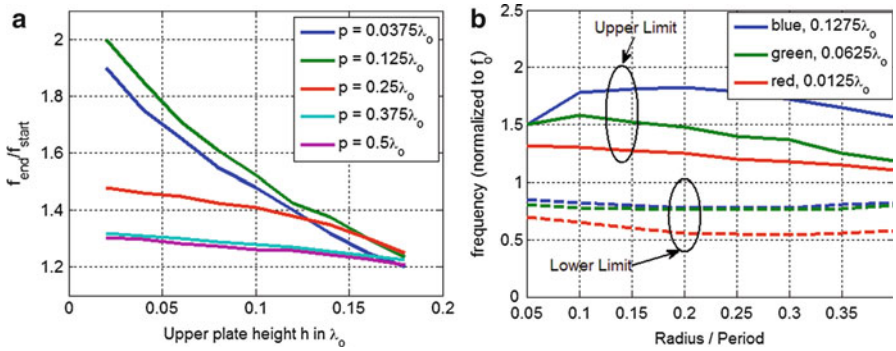


Fig. 10 Parametric effects on the stopband obtained by bed of nails, (**a**) relation between h and p, (**b**) relation between r/p and h

First of all, it has been found that the pin shape (circular or square) does not have any significant influence on the stopband (Rajo-Iglesias and Kildal 2011). If the height d and radius r of the pins are kept constant, for all values of p, the size of the stopband increases when the height h of the air gap decreases (Rajo-Iglesias and Kildal 2011). The start frequency of the stopband reduces when the period of the pins increases owing to increase in effective electrical length of the pins. On the other hand, the stop frequency is almost unaffected by the period as long as this is small enough. If the period becomes larger ($p \geq 0.25\lambda$), the upper limit is dramatically reduced owing to the propagation of new modes. The relative cutoff bandwidth (f_{end}/f_{start}) is summarized in Fig. 10a as a function of period p and air gap height h. Apart from the relation between p and h, the ratio r/p has also some influence on the stopband. The start and end frequencies of the stopband move in opposite directions when the radius-to-period ratio increases, reaching a maximum when the ratio is between 0.05 and 0.2. This is shown in Fig. 10b. The lattice constitutes the last investigated parameter for this geometry. From photonics it is known that the effect of this parameter in the stopband design cannot be neglected. So, in addition to the square lattice considered so far, a triangular lattice has also been studied in Rajo-Iglesias and Kildal (2011) varying the air gap height h and radius r. For the case of square lattice and triangular lattice, the results are not exactly the same, but they are similar enough to consider this parameter as having a secondary effect on defining the stopband properties of this structure. Even though any type of AMC can be used as a textured surface in gap waveguide structures, an important advantage of the bed of nail structure is that it is made from metal and no dielectric material is required. This makes the losses negligible at high frequencies.

Parallel-Plate Cutoff Realized by Mushroom-Type EBG

The mushroom-type EBGs shown in Fig. 11 are known to have the property of high-impedance surface (Sievenpiper et al. 1999). In this part, the parametric analysis for such mushroom-type EBG is presented. The parameters taken into account are air gap height h, substrate thickness d, the separation between the patches g, and radius of the vias r.

Among the parameters of this structure, initially analyzed parameter is the separation gap g. It is noted that the effect of g on the stopband is not very significant, except for the case where the gap height h is small. This is shown in Fig. 12a. The radius of the grounded vias r is another one of the parameters to be taken into account. Its effect on the stopband is presented in Fig. 12b. Actually, the radius r does not directly affect the stopband size but its position in the spectrum. The larger the radius r, the higher is stopband in the frequency spectrum. The next parameter to be considered is the thickness of the substrate d or the distance from the patch to the ground plane. The start frequency is completely shifted in frequency depending on the value of the substrate thickness. The thicker the substrate, the lower is the start frequency and this happens independently of the air gap height h. This effect is shown in Fig. 12c. Finally, the effect of the position of the grounded via in the patch as described in Fig. 11 has been studied.

DOI 10.1007/978-981-4560-75-7_130-1

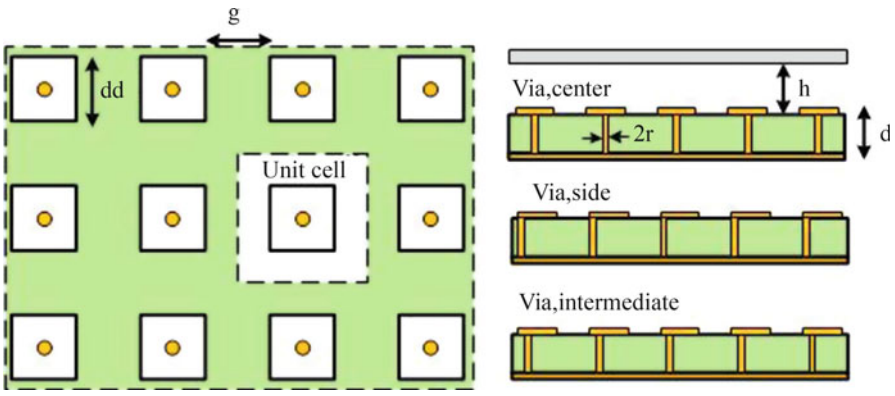


Fig. 11 Geometry of mushroom-type EBG (With permission from Razavi et al. 2014)

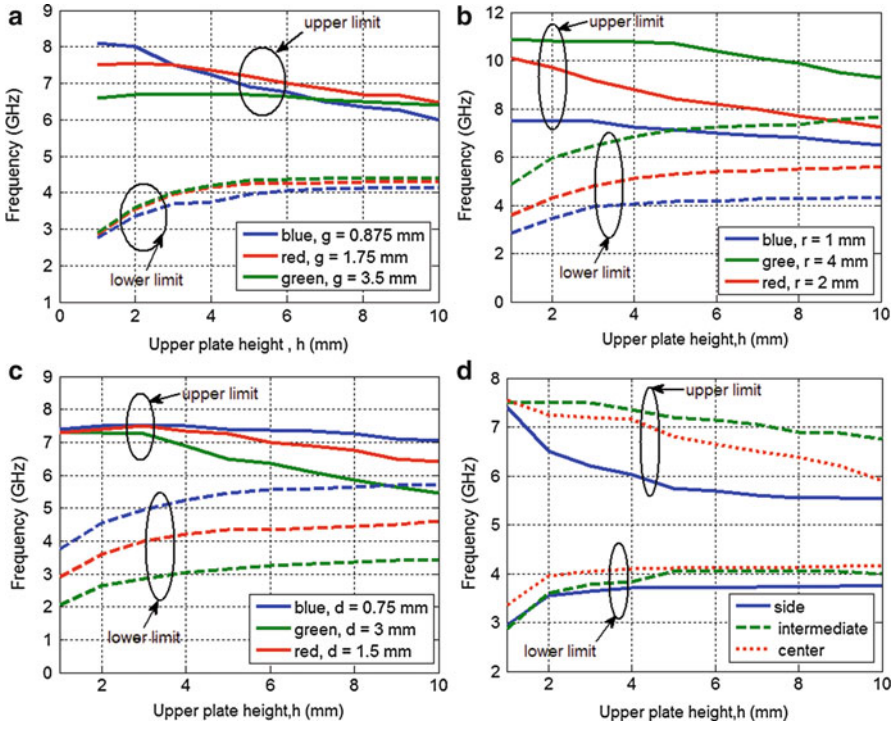


Fig. 12 Parametric effects on the parallel-plate stopband obtained by the mushroom-type EBG, (**a**) effect of *g*, (**b**) effect of *r*, (**c**) effect of *d*, and (**d**) the effect of the position of the grounded via (With permission from Razavi et al. 2014, Zaman et al. 2014)

Three positions for the via, centered, at the edge of the patch, and in an intermediate location between center and edge, have been selected. The results are presented in Fig. 12d. It is observed that, when the mushroom-type EBG surface is inside a parallel-plate structure, the displacement of the grounded via from the center to any other position causes a reduction on the stopband size.

Parallel-Plate Cutoff Realized by Bed of Springs

Apart from the commonly used bed of nails or mushroom-type EBG surface, some other periodic structures have been studied also for low-frequency application where the beds of nails or mushrooms

DOI 10.1007/978-981-4560-75-7_130-1

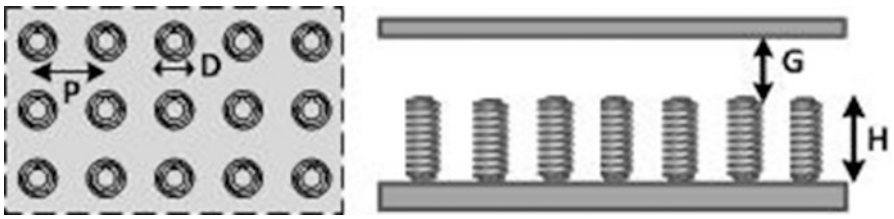


Fig. 13 Geometry of bed of springs

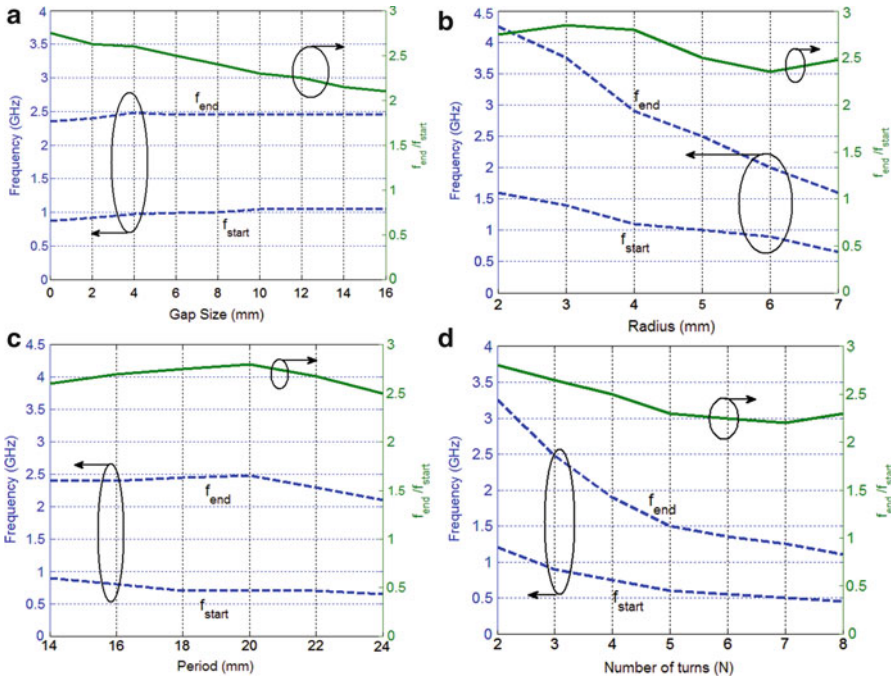


Fig. 14 Parametric effects on the parallel-plate stopband obtained by the bed of springs, (**a**) effect of *G*, (**b**) effect of *D*/*2*, (**c**) effect of *P*, and (**d**) the effect of *N* (With permission from Zaman and Kildal 2014, Zaman et al. 2014)

become too big in size and compact unit cells are needed particularly in applications such as packaging. That is why the bed of springs shown in Fig. 13 has been proposed in Rajo-Iglesias et al. (2012) as an alternative to the bed of nails at low frequencies. The period of a periodic helical structure can be of sub-wavelength, and most importantly due to the wire shape, the total height of the unit cell can also be small. The main parameters to be studied for this type of bed of springs are height of the spring *H*, height of air gap *G*, periodicity *P*, diameter of the spring *D*, and number of turns in each spring *N*, respectively. It is mentioned in Rajo-Iglesias et al. (2012) that the effect of the air gap height in this case is not so critical as compared to the case of bed of nails or mushrooms. This is due to lesser capacitive effect from the top of springs to upper plate, and this makes the gap a minor factor for the design. This is shown in Fig. 14a. The diameter of each spring *D* (keeping the period same) is now considered. The achieved results for this case are presented in Fig. 14b. The effect of this parameter is clear: a larger diameter shifts down all the modes in frequency as the total wire length is increasing, but the relative size of the stopband is not much affected as seen in the figure.

The period of the unit cell is also an important parameter as it determines the number of required unit cells for a given application. A summary of simulation results is presented in Fig. 14c. For this parameter

DOI 10.1007/978-981-4560-75-7_130-1

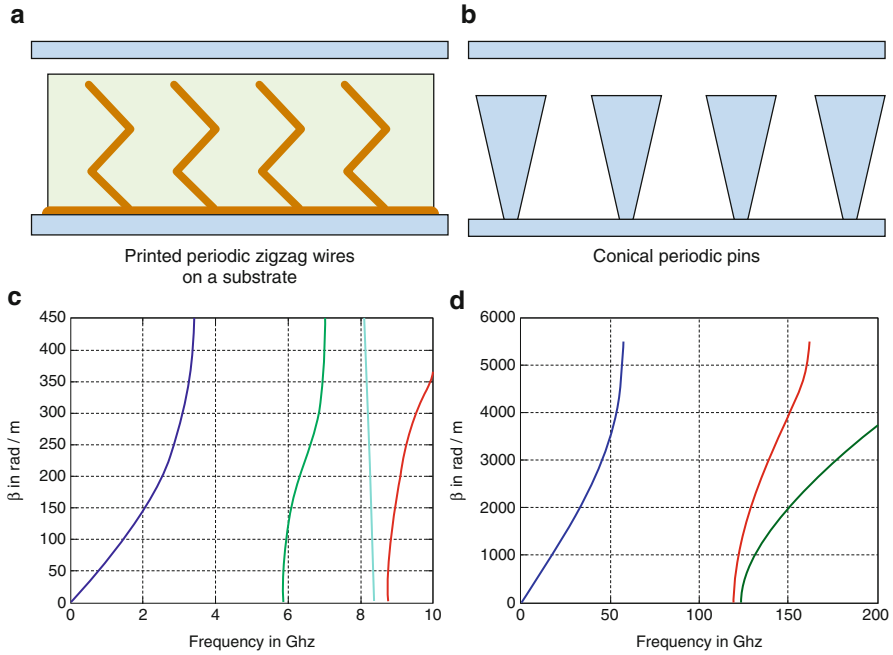


Fig. 15 Geometries of other types of periodic structures, (**a**) printed zigzag wires, (**b**) conical pins, (**c**) dispersion diagram of printed zigzag wires, and (**d**) dispersion diagram of conical pins (With permission from Zaman and Kildal 2014)

changing between 13 and 25 mm (which is small in terms of wavelength), the stopband is not greatly modified. When number of turns N is varied, keeping all the other parameters constant (including the height H), the results are summarized in Fig. 14d. As expected, increasing the number of turns, or, in other words, increasing the electrical length of the wire, all the modes have been moved to lower frequencies as presented in the graph.

Apart from the three periodic structures mentioned above, some other periodic structures have also been studied in Rajo-Iglesias et al. (2013) and Zaman et al. (2011). These are, namely, the printed zigzag wires and conical pins, and these periodic structures are shown in Fig. 15a and 15b, respectively. Both these periodic structures have certain advantages. The dispersion diagrams obtained for such periodic structures are also shown in Fig. 15c and 15d, respectively. As shown in dispersion diagrams, the printed zigzag wires are more suitable for lower frequency, and the conical pins are more suitable for high-frequency application. In terms of stopband size, the printed zigzag wires are very similar to that of the bed of springs. But it is much easier to fabricate the zigzag wires than the springs as they are simply printed on a substrate. For the conical pin case, the stopband size is actually 20 % bigger than that of circular metal pins described earlier. Also, the conical shapes are convenient to manufacture than the straight shapes by micromachining techniques at frequencies above 100 GHz (Rahiminejad et al. 2012).

Some Analytical Approaches Used for Analyzing the Gap Waveguide

From the analysis point of view, the gap waveguide structure is a bit different since the crucial element of this structure is a periodic surface which creates the stopband of the parallel-plate modes. The classical way to determine the fields and currents for such a structure would be the Floquet mode expansion combined with some general approach such as MoM, FEM, or FDTD. Also, a general 3D code could be utilized for the analysis of a finite gap waveguide structure. However, these approaches can be very complex and time-consuming if the source excites a spectrum of plane waves and if the structure is large.

DOI 10.1007/978-981-4560-75-7_130-1

Thus, it is desirable to develop faster analytical or semi-analytical methods which could be used as a first design tool by introducing some simplifications into rigorous approach. Apart from the emphasis on the analytical methods, there is also an immense need to investigate the accuracy of the available numerical ports in commercial EM solvers when applied to realistic ridge/groove gap waveguide microwave components.

Spectral Domain Approach

This sections present study of spectral domain Green's functions of parallel-plate waveguides, where one plate is a smooth metal surface and the other one is an artificial magnetic surface realized by one-dimensional (1D) or two-dimensional (2D) periodic structure. These two geometries are discussed in detail in Bosiljevac et al. (2010). The general idea here is to use some approximate boundary condition replacing the periodic texture and to derive Green's functions for simpler waveguides in the spectral domain. The derivation of Green's functions is obtained in plane-wave spectral domain, and plane-wave spectral domain is defined by

$$\tilde{E}\left(k_x, k_y, z\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E(x, y, z) e^{jk_x x} e^{jk_y y} dxdy \tag{1}$$

where "∼" denotes the two-dimensional Fourier transformation with k_x and k_y the spectral variables. Now, the spectral solutions of the Helmholtz differential equation for E_z and H_z components of the electromagnetic field will have the following form:

$$\tilde{E}_z\left(k_x, k_y, z\right) = A\cos\left(k_z z\right) + B\sin\left(k_z z\right) \tag{2a}$$

$$\tilde{H}_z\left(k_x, k_y, z\right) = C\cos\left(k_z z\right) + D\sin\left(k_z z\right) \tag{2b}$$

where the $e^{-jk_x x} e^{-jk_y y}$ variation is suppressed. Also, here ${k_z}^2 = {k_o}^2 - {k_x}^2 - {k_y}^2$ and A, B, C, and D are the unknowns which are to be determined by using boundary conditions. It is sufficient to determine only the z components of the electromagnetic field as all the other components can be found by using relations in Das and Pozar (1987). Now, the aforesaid asymptotic boundary condition for the considered surface is represented by the homogenized surface impedance/admittance in the plane-wave spectral domain. This remains valid in the limit when the periodicity of the surface approaches zero and the surface admittance components are defined as the following:

$$\tilde{Y}_{yx} = \frac{\tilde{H}_y}{\tilde{E}_x}; \quad \tilde{Y}_{xy} = -\frac{\tilde{H}_x}{\tilde{E}y} \tag{3}$$

1D and 2D Periodic-Type Structure

In this section, parallel-plate structure having 1D and 2D periodic structure at the bottom layer is analyzed. Fig. 16 shows these types of structures. While analyzing these structures, the periodicity at the bottom layer is considered to be much smaller than wavelength. Green's functions for such structures are derived in detail in Bosiljevac et al. (2010). These are also presented here for convenience of the reader.

For 1D case:

DOI 10.1007/978-981-4560-75-7_130-1

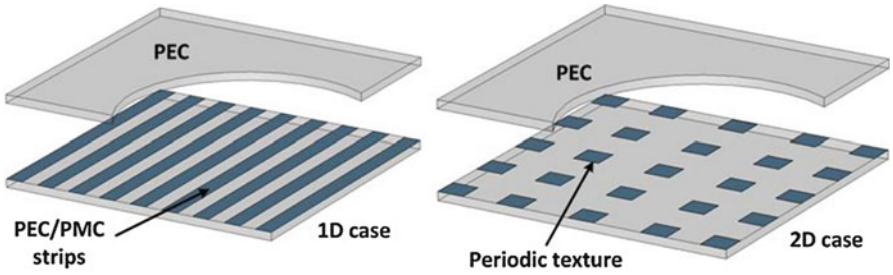


Fig. 16 Illustration of the 1D and 2D periodic-type structure (With permission from Pucci et al. 2014, Zaman and Kildal 2014, and Zaman et al. 2014)

$$\underset{\sim}{\tilde{G}}_{xz}^{HJ} = \tilde{H}_x = j\frac{k_y}{k_o^2 - k_y^2}\left[-\frac{k_x^2 \sin(k_z z)}{k_z \cos(k_z h)} + k_z \frac{\cos(k_z z)}{\sin(k_z h)}\right] \tag{3a}$$

$$\underset{\sim}{\tilde{G}}_{yz}^{HJ} = \tilde{H}_y = j\frac{1}{k_o^2 - k_y^2}\left[\left(k_o^2 - k_y^2\right)\frac{k_x \sin(k_z z)}{k_z \cos(k_z h)}\right] \tag{3b}$$

For 2D case:

$$\underset{\sim}{\tilde{G}}_{xz}^{HJ} = \frac{1}{\beta^2 D_{sw}}\left[C_1 k_x\left(\frac{\cos(k_z z)}{\sin(k_z h)} + \frac{\sin(k_z z)}{\cos(k_z h)}\right) + C_2 k_y \frac{\cos(k_z z)}{\sin(k_z h)} + C_3 \frac{k_y}{k_z}\frac{\sin(k_z z)}{\sin(k_z h)}\right] \tag{3c}$$

$$\underset{\sim}{\tilde{G}}_{yz}^{HJ} = \frac{1}{\beta^2 D_{sw}}\left[C_1 k_y\left(\frac{\cos(k_z z)}{\sin(k_z h)} + \frac{\sin(k_z z)}{\cos(k_z h)}\right) - C_2 k_x \frac{\cos(k_z z)}{\sin(k_z h)} - C_3 \frac{k_x}{k_z}\frac{\sin(k_z z)}{\sin(k_z h)}\right] \tag{3d}$$

where $C_1 = jk_x k_y k_z\left(\tilde{Y}_{yx} - \tilde{Y}_{xy}\right)$; $C_2 = -\beta^2 \eta_o k_o \tan(k_z h)\tilde{Y}_{xy}\tilde{Y}_{yx} - jk_z\left(k_x^2 \tilde{Y}_{yx} + k_y^2 \tilde{Y}_{xy}\right)$

$$C_3 = jk_o^2 \tan(k_z h)\left(k_x^2 \tilde{Y}_{xy} + k_y^2 \tilde{Y}_{yx}\right) - k_z \beta^2 \text{ and}$$

$$D_{sw} = -\tilde{Y}_{yx}\left(k_o^2 - k_x^2\right) - \tilde{Y}_{xy}\left(k_o^2 - k_y^2\right) + jk_o k_z\left[\eta_o \tan(k_z h)\tilde{Y}_{xy}\tilde{Y}_{yx} - \frac{1}{\eta_o}\cot(k_z h)\right]$$

Bed of Nail Structure

The bed of nail or the periodic pin structure is a favorable surface for use in gap waveguides as it can be made of metal only, and this structure can be easily milled. The topology of the surface is shown in Fig. 17. The homogenization method for such a structure is based on modeling the pin structure as uniaxial medium whose permittivity may be characterized by the permittivity tensor (Silveirinha et al. 2008):

$$\underset{\approx}{\varepsilon} = \varepsilon_o \varepsilon_r(\hat{x}\hat{x} + \hat{y}\hat{y} + \varepsilon_{zz}(\lambda, k_z, p, r)\hat{z}\hat{z}) \tag{4}$$

where p and r are the periodicity and radius of the pins, respectively. Using the TE-TM decomposition of the electric point source and by imposing the correct boundary conditions at all interfaces, the reflection coefficients for both polarizations can be derived (Silveirinha et al. 2008; Bosiljevac et al. 2010):

DOI 10.1007/978-981-4560-75-7_130-1

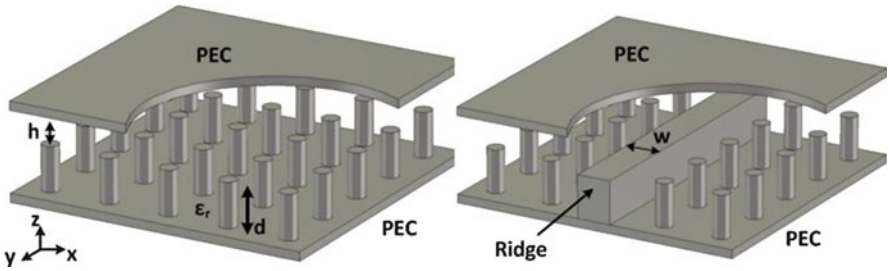


Fig. 17 Illustration of the bed of nail geometry and ridge gap waveguide with a top PEC plate (With permission from Zaman and Kildal 2014)

$$\Gamma^{TM} = -\frac{k_{\text{die}}k_p^2 \tan\,(k_h d) - \beta^2\gamma_{TM}\tanh\,(\gamma_{TM}d) + \varepsilon\gamma_o\left(k_p^2+\beta^2\right)}{k_{\text{die}}k_p^2 \tan\,(k_h d) - \beta^2\gamma_{TM}\tanh\,(\gamma_{TM}d) - \varepsilon\gamma_o\left(k_p^2+\beta^2\right)} \tag{5}$$

where k_{die} is the wavenumber in the metallic pin medium, k_p is the plasma wavenumber of the given pin lattice, $\beta^2 = {k_x}^2 + {k_y}^2$, ${\gamma_{\text{TM}}}^2 = {k_p}^2 + \beta^2 - {k_{\text{die}}}^2$, and ${\gamma_{\text{o}}}^2 = \beta^2 - {k_o}^2$.

For TE case:

$$\Gamma^{TE} = -\frac{\sqrt{k_h^2-\beta^2} - j\sqrt{k_o^2-\beta^2}\tan\left(d\sqrt{k_{\text{die}}^2-\beta^2}\right)}{\sqrt{k_h^2-\beta^2} + j\sqrt{k_o^2-\beta^2}\tan\left(d\sqrt{k_{\text{die}}^2-\beta^2}\right)} \tag{6}$$

After obtaining the reflection coefficient, the surface admittances of the pin structure are found as follows:

$$\tilde{Y}_{\text{surf}}^{TM} = \frac{k_o}{\eta_o k_z}\cdot\frac{\Gamma^{TM}+1}{\Gamma^{TM}-1}; \qquad \tilde{Y}_{\text{surf}}^{TE} = \frac{k_z}{\eta_o k_o}\cdot\frac{\Gamma^{TE}+1}{\Gamma^{TE}-1} \tag{7}$$

Once the surface admittances are derived, the spectral domain Green's functions can be determined by using the following boundary conditions:

$$\tilde{E}_t = 0 \quad \text{at the upper boundary} \quad \tilde{H}_y = \tilde{Y}_{\text{surf}}.\tilde{E}_x \tag{8a}$$

$$\tilde{H}_x = -\tilde{Y}_{\text{surf}}.\tilde{E}_y + \tilde{J}_y \qquad \text{at the lower boundary} \tag{8b}$$

The ridge has been introduced into the analysis via infinite transmission line current with propagation constant k_{eff}. The transmission line current is approximated as $J_y(x')e^{-jk_{\text{eff}}y'}$. The simplest assumption for $J_y(x')$ is to assume that this current is entirely longitudinal with a current distribution almost constant over the width of the ridge. The propagation constant of the quasi-TEM mode can be determined by forcing the suitable boundary condition at the ridge surface. The computed propagation constant is compared to the result obtained by commercially available CST microwave studio solver and is shown in Fig. 18a. The *Hy* component computed 0.5 mm above the ridge surface is also compared with the results obtained from CST microwave studio and is presented in Fig. 18b.

DOI 10.1007/978-981-4560-75-7_130-1

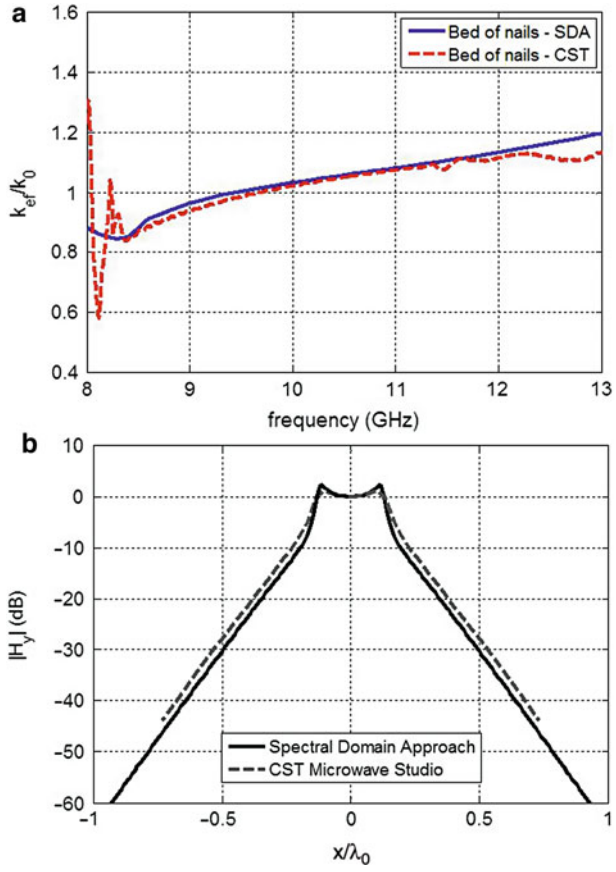


Fig. 18 (**a**) Propagation constant of fundamental mode of ridge gap waveguide and (**b**) H_y field distribution showing the lateral decay; both computed by spectral domain approach (SDA) and CST microwave studio

Modal Field Representation and Dispersion Equation for Ridge Gap Waveguide

In this approach, field expressions in the region above the ridge and the two side regions above the surrounding pin surfaces are used to establish the field continuity and match the modes at the interfaces among the regions. The main purpose of this approach is also to model the dispersion characteristics of the quasi-TEM mode in the ride gap waveguide geometry shown in Fig. 17. The fields associated with the quasi-TEM mode are mainly transverse to the direction of propagation, but small longitudinal components of the fields are also anticipated. The field is expected to have an exponential decay laterally away from the ridge into the side regions above the pins.

It is expected that a quasi-TEM modal field propagates along the ridge in the frequency band when the bed of nails acts as an AMC and a parallel-plate stopband is achieved. This mode will match the evanescent modes supported by the surrounding cutoff structure. In the pin region, only the first TM and TE modes (with respect to z) are considered to be present. However, to achieve the field confinement above the ridge region, the evanescent modes must decay laterally away from the ridge along direction x. Under the approximations used above, in the central region above the ridge, the assumed mode is the quasi-TEM mode which is propagating along the y axis. The fields for this quasi-TEM mode can be presented as follows (Polemi et al. 2011):

DOI 10.1007/978-981-4560-75-7_130-1

$$\left.\begin{aligned} E_z &= E_o \cos\left(\hat{k}_x x\right) e^{-jk_y y} \\ H_x &= -E_o \frac{k_y}{\xi k} \cos\left(\hat{k}_x x\right) e^{-jk_y y} \\ H_y &= -jE_o \frac{\hat{k}_x}{\xi k} \sin\left(\hat{k}_x x\right) e^{-jk_y y} \end{aligned}\right\} \tag{9}$$

where E_o is related to the incident power at the input port and $\hat{k}_x$ is the propagation constant along x.

The TM_z evanescent mode fields in the region above the metal pins can also be written as follows (Das and Pozar 1987):

$$\left.\begin{aligned} H_x &= jA_{TM} \frac{k_y}{\sqrt{k^2 - \tilde{k}_z^2}} \tilde{g}(x,y) \times \cos\left[\tilde{k}_z(z-h)\right] \\ H_y &= -A_{TM} \frac{\tilde{\alpha}_x}{\sqrt{k^2 - \tilde{k}_z^2}} \tilde{g}(x,y) \times \cos\left[\tilde{k}_z(z-h)\right] \\ E_x &= -jA_{TM} \frac{\xi}{k} \frac{\tilde{\alpha}_x \tilde{k_z}}{\sqrt{k^2 - \tilde{k}_z^2}} \tilde{g}(x,y) \times \sin\left[\tilde{k}_z(z-h)\right] \\ E_z &= -jA_{TM} \frac{\xi}{k} \sqrt{k^2 - \tilde{k}_z^2}.\tilde{g}(x,y) \times \cos\left[\tilde{k}_z(z-h)\right] \\ E_y &= A_{TM} \frac{\xi}{k} \frac{k_y \tilde{k_z}}{\sqrt{k^2 - \tilde{k}_z^2}} \tilde{g}(x,y) \times \sin\left[\tilde{k}_z(z-h)\right] \end{aligned}\right\} \tag{10}$$

where $\tilde{g}(x,y) = e^{-jk_y y} e^{-\tilde{\alpha}_x(|x|-w/2)}$;

$$k^2 = (-j\tilde{\alpha}_x)^2 + \tilde{k}_z^2 + k_y \tag{11}$$

In Eq. 10, A_{TM} is an unknown coefficient, while Eq. 11 is the dispersion relation for this mode where attenuation constant $\tilde{\alpha}_x$ and propagation constant k_y are also unknown. Apart from the TM_z mode, a TE_z mode is also assumed to be excited in the pin region, and the field equation for this mode is presented in Eq. 12. In Eq. 12, A_{TE} is an unknown coefficient, and the dispersion relation for this TE mode is shown in Eq. 13:

$$\left.\begin{aligned}
E_x &= jA_{TE}\frac{k_y}{\sqrt{k^2-\tilde{\tilde{k}}_z^2}}\tilde{\tilde{g}}(x,y)\times\ \sin\left[\tilde{\tilde{k}}_z(z-h)\right]\\
E_y &= -A_{TE}\frac{\tilde{\tilde{\alpha}}_x}{\sqrt{k^2-\tilde{\tilde{k}}_z^2}}\tilde{\tilde{g}}(x,y)\times\ \sin\left[\tilde{\tilde{k}}_z(z-h)\right]\\
H_x &= -jA_{TE}\frac{1}{\xi k}\frac{\tilde{\tilde{\alpha}}_x\tilde{\tilde{k}}_z}{\sqrt{k^2-\tilde{\tilde{k}}_z^2}}\tilde{\tilde{g}}(x,y)\times\ \cos\left[\tilde{\tilde{k}}_z(z-h)\right]\\
H_z &= jA_{TE}\frac{1}{\xi k}\sqrt{k^2-\tilde{\tilde{k}}_z^2}.\tilde{\tilde{g}}(x,y)\times\ \sin\left[\tilde{\tilde{k}}_z(z-h)\right]\\
H_y &= A_{TE}\frac{1}{\xi k}\frac{k_y\tilde{\tilde{k}}_z}{\sqrt{k^2-\tilde{\tilde{k}}_z^2}}.\tilde{\tilde{g}}(x,y)\times\ \cos\left[\tilde{\tilde{k}}_z(z-h)\right]
\end{aligned}\right\} \tag{12}$$

where $\tilde{\tilde{g}}(x,y)=e^{-jk_yy}e_x^{-\tilde{\tilde{\alpha}}}(|x|-w/2)$;

$$k^2=\left(-j\tilde{\tilde{\alpha}}_x\right)^2+\tilde{\tilde{k}}_z^2+k_y \tag{13}$$

The overall unknowns in the abovementioned equations are six, and these are A_{TM}, A_{TE}, $\tilde{\alpha}_x$, $\tilde{\tilde{\alpha}}_x$, $\hat{k}_x$, and k_y. So, six equations are needed to solve the problem. Three of them are the dispersion relations in the three regions mentioned above and shown below:

$$\left.\begin{aligned}
k_y^2 &= k^2-\tilde{k}_z^2+\tilde{\alpha}_x^2\\
k_y^2 &= k^2-\tilde{\tilde{k}}_z^2+\tilde{\tilde{\alpha}}_x^2\\
k_y^2 &= k^2-\hat{k}_x^2
\end{aligned}\right\} \tag{14}$$

To find the remaining equations, continuity of three field components across different regions is enforced, and the final dispersion equation for the ridge gap waveguide is obtained as follows:

$$\sqrt{k^2-k_y^2}\ \tan\left[\sqrt{k^2-k_y^2}\ w/2\right]\left(k^2-\tilde{k}_z^2\right)+\frac{k_y^2\tilde{k}_z^2-k^2\sqrt{k_y^2-k^2+\tilde{k}_z^2}\sqrt{k_y^2-k^2+\tilde{\tilde{k}}_z^2}}{\sqrt{k_y^2-k^2+\tilde{\tilde{k}}_z^2}}=0$$

Comparison Between the Modal Field Representation and Spectral Domain Green's Function

A comparison study between the two approaches mentioned above has been done in Bosiljevac et al. (2011). Both the ridge gap waveguide and groove gap waveguide have been studied analytically using the two approaches. Results obtained using both the approaches are compared with results from commercial solver CST microwave studio. Very good agreements have been observed among the results

DOI 10.1007/978-981-4560-75-7_130-1

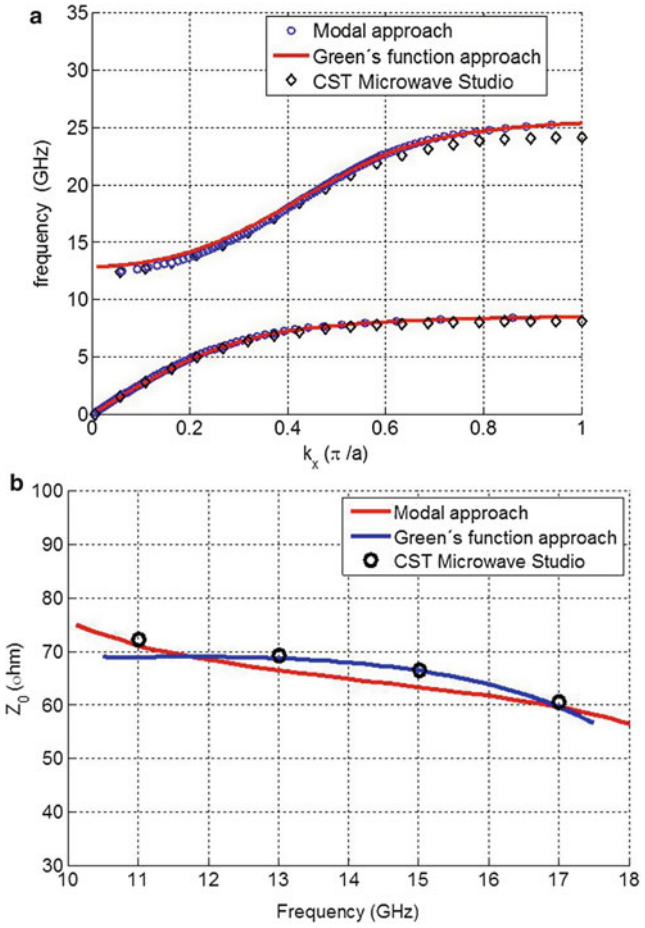


Fig. 19 (**a**) Dispersion diagram obtained by both methods and CST for ridge gap waveguide and (**b**) characteristic impedance calculated by both methods and CST for the same test geometry

obtained in most of the cases. In this section, some results regarding the comparison work are presented only for the ridge gap waveguide geometry. The ridge gap waveguide test geometry used for comparison has the following physical dimensions: pin heights 7.5 mm, air gap height 1 mm, pin size 1×1 mm^2, periodicity 2 mm, and ridge width 5 mm. The dispersion diagrams and characteristic impedances for this test geometry obtained by a different analysis approach are shown in Fig. 19a and 19b.

Numerical Analysis for Gap Waveguide Structures Based on CST and HFSS

Apart from developing rigorous analytical formulations for gap waveguide geometries, there is always a need to determine the correctness of using commercially available EM solvers for simulating gap waveguide components. There exists geometrical difference between the hollow waveguides and gap waveguide structures due to the periodicity of the textured surface. This section validates the use of commercial EM solvers like HFSS or CST, even for such periodic structures. By checking the similarities among hollow waveguides and gap waveguides, it is possible to come to a conclusion that commercial EM solvers can handle the gap waveguide structure with a certain level of accuracy (Raza et al. 2013).

First, the dispersion diagrams of the ridge/rectangular waveguides and the ridge/groove gap waveguides are obtained by using the eigenmode solver in CST microwave studio, where the periodic structure is assumed infinitely long in the direction of propagation. In this section, the material used for the simulation is the perfect electric conductor with zero surface roughness, and the geometry is shown in Fig. 20.

DOI 10.1007/978-981-4560-75-7_130-1

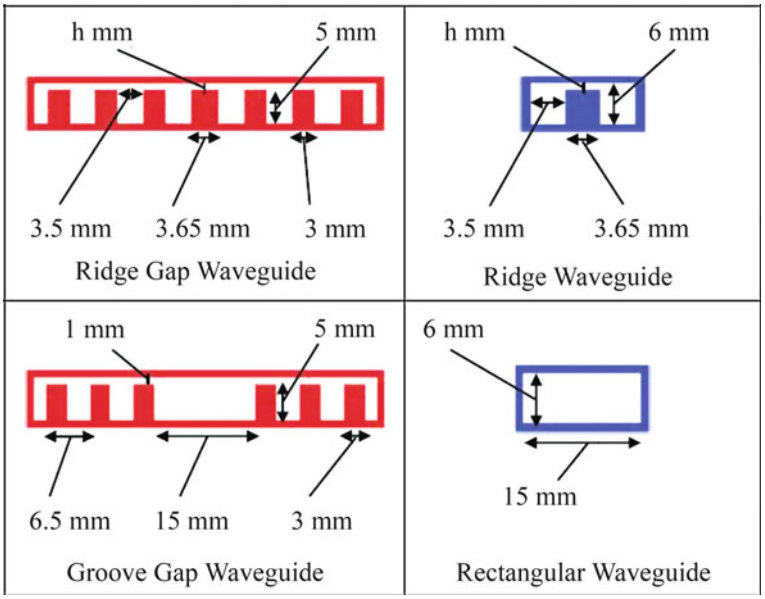


Fig. 20 Details of the geometries investigated for dispersion analysis (With permission from Pucci et al. 2014)

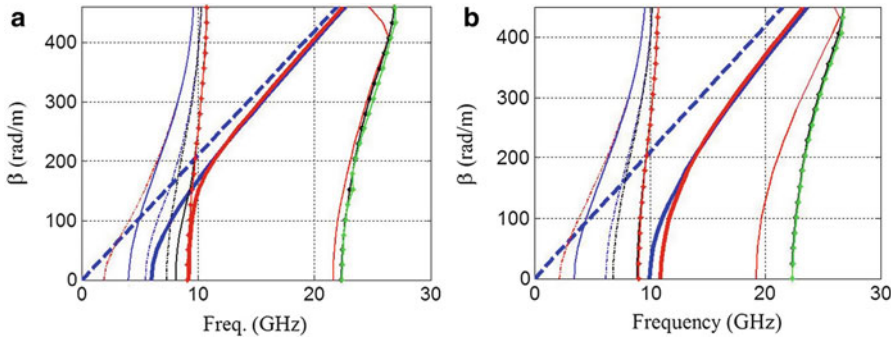


Fig. 21 (**a**) Dispersion diagram of ridge gap waveguide and hollow ridge waveguide (*dashed line*), (**b**) dispersion diagram of groove gap waveguide and hollow rectangular waveguide (*dashed line*)

The dispersion diagrams are shown in Fig. 21a and 21b, respectively, for hollow ridge/ridge gap waveguide and hollow rectangular/groove gap waveguide. From the simulated dispersion diagrams of the ridge gap waveguide and its equivalent ridge waveguide, it is seen that ridge gap waveguide has a dispersion curve (referred to as the ridge gap dispersion curve) very close to that of a TEM mode (the light line). More importantly, within the stopband, the ridge gap dispersion curve is very close to the dispersion curve of the fundamental mode in the equivalent conventional ridge waveguides. Similarly, when compared the dispersion diagrams between a groove gap waveguide and its equivalent hollow rectangular waveguide, the dispersion diagrams of the fundamental propagating modes in both the waveguides are similar within the stopband. In summary, the dispersion diagram of a ridge/groove gap waveguide can be approximated by that of its equivalent ridge/hollow waveguide, which can be easily obtained by either analytical expressions or empirical formulas.

Port Configuration for Simulating Gap Waveguide Components

A numerical port configuration has been studied in the situation where numerical ports are directly attached to the ridge/groove gap waveguide. The extensive study and obtained results are presented

DOI 10.1007/978-981-4560-75-7_130-1

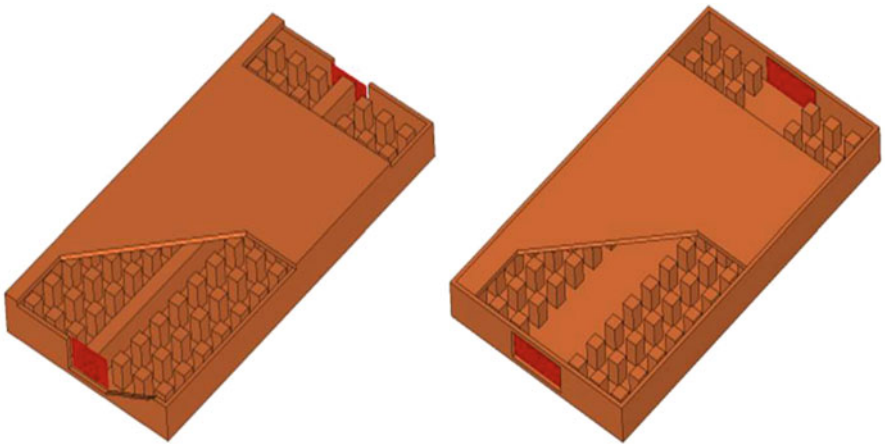


Fig. 22 Port configuration for simulating gap waveguide structures

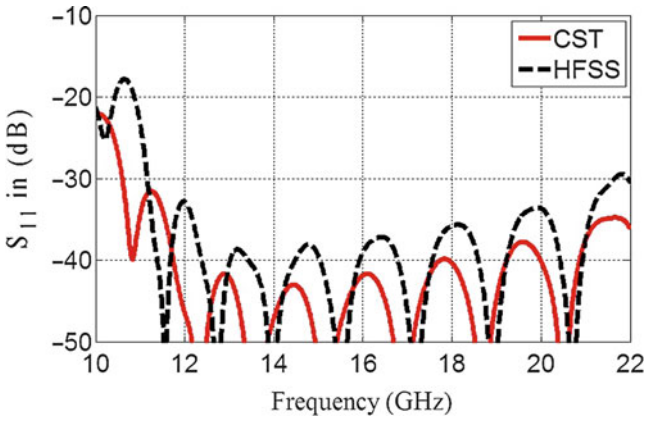


Fig. 23 S-parameters of the groove gap waveguide equivalent hollow waveguide ports in CST and HFSS

elaborately in Raza et al. (2013). However, here in this section, some results are presented only to show the level of port accuracy obtained by commercial EM solvers when the ports are directly placed at the gap waveguide structures as shown in Fig. 22. Ideally, with correct numerical ports, the S_{11} and S_{22} for such straight two-port waveguide would be quite low (in the order of −65 dB). But in this case, the hollow waveguide ports available in commercial EM solver, being the imperfect numerical ports, will give relatively high S_{11} and S_{22} and will induce errors in gap waveguide simulations. However, the level of S_{11} and S_{22} obtained both for ridge gap waveguide and groove gap waveguide is lower than −35 dB within the band of interest. This is shown in Figs. 23 and 24. As shown in the two plots, the matching between the gap waveguide and the equivalent waveguide port is still good, except at the lower end of the stopband. The periodic nature of the reflection coefficient comes from the two interfering reflections, from the ports at both ends of the waveguides. This means that the reflection coefficient from each port interface is 6 dB lower than the peaks of the combined reflection coefficient S_{11}. Thus, it is easy to conclude that the discontinuity effect at the numerical port over most of the stopband is quite small from practical viewpoint, and level wise it is in the same order of magnitude as that of measurement accuracy.

DOI 10.1007/978-981-4560-75-7_130-1

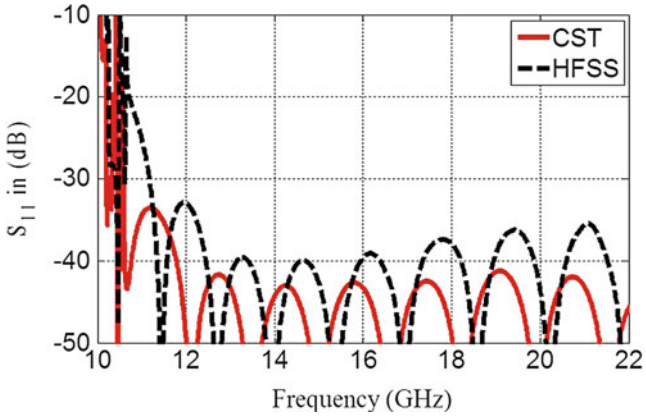


Fig. 24 S-parameters of the ridge gap waveguide equivalent hollow waveguide ports in CST and HFSS

Low-Loss Antennas Based on Gap Waveguide Technology

Wireless communication systems have evolved from cellular telephony with data rates of kilobits per second (kbps) to wireless local area networks (WLANs) and wireless personal area networks (WPANs) that communicate with megabits per second (Mbps). Also, with the advent of gigabit Wi-Fi and more and more commercial applications with higher data rate (Hansen 2011; Smulders 2002), high-gain and large-aperture antennas are becoming more and more important from system perspective.

Although high antenna efficiency can be obtained by using dielectric lens antennas or reflector antennas (Chantraine-Bares and Sauleau 2007; Mall and Waterhouse 2001; Wei et al. 2011), it is difficult to realize planar antennas with such configurations, as they essentially need focal spatial length. Microstrip array antennas can be lightweight, low cost, and low profile, but they suffer from high dielectric losses and ohmic losses at high-frequency bands (Borji et al. 2009). Also, the electrical performances of microstrip antennas depend on the substrate to be used. As a result, it is very difficult to optimize important antenna parameters such as bandwidth and efficiency simultaneously in case of microstrip antennas (Zwick et al. 2006). Furthermore, microstrip antennas are associated with problems such as leakage via surface waves and undesired radiation (Fonseca and Giarola 1984; Komanduri et al. 2013; Levine et al. 1989). Hence, realization of a high-efficiency high-gain microstrip antenna array is quite challenging especially at mmWave frequencies and above.

Slotted hollow waveguide planar array antennas are free from large feed network loss and can be applied to design high-efficiency high-gain or moderate-gain antennas (Hirokawa and Ando 2000; Hirokawa et al. 1992). However, the production cost of waveguide antennas is generally very high because they usually consist of metal blocks with complicated three-dimensional structures (Liu et al. 2009). Also, high-precision manufacturing techniques are needed to achieve good electrical contacts between the slotted metal plate and the bottom feed structure in such waveguide-fed slot array antennas. Consequently, conventional waveguide slot arrays have not been used commercially till today with the exception of few military applications and space applications.

Apart from the conventional microstrip antennas and waveguide slot antennas, substrate integrated waveguide (SIW)-based planar array antennas are also proposed to realize low-cost solutions (Xiao-Ping et al. 2010; Xu et al. 2009). This technology enables integration of active circuits together with the antennas. The losses in SIW are better than microstrip and coplanar structures. Still, losses may be of real concern, especially for high-gain (above 28 $\sim$ 30 dBi) configuration, due to the presence of dielectric material (Bozzi et al. 2011; Awida et al. 2011).

DOI 10.1007/978-981-4560-75-7_130-1

Based on the above discussions, it is implicit that existing antenna technologies have limitations with respect to manufacturing cost, efficiency, bandwidth, and mechanical simplicity, giving an opportunity for more research work on new technologies such as gap waveguide. As mentioned in section "Operating Principle of Gap Waveguides and Loss Analysis," different gap waveguide configurations can exist in practice, giving a unique opportunity to use multiple gap waveguide configurations to build antennas with specific performance requirement and cost. In this section, several antenna works based on gap waveguide technology have been described, and it is shown how different configurations of gap waveguide can be used in designing antennas.

Ridge Gap Waveguide Single-Layer Wideband Slot Array Antenna

In this section, a wideband 4 × 1 element linear slot array and a 2 × 2 element planar slot array designed at Ku-band using ridge gap waveguide technology are demonstrated. Both these antennas are designed for a fixed broadside beam and a 20 % relative bandwidth. The low-loss feed network needed for a high-gain array antenna can be built very easily with the help of ridge gap waveguide concept. In addition, radiating slots can be placed conveniently on the top smooth metal plate of ridge gap waveguide. It is possible to design slot antenna without having strict requirements of good electrical contacts between the slotted metal plate and the bottom feed structure. The details of the ridge gap slot array antenna work are published in Zaman and Kildal (2014).

The initial single slot antenna excited by ridge gap waveguide is shown in Fig. 25a. This single element is quite narrowband as is observed in simulated S_{11} results. Later on, this design has been improved with addition of a T-section ridge (Zaman and Kildal 2012, 2014). The length and the width of the slot element are also varied to have the S_{11} below −10 dB over a significant bandwidth. The improved element design and obtained reflection coefficient for the element is shown in Fig. 25b. The largest bandwidth is obtained when the first resonance from the slot itself and the second resonance from the feeding T-section appeared close to each other. These two resonance dips are clearly seen in the simulated S_{11} plots in Fig. 25b.

Another important component for antenna corporate feed network is the 3-dB power divider in the form of a T-junction. The T-junction usually consists of a quarter-wave transformer section and three 50-Ω lines. The ridge waveguide T-junction power divider and its simulated S-parameters are shown in Fig. 25c. The simulation results show that it is possible to design a wideband feed network for a complete array antenna based on this T-junction.

Linear Array Design

Based on the abovementioned slot element and the 3-dB power divider, a 4 × 1 linear slot array has been designed. At this point, the linear array is designed for a fixed beam in mind. So, the element spacing can be kept 0.80λ to avoid grating lobes, and all the slot elements are excited with equal amplitude and phase. The complete array is shown in Fig. 26a, and it is designed to operate between 11.5- and 14.5-GHz frequency bands. As shown in Fig. 26a, the bottom plate with ridge and texture pin structure holds the feeding network for the array. On top is placed a smooth metal plate with four radiating slots at a distance of 1 mm from the bottom plate. The dimension of the slot element is chosen in such a way that the slot length to width ratio S_w/S_L is kept smaller than 0.5. This ratio is important for the suppression of the cross-polarization. The thickness of the top metal plate is chosen to be 2 mm. A transition is needed from ridge gap waveguide to the standard transmission line for measurement of the entire antenna. This transition can be a simple interface consisting of steps in ridge which allows the quasi-TEM mode to couple to the coaxial mode of SMA connector with extended inner conductor. The measured results for this linear array are presented in Fig. 26b and 26c. As shown in the measured results, the measured S_{11} parameter is in good agreement with the simulated value, and the relative bandwidth is about 20 %. In case of the radiation patterns, the array has more directive beam and symmetric pattern in H-plane as expected. In

DOI 10.1007/978-981-4560-75-7_130-1

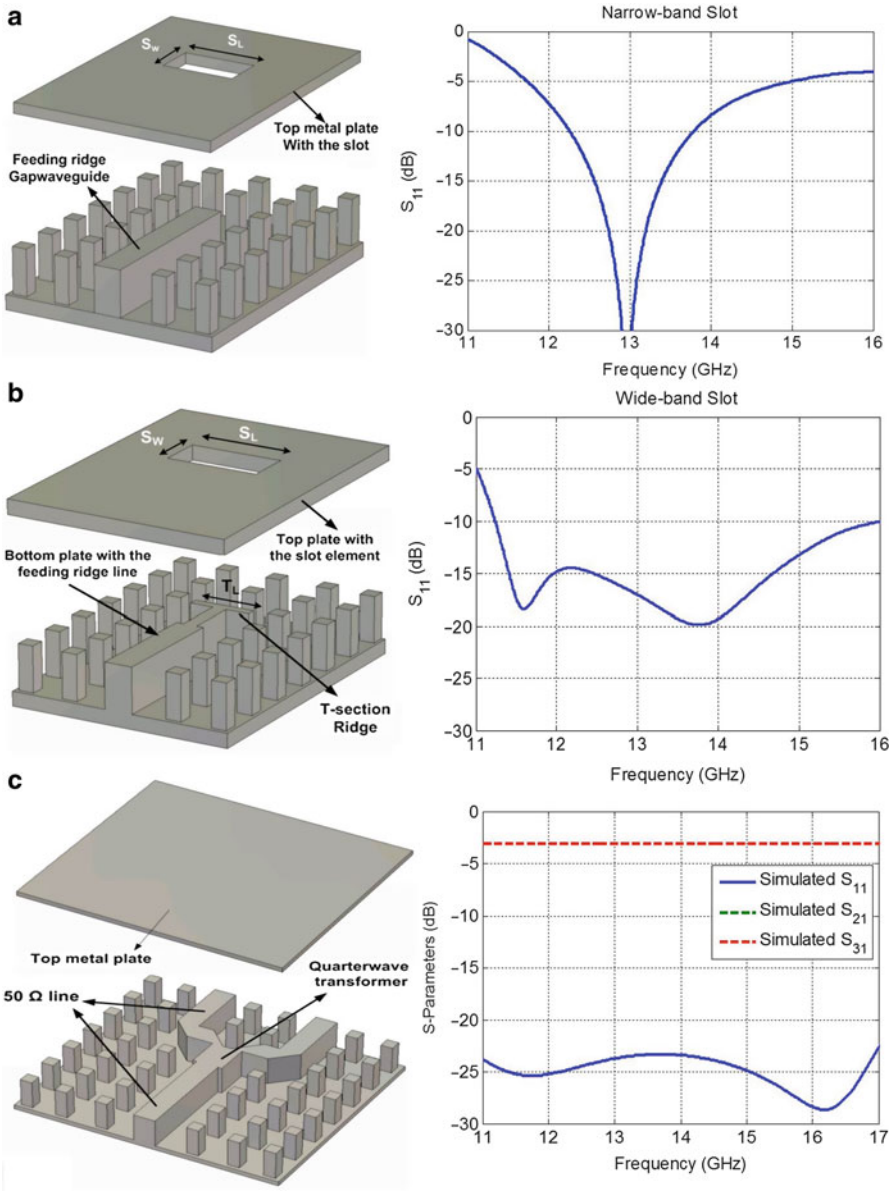


Fig. 25 (**a**) Narrowband slot with dimension $S_L = 11.5$ mm and $S_W = 5.85$ mm and its simulated reflection coefficient. (**b**) Wideband slot with T-section and its simulated reflection coefficient; $S_L = 11.75$ mm, $S_W = 5.85$ mm, $T_L = 8.25$ mm. (**c**) Simulated S-parameters for ridge gap waveguide T-junction

H-plane, the first sidelobe is 11.8 dB lower from the beam peak value which is also common for uniform arrays. On the other hand, in E-plane, the beam is wider and is not really symmetric as the slots are close to the edge on one side. In E-plane, there is a possibility of back radiation as the slots are close to the edge. To avoid this back radiation, few corrugations with quarter-wavelength depth at the lowest frequency of operation have been added.

Two-Dimensional Array Design

After designing the linear array, a simple 2 × 2 element planar array has been designed with ridge gap waveguide technology to operate over the band 12–15 GHz. The 2 × 2 element array is excited in phase and with equal amplitude by a corporate feeding. The element spacing is about 0.875λ at highest frequency of interest. The slot length S_L and slot width S_W are chosen to be 11.45 and 6.25 mm for this

DOI 10.1007/978-981-4560-75-7_130-1

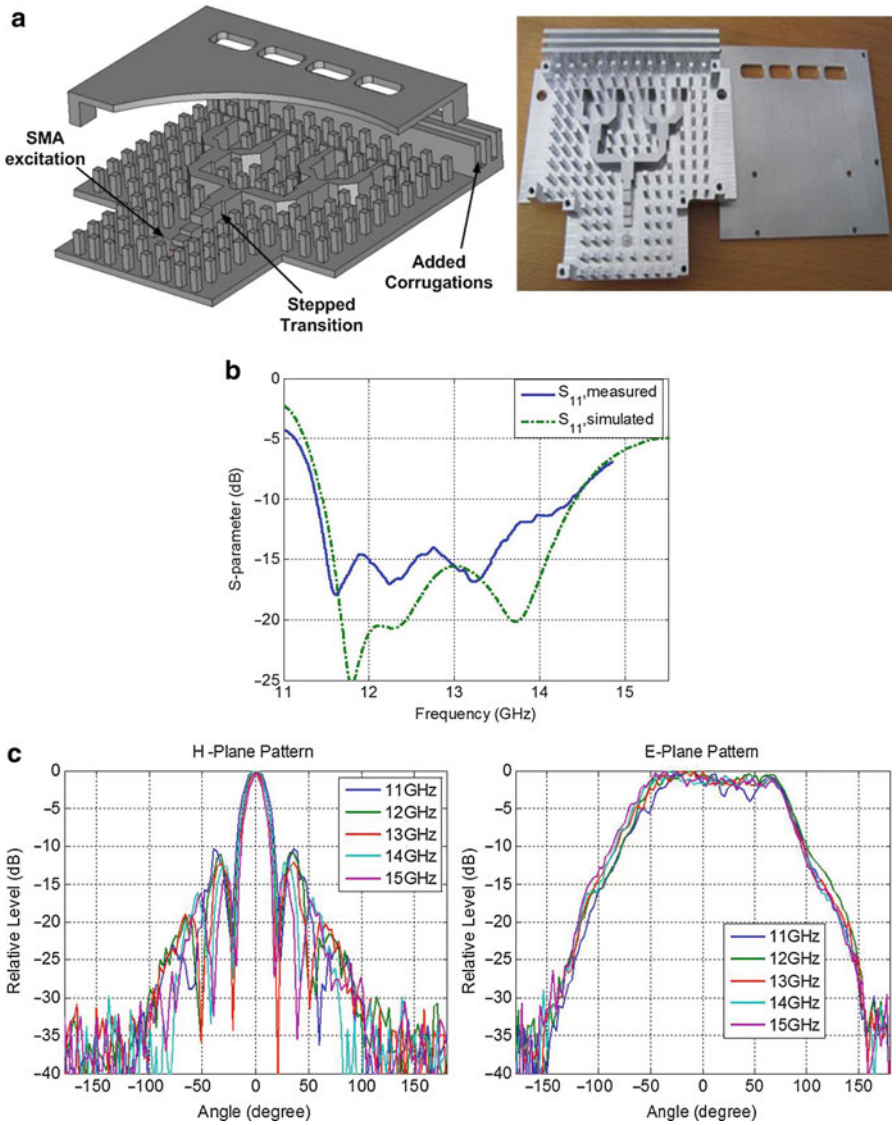


Fig. 26 (**a**) CST model of the linear array and manufactured array. (**b**) Measured and simulated S_{11} for the linear array. (**c**) Measured E- and H-plane patterns for the linear array

particular design. The antenna is excited with a standard Ku-band rectangular waveguide and is shown in Fig. 27a. In a two-dimensional array, it is more difficult to design the corporate feed network because the element spacing must be smaller than one wavelength in two orthogonal directions. For this reason, the feed network must be much more compact. So, in this design, the compact four-way power divider is designed without using a λ/4 transformer section. Instead, the widths of the two divided ridge sections are gradually tapered to make the impedance match (Zaman and Kildal 2014). Fig. 27b and 27c presents the measured reflection and measured principal plane radiation patterns, respectively. For the 2×2 array, the sidelobe is below -12.5 dB in H-plane for all the measured frequency points. In E-plane, the sidelobe levels remain below -10 dB at the lower frequencies, but they rise up to a level of -7.65 dB at the upper end of the frequency band. These high sidelobes are caused by the omnidirectional pattern of each individual slot in E-plane and the large element spacing (close to λ at higher frequency). Also, the array has an extent limited to only two elements in each direction, so we do not have a clear effect of the array factor. For a larger array, this first sidelobe will appear closer to the main lobe and will become lower.

DOI 10.1007/978-981-4560-75-7_130-1

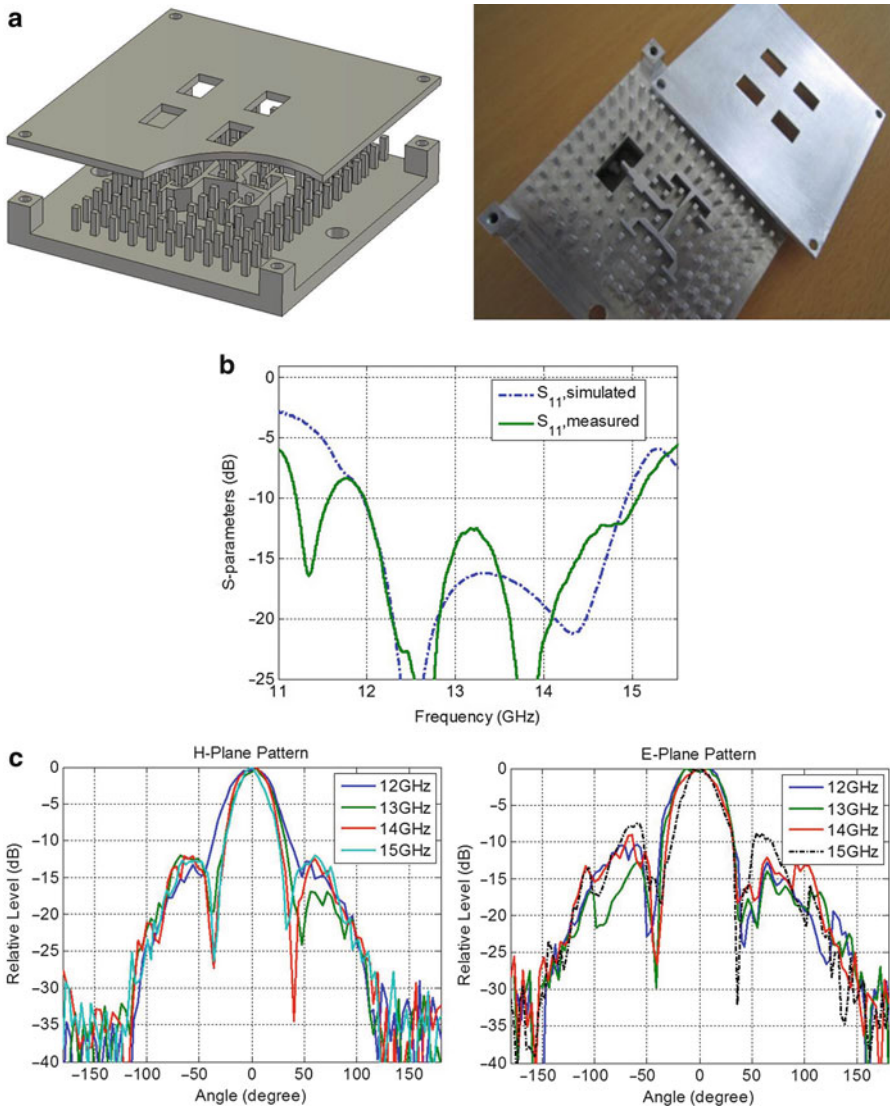


Fig. 27 (**a**) CST model of the 2 × 2 element array and manufactured array. (**b**) Measured and simulated S_{11} for the planar array. (**c**) Measured E- and H-plane patterns for the planar array

Groove Gap Waveguide Slot Arrays

As mentioned in section "Operating Principle of Gap Waveguides and Loss Analysis," among all the gap waveguide geometries, the groove gap waveguide offers lowest insertion loss, and the loss of this waveguide is very similar to that of conventional rectangular waveguide. So, it is of interest to design slot arrays based on groove gap waveguide, and very recently some work has been done on groove gap waveguide slot arrays (Valero-Nogueira et al. 2013; Martinez Giner et al. 2013). In this section, two groove gap linear arrays having slots in broad wall and narrow wall have been discussed. The field distributions inside the groove gap waveguide for these two cases are shown in Fig. 28.

The approach in designing the groove gap waveguide slot array is similar to the conventional approach. First, it is necessary to obtain the equivalent admittance of the slot in groove gap waveguide. But this is not a trivial problem, as in this case the structure to be analyzed is a periodic waveguide having nonperiodic obstacles such as slots. As a result, the direct use of the Bloch-Floquet theorem is not applicable here. However, an indirect approach is proposed in Valero-Nogueira et al. (2013) to obtain the S_{11} parameter associated to a slot in a groove gap waveguide. The indirect derivation of the reflection coefficient is

DOI 10.1007/978-981-4560-75-7_130-1

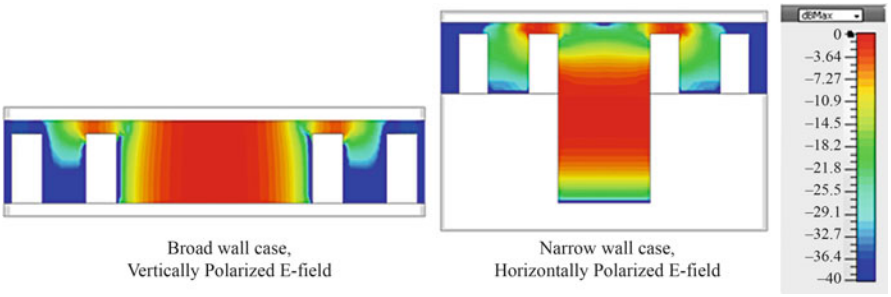


Fig. 28 The field distribution inside the groove gap waveguide for slots in broad wall or narrow wall

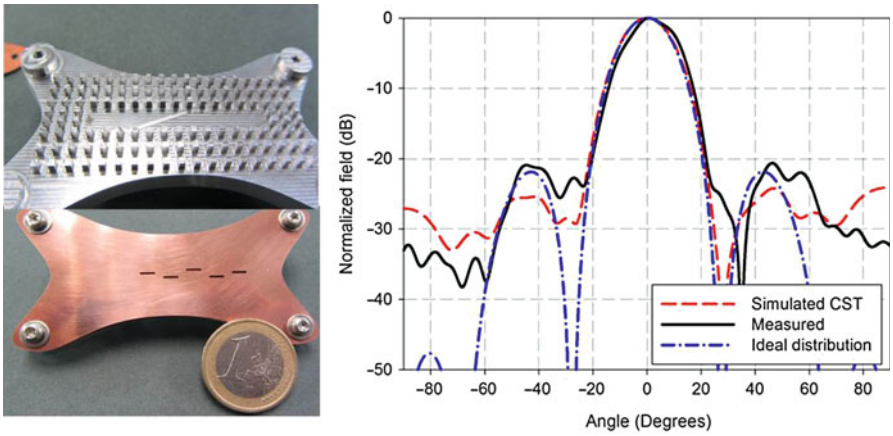


Fig. 29 Manufactured groove gap waveguide with slots in the broad wall and H-plane pattern for this slot array

performed starting from an aperture integral formulation applied to the slot. Using the approach mentioned in Valero-Nogueira et al. (2013), a small linear array having slots on the broad wall of groove gap waveguide has been designed at Ka-band. The manufactured slot array and measured radiation pattern for this slot array are shown in Fig. 29.

Apart from the slot array mentioned above, a resonant 8-element array has been designed in which the slots are in the narrow wall of the groove gap waveguide (Martinez Giner et al. 2013). Typically, when the slots are placed in the narrow wall, the slots are tilted or inclined. This degrades the cross-polar performance of the antenna. To avoid high cross-polar problem, the slots are kept untilted and a substrate etched with tilted parasitic dipoles has been used to excite the slots. This approach is similar to the approach mentioned in Hirokawa and Kildal (1997). The relative angle between the dipole and the slot determines the amount of coupling between these two elements. The schematic of the single element is shown in Fig. 30, and the picture of the manufactured array with radiation patterns is presented in Fig. 31a and 31b, respectively. As mentioned in Martinez Giner et al. (2013), the radiating elements in the array are positioned after every $\lambda_g/2$ position, at the point of each standing-wave maxima. The dipoles are also oriented in an alternating fashion to compensate for the alternate phase pattern of the standing-wave after every $\lambda_g/2$ length.

Horn Array Antenna Fed by Inverted Microstrip Gap Waveguide

This section presents the design of a 16-element quasi-planar dual-mode horn array fed by an inverted microstrip gap waveguide corporate distribution network. The final antenna is intended for 60-GHz application. But a scaled version of that antenna, a prototype working at X-band, is manufactured as a

DOI 10.1007/978-981-4560-75-7_130-1

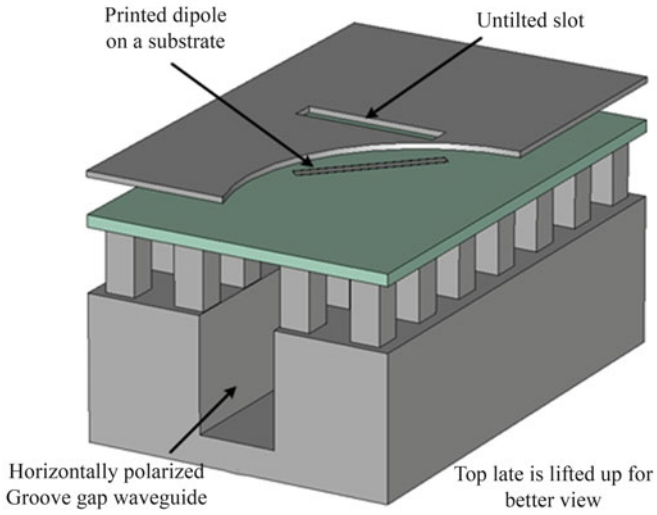


Fig. 30 Schematic of groove gap waveguide with untitled slot excited by parasitic dipole

proof of concept and is presented here. The detailed description of this work is presented in Pucci et al. (2014). The main objective of this work is to demonstrate inverted microstrip gap waveguide-based corporate feed network which works without unwanted leakage via substrate mode and distributes the power properly to a horn array. The emphasis has not been given to the design of the $2\lambda \times 2\lambda$ dual-mode horn element itself. The schematic of the single horn element is shown in Fig. 32. The single horn element is also excited by a T-shaped feed line, similar to the one used in Zaman and Kildal (2014). This makes it possible to achieve a larger bandwidth (over 10 %). In the complete array, horn elements are fed by a corporate feed network realized in microstrip gap waveguide. The feed network is designed by using conventional T-junction power dividers and quarter-wavelength impedance transformers. The input is provided from the central 50-Ω transmission line, which splits the power into two main branches, and then the power division continues until equal amount of power has traveled up to an individual radiator. The matching is obtained by using impedance transformers and by mitering the corner of the bends. The design details of the feed network are also mentioned in Pucci et al. (2014). The picture of the manufactured antenna with all the different parts is shown in Fig. 33.

For the manufactured 16-element planar dual-mode horn array fed by inverted microstrip gap waveguide distribution network, the realized gain is found to be around 24.5 dBi over the design frequency. The main loss in this antenna is found to be the power lost due to grating lobes. This particular antenna has inherently low aperture efficiency due to the high grating lobes originating from the size of the single horn element which is more than one wavelength. After calculating the grating lobe efficiency, it is found that $2 \sim 3$-dB power is lost due to high grating lobes over the frequency band of interest. Otherwise, the inverted microstrip gap waveguide-based feed network has relatively smaller losses, only in the order of $0.4 \sim 0.5$ dB. The radiation patterns and the realized gain plots are shown in Fig. 34a and 34b. However, the main advantage of this antenna lies in the fact that the substrate (on which the inverted microstrip gap waveguide feed network is printed) can be placed easily on top of the uniform bed of nail structure where the periodicity is constant. Such uniform pin structure can be milled with bigger milling tools, or several saw blades can be placed together to cut out such uniform block from bulk metal piece.

60-GHz Slot Array Element Using Two Double-Sided PCBs

In this section, a wideband 60-GHz planar slot array antenna is presented. The unit cell of the slot array is made of two double-sided PCB layers. Between the two PCBs, the bottom PCB is used for the corporate

DOI 10.1007/978-981-4560-75-7_130-1

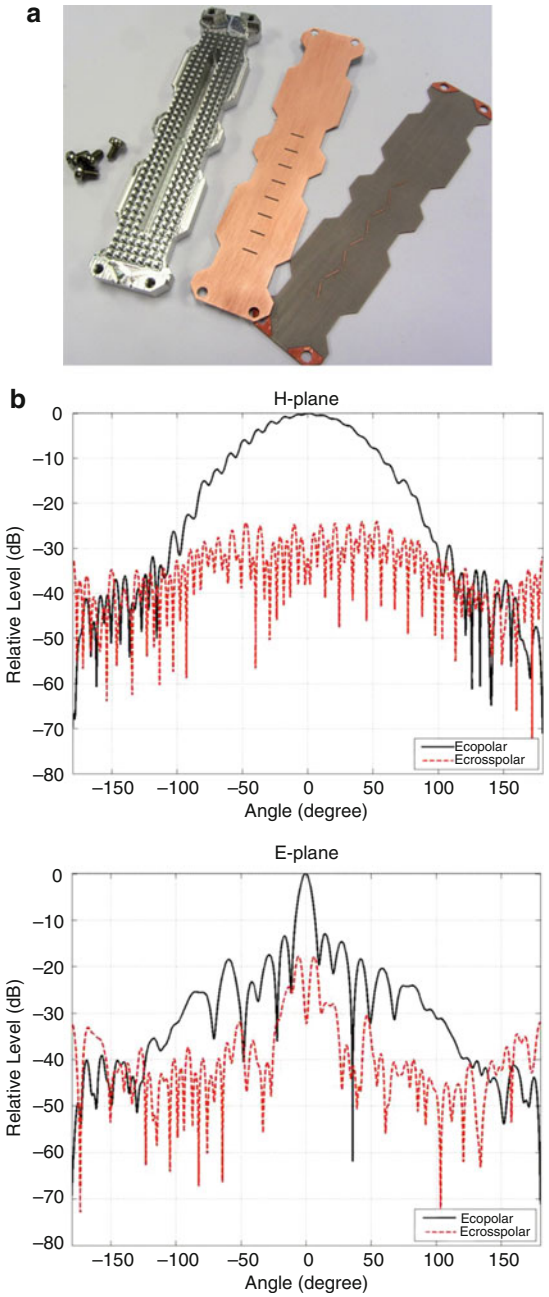


Fig. 31 (**a**) Manufactured groove gap slot array having slot in narrow wall. (**b**) Measured E- and H-plane radiation patterns for the slot array having slot in narrow wall

feed network of the antenna, and the feed network is based on microstrip-ridge gap waveguide. On the other hand, the upper PCB is a SIW cavity layer with internal walls. The ground plane of the upper PCB holds a coupling slot which is used to couple energy from microstrip-ridge gap waveguide feed network to the SIW cavity layer. After that, the energy is radiated from the SIW cavity to the air with the help of the radiating slots which are etched on the top metal layer of the upper PCB. There exists an air gap between the upper PCB and bottom PCB which is needed for the Q-TEM mode propagation along the microstrip-ridge gap waveguide, and this air gap is maintained to be 0.25 mm in this case. In this work, Rogers 5880 substrate ($\varepsilon_r = 2.2$) with thickness of 0.78 mm has been used as the upper PCB. On the other hand, Rogers 3003 substrate ($\varepsilon_r = 3.0$) with a thickness of 0.504 mm has been used as the bottom PCB. The schematic

DOI 10.1007/978-981-4560-75-7_130-1

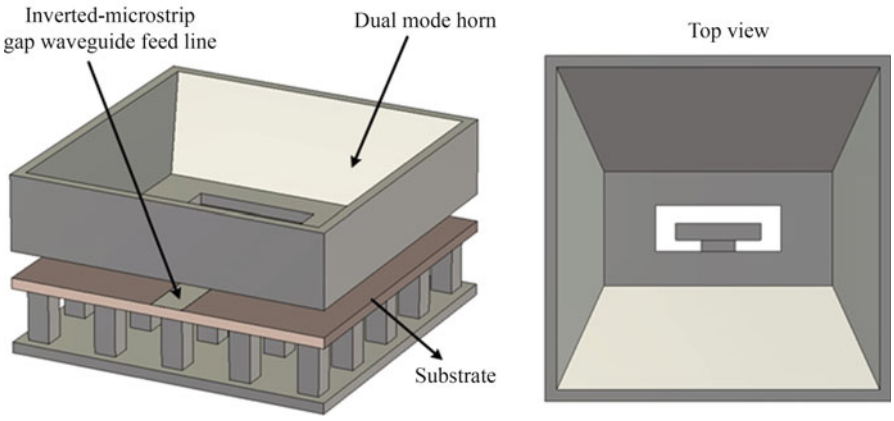


Fig. 32 Dual-mode horn element fed by inverted microstrip gap waveguide

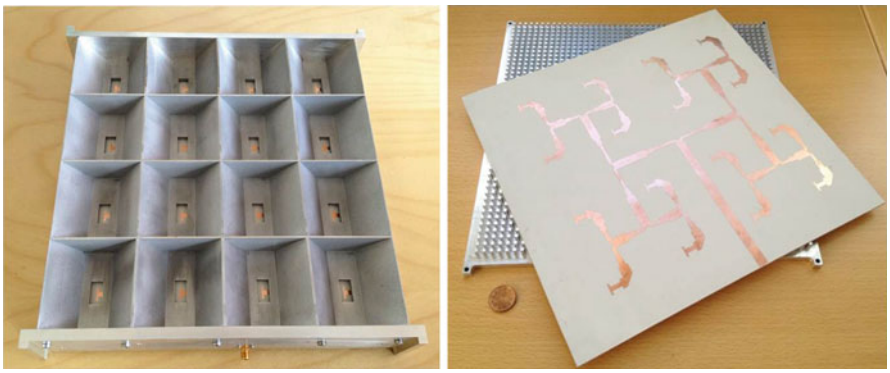


Fig. 33 Manufactured horn array with inverted microstrip gap waveguide feed network

of the 2 × 2 unit cell of the array is shown in Fig. 35a, and the details of the antenna are presented in Razavi et al. (2014). Actually, this work is the part of a larger project to realize a high-gain (∼35 dBi, 32 × 32 element) antenna based on PCB technology only.

In order to obtain the highest aperture efficiency and widest bandwidth, the 2 × 2 slot sub-array dimensions were optimized. It is well known that the slots (or other elements) in an antenna aperture should be uniformly spaced in orthogonal directions with spacing smaller than one wavelength in order to avoid grating lobes in large broadside arrays. This requirement may be relaxed if the element pattern has a null along the ground plane, such as in H-plane of slots. However, in E-plane, this requirement needs to be strictly followed, because slots have more uniform radiation pattern in E-plane. At the same time, the element spacing should be as big as possible in order to keep the total aperture size of the array large enough to have a required gain and directivity. In order to satisfy the above conditions, unit cell dimensions l and w as well as the inter-element slot spacing within the 2 × 2 slot sub-array have been optimized. It should be noted that the slot dimensions w_s and l_s as well as the coupling slot sizes w_c and l_c have been tuned at the same time to achieve good impedance matching over the operating bandwidth. SIW corrugations requiring width of d_c have been also introduced in this design between each 2 × 2 slot sub-array in E-plane, working as a soft surface. By this means, the mutual coupling has been suppressed between sub-arrays, and this made it easier to impedance match the S_{11} in the infinite array environment over 57–66-GHz band. At the same time, there will be a null in the sub-array pattern along the ground plane, and thereby the grating lobe problem is also reduced. The important sub-array dimensions and the SIW corrugations are shown in Fig. 35b. All the dimensional details are mentioned in Razavi et al. (2014).

DOI 10.1007/978-981-4560-75-7_130-1

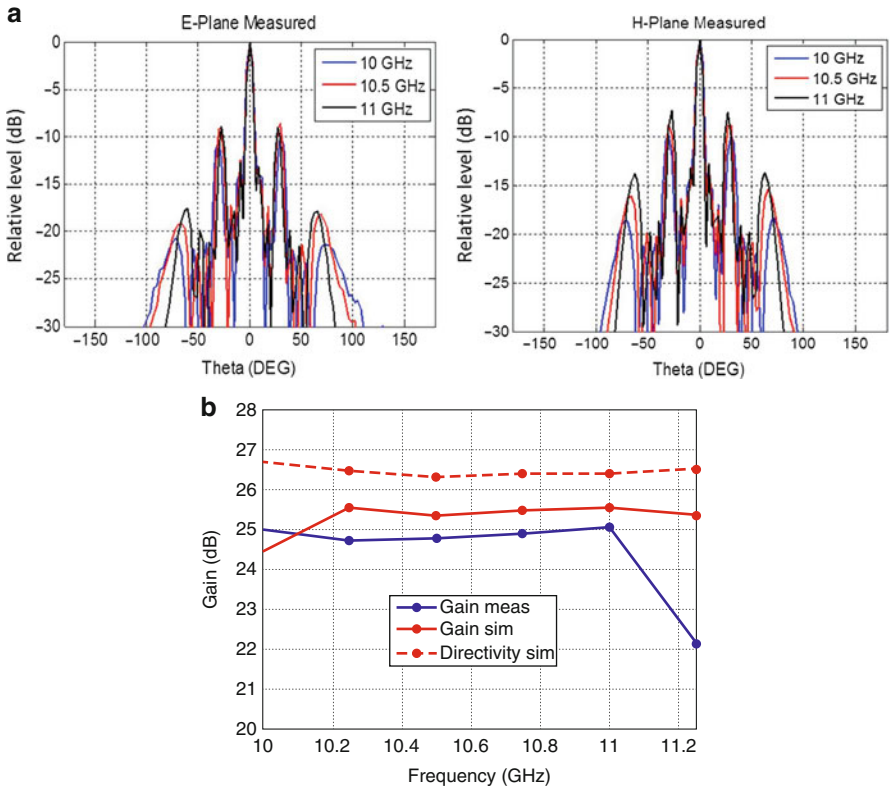


Fig. 34 (**a**) Measured E- and H-plane patterns for the horn array. (**b**) Measured gain and simulated gain for the horn array

After optimizing the 2 × 2 sub-array, a 4 × 4 element prototype antenna has been manufactured and measured to validate the concept. The picture of the manufactured prototype antenna and results for this antenna is shown in Fig. 36.

Transitions Between Gap Waveguide and Standard Transmission Line

The transitions between gap waveguide and other standard transmission lines play an important role and are integral modules related to gap waveguide components. In order to obtain maximum power transfer through a microwave subsystem, all components and modules constituting such a subsystem have to be matched. That is why transitions are needed to connect the gap waveguide passive components (i.e., antennas, filters, etc.) to a microwave subsystem and to have a good matching to standard microstrip line or standard rectangular waveguide. These transitions are also important from measurement point of view while assessing the performance of the designed gap waveguide components. In most of the cases, the rigorous equivalent circuit representations are not available for such transitions. So, optimum designs of these transitions are obtained by full-wave analysis using commercial full-wave solver. This chapter deals with standard microstrip to ridge gap waveguide transition, standard rectangular waveguide to ridge gap waveguide transitions, standard rectangular waveguide to microstrip-ridge gap waveguide transition, etc.

Microstrip to Ridge Gap Waveguide Transition

Successful integration of antenna and MMIC-based active RF circuit is quite critical in designing high-performance microwave modules at millimeter-wave frequencies. Many high-frequency TX/RX microwave modules contain MMIC chips, and they have typically microstrip line both at input and output.

DOI 10.1007/978-981-4560-75-7_130-1

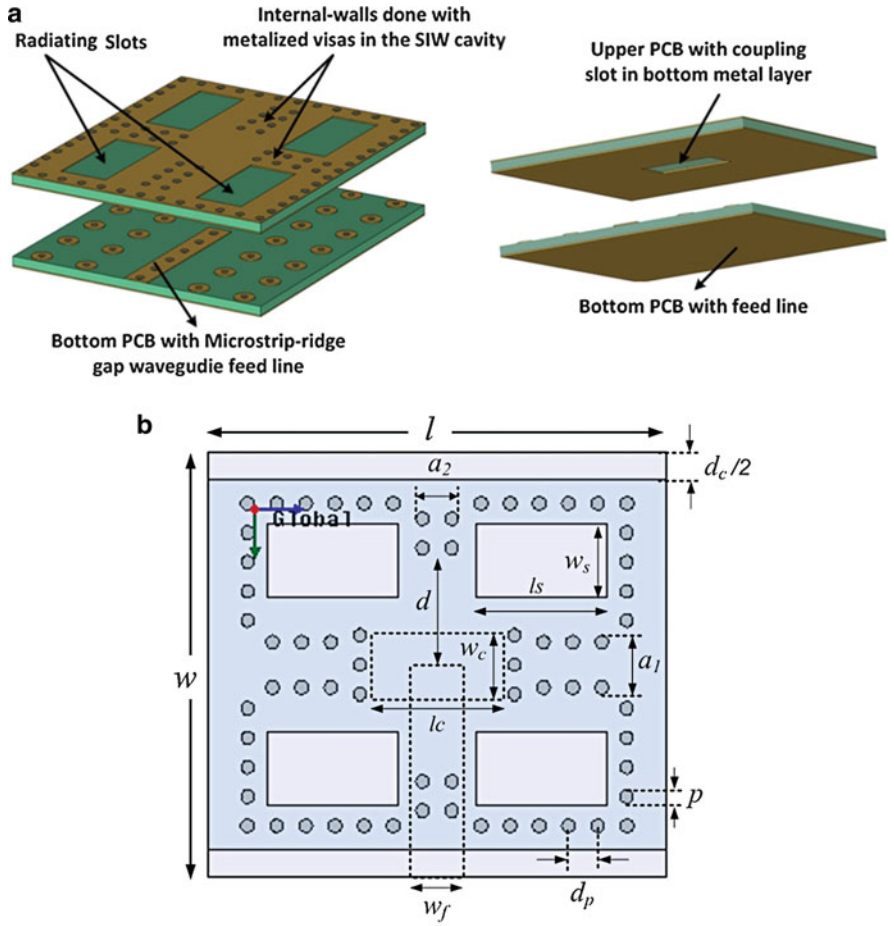


Fig. 35 (**a**) The 2 × 2 sub-array with SIW cavity layer and microstrip-ridge gap waveguide feed line. (**b**) Some parameters of the sub-array which needed to be optimized in an infinite array environment

Therefore, a good transition between microstrip line and gap waveguide is a key element for good integration of MMIC chips. Such a transition needs to be compact and should operate over as large bandwidth as possible. This can only be achieved by accurate impedance matching and mode transformation between these two different transmission lines. In this section, a standard microstrip line to ridge gap waveguide transition is mentioned.

The basics of this transition are based on the fact that the E-field of dominant mode in a microstrip line can be transformed easily to a standard ridge waveguide mode (Hui-Wen et al. 1994; Yunchi et al. 2010). In the case of ridge gap waveguide, it is even simpler as the dominant mode is also a quasi-TEM mode. This is shown in Fig. 37.

Due to such similar field distribution, only requirement for a good transition design lies on transforming E-fields in the dielectric to E-fields in the air. This can be done by tapering down the width of the ridge section in several steps (Zaman et al. 2013). In proposed design, the narrower ridge section is viewed as an extended section of the regular ridge section and is placed just above the microstrip line. The width of the ridge at this section is kept similar to that of the metal strip in microstrip line. In this approach, the tapered ridge section needs to be in electrical contact with the microstrip line and the substrate. This can be achieved by soldering, gluing, or simply pressing the ridge section down. Schematic of the designed transition between the microstrip line and ridge gap waveguide is shown in Fig. 38. A standard Chebyshev transformer based on several $\lambda_g/4$ sections of different widths can also be employed to improve the

DOI 10.1007/978-981-4560-75-7_130-1

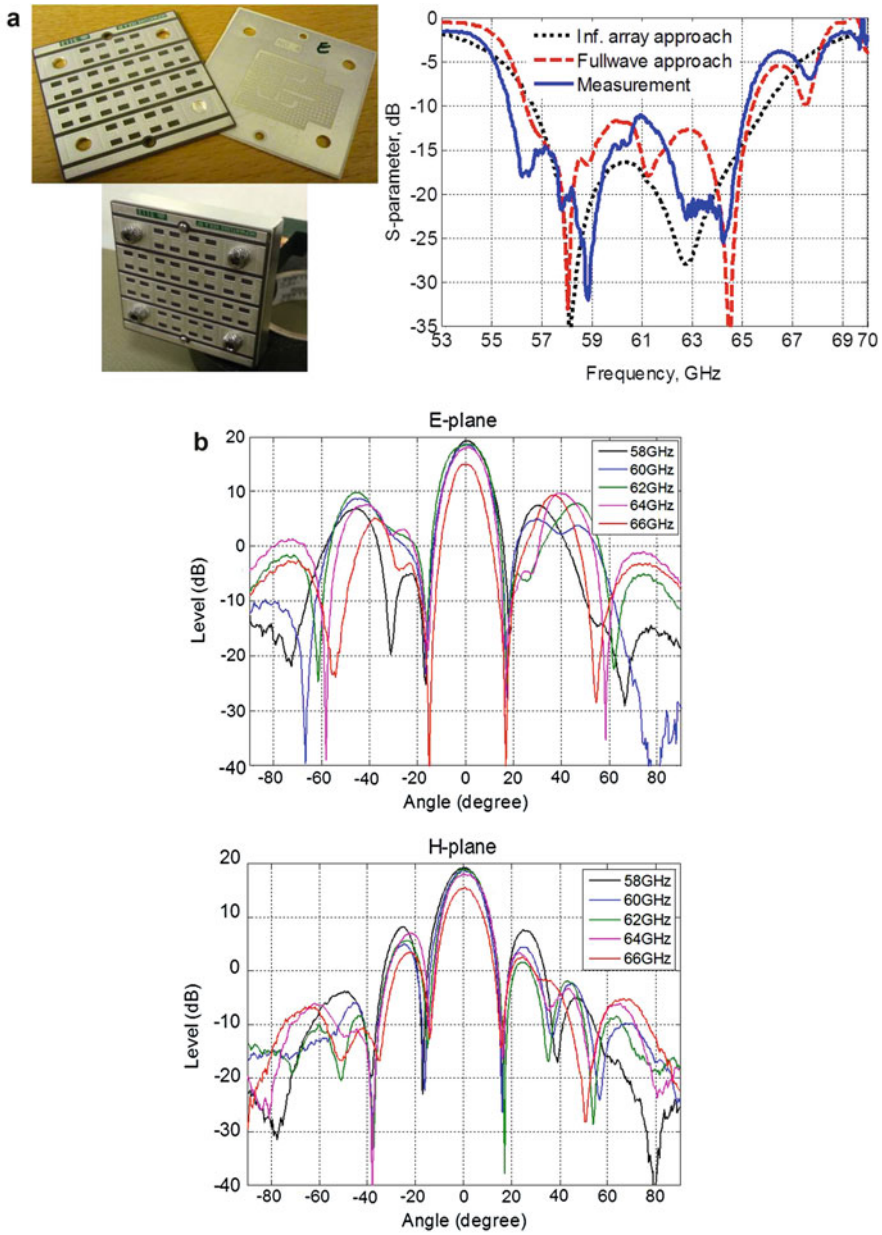


Fig. 36 (**a**) Manufactured prototype and the measured S_{11}. (**b**) Measured E-plane and H-plane radiation patterns over the band of interest

transition. But in this work, only one tapered section of the ridge has been used in a way to keep the overall transition compact. The manufactured transition and the measured results are shown in Fig. 39. The measured return loss is 14.15 dB for the 23–43-GHz band which means that a single transition will have even lower return loss over the entire Ka-band of interest. The maximum insertion loss over the same 23–43-GHz band is found to be 0.32 dB for the back-to-back transition including the losses of the ridge gap waveguide section meaning a loss smaller than 0.16 dB per single transition over the entire Ka-band (Zaman et al. 2013).

Apart from the transition mentioned above, another standard microstrip to ridge gap waveguide transition has been designed at W-band and presented in (Brazalez et al. 2012b). This transition is designed on the principle of electromagnetic coupling, meaning that there is no electrical contact

DOI 10.1007/978-981-4560-75-7_130-1

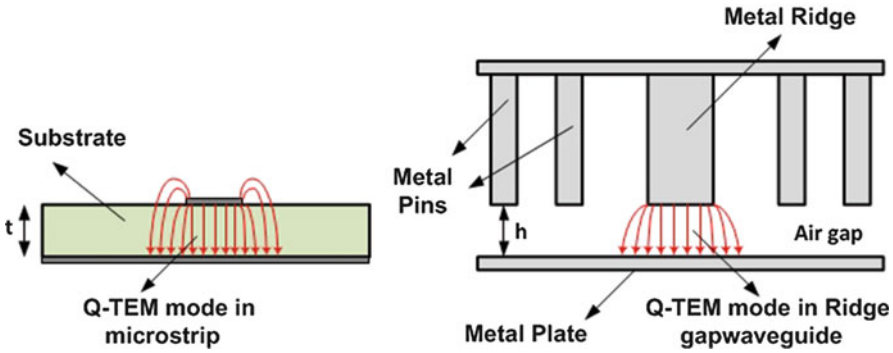


Fig. 37 E-field distributions of dominant mode of microstrip line and ridge gap waveguide

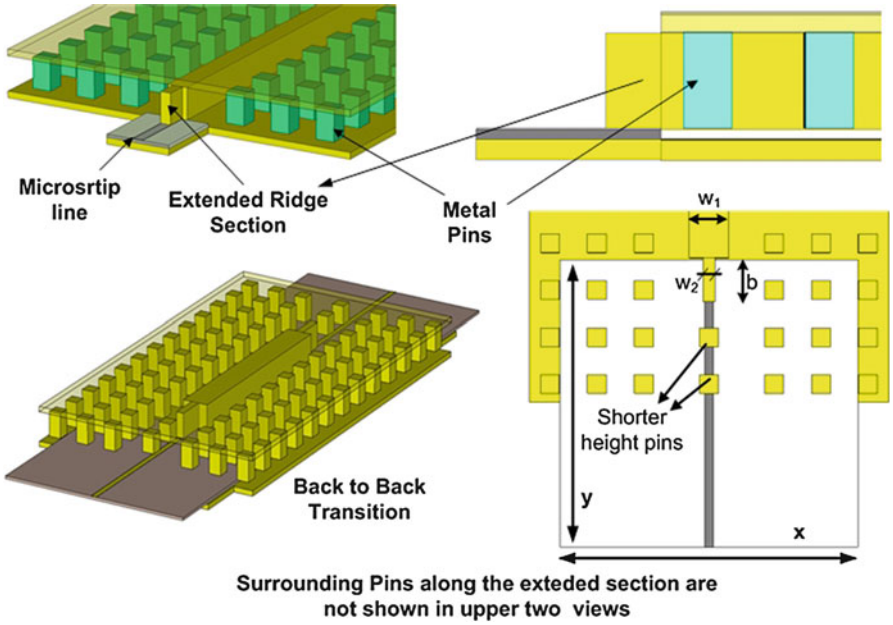


Fig. 38 Schematic details of the proposed transition, $W_1 = 2.65$ mm, $W_2 = 0.72$ mm, $x = y = 20$ mm, and $b = 2.75$ mm

whatsoever, between the typical microstrip line and ridge gap waveguide. The transition is based on overlapping λ/4 sections of transmission lines. There is a microstrip patch section and a ridge gap waveguide section which acts together as coupling sections. Strong coupling is achieved between the fringing EM field around the patch discontinuity and Q-TEM mode of ridge gap waveguide. The schematic of the transition is shown in Fig. 40a. The microstrip substrate is placed and stacked to the smooth metal plate of the ridge gap waveguide section. There exists an air gap of about 0.1 mm between the microstrip section and the ridge gap waveguide section. The ridge gap waveguide section which overlaps the microstrip patch has narrower width compared to the regular ridge section so as to improve the matching. The operating bandwidth of this proposed transition is about 25 %, which is smaller than that transition proposed earlier. But this transition has the advantage of having no direct contact between the microstrip section and the waveguide section. This can be a major advantage from mechanical point of view at W-band (∼100 GHz) or above. The critical dimensions for this transition are as follows: $h_1 = 0.2$ mm, $h_2 = 0.1$ mm, $W_1 = 0.35$ mm, $W_2 = 0.4$ mm, $W_{rid} = 0.55$ mm, $L_1 = 0.43$ mm, and $L_2 = 0.75$ mm. The simulated performance of this transition in back-to-back configuration is also presented in Fig. 40b.

DOI 10.1007/978-981-4560-75-7_130-1

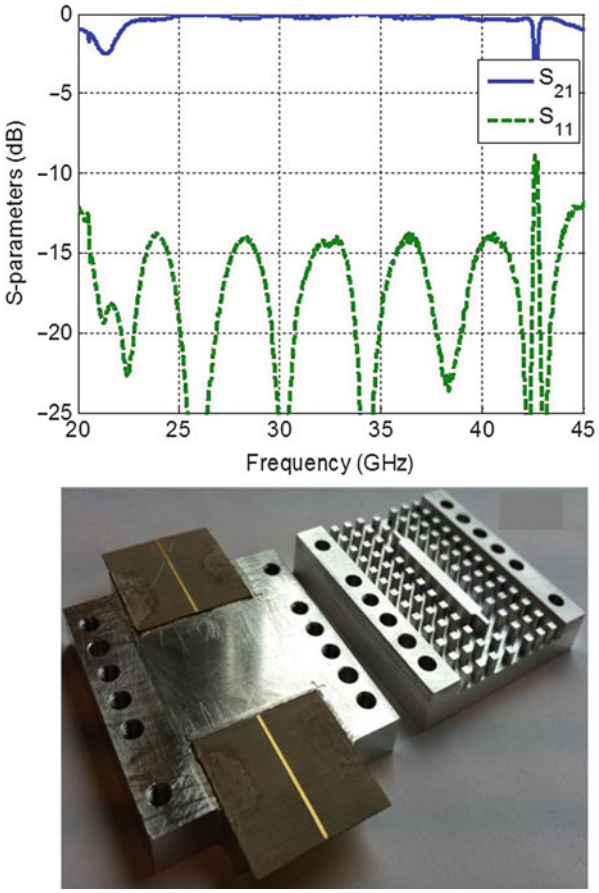


Fig. 39 Manufactured back-to-back transition and measured S-parameters for the transition

Rectangular to Ridge Gap Waveguide Transition

Two wideband transitions from ridge gap waveguide to rectangular waveguide have been designed recently. The first one is the inline transition where the height of the guiding ridge is reduced in several steps to match the height of the empty rectangular waveguide. In the second transition, the ridge gap waveguide is excited by a rectangular waveguide from the bottom. The first transition with stepped ridge sections is designed at Ka-band and is shown in Fig. 41 where the critical dimensions are as follows: $L_1 = 2.46$ mm, $L_2 = 2.27$ mm, $L_3 = 2.17$ mm, and $L_4 = 1.78$ mm and $S_1 = 0.22$ mm, $S_2 = 0.52$ mm, $S_3 = 0.76$ mm, $S_4 = 0.53$ mm, and $S_5 = 0.42$ mm. The starting point in designing this transition lies in the stepped design approach mentioned in Bornemann and Arndt (1987). After getting the initial design dimensions, the structure is numerically optimized to perform well over the entire Ka-band. The simulated performance of this back-to-back transition is also shown in Fig. 41.

The second transition is also a simple transition where the rectangular waveguide is feeding the ridge gap waveguide from the bottom. This type of transition is needed for measuring the ridge gap waveguide antenna where the feeding waveguide has to be placed below the antenna. The transition is designed at Ku-band and is shown in Fig. 42a, and all the critical dimensions for this transition are as follows: $P_L = 3.65$ mm, $P_w = 1.25$ mm, $P_s = 4.25$ mm, $R_w = 4$ mm, and $R = 2$ mm. The simulated performance of the single transition is shown in Fig. 42b. This transition has not been measured alone, but the same transition has been used in the ridge gap waveguide slot array work mentioned in Zaman and Kildal (2014).

DOI 10.1007/978-981-4560-75-7_130-1

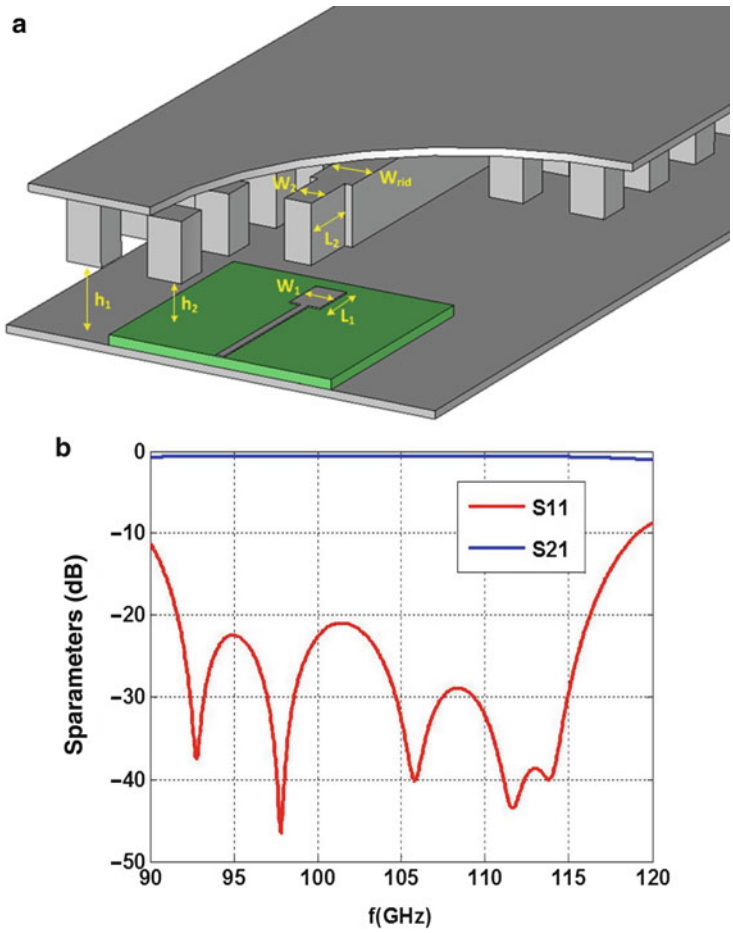


Fig. 40 (**a**) Schematic of the transition which does not need any electrical contact. (**b**) Simulated S-parameters of the transition which does not need any electrical contact

Coaxial to Ridge Gap Waveguide Transition

The first coaxial to ridge gap waveguide transition has been presented in Zaman et al. (2009). Conventional parallel-plate waveguides are easily excited by coaxial probes placed in the center of the broad face of rectangular waveguides with its inner conductor extending a distance into the guide (Williamson 1985; Heejin et al. 1998). In order to avoid the excitation in all directions, a short-circuiting wall or plunger is placed at a certain distance away from the probe. As the ridge gap waveguide is bit different from the conventional rectangular waveguides, the existing coaxial line to waveguide transitions needed to be modified or redesigned to make it appropriate for ridge gap waveguide. In this section, such a modified transition is presented, and the schematic of the transition is shown in Fig. 43a.

The transition is designed at Ku-band and it consists of a hole in the metal ridge having dimension of *de*. This allows the central conductor of the coaxial connector to launch EM field by electromagnetic coupling. Also, the transition has a back-short having a dimension of *b*. In this transition, the guiding metal ridge is short-circuited at both the ends. The critical dimensions for this transition are optimized by full-wave analysis, and these dimensions are as follows: $b = 5$ mm, $de = 3.4$ mm, $wd = 3.65$ mm, $s = 3.5$ mm, $p = 3.5$ mm, $h = 1$ mm, and $d = 5$ mm. The simulated performance for this transition in a back-to-back configuration is shown in Fig. 43b.

DOI 10.1007/978-981-4560-75-7_130-1

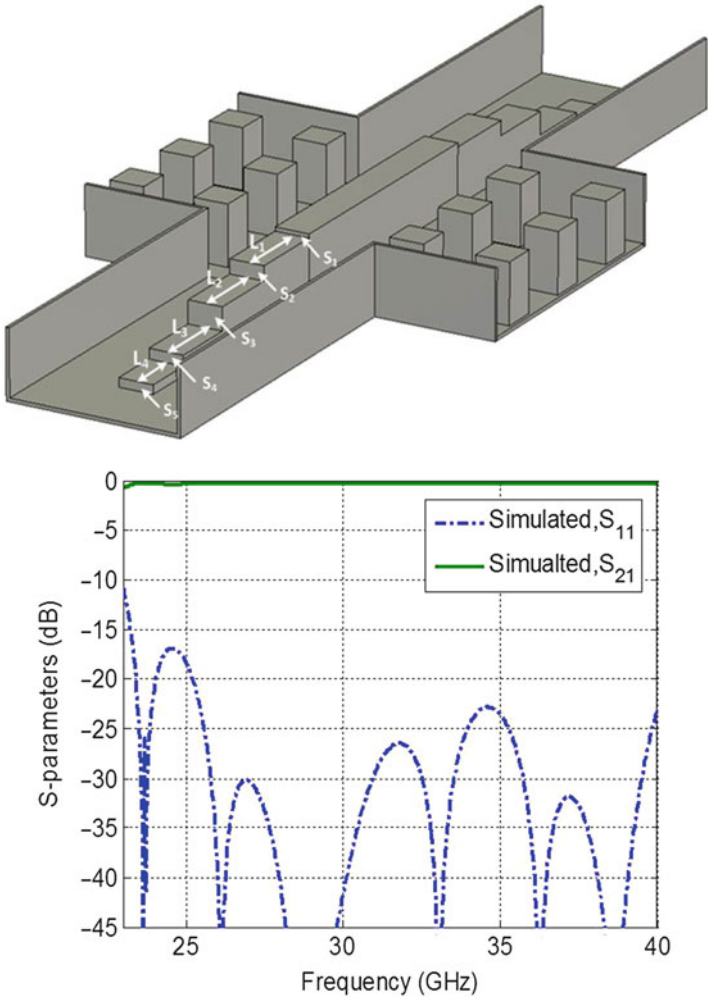


Fig. 41 Back-to-back transition with stepped ridge section (top metal plate not shown here) and the simulated S-parameters for this transition

Microstrip-Ridge Gap Waveguide to Rectangular Waveguide Transition

A simple microstrip-ridge gap waveguide to rectangular transition is described in this section. The vertical transition is designed bearing in mind that it would be used in a planar array having microstrip-ridge waveguide feed network mentioned in Razavi et al. (2014). In this transition, the rectangular waveguide is connected from the bottom to the microstrip-ridge gap waveguide. The schematic of the transition is shown in Fig. 44a. A λ/4 back-short is needed in most of the typical planar transmission line to rectangular waveguide transitions. But in the case of this transition design, there is not enough space for accommodating the required back-short, as this transition has been designed for a planar slot array. As a result, the transition operates over only 15 % relative bandwidth from 57 to 66 GHz. However, this 15 % bandwidth could be enough for many commercial applications as it covers the whole license-free band at 60 GHz. This transition has been manufactured and measured in a back-to-back configuration. Measured loss for a single transition is found to be 0.2 dB over the band of interest at 60 GHz. The detailed measurement results for this transition are presented in et al. (2014).

DOI 10.1007/978-981-4560-75-7_130-1

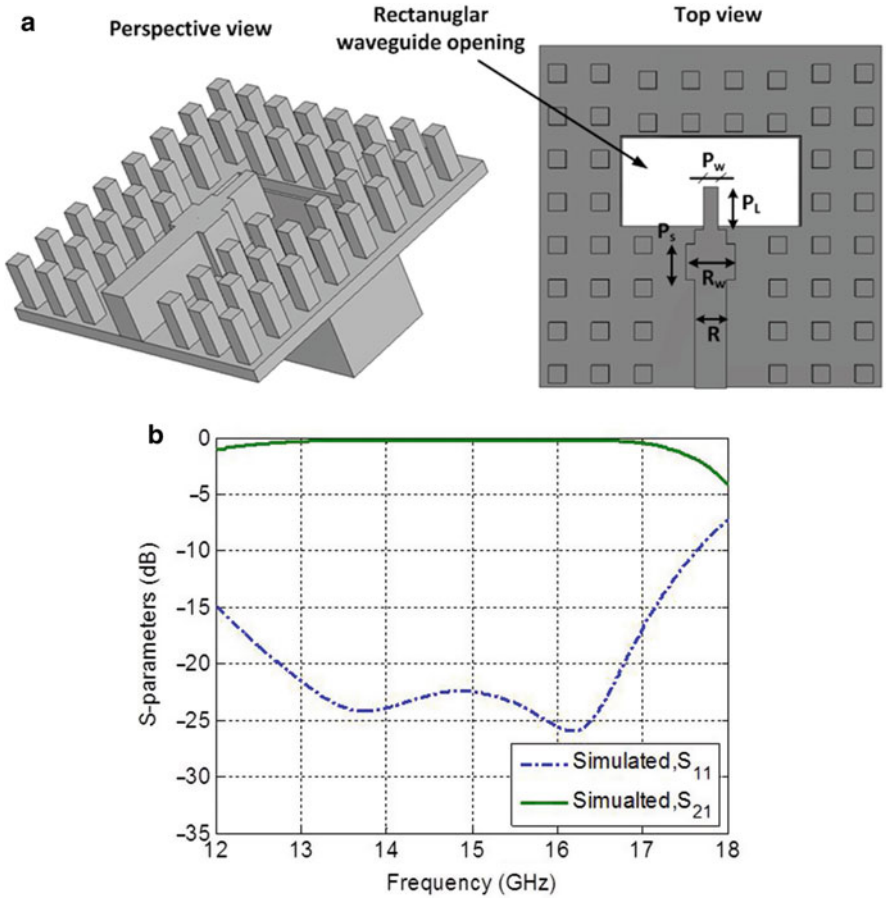


Fig. 42 (**a**) Ridge gap waveguide to rectangular waveguide transition (fed from bottom). (**b**) Simulated S-parameters for the ridge gap waveguide to rectangular waveguide transition (fed from bottom)

Gap Waveguide Narrowband High-Q Filters

High-Q filters are in demand in many wireless and satellite applications and are the most critical passive microwave components to be connected with the antenna feed systems. In a full-duplex communication, narrowband RF filters are required after high-power amplifiers (HPAs) and prior to low-noise amplifiers (LNAs). Usually, such full-duplex communication systems transmit and receive simultaneously. The transmitted power in the system is way too high in comparison to the received power. So, the transmitter filter must have very high attenuation in the receive band to stop the intermodulation noise and wideband thermal noise to be fed into the receiver. Also, the loss of the transmitter filter needs to be low due to the linearity and efficiency constraints of power amplifiers. In the receiver case, the noise figure of the receiver is dictated by the losses in the receiver filter as this filter sits before the LNA in the receiver chain. So, the receiver filter should also have a low insertion loss and very high selectivity in the transmit band. Thus, filters that are currently used in wireless applications have very stringent requirements in terms of return loss, insertion loss, and isolation. The current baseline design and manufacturing used today employ 3D structures such as coaxial, dielectric resonator or waveguide filters mainly to achieve high-Q value. Apart from the electrical performance requirement, the filters have to be mass producible and low cost. For this reason, H-plane iris-coupled filters and E-plane metal insert filters are extensively used today.

Both these types of filters have certain drawbacks at high frequency. In case of H-plane filters, filter assembly will exhibit a transitional resistance between the top metal plate and milled waveguide section.

DOI 10.1007/978-981-4560-75-7_130-1

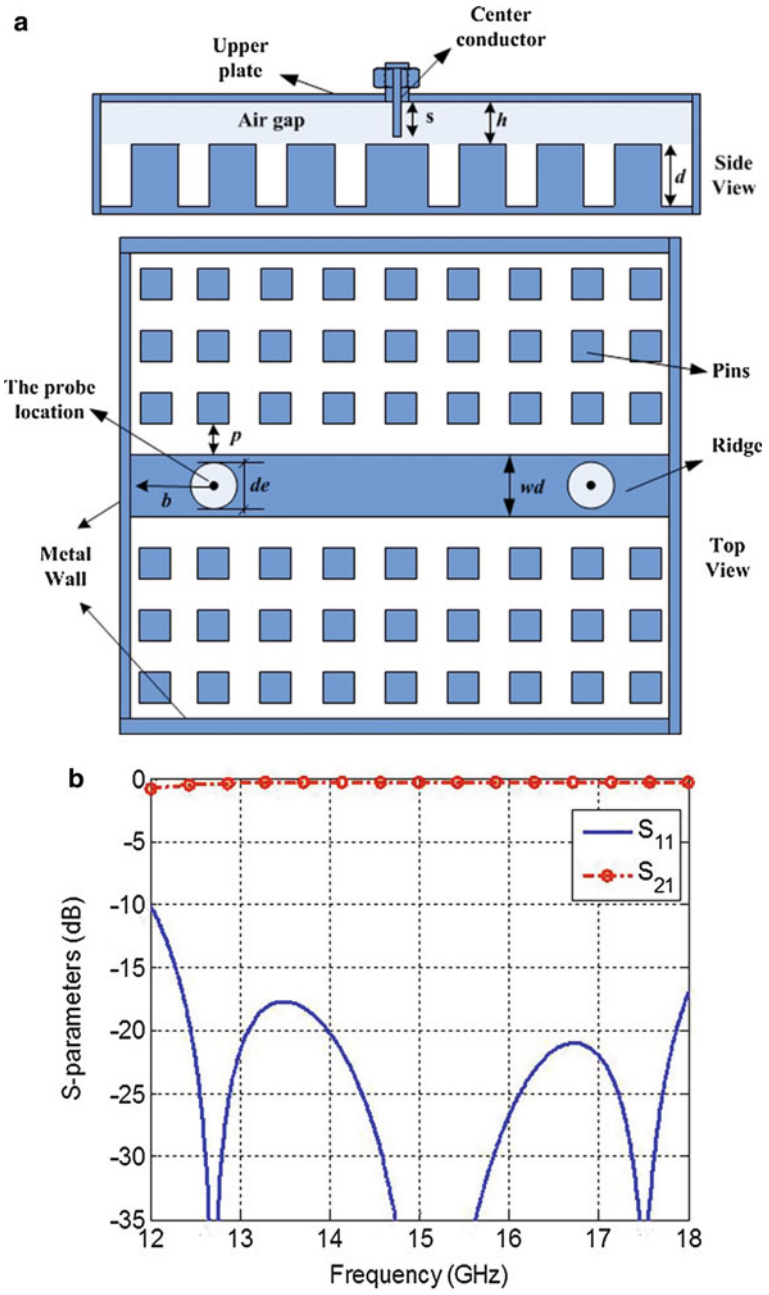


Fig. 43 (**a**) Schematic of ridge gap waveguide to coaxial line transition. (**b**) Simulated results for the back-to-back ridge gap waveguide to coaxial line transition

As this transition occurs at location of high current density, this will have to be taken care of by ensuring good electrical contact between the two parts; otherwise, significant amount of energy will leak through tiny gaps leading to Q value degradation of the resonators (Zaman et al. 2010b). Also, the milling process requires very small tool to manufacture rounded corners with smaller radius (Uher et al. 1993; Bornemann 2001). On the other hand, in case of E-plane filters, the electrical performance of filters is mainly determined by the pattern of the metal inserts. For example, the thickness of the metal insert needs to be thin enough (30–80 μm) for a moderate filter requirement at 38 GHz. The thickness of the metal insert will have to be even thinner at mmWave frequency range to comply with the filter requirements

DOI 10.1007/978-981-4560-75-7_130-1

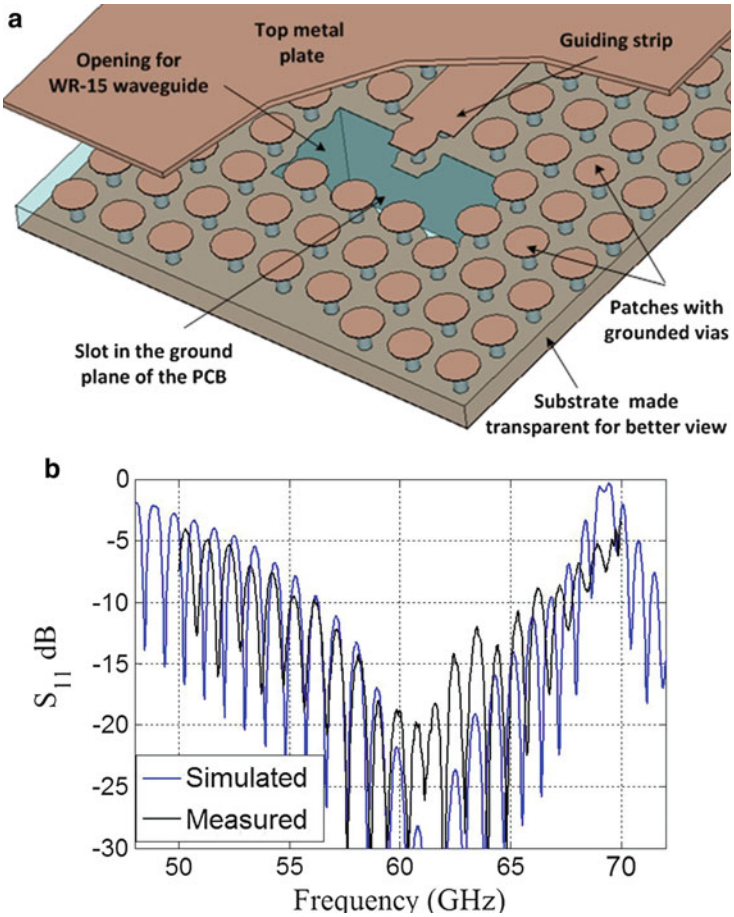


Fig. 44 (**a**) Schematic of microstrip-ridge gap waveguide to rectangular waveguide transition. (**b**) S-parameters for the microstrip-ridge gap waveguide to rectangular waveguide transition

(Ofli et al. 2005). Also, these metal waveguide types of filters are difficult to integrate in a single module with other passive components such as coupler or antennas and active RF components.

In this concern, there is a need for developing new types of high-Q filters which are more robust in terms of mechanical assembly and can be integrated with other modules of passive components such as antennas. The very low-loss groove gap waveguide-based high-Q filters can be of interest in this case. These high-Q filters can be built in an open parallel-plate structure surrounded by periodic metal pins without any side walls. The electrical contact between the parallel plates will not play a major role in the filter performance as the field leakage will be negligible after two or three rows of pins. Also, measured Q values for such groove gap resonator are similar to that of a rectangular waveguide (Pucci et al. 2013; see also Operating Principle of Gap Waveguides and Loss Analysis).

High-Q Groove Gap Waveguide Resonator and Narrowband Filters

The gap waveguide resonator is a resonator that has very controlled leakage of EM energy through the gap between two metal blocks or plates. This type of resonator basically allows a leakage up to a certain level and traps the rest of the EM energy within the structure. It has been mentioned previously in section "Operating Principle of Gap Waveguides and Loss Analysis" that, after two consecutive rows of pins, the field leakage goes down up to a level of −40 to −45 dB over significant bandwidth. This motivates the design of the high-Q resonator-based filters using two metal plates with much more manufacturing flexibility. There is no need for metal walls on each side of the resonator. Also, the electrical contact

DOI 10.1007/978-981-4560-75-7_130-1

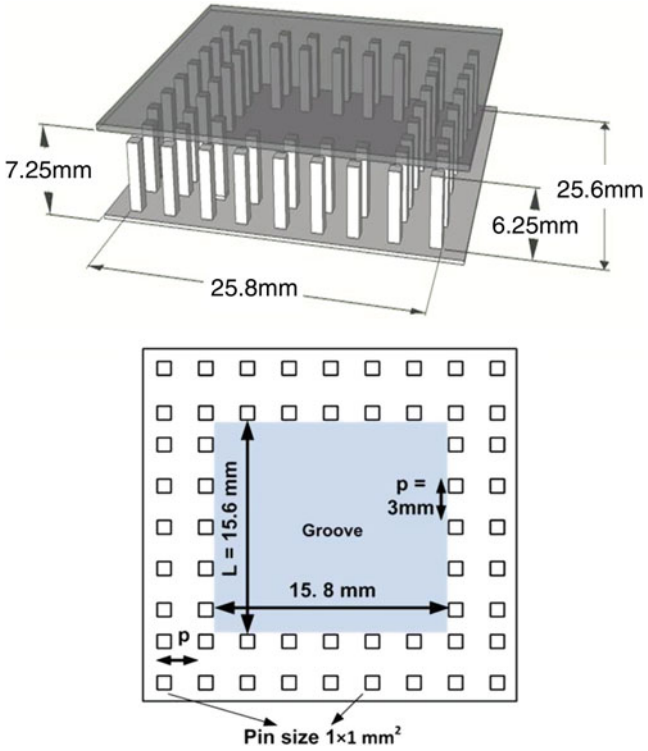


Fig. 45 Geometry of the groove gap resonator at Ku-band, perspective view and top view

between the top and bottom metal plate does not play a significant role in trapping the EM energy. This section describes a description of a groove gap resonator and several narrowband filter designs based on groove gap waveguide. Figure 45 shows the half-wavelength short-circuit groove gap waveguide resonator at Ku-band. The resonator also has only some texture of pins forming a groove in the lower metal plate. The pins are about λ/4 in height. The upper metal plate is a smooth plate and placed at a distance of 1 mm from the pin surface. It was chosen to have two rows of pins on each side of the resonator to make the field decay up to a level of about −45 dB.

The simulated Q value for this type of resonator with conductivity of aluminum is about 4605 which is 13 % lower than the simulated rectangular waveguide resonator at that frequency. Q factor of a practically realized groove gap resonator and a rectangular waveguide resonator is compared more in detail in Pucci et al. (2013). The field distribution inside the groove gap resonator is shown in Fig. 46.

Two coupling mechanisms are studied to achieve the required coupling between two adjacent groove gap waveguide resonators. These are shown in Fig. 47. In the case of ridge coupling, the width, length, and height of the ridge section can be varied to achieve coupling coefficient values varying over large range. For the other case, only the periodicity of the pin rows and distance between the pin rows separating the two resonators are varied. Very low values of coupling coefficients are possible to obtain which is required for very narrowband filters.

Based on these two coupling mechanisms, third-order and fifth-order Chebyshev band-pass filters have been designed at Ku-band. The excitation of the input and output resonators of the filters has been achieved by typical coaxial line using SMA connectors for simplicity of measurements with the help of available vector network analyzer at Ku-band. Simulated results for two third-order filters having 1 % relative bandwidth and passband ripple of 0.032 dB and 0.1 dB are shown in Fig. 48a. Out of these two filters, the filter with 0.1 dB passband ripple has been manufactured and tested. The measured results are presented in Fig. 48b. The details of this filter design can be found in Zaman et al. (2012b). As shown in

DOI 10.1007/978-981-4560-75-7_130-1

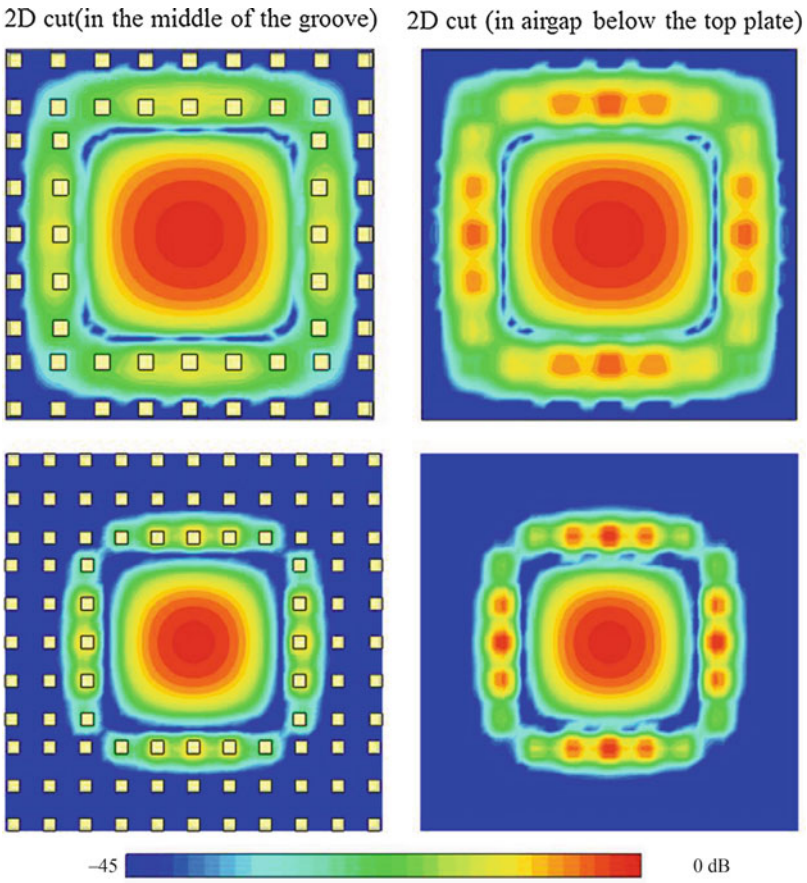


Fig. 46 Vertical E-field distribution for groove gap waveguide resonator

Fig. 48b, the filter response for the third-order filter agrees well with the measurement results. In this case, the measured mid-band insertion loss is about 0.31 dB more than the simulated cases. This extra loss is mainly attributed to the losses in SMA connectors and the degradation of the conductivity of aluminum due to surface roughness.

Ka-Band Filter Design for Commercial Diplexer Application and V-Band Filter Design

Usually, diplexer filters require very low insertion loss in the passband and high selectivity to reject frequencies close to passband. The purpose of this section is to demonstrate the potential of gap waveguide technology to design filters with very stringent commercial specifications. The key filter specifications for commercial 38-GHz diplexers are shown in Table 3.

The general band-pass filter design procedure utilizes the low-pass elements of the prototype filter to determine the required coupling coefficients between the adjacent resonators and the coupling to the external circuit, that is, to the source and load. For a filter with $N = 7$, the low-pass parameters are obtained from Swanson (2007). Once the low-pass parameters are known, the coupling coefficients (K) and external quality factor (Q_{ex}) are easily calculated by using the following equations:

$$K_{i,i+1} = \frac{\text{fractional bandwidth}}{\sqrt{g_i g_{i+1}}}; \qquad Q_{ex} = \frac{g_n g_{n+1}}{\text{fractional bandwidth}}$$

DOI 10.1007/978-981-4560-75-7_130-1

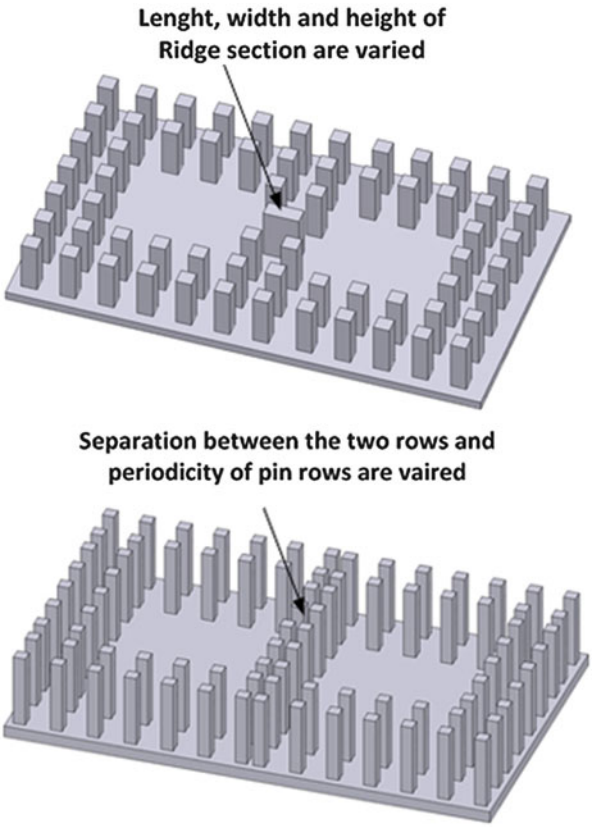


Fig. 47 Two types of coupling mechanism for groove gap waveguide resonators

To design the filter, the ridge coupling scheme presented previously has been chosen. The resonators are designed with pin dimension $0.7 \times 0.7 \times 2.3$ mm^3. The periodicity of the pins has been chosen as 2.1 mm. The manufactured seventh-order filter and the measured response of the filter are shown in Fig. 49. Details of this filter are presented in Alos et al. (2013). Two prototypes have been manufactured: one with aluminum without silver plating and the other one with silver plating. Both the prototypes have S_{11} better than -17 dB. The prototype without silver plating has more insertion loss as expected, and S_{21} has been found to be -1.5 dB. The other prototype with silver plating has an insertion loss of 1.0 dB. Both the filters needed the tuning screws to rectify the frequency shift induced due to mechanical tolerance errors. The performance of this groove gap filter has been compared with a commercially available rectangular filter working at 38 GHz. After comparison, the performance of the groove gap waveguide is found to be very similar to that of rectangular waveguide filter.

The next filter is a V-band filter designed in groove gap waveguide. The designed filter is a Chebyshev band-pass filter centered at 61 GHz, with a fractional bandwidth of 2.5 % (1.525 GHz), having a maximum ripple of 0.01 dB inside the passband. The filter is a fifth-order filter, and the implementation is carried out through cavity resonators coupled through inductive windows. The filters are fed by two 1.85 mm coaxial connectors. The picture of the manufactured filter and the measured performance is shown in Fig. 50. The details of the filter are found in Berenguer et al. (2014).

DOI 10.1007/978-981-4560-75-7_130-1

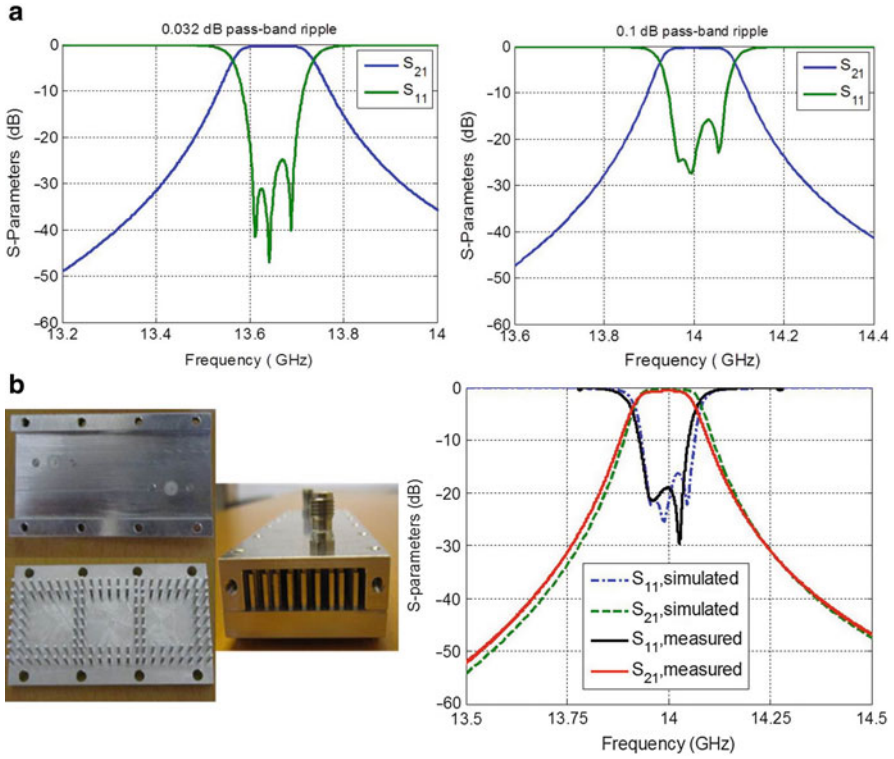


Fig. 48 (**a**) Two different groove gap waveguide third-order filter simulated responses. (**b**) Manufactured third-order filter and measured results for the same filter

Table 3 Specification for the 38-GHz diplexer filter

Passband	37.058–37.618 GHz
Stopband	38.318–38.878 GHz
Passband insertion loss	Max. 1.5 dB
Attenuation in the stopband	70 dB
Return loss	−16 dB

Gap Waveguide Packaging Solution for Microwave Modules

Most of mmWave microwave systems have to operate at outdoor, under harsh weather conditions, and therefore have to be shielded, i.e., *packaged* in a proper way to provide protection against mechanical stress and environmental conditions. Also, to comply with the smaller size and form factor requirements for mmWave microwave modules, large amount of electronic components are to be placed into a confined area. For such high-density microwave modules operating at mmWave frequency range, RF packaging is more and more essential in terms of isolation and interference suppression.

In a common microwave module, circuit components such as MMICs are placed on a carrier dielectric substrate which can carry the necessary interconnect lines, the passive components, etc. Enclosing the complete circuitry in a metal package usually degrades the RF circuit performance because of the onset of package resonances and more, in general, due to the interactions of electromagnetic fields (produced by the signal propagating in the electronic circuitry) with the package itself. Published studies in Tsuji et al. (1995) and Mesa et al. (2000) show that substantial power leakage and power loss exist on various printed circuit transmission lines, often related to surface waves in the dielectric substrate causing serious

DOI 10.1007/978-981-4560-75-7_130-1

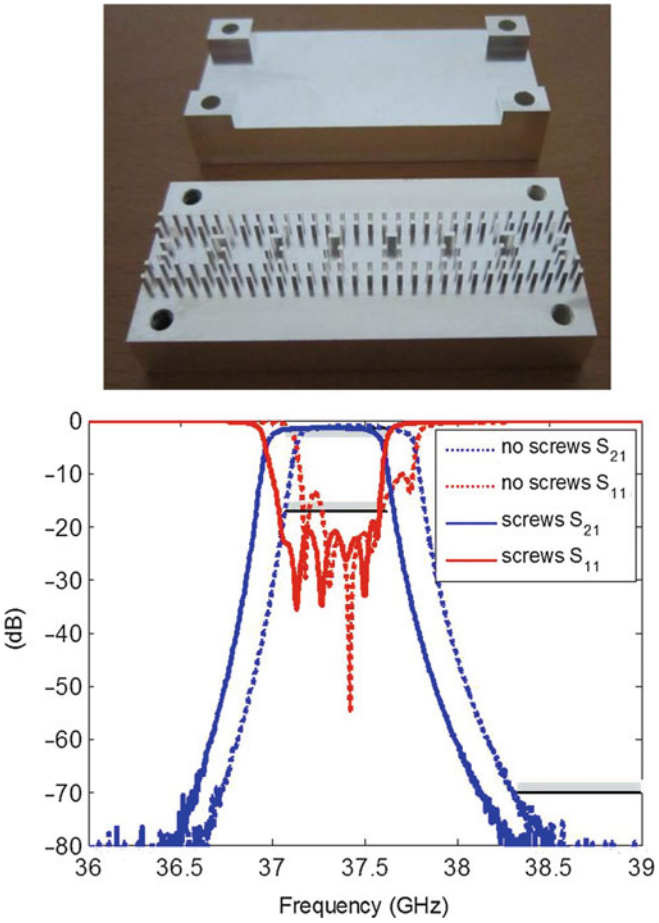


Fig. 49 Manufactured Ka-band seventh-order filter and measured results for the same filter

cross-talk and interference problems. Apart from this problem of surface wave or parallel-plate (Six et al. 2005) mode leakage and radiation, at mmWave frequencies, interconnects and transitions between different components or even the interconnecting signal lines may produce unwanted radiation or standing waves. This can easily get coupled (over the air) to the adjacent circuit elements resulting in interference and cross talk. Figure 51 displays some of the possible ways of electromagnetic field coupling and unwanted energy leakage between the adjacent circuit elements and also shows the conventional prevention techniques mentioned in Williams (1989), Izzat et al. (1996), and Dixon (2005). In practice, the efficiency of the different techniques available for reducing packaging problems and reduction of cross talk goes down with frequency; the higher is the frequency, the lower is usually the effectiveness of the techniques mentioned so far. So, there is a need for new packaging technology to be used at high frequency, and gap waveguide-based PMC packaging can be a real good solution to overcome the packaging-related problems.

Gap Waveguide Packaging Concept for Microwave Modules

The gap waveguide packaging solution is an extension of the basic gap waveguide geometry and is based on the concept shown in Fig. 52. Here, the parallel-plate cutoff is achieved by a PMC/AMC surface when the opposite surface is a perfect electric conductor (PEC) and as long as the spacing between these two surfaces is smaller than quarter wavelength. Once a substrate and the microstrip line are placed, the ground plane of the microstrip substrate acts as PEC surface, and the top lid with periodic texture surface acts as an AMC surface. In this case, the dominant microstrip mode can propagate along the strip, and field

DOI 10.1007/978-981-4560-75-7_130-1

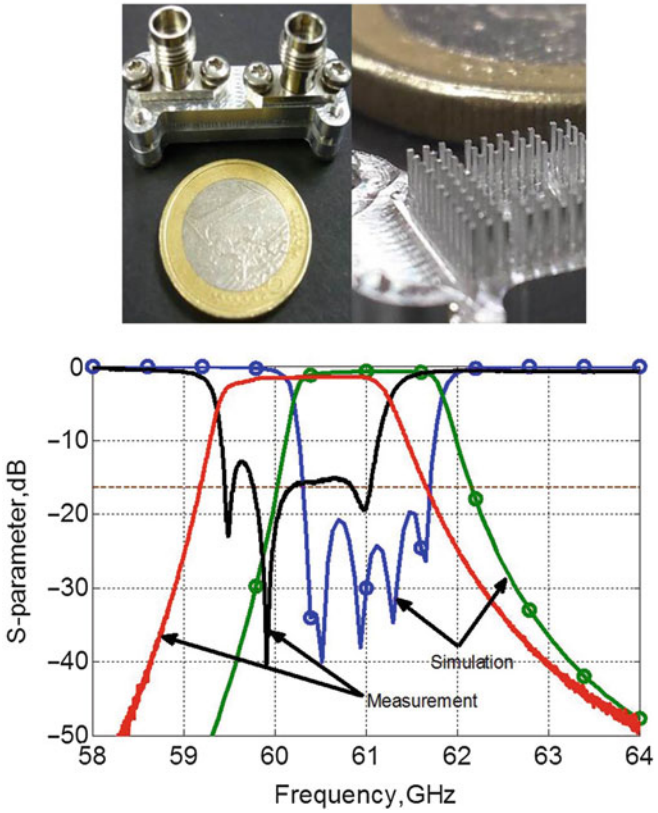


Fig. 50 Manufactured V-band fifth-order filter and measured results for the same filter

propagation in all other direction is stopped by the PEC-PMC parallel-plate cutoff condition. Thus, all the unwanted leakage or radiation via unwanted modes is suppressed within the frequency band when the PEC-PMC cutoff condition is enforced. The newly proposed packaging technique is easily scalable to high frequency and shows consistent performance over different microwave frequency bands extending from X-band up to the W-band (Zaman et al. 2012a; Rebollo et al. 2014).

Apart from the improvements in electrical performance, this new packaging technique offers an advantage while doing circuit design using full-wave simulations based on FTTD or FEM. With this new technique, the designer can reduce the computational domain by using an ideal PMC surface instead of using several matching layers above the circuit of interest at specific height. This saves computer time and requires no big modification afterward (when the circuits are practically realized) as the packaging effects are already considered during the full-wave design (Kishk et al. 2012). The application for gap waveguide packaging was first demonstrated successfully for a passive microstrip line in Rajo-Iglesias et al. (2013). Later, it was applied to a multi-port antenna feed network to improve the performance of well-known coupled line microstrip filter (Zaman et al. 2010a; Brazalez et al. 2012a). In all these cases, it was observed that unwanted cavity modes and surface waves were efficiently suppressed within the band of interest. In the abovementioned references, periodic pin structures have been used to realize the AMC. But recently, several other periodic structures such as periodic metal springs and periodic zigzag lines printed on a substrate have been used to obtain the AMC condition for such kind of work (Rajo-Iglesias et al. 2012, 2013). Some important features of gap waveguide packaging technique are listed below:

- As the energy coupling between the dominant microstrip mode and other leaky modes is reduced or suppressed, the overall insertion loss performance of the complete microwave circuit is also improved.

DOI 10.1007/978-981-4560-75-7_130-1

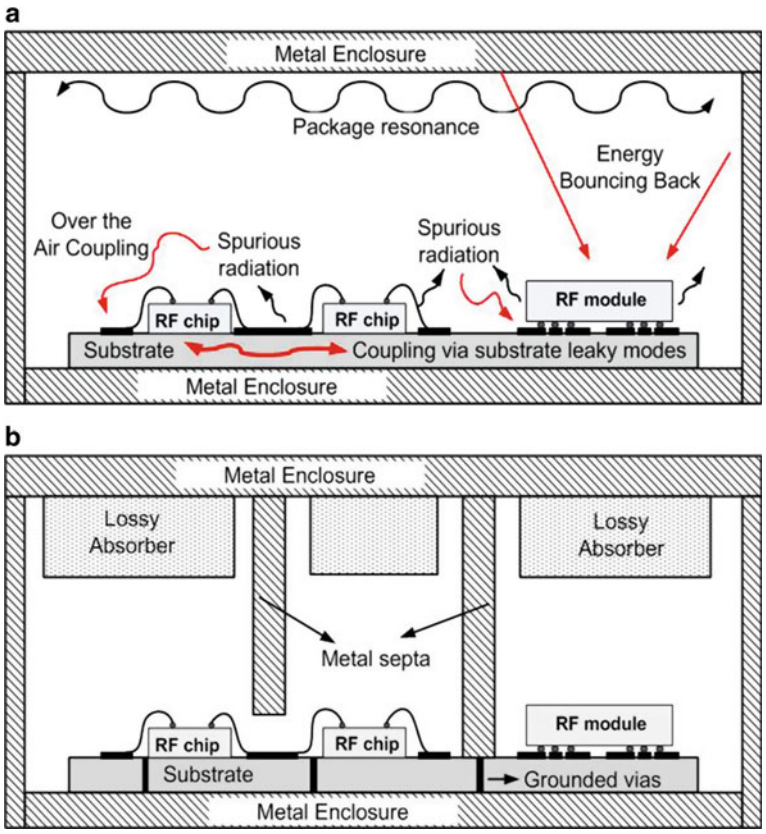


Fig. 51 (**a**) Various undesired coupling phenomena; (**b**) traditional prevention techniques

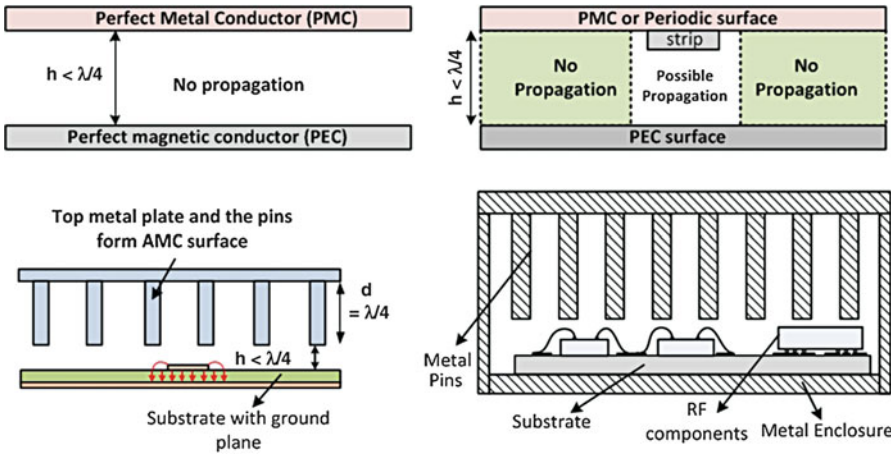


Fig. 52 Gap waveguide packaging concept

- There is no need for time-consuming postproduction trial and error process involving placement of lossy absorbing material within the package after the design has been completed.
- Isolation performance in this packaging technology is not dependent on very good metal contact between the circuit board and the metal enclosure, so to say; there is no need for proper grounding and use of conductive adhesive or gasket material for multi-compartment microwave modules.

DOI 10.1007/978-981-4560-75-7_130-1

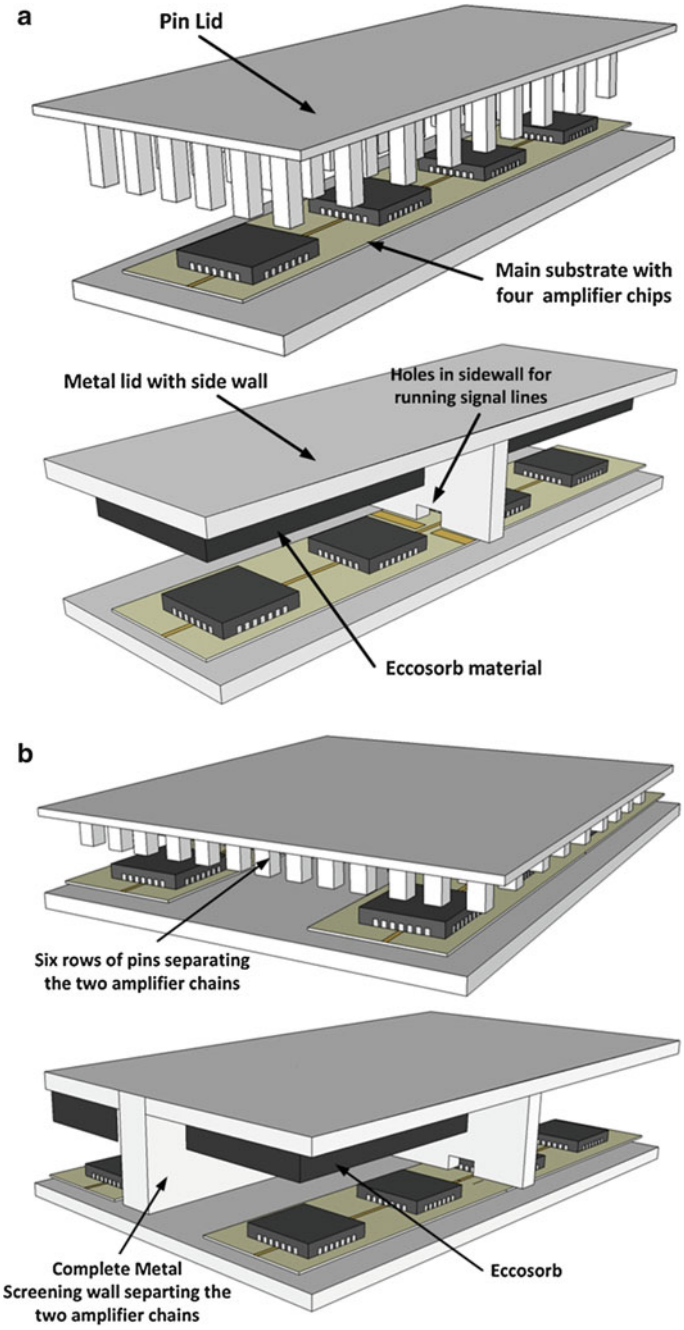


Fig. 53 (**a**) Test circuit for isolation evaluation of a single amplifier chain (side walls not shown), (**b**) test circuits for evaluation of side-by-side amplifier chain

Gap Waveguide Packaging for Ka-Band Active Amplifier Chains

Instability in an amplifier chain is mainly the result of unwanted feedback across the amplifier, and such feedback is normally induced by phenomena related to shielding imperfections such as leakage, spurious radiation, etc. This section explains how the newly proposed gap waveguide packaging technique has been applied on a Ka-band high-gain amplifier chain having four cascaded units. At first, one amplifier chain has been characterized by using gap waveguide packaging as well as the conventional absorber-based metal wall packaging solution. After that, isolation test has been performed for two active high-gain

DOI 10.1007/978-981-4560-75-7_130-1

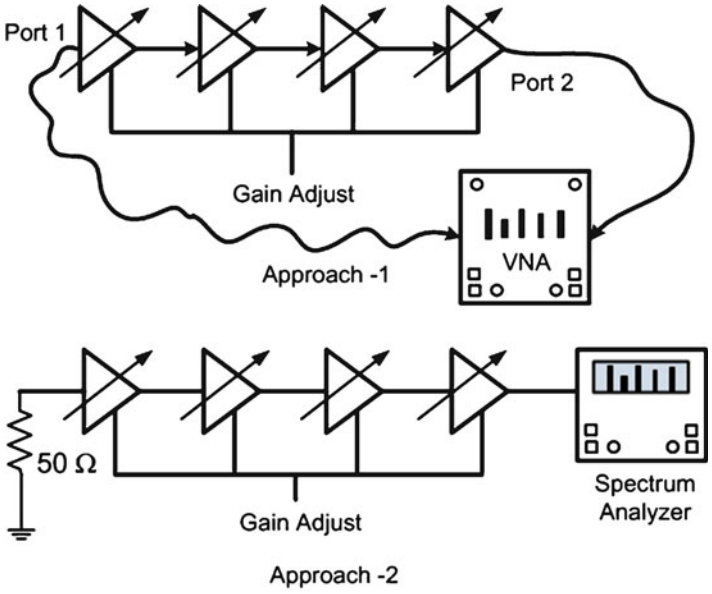


Fig. 54 Test setups for single chain isolation performance test

amplifier chains running in opposite directions so as to emulate the transmit/receive scenario in a real microwave module. Both the test cases are shown in Fig. 53.

Case A: Single Amplifier Chain Isolation

Two experimental approaches have been followed to test the isolation performance of the single chain amplifier: stable gain test and self-oscillation test, both being shown in Fig. 54. In the first approach, the amplifier chain under test is simply connected with Anritsu 36397C VNA, and a universal test fixture (UTF) from Rosenberger is used for launching the signal for S-parameter measurements. The gain in the amplifier chains has been increased by varying the control voltage until there has been an oscillation tendency. The second approach is a self-oscillation test where the input port of the first-stage amplifier is connected with a 50-Ω resistor (and no signal source) and the output of the final-stage amplifier is connected to a spectrum analyzer. Here also, the gain of the amplifier chain has been increased until resonance peaks are observed. The details of these experiments and obtained measurement results are presented in Zaman et al. (2014). In this section, only the results from the second approach have been presented in Fig. 55. From the single chain self-oscillation measurement results, it becomes obvious that no oscillation peaks appear for gap waveguide packaging case (pin lid packaging) even after 65 $\sim$ 70-dB total gain in the amplifier chain. On the other hand, for traditional metal wall packaging with absorbing materials, resonance peaks are observed after 35 $\sim$ 40 dB of total forward gain.

Case B: Side-by-Side Amplifier Chain Isolation

All full-duplex systems which activate both the transmitter and the receiver simultaneously require an excellent isolation between receive and transmit paths for preventing the transmit signal to cause saturation and intermodulation distortion in the sensitive receiver part. Particularly, the coupling between the final-stage transmit amplifier and the first-stage LNA of the receiver chain is of great importance due to the high difference of transmit and receive signal levels. In this work, such two amplifier chains have been placed side by side (to emulate the transmit and receive scenario), and sidewise isolation performance has been measured in the presence of gap waveguide shielding and conventional metal wall shielding. The measured results are presented in Fig. 56.

DOI 10.1007/978-981-4560-75-7_130-1

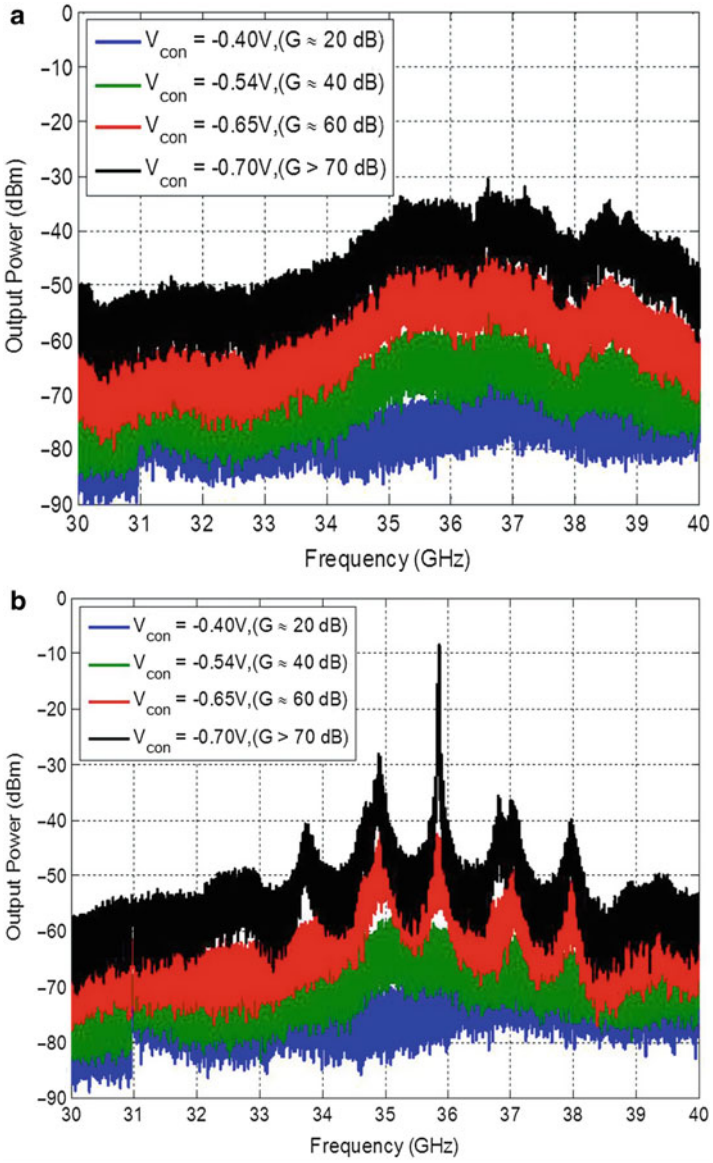


Fig. 55 (**a**) Self-oscillation analysis for single amplifier chain packaged with pin lid, (**b**) self-oscillation analysis for single amplifier chain packaged with conventional metal lid

For the test case of two amplifier chains placed side by side, the measurement results showed that an isolation around 70 dB in average was achieved for a six-row gap waveguide pin lid with isolation value ranging from 64 to 91 dB. Similar isolation trend was observed also for complete metal wall shielding case. For the case of gap waveguide lid having eight rows of pins, better isolation has been observed. The manufactured prototypes are shown in Fig. 57.

Conclusion

Wireless experts have devised all sorts of clever fixes to expand the capacity of today's LTE cellular network (4G) by means of multiple antennas, robust interference mitigation, smaller cells, and smarter

DOI 10.1007/978-981-4560-75-7_130-1

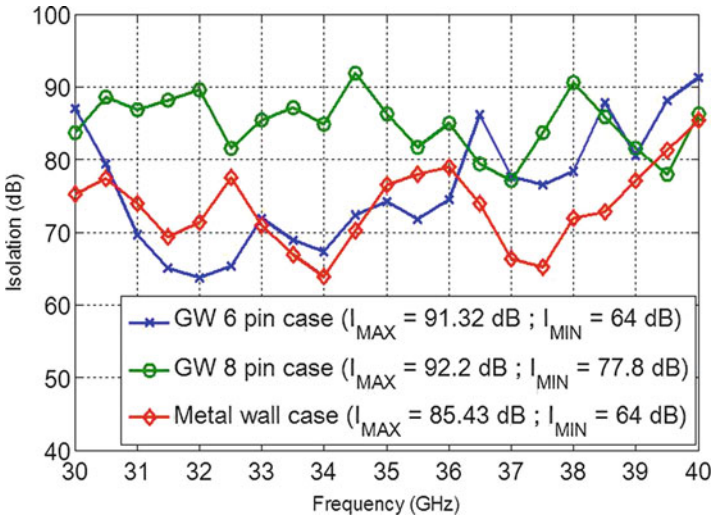


Fig. 56 Isolation between two closely spaced amplifier chains

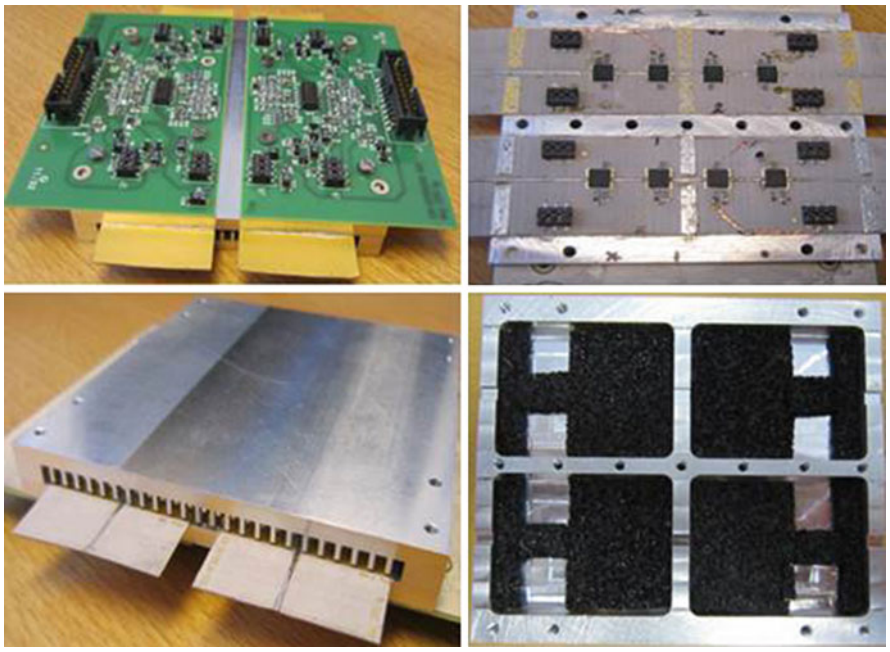


Fig. 57 Manufactured prototypes for testing the isolation in a single amplifier chain and two closely spaced amplifier chains

coordination between devices. But none of these solutions will sustain the oncoming traffic flow which is expected to boom within 4 to 6 years. Industry experts agree that the fifth-generation (5G) cellular technology will arrive by the end of this decade. The telecom operators will undoubtedly need new spectrum and will have to use millimeter-wave frequency band. Thus, the researchers are exploring more and more new concepts such as millimeter-wave massive MIMO where the base stations with very large number of antennas are envisioned to serve a set of relatively simpler co-channel user terminals. Digital beam forming combined with MIMO techniques is also considered to be more lucrative than conventional analog millimeter-wave phased arrays. These new systems will benefit from new waveguide and

electromagnetic (EM) packaging technologies, in order to ensure good performance with low power consumption, i.e., sustainability.

Cross-References

- ▶ Millimeter-Wave Antennas and Arrays
- ▶ MIMO Systems and Antennas for Terminals
- ▶ Substrate-Integrated Waveguide Antennas
- ▶ Transmission Lines
- ▶ Waveguide Slot Antennas and Arrays

References

Alos EA, Zaman AU, Kildal P (2013) Ka-band gap waveguide coupled-resonator filter for radio link diplexer application. Compon Packag Manuf Technol IEEE Trans 3(5):870–879. doi:10.1109/TCPMT.2012.2231140

Awida MH, Suleiman SH, Fathy AE (2011) Substrate-integrated cavity-backed patch arrays: a low-cost approach for bandwidth enhancement. Antennas Propag IEEE Trans 59(4):1155–1163. doi:10.1109/TAP.2011.2109681

Berenguer A, Baquero-Escudero M, Sanchez-Escuderos D, Bernardo-Clemente B, Boria-Esbert VE (2014) Low insertion loss 61 GHz narrow-band filter implemented with Groove Gap Waveguides. In: 44th European microwave conference, Rome, 2014

Borji A, Busuioc D, Safavi-Naeini S (2009) Efficient, low-cost integrated waveguide-fed planar antenna array for ku-band applications. Antennas Wirel Propag Lett IEEE 8:336–339. doi:10.1109/LAWP.2008.2004973

Bornemann J (2001) Design of waveguide filters without tuning elements for production-efficient fabrication by milling. In: APMC 2001. Asia-Pacific microwave conference, APMC - Taiwan, 3–6 Dec, 2001, vol 752, pp 759–762. doi:10.1109/APMC.2001.985481

Bornemann J, Arndt F (1987) Modal-S-matrix design of optimum stepped ridged and finned waveguide transformers. Microwave Theory Tech IEEE Trans 35(6):561–567. doi:10.1109/TMTT.1987.1133704

Bosiljevac M, Sipus Z, Kildal PS (2010) Construction of Green's functions of parallel plates with periodic texture with application to gap waveguides – a plane-wave spectral-domain approach. Microwaves Antennas Propag IET 4(11):1799–1810. doi:10.1049/iet-map.2009.0399, Barcelona

Bosiljevac M, Polemi A, Maci S, Sipus Z (2011) Analytic approach to the analysis of ridge and groove gap waveguides – comparison of two methods. In: Antennas and propagation (EUCAP), proceedings of the 5th European conference on, 11–15 Apr 2011, pp 1886–1889

Bozzi M, Georgiadis A, Wu K (2011) Review of substrate-integrated waveguide circuits and antennas. Microwaves Antennas Propag IET 5(8):909–920. doi:10.1049/iet-map.2010.0463

Brazalez AA, Zaman AU, Kildal PS (2012a) Improved microstrip filters using PMC packaging by lid of nails. Compon Packag Manuf Technol IEEE Trans 2(7):1075–1084. doi:10.1109/TCPMT.2012.2190931

Brazalez AA, Zaman AU, Kildal PS (2012b) Investigation of a Microstrip-to-Ridge Gap Waveguide transition by electromagnetic coupling. In: Antennas and Propagation Society international symposium (APSURSI), 2012 IEEE, 8–14 July 2012, pp 1–2. doi:10.1109/APS.2012.6349302

Chantraine-Bares B, Sauleau R (2007) Electrically-small shaped integrated lens antennas: a study of feasibility in Q-band. Antennas Propag IEEE Trans 55(4):1038–1044. doi:10.1109/TAP.2007.893377

Das NK, Pozar DM (1987) A generalized spectral-domain Green's function for multilayer dielectric substrates with application to multilayer transmission lines. Microwave Theory Tech IEEE Trans 35(3):326–335. doi:10.1109/TMTT.1987.1133646

Deslandes D, Ke W (2001) Integrated microstrip and rectangular waveguide in planar form. Microwave Wirel Compon Lett IEEE 11(2):68–70. doi:10.1109/7260.914305

Dixon P (2005) Cavity-resonance dampening. Microwave Mag IEEE 6(2):74–84. doi:10.1109/MMW.2005.1491270

Feng X, Ke W (2005) Guided-wave and leakage characteristics of substrate integrated waveguide. Microwave Theory Tech IEEE Trans 53(1):66–73. doi:10.1109/TMTT.2004.839303

Fonseca SDA, Giarola A (1984) Microstrip disk antennas, part II: the problem of surface wave radiation by dielectric truncation. Antennas Propag IEEE Trans 32(6):568–573. doi:10.1109/TAP.1984.1143367

Fujii S, Tsunemitsu Y, Yoshida G, Goto N, Zhang M, Hirokawa J, Ando M (2008) A wideband single-layer slotted waveguide array with an embedded partially corporate feed. In: Proceedings of the international symposium on antennas and propagation TP-C27-5, Sandiego

Gahete Arias C, Baquero Escudero M, Valero Nogueira A, Vila Jimenez A (2013) Test-fixture for suspended-strip gap-waveguide technology on ka-band. Microwave Wirel Compon Lett IEEE 23(6):321–323. doi:10.1109/LMWC.2013.2258000

Hansen CJ (2011) WiGiG: multi-gigabit wireless communications in the 60 GHz band. Wirel Commun IEEE 18(6):6–7. doi:10.1109/MWC.2011.6108325

Heejin K, Piljun P, Jaehoon C, Kyungwan Y, Jin-Dae K (1998) The design and analysis of a Ka-band coaxial to waveguide transition. In: Antennas and propagation society international symposium, Atlanta. IEEE, 21–26 June 1998, vol 521, pp 524–527. doi:10.1109/APS.1998.699193

Hirokawa J, Ando M (2000) Efficiency of 76-GHz post-wall waveguide-fed parallel-plate slot arrays. Antennas Propag IEEE Trans 48(11):1742–1745. doi:10.1109/8.900232

Hirokawa J, Kildal PS (1997) Excitation of an untilted narrow-wall slot in a rectangular waveguide by using etched strips on a dielectric plate. Antennas Propag IEEE Trans 45(6):1032–1037. doi:10.1109/8.585752

Hirokawa J, Ando M, Goto N (1992) Waveguide-fed parallel plate slot array antenna. Antennas Propag IEEE Trans 40(2):218–223. doi:10.1109/8.127406

Hui-Wen Y, Abdelmonem A, Ji-Fuh L, Zaki KA (1994) Analysis and design of microstrip-to-waveguide transitions. Microwave Theory Tech IEEE Trans 42(12):2371–2380. doi:10.1109/22.339769

Izzat N, Hilton GH, Railton CJ, Meade S (1996) Use of resistive sheets in damping cavity resonance. Electron Lett 32(8):721–722. doi:10.1049/el:19960535

Ke W (2006) Towards system-on-substrate approach for future millimeter-wave and photonic wireless applications. In: Microwave conference, 2006. APMC 2006. Asia-Pacific, Yokohama, 12–15 Dec 2006, pp 1895–1900. doi:10.1109/APMC.2006.4429778

Kehn MNM, Kildal PS, Skobelev SP (2004) Miniaturized dielectric-loaded rectangular waveguides for use in multi-frequency arrays. In: Antennas and Propagation Society international symposium, California. IEEE, 20–25 June 2004, vol 801, pp 803–806. doi:10.1109/APS.2004.1329792

Kehn MNM, Nannetti M, Cucini A, Maci S, Kildal PS (2006) Analysis of dispersion in dipole-FSS loaded hard rectangular waveguide. Antennas Propag IEEE Trans 54(8):2275–2282. doi:10.1109/TAP.2006.879198

Kildal PS (1990) Artificially soft and hard surfaces in electromagnetics. Antennas Propag IEEE Trans 38(10):1537–1544. doi:10.1109/8.59765

Kildal PS (2009) Three metamaterial-based gap waveguides between parallel metal plates for mm/submm waves. In: Antennas and propagation, 2009. EuCAP 2009. 3rd European conference on, Berlin, 23–27 Mar 2009, pp 28–32

Kildal PS, Kehn MNM (2010) The ridge gap waveguide as a wideband rectangular hard waveguide. In: Antennas and propagation (EuCAP), 2010 proceedings of the fourth European conference on Barcelona, 12–16 Apr 2010, pp 1–4

Kildal PS, Lier E (1988) Hard horns improve cluster feeds of satellite antennas. Electron Lett 24(8):491–492

Kildal PS, Kishk AA, Tengs A (1996) Reduction of forward scattering from cylindrical objects using hard surfaces. Antennas Propag IEEE Trans 44(11):1509–1520. doi:10.1109/8.542076

Kildal PS, Kishk A, Sipus Z (2007) RF invisibility using metamaterials: Harry Potters Cloak or The Emperors New Clothes? In: Antennas and Propagation Society international symposium, 2007 IEEE, 9–15 June 2007, pp 2361–2364. doi:10.1109/APS.2007.4396006

Kildal PS, Alfonso E, Valero-Nogueira A, Rajo-Iglesias E (2009) Local metamaterial-based waveguides in gaps between parallel metal plates. Antennas Wirel Propag Lett IEEE 8:84–87. doi:10.1109/LAWP.2008.2011147

Kildal PS, Zaman AU, Rajo-Iglesias E, Alfonso E, Valero-Nogueira A (2011) Design and experimental verification of ridge gap waveguide in bed of nails for parallel-plate mode suppression. Microwaves Antennas Propag IET 5(3):262–270. doi:10.1049/iet-map.2010.0089

Kimura Y, Hirano T, Hirokawa J, Ando M (2001) Alternating-phase fed single-layer slotted waveguide arrays with chokes dispensing with narrow wall contacts. Microwaves Antennas Propag IEE Proc 148(5):295–301. doi:10.1049/ip-map:20010645

Kirino H, Ogawa K (2012) A 76 GHz multi-layered phased array antenna using a non-metal contact metamaterial waveguide. Antennas Propag IEEE Trans 60(2):840–853. doi:10.1109/TAP.2011.2173112

Kishk A, Uz Zaman A, Kildal P-S (2012) Numerical prepackaging with PMC lid – efficient and simple design procedure for microstrip circuits including the packaging. ACES Appl Comput Soc J 27(5):389–398

Komanduri VR, Jackson DR, Williams JT, Mehrotra AR (2013) A general method for designing reduced surface wave microstrip antennas. Antennas Propag IEEE Trans 61(6):2887–2894. doi:10.1109/TAP.2013.2254441

Levine E, Malamud G, Shtrikman S, Treves D (1989) A study of microstrip array antennas with the feed network. Antennas Propag IEEE Trans 37(4):426–434. doi:10.1109/8.24162

Lier E (2010) Review of soft and hard horn antennas, including metamaterial-based hybrid-mode horns. Antennas Propag Mag IEEE 52(2):31–39. doi:10.1109/MAP.2010.5525564

Lier E, Kildal PS (1988) Soft and hard horn antennas. Antennas Propag IEEE Trans 36(8):1152–1157. doi:10.1109/8.7229

Liu D, Gaucher B, Pfeiffer U, Grzyb J (2009) Advanced millimeter-wave technologies: antennas, packaging and circuits. Wiley, Chichester/Hoboken. ISBN 9780470996171

Mall L, Waterhouse RB (2001) Millimeter-wave proximity-coupled microstrip antenna on an extended hemispherical dielectric lens. Antennas Propag IEEE Trans 49(12):1769–1772. doi:10.1109/8.982458

Martinez Giner S, Valero-Nogueira A, Herranz Herruzo JI, Baquero Escudero M (2013) Excitation of untilted narrow-wall slot in groove gap waveguide by using a parasitic dipole. In: Antennas and propagation (EuCAP), 2013 7th European conference on, Gothenburg, 8–12 Apr 2013, pp 3082–3085

McKinzie WE, Alexopoulos N (1992) Leakage losses for the dominant mode of conductor-backed coplanar waveguide. Microwave Guided Wave Lett IEEE 2(2):65–66. doi:10.1109/75.122412

Mesa F, Oliner AA, Jackson DR, Freire MJ (2000) The influence of a top cover on the leakage from microstrip line. Microwave Theory Tech IEEE Trans 48(12):2240–2248. doi:10.1109/22.898970

Miura Y, Hirokawa J, Ando M, Shibuya Y, Yoshida G (2011) Double-layer full-corporate-feed hollow-waveguide slot array antenna in the 60-GHz band. Antennas Propag IEEE Trans 59(8):2844–2851. doi:10.1109/TAP.2011.2158784

Nishikawa K, Sugitani S, Inoue K, Ishii T, Kamogawa K, Piernas B, Araki K (2001) Low-loss passive components on BCB-based 3D MMIC technology. In: Microwave symposium digest, 2001 I.E. MTT-S international, Phonix, 20–24 May 2001, vol 1883, pp 1881–1884. doi:10.1109/MWSYM.2001.967275

Ofli E, Vahldieck R, Amari S (2005) Novel E-plane filters and diplexers with elliptic response for millimeter-wave applications. Microwave Theory Tech IEEE Trans 53(3):843–851. doi:10.1109/TMTT.2004.842506

Polemi A, Maci S, Kildal PS (2011) Dispersion characteristics of a metamaterial-based parallel-plate ridge gap waveguide realized by bed of nails. Antennas Propag IEEE Trans 59(3):904–913. doi:10.1109/TAP.2010.2103006

Pozar DM (1983) Considerations for millimeter wave printed antennas. Antennas Propag IEEE Trans 31(5):740–747. doi:10.1109/TAP.1983.1143124

Pucci E, Rajo-Iglesias E, Kildal PS (2012) New microstrip gap waveguide on mushroom-type EBG for packaging of microwave components. Microwave Wirel Compon Lett IEEE 22(3):129–131. doi:10.1109/LMWC.2011.2182638

Pucci E, Zaman AU, Rajo-Iglesias E, Kildal PS, Kishk A (2013) Study of Q-factors of ridge and groove gap waveguide resonators. Microwaves Antennas Propag IET 7(11):900–908. doi:10.1049/iet-map.2013.0081

Pucci E, Rajo-Iglesias E, Vazquez-Roy JL, Kildal PS (2014) Planar dual-mode horn array with corporate-feed network in inverted microstrip gap waveguide. Antennas Propag IEEE Trans 62(7):3534–3542. doi:10.1109/TAP.2014.2317496

Rahiminejad S, Zaman AU, Pucci E, Raza H, Vassilev V, Haasl S, Lundgren P, Kildal PS, Enoksson P (2012) Micromachined ridge gap waveguide and resonator for millimeter-wave applications. Sens Actuators A Phys 186:264–269. doi:10.1016/j.sna.2012.02.036

Rajo-Iglesias E, Kildal PS (2010) Groove gap waveguide: a rectangular waveguide between contactless metal plates enabled by parallel-plate cut-off. In: Antennas and propagation (EuCAP), 2010 proceedings of the fourth European conference on, Barcelona, 12–16 Apr 2010, pp 1–4

Rajo-Iglesias E, Kildal PS (2011) Numerical studies of bandwidth of parallel-plate cut-off realised by a bed of nails, corrugations and mushroom-type electromagnetic bandgap for use in gap waveguides. Microwaves Antennas Propag IET 5(3):282–289. doi:10.1049/iet-map.2010.0073

Rajo-Iglesias E, Kildal PS, Zaman AU, Kishk A (2012) Bed of springs for packaging of microstrip circuits in the microwave frequency range. Compon Packag Manuf Technol IEEE Trans 2(10):1623–1628. doi:10.1109/TCPMT.2012.2207957

Rajo-Iglesias E, Pucci E, Kishk AA, Kildal P (2013) Suppression of parallel plate modes in low frequency microstrip circuit packages using lid of printed zigzag wires. Microwave Wirel Compon Lett IEEE 23(7):359–361. doi:10.1109/LMWC.2013.2265257

Raza H, Jian Y, Kildal PS, Alfonso E (2013) Resemblance between gap waveguides and hollow waveguides. Microwaves Antennas Propag IET 7(15):1221–1227. doi:10.1049/iet-map.2013.0178

Raza H, Yang J, Kildal PS, Alfonso Alos E (2014) Microstrip-ridge gap waveguide: study of losses, bends, and transition to WR-15. Microwave Theory Tech IEEE Trans 62(9):1943–1952. doi:10.1109/TMTT.2014.2327199

Razavi A, Kildal P-S, Liangliang X, Alfonso E, Chen H (2014) 2x2-slot element for 60GHz planar array antenna realized on two doubled-sided PCBs using SIW cavity and EBG-type soft surface fed by microstrip-ridge gap waveguide. Antennas Propag IEEE Trans 99:1–1. doi:10.1109/TAP.2014.2331993

Rebollo A, Gonzalo R, Ederra I (2014) Optimization of a pin surface as a solution to suppress cavity modes in a packaged W-band microstrip receiver. Compon Packag Manuf Technol IEEE Trans 4(6):975–982. doi:10.1109/TCPMT.2014.2312252

Schurig D, Mock JJ, Justice BJ, Cummer SA, Pendry JB, Starr AF, Smith DR (2006) Metamaterial electromagnetic cloak at microwave frequencies. Science 314(5801):977–980. doi:10.1126/science.1133628

SeHyun P, Tsunemitsu Y, Hirokawa J, Ando M (2006) Center feed single layer slotted waveguide array. Antennas Propag IEEE Trans 54(5):1474–1480. doi:10.1109/TAP.2006.874310

Shafique MF, Robertson ID (2011) Laser prototyping of multilayer LTCC microwave components for system-in-package applications. Microwaves Antennas Propag IET 5(8):864–869. doi:10.1049/iet-map.2010.0352

Sievenpiper D, Lijun Z, Broas RFJ, Alexopolous NG, Yablonovitch E (1999) High-impedance electromagnetic surfaces with a forbidden frequency band. Microwave Theory Tech IEEE Trans 47(11):2059–2074. doi:10.1109/22.798001

Silveirinha MG, Fernandes CA, Costa JR (2008) Electromagnetic characterization of textured surfaces formed by metallic pins. Antennas Propag IEEE Trans 56(2):405–415. doi:10.1109/TAP.2007.915442

Sipus Z, Merkel H, Kildal PS (1997) Green's functions for planar soft and hard surfaces derived by asymptotic boundary conditions. Microwaves Antennas Propag IEE Proc 144(5):321–328. doi:10.1049/ip-map:19971335

Six G, Prigent G, Rius E, Dambrine G, Happy H (2005) Fabrication and characterization of low-loss TFMS on silicon substrate up to 220 GHz. Microwave Theory Tech IEEE Trans 53(1):301–305. doi:10.1109/TMTT.2004.839915

Skobelev SP, Kildal PS (2005) Mode-matching modeling of a hard conical quasi-TEM horn realized by an EBG structure with strips and vias. Antennas Propag IEEE Trans 53(1):139–143. doi:10.1109/TAP.2004.840417

Smulders P (2002) Exploiting the 60 GHz band for local wireless multimedia access: prospects and future directions. Commun Mag IEEE 40(1):140–147. doi:10.1109/35.978061

Sotoudeh O, Kildal PS, Ingvarson P, Skobelev SP (2006) Single- and dual-band multimode hard horn antennas with partly corrugated walls. Antennas Propag IEEE Trans 54(2):330–339. doi:10.1109/TAP.2005.863389

Swanson DG (2007) Narrow-band microwave filter design. Microwave Mag IEEE 8(5):105–114. doi:10.1109/MMM.2007.904724

Tsugawa T, Sugio Y, Yamada Y (1997) Circularly polarized dielectric-loaded planar antenna excited by the parallel feeding waveguide network. Broadcast IEEE Trans 43(2):205–212. doi:10.1109/11.598371

Tsuji M, Shigesawa H, Oliner AA (1995) Simultaneous propagation of bound and leaky dominant modes on printed-circuit lines. IEEE Trans Microwave Theory Tech 43(12):3007–3019

Tze-Min S, Chi-Feng C, Huang T-Y, Wu R-B (2007) Design of vertically stacked waveguide filters in LTCC. Microwave Theory Tech IEEE Trans 55(8):1771–1779. doi:10.1109/TMTT.2007.902080

Uher J, Bornemann J, Rosenberg U (1993) Waveguide components for antenna feed systems: theory and CAD. Artec House, Norwood

Valero-Nogueira A, Alfonso E, Herranz JI, Kildal PS (2009) Experimental demonstration of local quasi-TEM gap modes in single-hard-wall waveguides. Microwave Wirel Compon Lett IEEE 19(9):536–538. doi:10.1109/LMWC.2009.2027051

Valero-Nogueira A, Baquero M, Herranz JI, Domenech J, Alfonso E, Vila A (2011) Gap waveguides using a suspended strip on a bed of nails. Antennas Wirel Propag Lett IEEE 10:1006–1009. doi:10.1109/LAWP.2011.2167591

Valero-Nogueira A, Herranz-Herruzo JI, Baquero M, Hernandez-Murcia R, Rodrigo V (2013) Practical derivation of slot equivalent admittance in periodic waveguides. Antennas Propag IEEE Trans 61(4):2321–2324. doi:10.1109/TAP.2012.2231934

Wei W, Yang J, Ostling T, Schafer T (2011) New hat feed for reflector antennas realised without dielectrics for reducing manufacturing cost and improving reflection coefficient. Microwaves Antennas Propag IET 5(7):837–843. doi:10.1049/iet-map.2010.0181

Williams DF (1989) Damping of the resonant modes of a rectangular metal package [MMICs]. Microwave Theory Techn IEEE Trans 37(1):253–256. doi:10.1109/22.20046

Williamson AG (1985) Coaxially fed hollow probe in a rectangular waveguide. Microwaves Antennas Propag IEE Proc H 132(5):273–285. doi:10.1049/ip-h-2.1985.0051

Xiao-Ping C, Ke W, Liang H, Fanfan H (2010) Low-cost high gain planar antenna array for 60-GHz band applications. Antennas Propag IEEE Trans 58(6):2126–2129. doi:10.1109/TAP.2010.2046861

Xu JF, Hong W, Chen P, Wu K (2009) Design and implementation of low sidelobe substrate integrated waveguide longitudinal slot array antennas. Microwaves Antennas Propag IET 3(5):790–797. doi:10.1049/iet-map.2008.0157

Yunchi Z, Ruiz-Cruz JA, Zaki KA, Piloto AJ (2010) A waveguide to microstrip inline transition with very simple modular assembly. Microwave Wirel Compon Lett IEEE 20(9):480–482. doi:10.1109/LMWC.2010.2056358

Zaman AU, Kildal PS (2012) Slot antenna in ridge gap waveguide technology. In: Antennas and propagation (EUCAP), 2012 6th European conference on, Prague, 26–30 Mar 2012, pp 3243–3244. doi:10.1109/EuCAP.2012.6206129

Zaman AU, Kildal PS (2014) Wide-band slot antenna arrays with single-layer corporate-feed network in ridge gap waveguide technology. Antennas Propag IEEE Trans 62(6):2992–3001. doi:10.1109/TAP.2014.2309970

Zaman AU, Rajo-Iglesias E, Alfonso E, Kildal PS (2009) Design of transition from coaxial line to ridge gap waveguide. In: Antennas and Propagation Society international symposium, 2009. APSURSI '09, North Charleston. IEEE, 1–5 June 2009, pp 1–4. doi:10.1109/APS.2009.5172186

Zaman AU, Jian Y, Kildal PS (2010a) Using lid of pins for packaging of microstrip board for descrambling the ports of eleven antenna for radio telescope applications. In: Antennas and Propagation Society international symposium (APSURSI), 2010 IEEE, Toronto, 11–17 July 2010, pp 1–4. doi:10.1109/APS.2010.5561211

Zaman AU, Kildal PS, Ferndahl M, Kishk A (2010b) Validation of ridge gap waveguide performance using in-house TRL calibration kit. In: Antennas and propagation (EuCAP), 2010 proceedings of the fourth European conference on, Barcelona, 12–16 Apr 2010, pp 1–4

Zaman AU, Vassilev V, Kildal PS, Kishk A (2011) Increasing parallel plate stop-band in gap waveguides using inverted pyramid-shaped nails for slot array application above 60 GHz. In: Antennas and propagation (EUCAP), proceedings of the 5th European conference on, 11–15 Apr 2011, pp 2254–2257

Zaman AU, Ellis MS, Kildal PS (2012a) Metamaterial based packaging method for improved isolation of circuit elements in microwave modules. In: Microwave integrated circuits conference (EuMIC), 2012 7th European, Amsterdam, 29–30 Oct 2012, pp 834–837

Zaman AU, Kildal PS, Kishk AA (2012b) Narrow-band microwave filter using high-Q groove gap waveguide resonators with manufacturing flexibility and no sidewalls. Compon Packag Manuf Technol IEEE Trans 2(11):1882–1889. doi:10.1109/TCPMT.2012.2202905

Zaman AU, Vukusic T, Alexanderson M, Kildal PS (2013) Design of a simple transition from microstrip to ridge gap waveguide suited for MMIC and antenna integration. Antennas Wirel Propag Lett IEEE 12:1558–1561. doi:10.1109/LAWP.2013.2293151

Zaman AU, Alexanderson M, Vukusic T, Kildal PS (2014) Gap waveguide PMC packaging for improved isolation of circuit components in high-frequency microwave modules. Compon Packag Manuf Technol IEEE Trans 4(1):16–25. doi:10.1109/TCPMT.2013.2271651

Zwick T, Duixian L, Gaucher BP (2006) Broadband planar superstrate antenna for integrated millimeterwave transceivers. Antennas Propag IEEE Trans 54(10):2790–2796. doi:10.1109/TAP.2006.882167

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_133-1

Impedance Matching and BALUNs

Homayoon Oraizi*
College of Electrical Engineering, Iran University of Science and Technology, Tehran, Tehran, Iran

Abstract

In this chapter, the subject matters of impedance transformation and BALUNs are studied, which should be designed for inclusion in the input ports of transmitter and receiver antennas. The Smith chart is commonly used for the design of lumped and distributed circuits for impedance matching in the technical journals and textbooks. Its explanations and usage are readily accessible in the available textbooks. However, the method of least squares is developed here for the design and optimization of various circuits of impedance transformers for microwave and higher-frequency circuits.

The topics discussed here include the following:

The concept of impedance, transmission lines, power gains, varieties of matching networks, impedance transformer design by the method of least squares, the quarter-wave line, theory of small reflections, multi-section transformers, design of step-line transformers, design of taper lines, devices and components for impedance matching, and BALUNs including waveguide and planar circuit implementations.

Keywords

Matched load; Matched source; Maximum load power; Power gains; Impedance transformer; Impedance matching; Method of least squares; Quarter-wave line; Small reflections; Multi-section transformers; Step-line transformers; Taper lines; BALUNs

Introduction

Impedance matching networks are placed at the source and load sides of a network to achieve various system requirements, such as maximum power transfer, maximization of power handling capacity, minimum reflection or reflectionless operation, minimization of noise interference, minimization of power loss, linearization of frequency response, and most importantly matching and transformation of impedance levels. Discrete and lumped components may be used at frequencies lower than 1 GHz, but distributed circuits and transmission lines should be used at microwave frequencies higher than 1 GHz. Analytical methods and graphical procedures such as the Smith chart may be used for the design of impedance transformation networks, which may both be implemented by computer programs. Solutions by analytical equations are quite precise, whereas solution by graphical procedure may give better insight into the physical behavior of circuits. Mixed matching networks combining lumped elements and distributed circuits may also be designed for impedance transformations.

There are different engineering factors to be considered for the selection of a matching circuit, such as DC biasing, stability, frequency response, quality factor, and also availability of discrete components with required values and substrates with suitable dielectric constants. The design of matching circuits for

*Email: h_oraizi@iust.ac.ir

DOI 10.1007/978-981-4560-75-7_133-1

oscillators and amplifiers requires high- and low-quality factors, respectively. L circuits cannot provide high Q; rather higher-order circuits are needed, which have more elements, such as T and π-networks.

The radiation resistance, antenna radiation efficiency, radiation pattern, and gain of an antenna are not affected by its impedance mismatch. Matching serves to transfer the available power from the transmission line into the antenna, reduces transmission loss, and minimizes standing voltage on the line.

What Is the Concept of Impedance, Resistance, and Reactance?

Impedance as an engineering concept was originally defined as resistance for direct current (DC) electric circuits, but it was then extended and generalized to other electromagnetic phenomena and other engineering problems (Valkenburg 1991). In a similar manner, current has also been generalized from the original motion of charged particles to include conduction $\left(\tilde{J} = \sigma\tilde{E}\right)$, convection $\left(\tilde{J} = \rho\tilde{v}, \tilde{J} = k\tilde{v}, I = \lambda\tilde{v}\right)$, electric displacement $\left(\tilde{J}_D = \frac{\partial\tilde{D}}{\partial t}\right)$, magnetic displacement $\left(\tilde{M}_D = \frac{\partial\tilde{B}}{\partial t}\right)$, electric current sources (J^i, J_s^i, I^i), and magnetic voltage sources (M^i, M_s^i, K^i). Ohm's law, considered here as a constitutive relation for conductive media, is expressed as

$$\tilde{J} = \sigma\tilde{E} \tag{1}$$

where $\tilde{J}$ is the electric volume current density (amperes/meter^2), σ is the conductivity of the medium (Simens/meter), and $\tilde{E}$ is the electric field intensity (volts/meter). For linear, homogeneous, and isotropic matter (simple matter), σ is not a function of E and coordinate axes and has identical behavior in all directions.

Consider now a cylindrical object made of linear, homogeneous, and isotropic conductive matter (with conductivity σ) having a uniform cross section (S) and length (ℓ). The voltage drop across its length is

$$v = \int_c \vec{E} \cdot \vec{d\ell} = E\ell = \frac{J\ell}{\sigma} \tag{2}$$

The current through it along its axis is

$$i = \int_S \vec{J} \cdot \hat{n} ds = JS \tag{3}$$

Then, Ohm's law as the relation between v and i is expressed as

$$v = \left(\frac{\ell}{\sigma S}\right) i = Ri \tag{4}$$

where the resistance of the conductive cylinder is

$$R = \frac{\ell}{\sigma S} \tag{5}$$

Equation (4) is the governing relation for resistors. The resistance (R) is defined as the impediment of resistor toward the motion of charged particles (here, free electrons in the metal) due to the voltage

DOI 10.1007/978-981-4560-75-7_133-1

difference across the resistor. The voltage difference is actually defined as the difference of potential energy per unit charge:

$$\Delta v = \frac{\Delta W}{q} = \frac{\vec{F} \cdot \vec{\Delta \ell}}{q} = \vec{E} \cdot \vec{\Delta \ell}$$

Similarly, consider a capacitor, which is a component to store charge and electric energy. The constitutive relation for dielectric media is

$$\tilde{D} = \varepsilon \tilde{E} \tag{6}$$

where $\tilde{D}$ is the electric flux density (Coulombs/meter^2), ε is the permittivity (Farad/meter), and $\tilde{E}$ is the electric field intensity (volts/meter or Newton/Coulomb).

Consider now a capacitor made of two parallel perfect electric conductor plates having area (S) and separated by (d) filled by a linear, homogeneous, and isotropic dielectric matter (ε). According to Gauss' law and definition of voltage difference expressed in Eq. (2) (assuming uniform electric field and ignoring fringing effects),

$$q = \oint_S \tilde{D} \cdot \hat{n} ds = \oint_S \varepsilon \tilde{E} \cdot \hat{n} ds = \varepsilon S E \tag{7}$$

$$v = \int \tilde{E} \cdot \vec{d\ell} = Ed \tag{8}$$

The definition of capacitance is then used as

$$C = \frac{q}{v} = \frac{\varepsilon S}{d} \text{ (Farad or Coulombs/volt)} \tag{9}$$

Note that q is the total charge induced on one plate.

Considering the definition of electric current as the passage of charge increment (Δq) during a time interval (Δt) through a point,

$$i = \frac{dq}{dt} \tag{10}$$

Also

$$i = \frac{d}{dt}(Cv) = C\frac{dv}{dt} \tag{11}$$

where capacitance C is here assumed time invariant. Also

$$v = \frac{1}{C} \int_{-\infty}^{t} i(t) dt \tag{12}$$

DOI 10.1007/978-981-4560-75-7_133-1

Equations (11) and (12) are the governing relations for capacitors.

Consider now a coil of N turns wound on a cylindrical core having a uniform cross section of area (S) and length (l) filled by a linear, homogeneous, and isotropic magnetic medium having the constitutive relation

$$\tilde{B} = \mu\tilde{H} \tag{13}$$

The density of turns in the coil is

$$n = \frac{N}{l}\,(\text{turns/meter})$$

The magnetic flux density B(Tesla, Weber/meter^2) passing through it is

$$\tilde{B} = \mu n i \tag{14}$$

and the magnetic flux ϕ(Weber) passing through it is

$$\phi = \int_S \tilde{B} \cdot \hat{n} ds = \tilde{B}S = \mu n S i \tag{15}$$

The induced voltage in the solenoid is equal to the time derivative of the flux linkage in it due to Faraday's law:

$$v = \frac{\partial \phi}{\partial t} = \mu n S \frac{di}{dt} \tag{16}$$

Here, Lenz's law determines the sense of induced voltage.

On the other hand, inductance L(Henry) is defined as

$$L = \frac{\phi}{i} \tag{17}$$

Consequently, the induced voltage across the inductor is

$$v = \frac{d\phi}{dt} = L\frac{di}{dt} \tag{18}$$

where the self-inductance of the winding is

$$L = \mu n S \tag{19}$$

The following correspondences are now observed among the circuit elements, namely, resistor (R), capacitor (C), and inductor (L):

DOI 10.1007/978-981-4560-75-7_133-1

$$
\begin{aligned}
& v = Ri \quad , \quad i = Gv \\
& q = Cv \ , \ i = \frac{dq}{dt} = C\frac{dv}{dt} \quad , \quad v = \frac{1}{C}\int_{-\infty}^{t} idt, \quad \mathrm{q} = \int_{-\infty}^{t} idt \\
& \phi = Li \quad , \ v = L\frac{di}{dt} \qquad\quad , \quad i = \frac{1}{L}\int_{-\infty}^{t} vdt
\end{aligned}
\tag{20}
$$

The energy and power of the resistor, inductor, and capacitor are

$$
\begin{aligned}
& P_R = \frac{dW}{dt} = vi = Ri^2 \quad , \quad W_R = \int_{-\infty}^{t} vidt = \int_{-\infty}^{t} Ri^2 dt \\
& P_C = Cv\frac{dv}{dt} \qquad\quad , \quad W_C = \int_{-\infty}^{t} Cv\frac{dv}{dt}dt = \frac{1}{2}Cv^2 \\
& P_L = Li\frac{di}{dt} \qquad\quad , \quad W_L = \int_{-\infty}^{t} Li\frac{di}{dt}dt = \frac{1}{2}Li^2
\end{aligned}
\tag{21}
$$

For harmonic sources, the governing equations $\left(with\ \frac{\partial}{\partial t} = j\omega\right)$ for the resistors, inductors, and capacitors are

$$
\begin{aligned}
& V = RI \qquad\quad , \ P_R = \frac{1}{2}R|I|^2 \\
& I = (j\omega C)V = jB_C V \quad , \quad P_C = \frac{1}{2}B_C|V|^2 \ , \ W_C = \frac{1}{4}CV^2 \\
& V = (j\omega L)I = jX_L I \quad , \quad P_L = \frac{1}{2}X_L|I|^2 \ , \ W_L = \frac{1}{4}LI^2
\end{aligned}
\tag{22}
$$

The input impedance of a general one-port network may be indicated as

$$Z(\omega) = R(\omega) + jX(\omega) \tag{23}$$

which is composed of real and imaginary parts denoted as the resistive and reactive parts, respectively.

However, the concept of impedance has evolved from the basic DC electric circuits to other electromagnetic phenomena.

The intrinsic impedance of a medium is

$$\eta = \sqrt{\hat{z}/\hat{y}} \tag{24}$$

where impedivity is defined as

$$\hat{z} = j\omega\hat{\mu} = j\omega(\mu' - j\mu'') = \omega\mu'' + j\omega\mu' \ \ (\Omega/m) \tag{25}$$

and admittivity as

DOI 10.1007/978-981-4560-75-7_133-1

$$\hat{y} = \sigma + j\omega\hat{\varepsilon} = \sigma + j\omega(\varepsilon' - j\varepsilon'') = \sigma + \omega\varepsilon'' + j\omega\varepsilon' \quad (S/m) \tag{26}$$

The plane wave in a homogeneous medium is expressed as

$$\vec{H} = \frac{1}{\eta}\hat{n} \times \vec{E} \tag{27}$$

where $\vec{E}$ and $\vec{H}$ as the electric and magnetic field intensities are perpendicular to the direction of propagation in space $(\hat{n})$.

The transverse wave components in the cross section of a uniform waveguide are related by the wave impedance (Z_w).

$$\vec{H}_t = \frac{1}{Z_w}\hat{u}_z \times \vec{E}_t \tag{28}$$

where Z_w is Z_{TEM}, Z_{TE}, or Z_{TM} for the TEM, TE, or TM modes, respectively.

The forward and backward traveling wave voltages and currents on a transmission line are related by the characteristic impedance (Z_0) of the line, namely,

$$\frac{V^+}{I^+} = \frac{V^-}{I^-} = Z_0 \tag{29}$$

The input impedance in one direction on a line is actually defined by the equation of impedance transformation

$$Z_{in} = \frac{V}{I} = \frac{V^+ + V^-}{I^+ + I^-} \tag{30}$$

where the total voltage and current are the sum of forward and backward traveling voltages and currents, respectively.

The impedance matching from the perspective of electromagnetic fields is the matching of tangential electric fields and tangential magnetic fields at a boundary surface (having no surface currents), whereas from the perspective of circuit theory, it is the matching of impedances based on some specifications.

Connection of the Source to Load Through a Transmission Line

Consider now the case of a source connected to a load through a transmission line and study the behavior of power transfers between them (Rizzi 1988; Pozar 2011; Collin 2000; Gonzalez 1997). The length, characteristic impedance, and propagation constant of the line are l, Z_0, and $\gamma = \alpha + j\beta$ (with attenuation constant α and phase constant β), respectively. The source and load impedances are Z_S and Z_L.

The voltage and current on the line may be written as the sum of forward and backward traveling waves (Fig. 1).

$$V(z) = V_0^+(e^{-\gamma z} + \Gamma_L e^{\gamma z}) \tag{31}$$

DOI 10.1007/978-981-4560-75-7_133-1

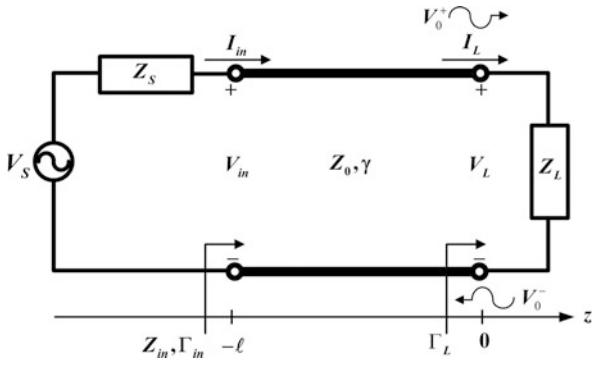


Fig. 1 The transmission line connecting a source to a load

$$I(z) = \frac{V_0^+}{Z_0}(e^{-\gamma z} - \Gamma_L e^{\gamma z}) \tag{32}$$

where the reflection coefficient at the load is defined as

$$\Gamma_L = \frac{V_0^-}{V_0^+} = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{33}$$

Using Eqs. (31), (32), and (33), the equation of impedance transformation on the line is obtained:

$$Z_{in} = \frac{V(z)}{I(z)} = Z_0 \frac{Z_L + Z_0 \tanh(\gamma l)}{Z_0 + Z_L \tanh(\gamma l)} \tag{34}$$

For lossless lines $(\alpha = 0)$

$$Z_{in} = \frac{V(z)}{I(z)} = Z_0 \frac{Z_L + jZ_0 \tan(\beta l)}{Z_0 + jZ_L \tan(\beta l)} \tag{35}$$

The two boundary conditions at the load and source sides of the line are

$$V_L = I_L Z_L \text{ and } V_S = I_{in}(Z_S + Z_{in}) \tag{36}$$

The voltage at the source end of the line is

$$V(z = -l) = V_{in} = I_{in} Z_{in} = V_S \frac{Z_{in}}{Z_{in} + Z_S} = V_0^+ \left(e^{\gamma l} + \Gamma_L e^{-\gamma l}\right) \tag{37}$$

Then, the forward traveling voltage amplitude is

$$V_0^+ = V_S \frac{Z_{in}}{Z_{in} + Z_S} \frac{1}{e^{\gamma l} + \Gamma_L e^{-\gamma l}} \tag{38}$$

This equation may be written as

$$V_0^+ = V_S \frac{Z_0}{Z_0 + Z_S} \frac{e^{-\gamma l}}{(1 - \Gamma_L \Gamma_S e^{-2\gamma l})} \tag{39}$$

by using Eq. (34) and the reflection coefficient at the source defined as

$$\Gamma_S = \frac{Z_S - Z_0}{Z_S + Z_0} \tag{40}$$

The voltage amplitude V_0^+ may be used in Eqs. (31) and (32) to obtain the voltage and current on the line. The standing wave ratio on the lossless line is

$$SWR = \frac{V_{\max}}{V_{\min}} = \frac{1 + |\Gamma_L|}{1 - |\Gamma_L|} \tag{41}$$

The power delivered to the load on a lossless line is

$$P_L = \mathrm{Re}\left(\frac{1}{2} V_{in} I_{in}^*\right) = \mathrm{Re}\left(\frac{1}{2} V_L I_L^*\right) = \frac{1}{2}|V_{in}|^2 \mathrm{Re}\left(\frac{1}{Z_{in}^*}\right) = \frac{1}{2}|V_S|^2 \left|\frac{Z_{in}}{Z_{in} + Z_S}\right|^2 \mathrm{Re}\left(\frac{1}{Z_{in}}\right) \tag{42}$$

which may be written as

$$P_L = \frac{1}{2}|V_S|^2 \frac{R_{in}}{(R_{in} + R_S)^2 + \left(X_{in} + {}^{\mathrm{in}}X_S\right)^2} \tag{43}$$

for $Z_{in} = R_{in} + jX_{in}$ and $Z_S = R_S + jX_S$.

Three cases of load impedance may be considered for a fixed source impedance (Z_S).

Matched Load

For the case of the load matched to the line, the relevant parameters are $Z_L = Z_0$, $\Gamma_L = 0$, $SWR = 1$, $Z_{in} = Z_0$, $R_{in} = Z_0$, and $X_{in} = 0$. Therefore, the load power is obtained from Eq. (42) as

$$P_L = \frac{1}{2}|V_S|^2 \frac{Z_0}{\left(Z_0 + {}^{0}R_S\right)^2 + {X_S}^2} \tag{44}$$

Matched Source

For the case of the source matched to the loaded line (Z_L) by choosing appropriate values for Z_L, βl, and Z_0, so that $Z_{in} = Z_S$, $R_{in} = R_S$, $\Gamma_{in} = 0$, and $\Gamma_L \neq 0$. The load power is then given from Eq. (42):

$$P_L = \frac{1}{8}|V_s|^2 \frac{R_S}{\left(R_S^2 + {X_S}^2\right)} \tag{45}$$

Observe that the load power in the case of matched source may be less than the case of matched load.

Maximum Load Power

The condition for the maximum power transfer from the source to the load is then determined. For a fixed source impedance, but variable input impedance $Z_{in} = R_{in} + jX_{in}$, the required conditions are

DOI 10.1007/978-981-4560-75-7_133-1

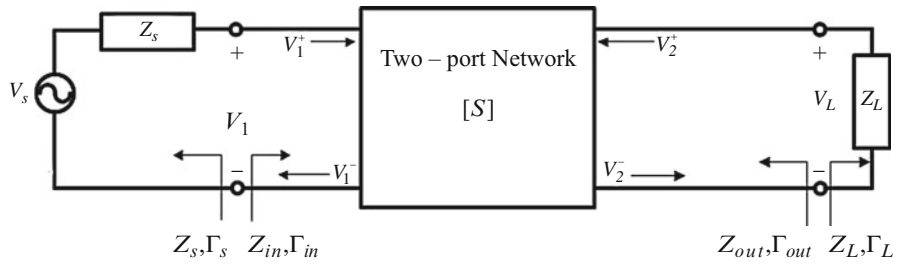


Fig. 2 A two-port network connecting a source to a load

$$\frac{\partial P_L}{\partial R_{in}} = 0 \text{ and } \frac{\partial P_L}{\partial X_{in}} = 0 \tag{46}$$

which lead to two relations $R_{in} = R_S$ and $X_{in} = -X_S$, namely,

$$Z_{in} = Z_S^* \tag{47}$$

Consequently, the conjugate matching provides the condition for maximum power transfer to the load, which is equal to (from Eq. (42))

$$P_L = \frac{1}{8}|V_s|^2 \frac{1}{R_S} \tag{48}$$

The reflection coefficients in this case, namely, Γ_L, Γ_S, and Γ_{in}, are not necessarily zero. Therefore, for maximum power transfer, there may be reflections on the line, which on a mismatched line may actually add up to increase the power transfer to the load.

Observe that the two cases of reflectionless matching $(Z_{in} = Z_S)$ and conjugate matching $\left(Z_{in} = Z_S^*\right)$ do not provide a circuit with optimum efficiency. The power dissipated in the source impedance is

$$P_{ds} = \frac{1}{2}R_S|I_S|^2 = \frac{1}{2}R_S|V_S|^2 \frac{1}{(R_{in} + R_S)^2 + \left(X_{in} + {}^{\text{in}}X_S\right)^2} \tag{49}$$

Now, for the case of the line matched to the source and load (no reflections), the conditions are $Z_L = {}^{L}Z_S = Z_0$ (and $X_S = X_L = X_{in} = 0$, $R_{in} = R_S = R_L = Z_0$)

$$P_{ds} = \frac{1}{2}R_S|V_S|^2 \frac{1}{4R_S^2} = \frac{|V_S|^2}{8R_S} \tag{50}$$

Therefore, half the power is lost in the source impedance (since in this case the source power is $P_S = |V_S|^2/4R_S$), resulting in an efficiency of 50 %.

Power Gains of Two-Port Networks

Consider a two-port network connected to a source and a load, as shown in Fig. 2 (Rizzi 1988; Pozar 2011; Collin 2000; Gonzalez 1997).

Three types of gain are distinguished:

DOI 10.1007/978-981-4560-75-7_133-1

1. Power gain

$$G = \frac{P_L}{P_{in}} = \frac{\text{Load Power}}{\text{Input power to network}} \tag{51}$$

G is independent of Z_S.

2. Available power gain

$$G_A = \frac{P_{avn}}{P_{avs}} = \frac{\text{Power available from network}}{\text{Power available from source}} \tag{52}$$

G_A assumes conjugate matching of both the source and the load. It depends on Z_S but not Z_L.

3. Transducer power gain

$$G_T = \frac{P_L}{P_{avs}} = \frac{\text{Load Power}}{\text{Power available from source}} \tag{53}$$

The definitions of power gains depend on the method of source and load matching. For example, if they are both conjugate matched, then maximum gain is achieved and $G = G_A = G_T$.

Assuming a reference characteristic impedance (Z_0), the reflection coefficients Γ_L and Γ_S are given in Eqs. (33) and (39), respectively. Since the input port of network is generally mismatched, its input reflection coefficient (Γ_{in}) (see Fig. 2) is then derived.

Using the scattering parameters and $\Gamma_L = V_2^+/V_2^-$

$$V_1^- = S_{11}V_1^+ + S_{12}V_2^+ = S_{11}V_1^+ + S_{12}\Gamma_L V_2^- \tag{54a}$$

$$V_2^- = S_{21}V_1^+ + S_{22}V_2^+ = S_{21}V_1^+ + S_{22}\Gamma_L V_2^- \tag{54b}$$

$$\Gamma_{in} = \frac{Z_{in} - Z_0}{Z_{in} + Z_0} = \frac{V_1^-}{V_1^+} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \tag{55}$$

Similarly, for the output reflection coefficient with $\Gamma_S = \frac{V_1^+}{V_1^-}$

$$\Gamma_{out} = \frac{V_2^-}{V_2^+} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S} \tag{56}$$

At the input port of network

$$V_1 = V_S\frac{Z_{in}}{Z_{in} + Z_S} = V_1^+ + V_1^- = V_1^+(1 + \Gamma_{in}) \tag{57}$$

Therefore,

$$V_1^+ = \frac{V_S}{2}\frac{1 - \Gamma_S}{1 - \Gamma_S\Gamma_{in}} \tag{58}$$

where $Z_{in} = Z_0(1+\Gamma_{in})/(1-\Gamma_{in})$ from Eq. (54b) is used.

The average power delivered to the network is

$$P_{in} = \frac{1}{2Z_0}\left|V_1^+\right|^2\left(1-|\Gamma_{in}|^2\right) = \frac{|V_S|^2}{8Z_0}\frac{|1-\Gamma_S|^2}{|1-\Gamma_S\Gamma_{in}|^2}\left(1-|\Gamma_{in}|^2\right) \tag{59}$$

The load power is

$$P_L = \frac{1}{2Z_0}\left|V_2^-\right|^2\left(1-|\Gamma_L|^2\right) \tag{60}$$

Substitute V_2^- from Eq. (54a) and V_1^+ from Eq. (57) in Eq. (59) to determine the load power

$$P_L = \frac{|V_S|^2}{8Z_0}|S_{21}|^2\frac{|1-\Gamma_S|^2}{|1-\Gamma_S\Gamma_{in}|^2}\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} \tag{61}$$

The power gain is now

$$G = \frac{P_L}{P_{in}} = |S_{21}|^2\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2}\frac{1}{1-|\Gamma_{in}|^2} \tag{62}$$

The maximum deliverable power to the network (P_{in}) is equal to the available power from the source (P_{avs}), which is achieved for the case of conjugate matching between the source and the input of network, namely, $Z_{in} = Z_S^*$ or $\Gamma_{in} = \Gamma_S^*$. Then, from Eq. (58)

$$P_{avs} = P_{in}\big|_{\Gamma_{in}=\Gamma_S^*} = \frac{|V_S|^2}{8Z_0}\frac{|1-\Gamma_S|^2}{\left(1-|\Gamma_S|^2\right)} \tag{63}$$

The maximum deliverable power to the load (P_L) is equal to the available power from the network (P_{avn}), which is achieved for conjugate matching between the output of network and the load, namely, $\Gamma_L = \Gamma_{out}^*$ and $Z_L = Z_{out}^*$. Therefore, the following relation is obtained from Eq. (60):

$$P_{avn} = P_L\big|_{\Gamma_L=\Gamma_{out}^*} = \frac{|V_S|^2}{8Z_0}|S_{21}|^2\frac{|1-\Gamma_S|^2}{|1-\Gamma_S\Gamma_{in}|^2}\frac{1-|\Gamma_{out}|^2}{\left|1-S_{22}\Gamma_{out}^*\right|^2} \tag{64}$$

In this equation for P_{avn}, the input reflection coefficient Γ_{in} must be evaluated for $\Gamma_L = \Gamma_{out}^*$ and $Z_L = Z_{out}^*$.

The following equation is now derived:

$$|1-\Gamma_S\Gamma_{in}|^2\big|_{\Gamma_L=\Gamma_{out}^*} = \frac{|1-S_{11}\Gamma_S|^2\left(1-|\Gamma_{out}|^2\right)^2}{\left|1-S_{22}\Gamma_{out}^*\right|^2} \tag{65}$$

The following quantities are evaluated from Eqs. (54b) and (55):

DOI 10.1007/978-981-4560-75-7_133-1

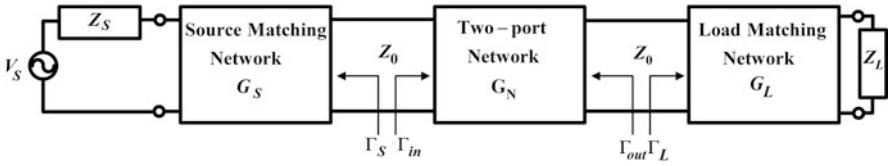


Fig. 3 A two-port network with source and load matching networks

$$\Gamma_{in} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} = \frac{S_{11} - \Gamma_L\Delta}{1 - S_{22}\Gamma_L} \tag{66}$$

$$\Gamma_{out} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S} = \frac{S_{22} - \Gamma_S\Delta}{1 - S_{11}\Gamma_S} = \Gamma_L^* \tag{67}$$

where $\Delta = S_{11}S_{22} - S_{12}S_{21}$. Substitute Γ_{in} from Eq. (65) and Γ_S from Eq. (66) in the left-hand side of (64) to verify it. Consequently, Eq. (63) reduces to

$$P_{avn} = \frac{|V_S|^2}{8Z_0}|S_{21}|^2 \frac{|1 - \Gamma_S|^2}{|1 - S_{11}\Gamma_S|^2\left(1 - |\Gamma_{out}|^2\right)} \tag{68}$$

The powers P_{avs} and P_{avn} are expressed in terms of source voltage (V_S), which is independent of the input and load impedances.

Using these relations, the available (G_A) and transducer (G_T) power gains are

$$G_A = \frac{P_{avn}}{P_{avs}} = |S_{21}|^2 \frac{\left(1 - |\Gamma_S|^2\right)}{|1 - S_{11}\Gamma_S|^2\left(1 - |\Gamma_{out}|^2\right)} \tag{69}$$

$$G_T = \frac{P_L}{P_{avs}} = |S_{21}|^2 \frac{\left(1 - |\Gamma_S|^2\right)\left(1 - |\Gamma_L|^2\right)}{|1 - \Gamma_S\Gamma_{in}|^2|1 - S_{22}\Gamma_L|^2} \tag{70}$$

When both the input and the output ports are matched for zero reflection, namely, $\Gamma_S = \Gamma_L = 0$, then the transducer power gain reduces to

$$G_T = |S_{21}|^2 \tag{71}$$

For a unilateral transducer (or a nonreciprocal amplifier), $S_{12} = 0$, which leads to $\Gamma_{in} = S_{11}$ from Eq. (54b). Therefore, the transducer power gain is obtained by letting $\Gamma_{in} = S_{11}$ in Eq. (69).

However, matching networks are usually required at the source and load sides of the two-port network to transform its input port to the source impedance and its output port to the load impedance, as shown in Fig. 3.

The transducer power gain (G_T) is appropriate for such a system consisting of the series connection of the source matching circuit, the two-port network, and the load matching circuit, where effective power gain factors are defined for each one. Equation (69) for G_T may be factorized as

$$G_T = G_S G_N G_L \tag{72a}$$

DOI 10.1007/978-981-4560-75-7_133-1

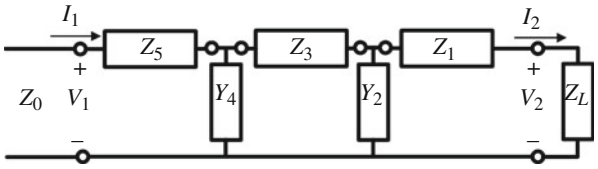


Fig. 4 A general matching network

where

$$G_S = \frac{\left(1 - |\Gamma_S|^2\right)}{|1 - \Gamma_S \Gamma_{in}|^2} \tag{72b}$$

$$G_N = |S_{21}|^2 \tag{72c}$$

$$G_L = \frac{\left(1 - |\Gamma_L|^2\right)}{|1 - S_{22}\Gamma_L|^2} \tag{72d}$$

Varieties of Matching Networks

A general matching network may be represented as shown in Fig. 4, although more sections may be added to increase its bandwidth (Rizzi 1988; Pozar 2011; Collin 2000; Gonzalez 1997; Misra 2001; Ludwig and Bretchko 2000).

The series blocks (Z_1, Z_3, Z_5) and parallel blocks (Y_2, Y_4) may be series or parallel connections of L and C discrete components. They may also be a transmission line of length (l_i), characteristic impedance Z_{0i}, and propagation constant γ_i. The transmission line may be open circuited or short circuited. The different types of blocks are given in Table 1.

The series blocks (Z_1, Z_3, and Z_5) may allow short- or open-circuited stubs as some reactive or susceptive element and also any combination of the other circuit elements. The parallel blocks (Y_2 and Y_4) allow transmission line sections and also any combination of the other circuit elements. Therefore, many circuits may be devised for any impedance matching network as drawn in Fig. 4, of which the blocks are made of any combination of the circuit elements given in Table 1.

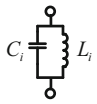
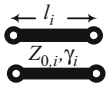
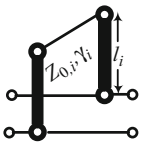
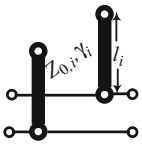
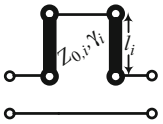
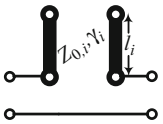
Many types of circuits have been devised for impedance matching networks, which are actually special cases of the general network in Fig. 4. For the case of single-frequency operation, the series LC and parallel LC circuits actually reduce to a single L or C component. For the design of impedance matching network for a specified bandwidth, both series and parallel LC circuits are required in its series and shunt sections.

Consider now the matching network usually denoted as L-network as shown in Fig. 5a, b.

For Fig. 5a, the blocks Z_3, Y_4, and Z_5 are removed, but blocks Z_1 and Y_2 are retained as denoted in Fig. 4. For Fig. 5b, blocks Z_1, Y_4, and Z_5 are removed, but blocks Y_2 and Z_3 are retained. For single-frequency operation, four different circuits are distinguished for Fig. 5a and also for Fig. 5b as shown in Fig. 6a, b, respectively.

DOI 10.1007/978-981-4560-75-7_133-1

Table 1 Various components for matching networks $\left(\omega_{0,i} = 1/\sqrt{L_i C_i}\right)$

		Z	Y	Series [T]	Parallel [T]
Series *LC*		$\frac{1-\left(\omega/\omega_{0,i}\right)^2}{j\omega C_i}$	$\frac{j\omega C_i}{1-\left(\omega/\omega_{0,i}\right)^2}$	$\begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix}$
Parallel *LC*		$\frac{j\omega L_i}{1-\left(\omega/\omega_{0,i}\right)^2}$	$\frac{1-\left(\omega/\omega_{0,i}\right)^2}{j\omega L_i}$	$\begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix}$
Transmission line		–	–	$\begin{bmatrix} \cos\beta_i l_i & Z_{0,i}\sin\beta_i l_i \\ Y_{0,i}\sin\beta_i l_i & \cos\beta_i l_i \end{bmatrix}$	–
Shorted parallel stub		–	$-jY_{0,i}\cot\beta_i l_i$	–	$\begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix}$
Open parallel stub		–	$jY_{0,i}\tan\beta_i l_i$	–	$\begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix}$
Shorted series stub		$jZ_{0,i}\tan\beta_i l_i$	–	$\begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix}$	–
Open series stub		$-jZ_{0,i}\cot\beta_i l_i$	–	$\begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix}$	–

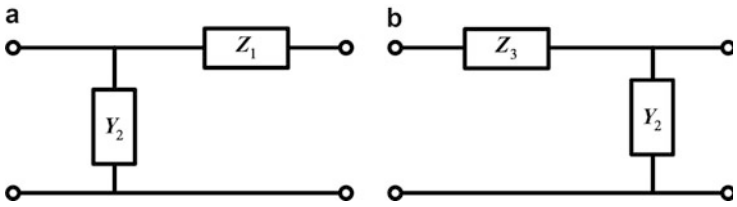


Fig. 5 Matching L-networks

DOI 10.1007/978-981-4560-75-7_133-1

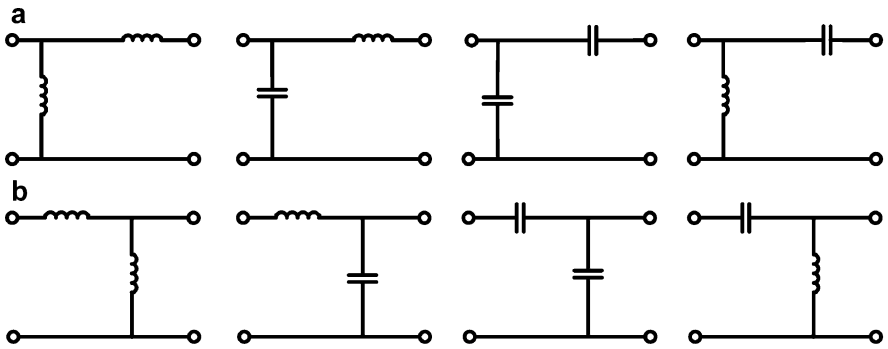


Fig. 6 Several representative impedance matching L-networks

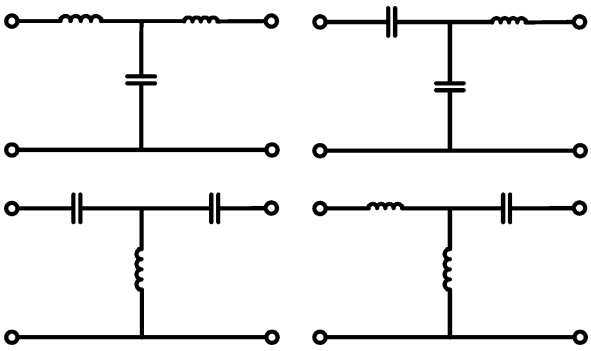


Fig. 7 Several representative impedance matching T-networks

The circuit in Fig. 6a may be used to match the load impedances located outside the circle of unit resistance in the Smith chart, and those in Fig. 6b are used for matching the load impedances lying inside this circle. The values of circuit components may be obtained by algebraic methods and also by graphical methods on the Smith chart.

The details of these methods are available in the textbooks on microwave engineering, where the reader may refer to Rizzi (1988) and Ludwig and Bretchko (2000). However, the method of least squares is used here for their designs.

Consider next the matching circuit usually denoted as T-network as shown in Fig. 7. This circuit is obtained by retaining blocks Z_1, Y_2, and Z_3 but removing blocks Y_4 and Z_5. The blocks Z_1, Y_2, and Z_3 may represent any combination of series and parallel LC circuits. Representative T-networks are drawn in Fig. 7. The circuits may be symmetric or asymmetric.

Consider then the matching circuit commonly called π-networks, as shown in Fig. 8. This circuit is obtained by the network in Fig. 4 by retaining blocks Y_2, Z_3, and Y_4 but removing blocks Z_1 and Z_5. The blocks Y_2, Z_3, and Y_4 may be composed of any combinations of series and parallel LC circuits. Several π-networks are drawn in Fig. 8.

Many other matching networks composed of series and parallel circuits may be devised in the form of the network in Fig. 4. Their general design may be based on the method of least squares to be presented shortly.

Several matching networks may be devised by the implementation of transmission lines. The single-stub tuner is obtained by keeping blocks Z_1 and Y_2 and removing Z_3, Y_4, and Z_5. Block Z_1 is a transmission line section and block Y_2 is a short-circuit or an open-circuit stub, as shown in Fig. 9a. Another

DOI 10.1007/978-981-4560-75-7_133-1

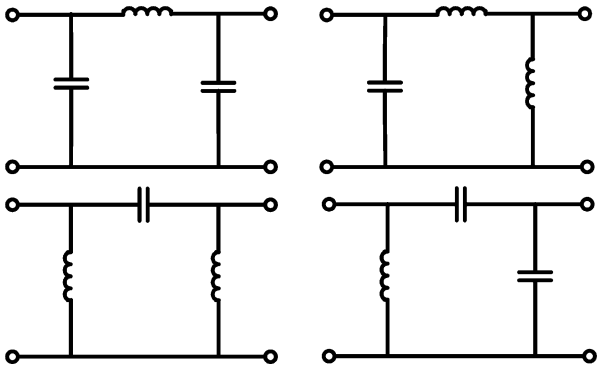


Fig. 8 Several representative impedance matching $R_{in} = R_S$-networks

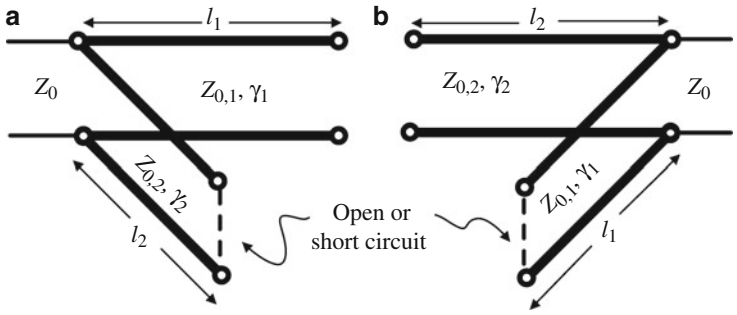


Fig. 9 Two configurations of single-stub tuning

configuration is obtained by keeping blocks Y_2 and Z_3 but removing blocks Z_1, Y_4, and Z_5. Block Y_2 is a short- or an open-circuited stub and Z_3 is a transmission line section, as shown in Fig. 9b.

The double-stub tuner is obtained by keeping all the blocks of the network in Fig. 4. Blocks Z_1, Z_3, and Z_5 are transmission line sections, whereas blocks Y_2 and Y_4 are short- or open-circuited stubs, as shown in Fig. 10.

Triple-stub tuning circuits may also be devised, which are composed of three open- or short-circuited stubs.

The design of single-stub and double-stub tuners may be performed by analytical methods using the equation of impedance transformation and also by the geometrical methods using the Smith chart. However, the method of least squares is used here for their designs.

Other special cases of the matching network shown in Fig. 4 may also be considered. For example, consider the case of assuming block Z_1 as the series connection of a transmission line section and a reactive element as shown in Fig. 11, where in Fig. 4, blocks Z_1 and Z_3 are kept, but blocks Y_2, Y_4, and Z_5 are removed.

The series reactive element in Fig. 11 may be replaced by a short- or open-circuited stub as shown in Fig. 12.

Next, consider the case where blocks Z_1 and Y_2 are kept, but blocks Z_3, Y_4, and Z_5 are removed. Block Z_1 is considered as a transmission line section, where block Y_2 is taken as a discrete element, as shown in Fig. 13. Block Z_n may be considered as a parallel stub of transmission line, which was mentioned earlier.

DOI 10.1007/978-981-4560-75-7_133-1

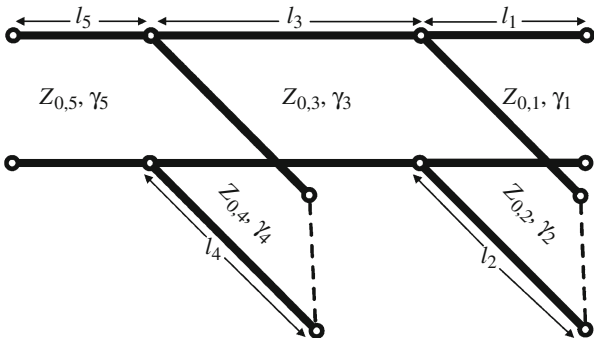


Fig. 10 Double-stub tuning

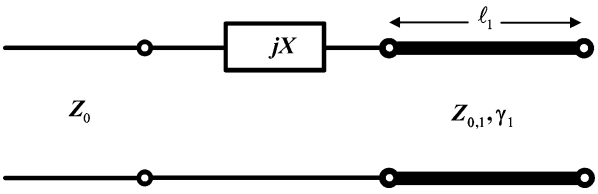


Fig. 11 Series reactive matching network

Impedance Transformer Design by the Method of Least Squares

The method of least squares is now developed for the design of matching circuit as proposed in the general network depicted in Fig. 4 (Oraizi 2001, 2006).

The total transmission matrix of the whole network is

$$[T] = [T_5][T_4][T_3][T_2][T_1] \tag{73}$$

where each transmission matrix $[T_i]$ may be obtained from the appropriate one given in Table 1 for any adopted matching circuit. Now, the input voltage and current are related to the output ones by the network transmission matrix

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = [T] \begin{bmatrix} V_2 \\ I_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix} \tag{74}$$

or

$$V_1 = AV_2 + BI_2 \tag{75}$$

$$I_1 = CV_2 + DI_2 \tag{76}$$

The input impedance of the circuit $\left(Z_{in} = \frac{V_1}{I_1}\right)$ may be written in term of the load impedance $\left(Z_L = \frac{V_2}{I_2}\right)$. Then, dividing the two equations in Eqs. (74) and (75) and replacing Z_L, the following relation is obtained:

$$Z_{in} = \frac{V_1}{I_1} = \frac{AZ_L + B}{CZ_L + D} \tag{77}$$

DOI 10.1007/978-981-4560-75-7_133-1

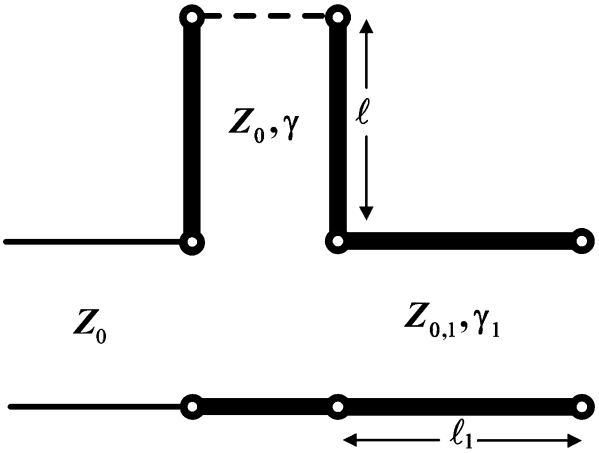


Fig. 12 Matching network using series short- or open-circuited reactive stubs

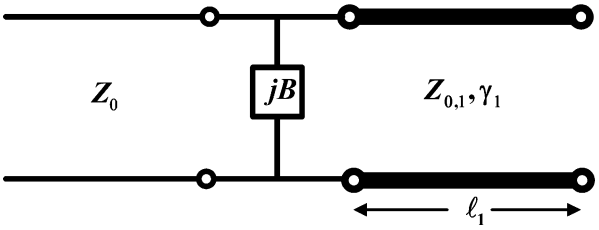


Fig. 13 Matching network using parallel short- or open-circuited reactive stubs

The matching circuit may be designed for reflectionless condition at the input (namely, $Z_0 = Z_{in}$) or for conjugate matching for maximum power transfer (namely, $Z_0 = Z_{in}^*$). Then, an error function may be constructed as

$$error = \sum_k \left| Z_{in,k} - Z_0^* \right|^2 = \sum_k \left| \frac{A_k Z_{L,k} + B_k}{C_k Z_{L,k} + D_k} - Z_0^* \right|^2 \tag{78}$$

where the frequency interval between the limit frequencies (f_1 to f_2) is divided into K intervals with $f_k = f_1 + k\Delta f = f_1 + k\frac{f_2 - f_1}{K}, k = 0, 1, 2, \ldots, K$. If the characteristic impedance of the input line (Z_0) is real, then the conditions for reflectionless line and conjugate matching are the same. For the design of a single-frequency matching circuit, there is no summation on (K + 1) frequencies in Eq. (77). The error function depends on the geometrical dimensions and electrical characteristics of the circuit elements selected from Table 1, for various blocks in Fig. 4. Observe that the matching network may also be implemented for the case of Z_0 as a function of frequency.

The minimization of the error function determines all parameters of the different sections of the matching circuit. If the error function is a quadratic function of the variables, then it has a unique minimum. If it is not quadratic, then it has several minima. A minimization algorithm needs to be used, such as the steepest descent, conjugate gradient, any evolutionary algorithm (such as genetic algorithm, particle swarm optimization, ant colony, or invasive weed method), or any combination of such algorithms.

DOI 10.1007/978-981-4560-75-7_133-1

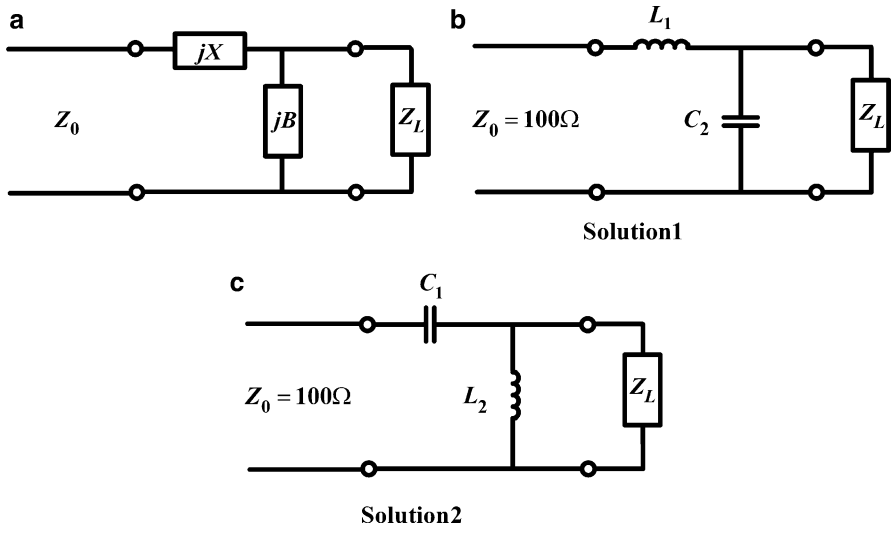


Fig. 14 Example 1. (**a**) Block diagram; (**b**) solution 1; (**c**) solution 2

Table 2 Input data of Example 1 for lumped element transformer

Example	Type of first element from source	Number of elements	Freq (GHz)	Error value margin	Z_s (Ω)	Z_L (Ω)
1	Series	2	0.5	0.001	100	200-100j

Example 1

Design a lumped element transformer as shown in Fig. 14 according to the specified data in Table 2. Since $R_L > Z_0$, the circuit in Fig. 14 should be used. The output parameters are listed in Table 3, which are obtained by the computer programs. The reflection coefficient at the input of circuits for the two solutions is shown in Fig. 15.

Next, design several distributed transmission line transformers according to the specified input data in Table 4. The output parameters are listed in Table 5, which are obtained by computer programs.

Example 2

Design of a single-stub tuning circuit with parallel open-circuit stub shown in Fig. 16. Its frequency response is drawn in Fig. 17.

Example 3

Design of a single-stub tuning circuit with series open-circuit stub shown in Fig. 18. Its frequency response is drawn in Fig. 19.

Example 4

Design of a double-stub tuning circuit with parallel short-circuit stubs shown in Fig. 20. Its frequency response is drawn in Fig. 21.

DOI 10.1007/978-981-4560-75-7_133-1

Table 3 Output parameters of Example 1 for lumped element transformer

Example		X_1	B_2	Z_s (Ω)		Z_L (Ω)	
				R_s	L_s (nH)	R_L	C_L (pF)
1	Solution 1	1.2247 $\rightarrow L_1 = 38.9846$ (nH)	0.28996 $\rightarrow C_2 = 0.9229$ (pF)	100	0	200	3.1831
	Solution 2	−1.2247 $\rightarrow C_1 = 2.5991$ (pF)	−0.6899 $\rightarrow L_2 = 46.1388$ (nH)				

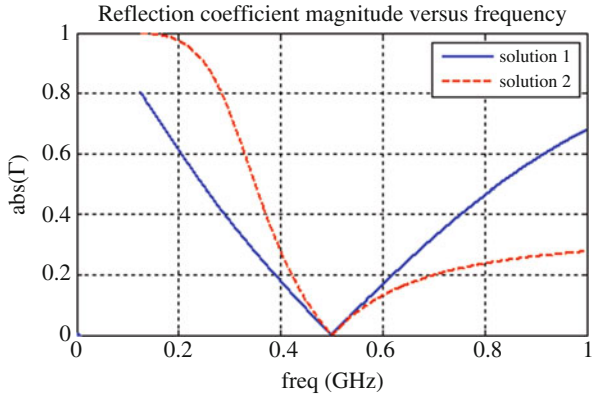


Fig. 15 Example 1. Reflection coefficient magnitude versus frequency

Example 5

Design of various tuning circuits (single-stub tuning with parallel short-circuit stub, single-stub tuning with series short-circuit stub, and double-stub tuning circuit with parallel short-circuit stubs) shown in Fig. 22. Their frequency responses are drawn in Fig. 23.

The reflection coefficients of the various tuning circuits are drawn in Fig. 23 for comparison. Observe that the short-circuit series stub tuning circuit has the broadest bandwidth at the specified operation frequency. The reason is that the length of its stub is shortest among the tuners.

A Quarter-Wave Line as a Basic Unit of Impedance Transformers

Multi-section impedance transformers are used to match two systems of different impedances, such as two transmission lines with different characteristic impedances, two waveguides with different wave impedances, and matching of a radiating system (such as lenses) to free space (Rizzi 1988; Pozar 2011; Collin 2000). The lengths of the sections of transformer may be arbitrary or equal to a quarter wavelength, in which case it is called a quarter-wave transformer.

Quarter-Wave Transformer

Consider a transmission line of length ℓ, characteristic impedance Z_1, and phase constant β_l connecting a load resistor R_L to an impedance Z_0, as shown in Fig. 24.

The input impedance to the line is

DOI 10.1007/978-981-4560-75-7_133-1

Table 4 Input data of Examples 2–5 for transmission line transformers

Example	2	3	4	5 c	b	a
Number of element	2	2	3	3	2	2
Freq (GHz)	2	2	2	f_0	f_0	f_0
Error value margin	0.001	0.001	0.001	0.001	0.001	0.001
Z_s (Ω)	50	50	50	100	100	100
Z_L (Ω)	10 + 15j	100 + 80j	60−80j	200	200	200
Type of odd elements from source	P & O	S & O	[P & S, P & S]	[P & S, S & S]	S & S	P & S
Z_0 of odd element	50	50	[50,50]	[100,10]	100	100
LB of stubs length	0.01λ	0.01λ	[0.01,0.01] λ	[0.01,0.01] λ	0.01λ	0.01λ
UB of stubs length	0.5λ	0.5λ	[0.5,0.5] λ	[0.5,0.5] λ	0.5λ	0.5λ
Z_0 of even element	50	50	50	100	100	100
LB of TL length	0.01λ	0.01λ	0.125λ	0.125λ	0.01λ	0.01λ
UB of TL length	0.5λ	0.5λ	0.125λ	0.125λ	0.5λ	0.5λ

Table 5 Output parameters of Examples 2–5 for transmission line transformers

Example			Width of source (*mm*)	*L*	Z_s (Ω)		Z_L (Ω)		Error value
					R_s	L_s (*nH*)	R_L	L_L (*nH*)	
2		Solution 1	4.916	[0.3526,0.3874] λ_g =[52.9042,58.1025]	50	0	15	0.79577	9.9473e-8
		Solution 2		[0.1473,0.0440] λ_g =[22.0996,6.6053]					9.7034e-9
3		Solution 1	4.916	[0.1023,0.4633] λ_g =[15.353,69.509]	50	0	100	6.362	2.9047e-8
		Solution 2		[0.3976,0.1197] λ_g =[59.6489,17.9604]					4.5631e-8
4		Solution 1	4.916	[0.4542,0.125,0.2319] λ_g = [68.1423,18.75,34.7867]	50	0	60	0.99472	2.9533e-6
		Solution 2		[0.0099,0.125,0.2319] λ_g = [14.9663,18.75,34.7867]					4.4364e-12
5	a	Solution 1	1.6271	[0.3479,0.3479] λ_g	100	0	200	0	1.0882e-9
		Solution 2		[0.15120,0.1520] λ_g					6.1842e-9
	b	Solution 1	1.6271	[0.4020,0.4020] λ_g	100	0	200	0	4.1523e-8
		Solution 2		[0.0979,0.0979] λ_g					2.4558e-8
	c	Solution 1	1.6271	[0.4441,0.125,0.4217] λ_g	100	0	200	0	2.3358e-6
		Solution 2		[0.1494,0.125,0.27119] λ_g					1.3407e-8

$$Z_{in} = Z_1 \frac{R_L + jZ_1 t}{Z_1 + jR_L t} \tag{79}$$

where $t = \tan \beta l$. The reflection coefficient at the line input is obtained by the substitution of Eq. (78) in

DOI 10.1007/978-981-4560-75-7_133-1

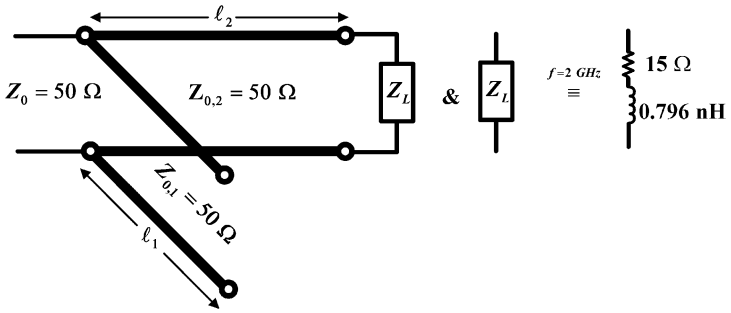


Fig. 16 Example 2. Single-stub tuning circuit with parallel open-circuit stub

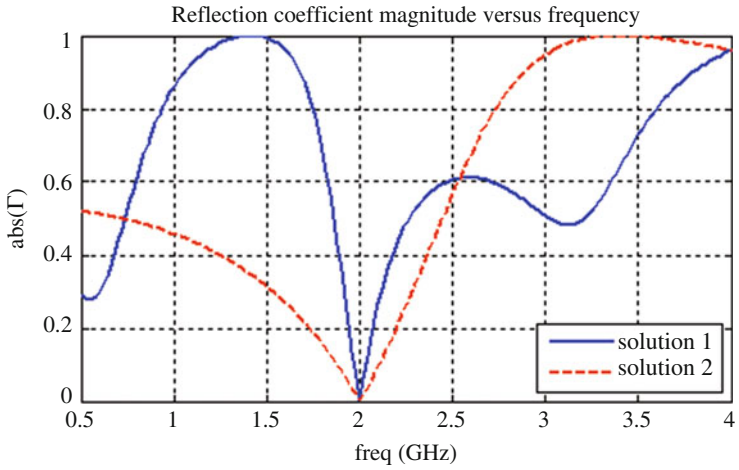


Fig. 17 Example 2. Reflection coefficient magnitude versus frequency

$$\Gamma_{in} = \frac{Z_{in} - Z_0}{Z_0 - Z_{in}} = \frac{Z_1(R_L - Z_0) + jt\left(Z_1^2 - Z_0 R_L\right)}{Z_1(R_L + Z_0) + jt\left(Z_1^2 + Z_0 R_L\right)} \tag{80}$$

If the length of line is equal to a quarter wave at the center frequency (f_0) and the load R_L is matched through the line to Z_0, then $Z_1 = \sqrt{Z_0 R_L}$ and

$$\Gamma_{in} = \frac{R_L - Z_0}{R_L + Z_0 + j2t\sqrt{Z_0 R_L}} \tag{81}$$

The magnitude of Γ_{in} becomes

$$|\Gamma_{in}| = \frac{1}{\left[1 + \frac{4Z_0 R_L \sec^2\theta}{(R_L - Z_0)^2}\right]^{1/2}} \tag{82}$$

For frequencies near the design frequency (f_0), the following relations are valid for a TEM line:

DOI 10.1007/978-981-4560-75-7_133-1

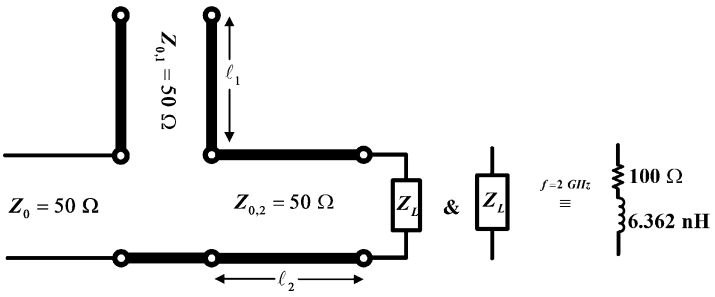


Fig. 18 Example 3. Single-stub tuning circuit with series open-circuit stub

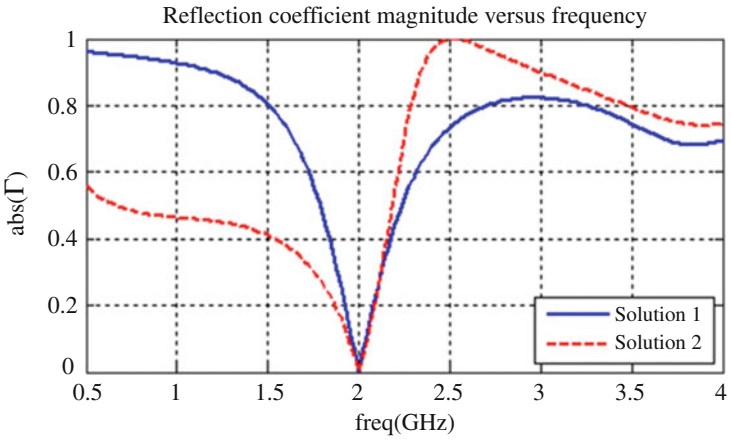


Fig. 19 Example 3. Reflection coefficient magnitude versus frequency

$$\theta = \beta l = \frac{2\pi}{\lambda}\frac{\lambda_0}{4} = \frac{\pi f}{2f_0} \approx \frac{\pi}{2} \quad \text{and} \quad \sec^2\theta \gg 1 \tag{83a}$$

$$|\Gamma_{in}| \approx \frac{|R_L - Z_0|}{2\sqrt{Z_0 R_L}} |\cos\theta| \quad , \quad \theta \approx \frac{\pi}{2} \tag{83b}$$

The input reflection coefficient $|\Gamma_{in}|$ is drawn versus θ in Fig. 25.

For a specified value of reflection coefficient Γ_m, Eq. (81) gives the angle θ_m

$$\cos\theta_m = \frac{\Gamma_m}{\sqrt{1-\Gamma_m^2}} \frac{2\sqrt{Z_0 R_L}}{|R_L - R_0|} \tag{84}$$

Now, define the bandwidth of the transformer referring to Fig. 25 as

$$\Delta\theta = 2\left(\frac{\pi}{2} - \theta_m\right) \tag{85}$$

Then, the frequency of the lower end of the band is

$$f_m = \frac{2}{\pi} f_0 \theta_m \tag{86}$$

DOI 10.1007/978-981-4560-75-7_133-1

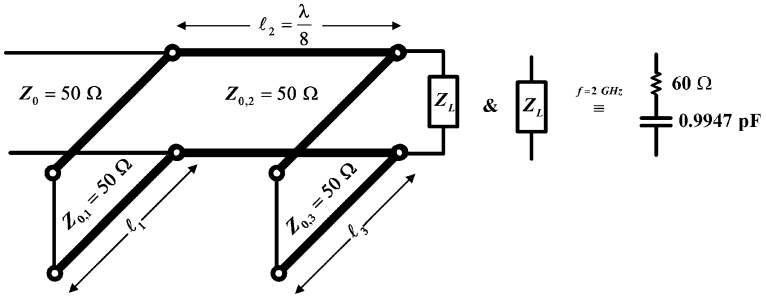


Fig. 20 Example 4. Double-stub tuning circuit with parallel short-circuit stubs

Consequently, the fractional bandwidth is

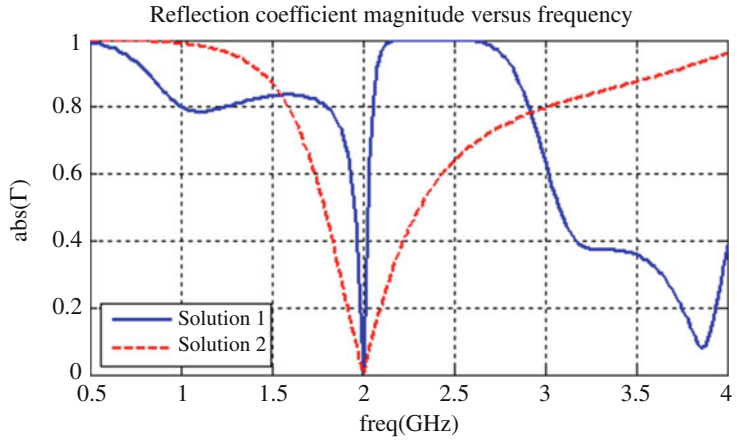


Fig. 21 Example 4. Reflection coefficient magnitude versus frequency

$$\frac{\Delta f}{f_0} = \frac{2(f_0 - f_m)}{f_0} = 2 - \frac{4}{\pi}\theta_m \tag{87}$$

$$= 2 - \frac{4}{\pi}\cos^{-1}\left[\frac{\Gamma_m}{\sqrt{1-\Gamma_m^2}}\frac{2\sqrt{Z_0 R_L}}{|R_L - Z_0|}\right] \tag{88}$$

Observe that the transformer bandwidth increases as R_L approaches Z_0, that is, for the lesser mismatched cases.

The above analysis assumes linear dependence of β on frequency, as for TEM modes. For nonlinear lines, the propagation constant is a nonlinear function of frequency, and the wave impedance is also a function of frequency. The analysis will thus become complicated. However, such frequency dependencies do not affect the analysis for narrowband performance. Furthermore, discontinuities at the junction of transmission lines were ignored in the above solution. Such effects may be considered by adjusting the length of line sections. The effect of discontinuities will be taken into account in the subsequent treatment.

DOI 10.1007/978-981-4560-75-7_133-1

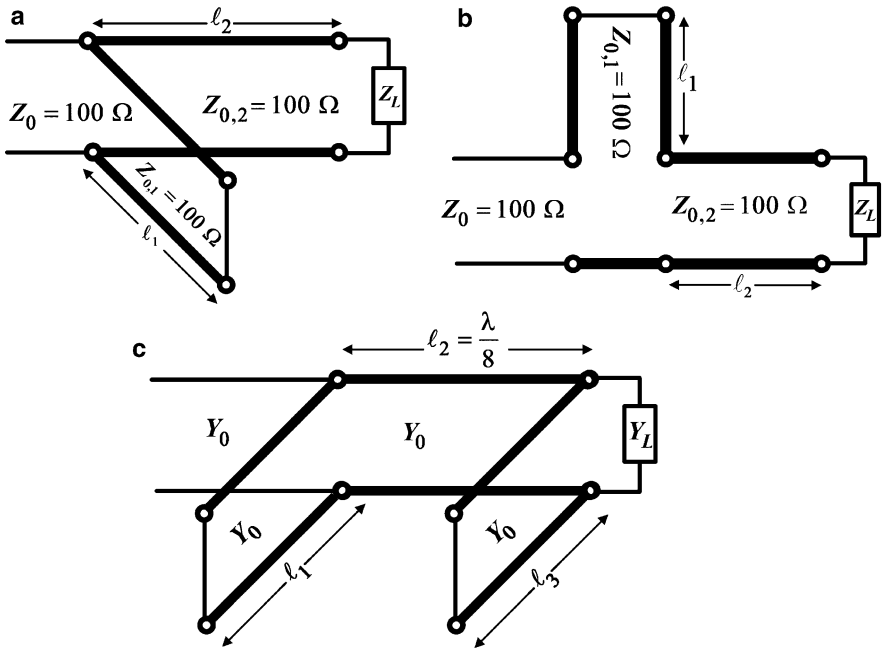


Fig. 22 Example 5. (**a**) Single-stub tuning circuit with parallel short-circuit stub; (**b**) single-stub tuning circuit with series short-circuit stub; (**c**) double-stub tuning circuit with parallel short-circuit stubs

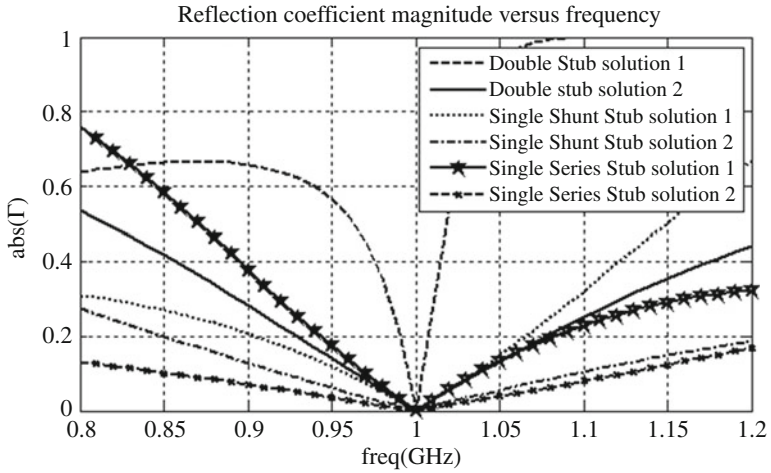


Fig. 23 Example 5. Reflection coefficient magnitude versus frequency ($f_0 = 1$ GHz)

Theory of Small Reflections

Consider a single-section transformer of length ℓ and characteristic impedance Z_2 connecting load impedance Z_L to impedance Z_1, as shown in Fig. 26a (Pozar 2011; Collin 2000). A simple formula is derived for the total reflection at its input (Γ) in terms of partial reflections at its input (Γ_1) and output (Γ_3).

The partial reflection and transmission coefficients at its input and output ports as indicated in Fig. 16b are

DOI 10.1007/978-981-4560-75-7_133-1

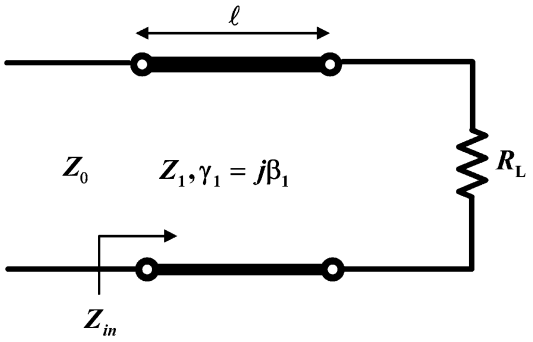


Fig. 24 A transmission line (Z_1) connecting a load (R_L) to a line of impedance (Z_0)

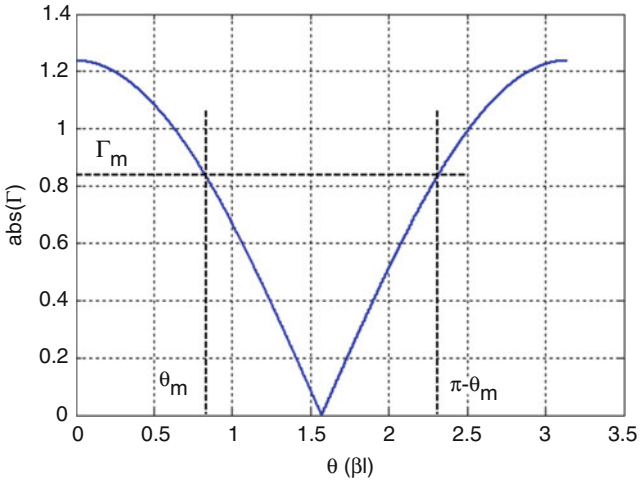


Fig. 25 Frequency response of a single-section quarter-wave transformer

$$\Gamma_1 = -\Gamma_2 = \frac{Z_2 - Z_1}{Z_2 + Z_1} \tag{89a}$$

$$\Gamma_3 = \frac{Z_L - Z_2}{Z_L + Z_2} \tag{89b}$$

$$T_{21} = 1 + \Gamma_1 = \frac{2Z_2}{Z_2 + Z_1} \tag{89c}$$

$$T_{12} = 1 + \Gamma_2 = \frac{2Z_1}{Z_2 + Z_1} \tag{89d}$$

The total input reflection coefficient (Γ) is obtained by the impedance method using the impedance transformation equation for Z_{in}. However, the multiple reflection method is followed here, as indicated in Fig. 26b. The total reflection at the input of the line section is the sum of all multiple reflections at its input and output ends, after they emerge from its input,

DOI 10.1007/978-981-4560-75-7_133-1

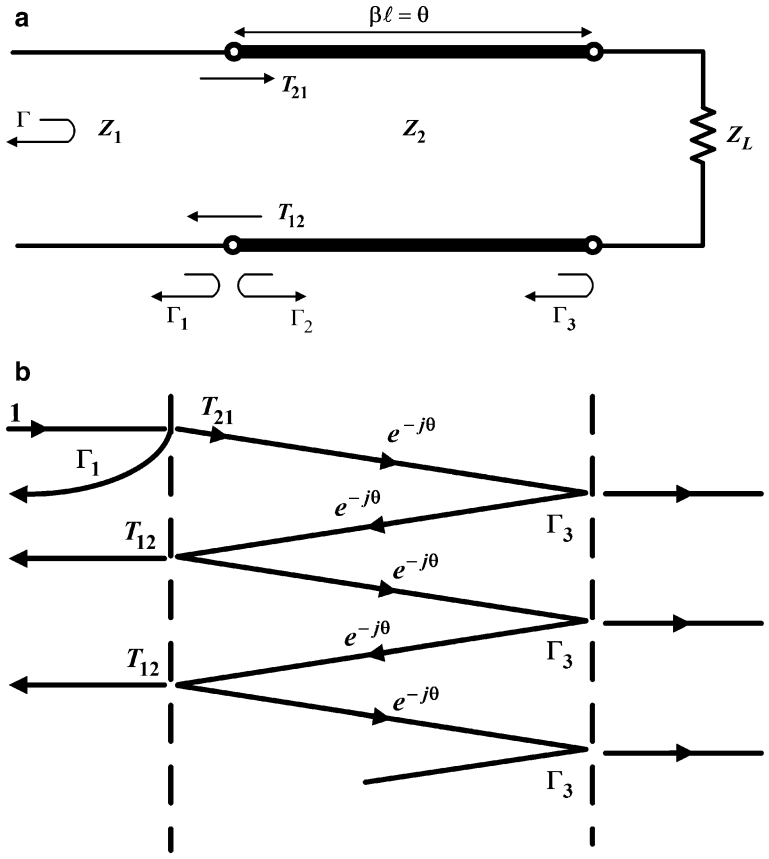


Fig. 26 Partial reflections and transmission on a terminated line section. (**a**) Transmission line; (**b**) partial multiple reflections

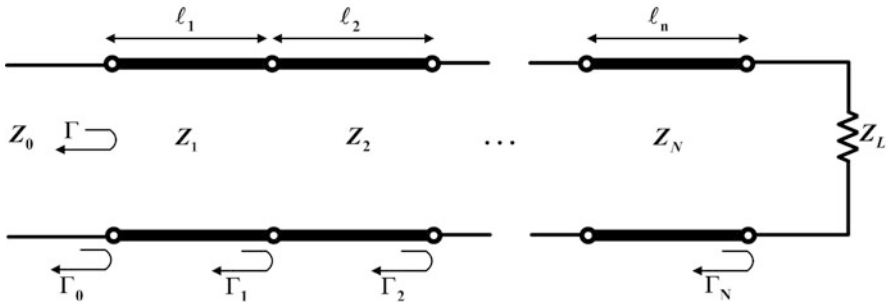


Fig. 27 A general multi-section impedance transformer

$$\Gamma = \Gamma_1 + T_{21}T_{12}\Gamma_3 e^{-j2\theta} + T_{21}T_{12}\Gamma_3^2\Gamma_2 e^{-j4\theta} + \cdots = \Gamma_1 + T_{21}T_{12}\Gamma_3 e^{-j2\theta}\sum_{i=0}^{\infty}\left[\Gamma_2\Gamma_3 e^{-j2\theta}\right]^n \tag{90}$$

Since $|\Gamma_2\Gamma_3| \leq 1$, use the formula for the sum of geometric series to obtain

$$\Gamma = \Gamma_1 + \frac{T_{21}T_{12}\Gamma_3 e^{-j2\theta}}{1 - \Gamma_2\Gamma_3 e^{-j2\theta}} \tag{91}$$

DOI 10.1007/978-981-4560-75-7_133-1

Using Eq. (89), the following equation is obtained:

$$\Gamma = \frac{\Gamma_1 + \Gamma_3 e^{-j2\theta}}{1 + \Gamma_1 \Gamma_3 e^{-j2\theta}} \tag{92}$$

Ignoring the discontinuities on the line and with $|\Gamma_1 \Gamma_3| \ll 1$, the key formula for the case of small reflections on the line is obtained:

$$\Gamma \cong \Gamma_1 + \Gamma_3 e^{-j2\theta} \tag{93}$$

It states that the resultant reflection at the input of the line is equal to the reflection at the input discontinuity plus that of the load including the phase shift due to the signal traveling up and down the line. This relation is the basis of the design of multi-section impedance transformers.

If the length of the line is equal to a quarter-wavelength $\left(\frac{\lambda}{4}\right)$ of the design frequency, then $\theta = \beta l = \frac{2\pi}{\lambda}\frac{\lambda}{4} = \frac{\pi}{2}$ and $e^{-j2\theta} = -1$. It may be substituted in Eqs. (89d) and (90) to obtain its reflection coefficient. If $Z_2 = \sqrt{Z_1 Z_L}$, then $\Gamma = 0$ as expected.

Multi-section Transformer

Consider the cascade connection of N line sections of lengths ℓ_i and electrical lengths $\theta_i = \beta_i \ell_i$, characteristic impedances $Z_{0,i}$, and propagation constants $\gamma_i = \alpha_i + j\beta_i$ shown in Fig. 27 (Rizzi 1988; Pozar 2011; Collin 2000; Gonzalez 1997; Misra 2001; Ludwig and Bretchko 2000). The equal-length (commensurate) lines ($l_i = l$), lossless lines ($\alpha_i = 0$), and dispersionless lines $\left(\beta = \frac{\omega}{\upsilon}\right)$ are all special cases of the general transformer. The resultant reflection coefficient at the line input is then derived.

Consider the partial reflection coefficients at the junctions of line sections.

$$\Gamma_0 = \frac{Z_1 - Z_0}{Z_1 + Z_0} \tag{94a}$$

$$\Gamma_n = \frac{Z_{n+1} - Z_n}{Z_{n+1} + Z_n} \tag{94b}$$

$$\Gamma_N = \frac{Z_L - Z_N}{Z_L + Z_N} \tag{94c}$$

The characteristic impedances depend on the propagation modes of the waveguide.

The small reflection formula is invoked to compute the overall input reflection coefficient

$$\begin{aligned} \Gamma(\theta) &= \Gamma_0 + \Gamma_1 e^{-j2\theta_1} + \Gamma_2 e^{-j2(\theta_1+\theta_2)} + \cdots + \Gamma_N e^{-j2(\theta_1+\theta_2+\cdots\theta_N)} \\ &= \sum_{n=0}^{N} \Gamma_n e^{-j2\sum_{m=0}^{n}\theta_m} = \sum_{n=0}^{N} \frac{Z_{n+1} - Z_n}{Z_{n+1} + Z_n} \exp\left(-j2\sum_{m=0}^{n}\beta_m l_m\right) \end{aligned} \tag{95}$$

where $\theta_0 = 0$ and $\theta_m = \beta_m l_m = \omega\sqrt{\mu_m \varepsilon_m} l_m$ for the TEM mode. However, for non-TEM modes, the characteristic impedances (Z_n) and phase constants (β_n) are nonlinear functions of frequency.

Consequently, the specified bandwidth from f_1 to f_2 may be divided into K discrete frequencies. An error function may be constructed in terms of the reflection coefficients

$$\text{error} = \sum_{k=1}^{K} |\Gamma_k|^2 \tag{96}$$

There are various formulas in the literature for the characteristic impedance Z_n and phase constant β_n of various transmission lines in terms of their geometrical dimensions and constitutive matter, such as striplines, microstrips, and waveguides. By using such formulas, the error function may be minimized to determine the geometrical dimensions of the transformer.

Now, simplify the transformer design and assume that it is commensurate and the length of line sections is equal to a quarter wavelength $\left(\frac{\lambda}{4}\right)$ and $\beta l = \frac{\pi}{2}$. The error function reduces to the following relation without explicitly considering the frequency bandwidth.

$$\text{error} = \left| \sum_{n=0}^{N} \frac{Z_{n+1} - Z_n}{Z_{n+1} + Z_n} \exp\left(-j\pi \sum_{m=0}^{n} m \right) \right|^2 \tag{97}$$

It is then minimized to obtain Z_n and eventually determine the geometrical dimensions of the transformer.

The binomial and Chebyshev multi-section matching transformers are then developed. Let us assume that Z_L is resistive and real, and Z_n increases or decreases monotonically. Therefore, all Γ_n are real and have the same sign ($\Gamma_n > 0$ for $Z_L > Z_0$ or $\Gamma_n < 0$ for $Z_L < Z_0$). The input reflection coefficient is then from Eq. (94c):

$$\Gamma(\theta) = \Gamma_0 + \Gamma_1 e^{-j2\theta} + \Gamma_2 e^{-j4\theta} + \cdots + \Gamma_N e^{-j2N\theta} \tag{98}$$

Now, assume that Γ_n are symmetrical (but not Z_n), namely, $\Gamma_i = \Gamma_{N-i}$ for $i = 0, 1, \ldots, N$. This leads to a finite Fourier cosine series as

$$\Gamma(\theta) = 2e^{-jN\theta}\left[\Gamma_0 \cos N\theta + \Gamma_1 \cos(N-2)\theta + \cdots + \Gamma_n \cos(N-2n)\theta + \cdots + \frac{1}{2}\Gamma_{\frac{N}{2}}\right] \quad \text{for N even} \tag{99a}$$

$$\Gamma(\theta) = 2e^{-jN\theta}\left[\Gamma_0 \cos N\theta + \Gamma_1 \cos(N-2)\theta + \cdots + \Gamma_n \cos(N-2n)\theta + \cdots + \frac{1}{2}\Gamma_{\frac{(N-1)}{2}} \cos\theta\right] \quad \text{for N odd} \tag{99b}$$

Observe that any desired response may be realized for $\Gamma(\theta)$, because it is expressed in a Fourier cosine series, which may express an arbitrary function, if enough terms are included in it.

A binomial transformer is then obtained, which has a maximally flat response in its passband, provided that its $(N-1)$ derivatives are set equal to zero. Therefore, set

$$\Gamma(\theta) = A\left(1 + e^{-j2\theta}\right)^N \tag{100a}$$

$$|\Gamma(\theta)| = 2^N |A| |\cos\theta|^N \tag{100b}$$

Observe that $|\Gamma(\theta)| = 0$ and $\frac{d^n|\Gamma(\theta)|}{d\theta^n} = 0$ at $\theta = \frac{\pi}{2}$ (for $n = 0, 1, \cdots, N-1$), which corresponds to the center frequency f_0, where $\ell = \frac{\lambda}{4}$ and $\theta = \beta l = \frac{\pi}{2}$. Now, for $f \to 0, \theta \to 0$,

$$A = 2^{-N}\left|\frac{Z_L - Z_0}{Z_L + Z_0}\right| \tag{101}$$

Expand Eq. (99b) and equate it to Eq. (97). Then,

$$\Gamma(\theta) = A\left(1 + e^{-j2\theta}\right)^N = A\sum_{n=0}^{N} C_n^N e^{-j2n\theta} = \sum_{n=0}^{N} \Gamma_n e^{-j2n\theta} \tag{102}$$

Therefore, equate the coefficients in Eq. (101) to obtain

$$\Gamma_n = AC_n^N = 2^{-N}\left|\frac{Z_L - Z_0}{Z_L + Z_0}\right| \frac{N!}{n!(N-n)!} \tag{103a}$$

$$\frac{Z_{n+1} - Z_n}{Z_{n+1} + Z_n} = 2^{-N}\left|\frac{Z_L - Z_0}{Z_L + Z_0}\right| \frac{N!}{n!(N-n)!},\ n = 0, 1, \cdots, N-1 \tag{103b}$$

Finally, the characteristic impedances Z_1 to Z_n may be evaluated. Evaluation of Z_n may be determined by more exact methods, which may be for $Z_L > Z_0$ or in reverse $Z_L < Z_0$.

For the evaluation of the transformer bandwidth, the maximum value of reflection coefficient is set as Γ_m in Eq. (100a) to get $\theta_m < \frac{\pi}{2}$, which designates the lower edge of the bandwidth

$$\theta_m = \cos^{-1}\left[\frac{1}{2}\left(\frac{\Gamma_m}{A}\right)^{\frac{1}{N}}\right] \tag{104}$$

The fractional bandwidth is calculated by Eq. (87):

$$\frac{\Delta f}{f_0} = 2 - \frac{4}{\pi}\cos^{-1}\left[\frac{1}{2}\left(\frac{\Gamma_m}{A}\right)^{\frac{1}{N}}\right] \tag{105}$$

Next, design a Chebyshev transformer, which has an equal ripple response in its passband. The reflection coefficient $\Gamma(\theta)$ in Eq. (99) is set proportional to the Chebyshev polynomial $T_N(\sec\ \theta_m\ \cos\ \theta)$, where N is the number of sections.

$$\begin{aligned}\Gamma(\theta) &= 2e^{-jN\theta}[\Gamma_0 \cos N\theta + \Gamma_1 \cos(N-2)\theta + \cdots + \Gamma_n \cos(N-2n)\theta + \cdots] \\ &= \mathrm{A}e^{-jN\theta}T_N(\sec\theta_m \cos\theta)\end{aligned} \tag{106}$$

The last term in the series in Eq. (105) is equal to $\frac{1}{2}\Gamma_{N/2}$ and $\Gamma_{(N-1)/2}\cos\theta$ for N even and odd, respectively. The constant A is obtained by setting $\theta = 0$ in Eq. (105):

$$A = \frac{Z_L - Z_0}{Z_L + Z_0}\frac{1}{T_N(\sec\theta_m)} \tag{107}$$

DOI 10.1007/978-981-4560-75-7_133-1

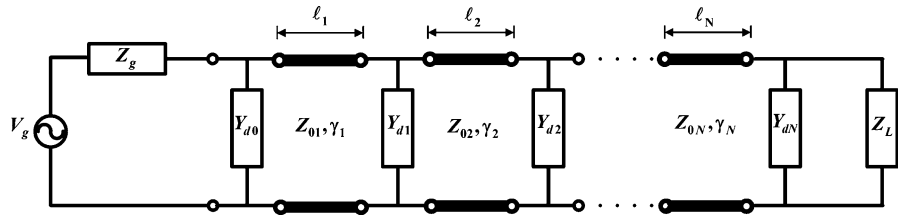


Fig. 28 An N-section step-line impedance transformer

Specifying the maximum reflection coefficient in the band (Γ_m), then

$$\Gamma_m = A \tag{108}$$

since the maximum value of T_N in the band is equal to unity. Therefore, θ_m may be obtained by Eq. (106) and

$$\sec\theta_m = \cosh\left[\frac{1}{N}\cosh^{-1}\left(\frac{1}{\Gamma_m}\left|\frac{Z_L - Z_0}{Z_L + Z_0}\right|\right)\right] \tag{109}$$

since $T_n(x) = \cosh\left(n\cosh^{-1}x\right)$ for $|x| > 1$

The fractional bandwidth is calculated by Eq. (87):

$$\frac{\Delta f}{f_0} = 2 - \frac{4\theta_m}{\pi}$$

Now, the Chebyshev polynomial may be expanded and equated to $\Gamma(\theta)$ in Eq. (105) to calculate Γ_n and eventually determine Z_n by Eq. (94). Exact evaluation of Z_n for Chebyshev transformers is computed and given in extensive tables (Matthaei et al. 1964).

Design of Step-Line Transformers by the Method of Least Squares

A nonuniform transmission line is approximated here by a cascade of line sections which is shown to be possible (Giguere 1972; Oraizi 2001; Oraizi and Seyyed-Esfahlan 2011). Therefore, if a nonuniform line theoretically exists for a particular matching situation, its step-line approximation may be obtained.

Now, consider a step-line transformer composed of the cascade connection of N unequal-length (non-commensurate) or equal-length (commensurate) line sections for the matching of two complex impedances (e.g., source (Z_g) and load (Z_L) impedances) over a frequency bandwidth from f_1 to f_2 as shown in Fig. 28. Step discontinuities, such as width change of striplines and microstrips and diameter change of center conductors of coaxial lines (Y_{di}), are also considered. Discontinuities at the source (Y_{d0}) and load (Y_{dN}) sides of the transformer are also shown.

Common design procedures for transformers are for matching two constant resistances (with a given resistance ratio R) over a specified frequency bandwidth (with a given fractional bandwidth ω_g) (Matthaei et al. 1964).

Assume that the ith line section has length ℓ_i, propagation constant γ_i, and characteristic impedance Z_{0i}. The admittance of the step discontinuity is denoted by Y_{di}. The specified bandwidth from f_1 to f_2 is divided into K discrete frequencies. The kth frequency is $f_k = f_1 + k\Delta,\ \Delta = (f_2 - f_1)/K$ for $k = 0, 1, \ldots, K$.

DOI 10.1007/978-981-4560-75-7_133-1

Consequently, the transmission matrices of the ith line and discontinuity at the kth frequency are, respectively,

$$\left[P_{i,k}\right] = \begin{bmatrix} \cosh\gamma_{i,k}l_i & Z_{0,i}\sinh\gamma_{i,k}l_i \\ Y_{0,i}\sinh\gamma_{i,k}l_i & \cosh\gamma_{i,k}l_i \end{bmatrix}, \quad i = 1, 2, \ldots, N \tag{110a}$$

$$\left[Q_{i,k}\right] = \begin{bmatrix} 1 & 0 \\ Y_{d,i} & 1 \end{bmatrix}, \quad i = 0, 1, 2, \ldots, N \tag{110b}$$

The combined transmission matrix of the ith section at kth frequency is

$$\left[T_{i,k}\right] = \left[P_{i,k}\right]\left[Q_{i,k}\right], i = 1, 2, \ldots, N \tag{111}$$

The total transmission matrix of the step-line at f_k is then

$$[T_k] = \begin{bmatrix} A_k & B_k \\ C_k & D_k \end{bmatrix} = \left[Q_{o,k}\right] \prod_{i=1}^{N} \left[T_{i,k}\right] = \left[Q_{o,k}\right] \prod_{i=1}^{N} \left[P_{i,k}\right]\left[Q_{i,k}\right] \tag{112}$$

The input voltage and current (V_g, I_g) are expressed in terms of the load voltage and current (V_L, I_L) as

$$\begin{bmatrix} V_g \\ I_g \end{bmatrix} = \begin{bmatrix} A_k & B_k \\ C_k & D_k \end{bmatrix} \begin{bmatrix} V_L \\ I_L \end{bmatrix} \tag{113}$$

The input impedance of the step-line for $V_L = Z_{L,k}I_L$ at f_k is then

$$Z_{in,k} = \frac{V_g}{I_g} = \frac{A_k Z_{L,k} + B_k}{C_k Z_{L,k} + D_k} \tag{114}$$

Two cases of impedance matching may be considered, namely, reflectioless line ($\Gamma_{in} = 0$ or in particular $Z_{in,k} = Z_{g,k}$) and maximum power transfer $\left(Z_{in,k} = Z^*_{g,k}\right)$. An error function is constructed here for the former case over the specified bandwidth:

$$\varepsilon = \sum_{k=1}^{K} \left(Z_{in,k} - Z_{g,k}\right)\left(Z^*_{in,k} - Z^*_{g,k}\right) \tag{115}$$

This error function depends on the lengths of line sections (l_i), characteristic impedances ($Z_{i,k}$), and propagation constants ($\gamma_{i,k}$). The propagation constants $\gamma_{i,k}$ for TEM modes are simple functions of effective dielectric constants of substrates and materials in the waveguide. Eventually, $Z_{i,k}$ and $\gamma_{i,k}$ are

Table 6 Input data of Examples 6–8 for step-line transformer

Example	N	K	f_l (GHz)	f_u (GHz)	h (mm)	ε_r	Z_g (Ω)	Z_L (Ω)		Error value margin
								R_L	X_L	
6	5	20	1	3	0.508	2.2	50	100	0	0.01
7	3	25	2	3	0.508	3.55	75	30	10	0.01
8	3	10	2	3	0.508	4.4	50	100	−10	0.001

Table 7 Output parameters of Examples 6–8 for step-line transformer

Example	Width (*mm*)		Impedance and width (mm) of each section		Length of each section (mm)	Z_s (Ω)		Z_L (Ω)		Error value
	Source	Load				R_s	L_s (*nH*)	R_L	L_L C_L	
6	1.57	0.46	Z_0	[51.88,57.83,68.96,83.16,94.85]	27.08	50	0	100	0	0.001
			W	[1.48,1.25,0.94,0.67,0.51]						
7	0.56	–	Z_0	[57.78,56,84,32.55]	9.75	75	0	30	0.64 (*nH*)	0.008
			W	[0.89,0.92,2.14]						
8	0.97	–	Z_0	[57.2,67.85,86.56]	13.5	50	0	100	6.37 (*pF*)	0
			W	[0.77,0.56,0.33]						

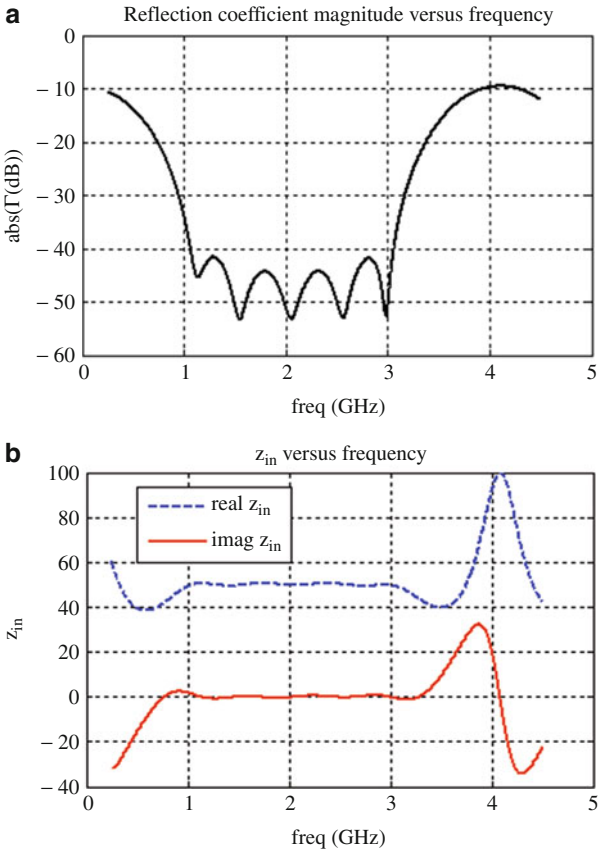


Fig. 29 Example 6. (**a**) Reflection coefficient magnitude versus frequency; (**b**) input impedance versus frequency

functions of the geometrical dimensions and material parameters of the waveguide. However, first the lengths of line sections (l_i) and their characteristic impedances ($Z_{0,i}$) are determined which may then be used to compute their geometrical dimensions.

The minimization of the error function may be performed by several methods, such as the steepest descent and conjugate gradient methods, which need the computation of derivatives of the error function. Their details for this problem are developed in reference (Oraizi 2001). However, for simplicity, the minimization routines in the MATLAB software are used here.

Furthermore, the details of the design of coaxial transformer considering the step changes of its inner conductor are provided in reference (Oraizi 2001). It has been shown that the step discontinuity

DOI 10.1007/978-981-4560-75-7_133-1

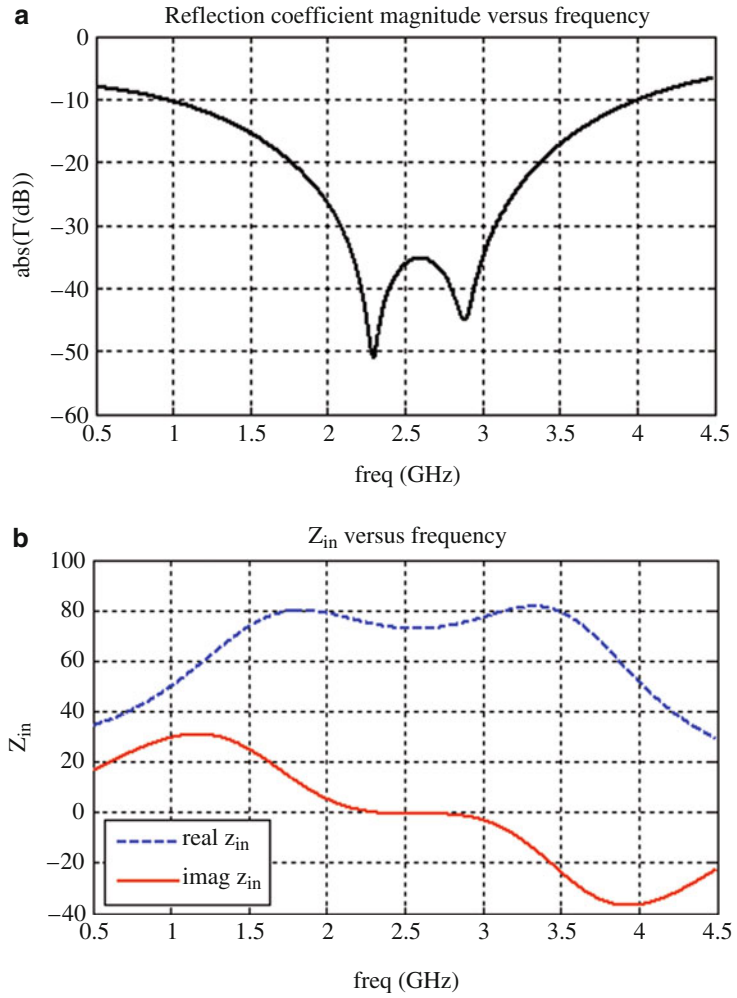


Fig. 30 Example 7. (**a**) Reflection coefficient magnitude versus frequency; (**b**) input impedance versus frequency

admittance of coaxial center conductor, step-line, and microstrip line sections has no appreciable effect on the transformer design, provided the line steps are quite small. The step discontinuities are ignored here for the transformer designs.

Next, design several examples of step-line transformers according to the specified input data expressed in Table 6. The initial length of step-line transformer is selected equal to the quarter wavelength (λ_g) of lower frequency limit of the specified bandwidth for the chosen substrate. The output parameters are listed in Table 7, which are obtained by computer programs. The widths of microstrip lines are obtained from the characteristic impedances of line sections together with the parameters of substrates by the approximate formulas in the literature (Edwards and Steer 2000).

Example 6

Design a 5-section step-line transformer. The magnitude of reflection coefficient and input impedance are drawn in Fig. 29a, b, respectively.

Example 7

Design a 3-section step-line transformer. The magnitude of reflection coefficient and input impedance are drawn in Fig. 30a, b, respectively.

DOI 10.1007/978-981-4560-75-7_133-1

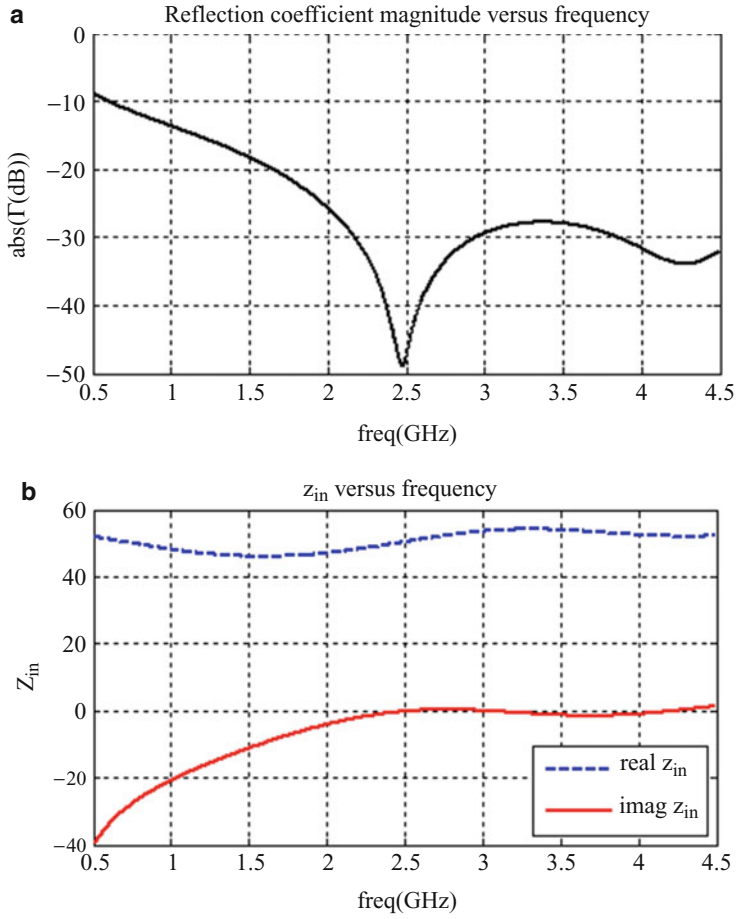


Fig. 31 Example 8. (**a**) Reflection coefficient magnitude versus frequency; (**b**) input impedance versus frequency

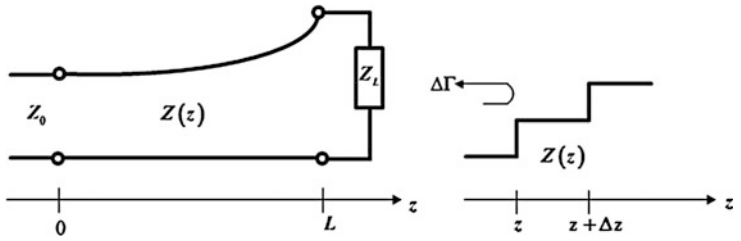


Fig. 32 The taper line. (**a**) Geometrical configuration of taper; (**b**) infinitesimal step of taper

Example 8

Design a 3-section step-line transformer. The magnitude of reflection coefficient and input impedance are drawn in Fig. 31a, b, respectively.

Taper Lines

As the number of sections of a step-line transformer increases and the length of each section decreases, the transformer shape becomes a continuous taper with a fixed length (Pozar 2011; Collin 2000). Consider a

DOI 10.1007/978-981-4560-75-7_133-1

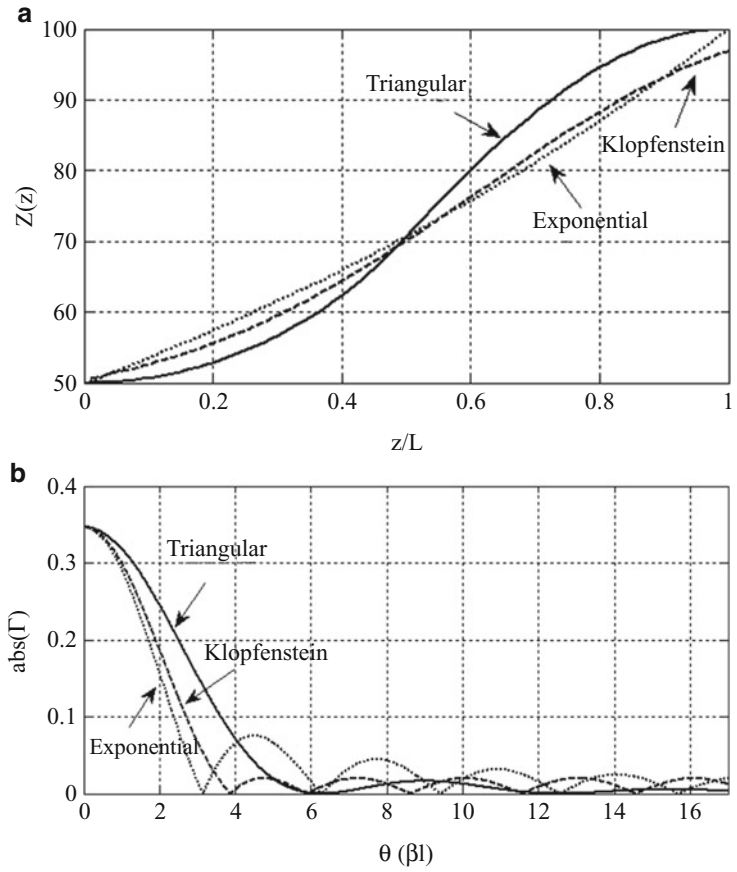


Fig. 33 Exponential, triangular, and Klopfenstein tapers ($Z_s = 50\Omega, \quad Z_L = 100\Omega$). (**a**) Characteristic impedance; (**b**) input reflection coefficient magnitude

tapered line of continuous characteristic impedance function $Z(z)$, propagation constant γ, and length L, connecting a source impedance Z_0 to a load impedance Z_L, as shown in Fig. 32a.

The taper line transformer provides a high-pass filter response, whereas the step-line transformer gives a band-pass filter characteristic.

Since the geometrical configuration of a taper continuously changes and the characteristic impedances of two consecutive differential lengths of line change infinitesimally, the theory of small reflections may be used to determine the reflection coefficient at the input of taper line.

Accordingly, consider an infinitesimal line section of length Δz with an impedance increment of ΔZ between two consecutive steps. The incremental reflection coefficient at point z on the line is then

$$\Delta\Gamma = \frac{Z + \Delta Z - Z}{Z + \Delta Z + Z} \approx \frac{\Delta Z}{2Z}$$

In the limit

$$d\Gamma = \frac{dZ}{2Z} = \frac{1}{2}\frac{d}{dz}\left(\ln\frac{Z}{Z_0}\right)dz$$

The characteristic impedance $Z(z)$ may be normalized with respect to Z_0, namely, the impedance of input line.

DOI 10.1007/978-981-4560-75-7_133-1

According to the theory of small reflections on the taper line, the total reflection coefficient at its input is equal to the sum of all small reflections along its length modified by the phase of wave on its round trip to and from point z on the line of length L. Therefore,

$$\Gamma(\omega) = \frac{1}{2}\int_{z=0}^{L} e^{-2\gamma z}\frac{d}{dz}\left(\ln\frac{Z(z)}{Z_0}\right)dz \tag{116}$$

Consequently, for a known function of characteristic impedance $Z(z)$, the input reflection coefficient may be found as a function of frequency. Conversely, for a desired specified reflection coefficient $\Gamma(\omega)$, the characteristic impedance may be determined by this integral equation. However, its analytical solution is quite involved. Shortly, the design of impedance transformers will be implemented by solving the integral equation numerically by the method of least squares.

For now, taper lines are designed by assuming various functions for their characteristic impedance and investigate their impedance matching responses.

Exponential Taper

Assume an exponential taper for the line characteristic impedance as depicted in Fig. 33a:

$$\frac{Z_E(z)}{Z_0} = \exp(az), \quad 0 < z < L \tag{117}$$

Constant a is evaluated as

$$a = \frac{1}{L}\ln\left(\frac{Z_L}{Z_0}\right) \tag{118}$$

For $Z_E(z=0) = Z_0$ and $Z_E(z=L) = Z_L$

The characteristic impedance $\frac{Z_E(z)}{Z_0}$ is drawn in Fig. 33a.

If the phase constant is independent of frequency (valid for TEM lines), the reflection coefficient is determined by substituting Eq. (116) in Eq. (115) using Eq. (117).

$$\Gamma_E(\beta L) = \frac{1}{2}\ln\left(\frac{Z_L}{Z_0}\right)\exp(-j\beta L)\frac{\sin\beta L}{\beta L} \tag{119}$$

The magnitude of reflection coefficient $|\Gamma_E|$ is drawn in Fig. 33b. Its zeroes occur at $\beta L = n\pi$, for $n = 1, 2, \ldots$.

Its magnitude decreases with increasing taper length L, which should be longer than $\frac{\lambda}{2}$ (namely $\beta L > \pi$) for an acceptable reflection coefficient.

Triangular Taper

Consider a triangular taper for $\frac{d}{dz}\ln\left(\frac{Z}{Z_0}\right)$ as

DOI 10.1007/978-981-4560-75-7_133-1

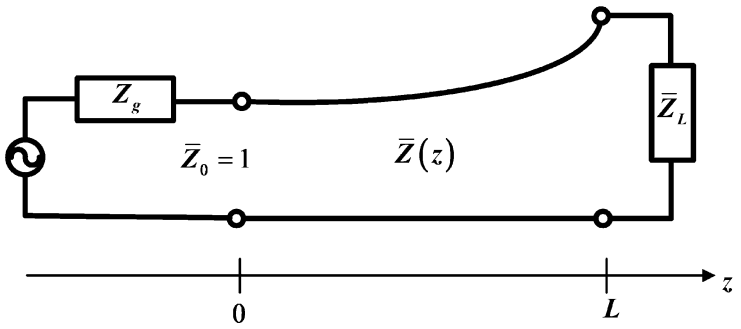


Fig. 34 Taper line impedance transformer

$$\frac{d}{dz}\left(\ln\frac{Z_T}{Z_0}\right) = \begin{cases} \dfrac{4z}{L^2}\ln\left(\dfrac{Z_L}{Z_0}\right) & , \ 0 \le z \le \dfrac{L}{2} \\ \left(\dfrac{4}{L} - \dfrac{4z}{L^2}\right)\ln\left(\dfrac{Z_L}{Z_0}\right) & , \ \dfrac{L}{2} \le z \le L \end{cases} \tag{120}$$

$$\frac{Z_T(z)}{Z_0} = \begin{cases} \exp\left[2\left(\frac{z}{L}\right)^2\right]\ln\left(\dfrac{Z_L}{Z_0}\right) & , \ 0 \le z \le \dfrac{L}{2} \\ \exp\left[\dfrac{4z}{L} - \dfrac{2z^2}{L^2} - 1\right]\ln\left(\dfrac{Z_L}{Z_0}\right) & , \ \dfrac{L}{2} \le z \le L \end{cases} \tag{121}$$

The characteristic impedance $\frac{Z_T}{Z_0}$ is drawn in Fig. 33a.

The reflection coefficient is obtained by substituting Eq. (119) in Eq. (115) and evaluating the integral

$$\Gamma_T(\beta L) = \frac{1}{2}\exp(-j\beta L)\ln\left(\frac{Z_L}{Z_0}\right)\left[\frac{\sin(\beta L/2)}{(\beta L/2)}\right]^2 \tag{122}$$

The magnitude of reflection coefficients $|\Gamma_T|$ is drawn in Fig. 33b. Its zeroes occur at $2n\pi$ for $n = 1, 2, \ldots$. Observe that the magnitude of reflection coefficient of triangular taper is lower than that of the exponential taper for $\beta L > 2\pi$.

Klopfenstein Taper

It has been shown that the Klopfenstein impedance taper gives the minimum reflection for a given taper length (longer than a critical value), and conversely it gives the shortest taper length for a given value of reflection coefficient. The details of its derivation and design procedure are given in the literature (Klopfenstein 1956). Its characteristic impedance $\frac{Z_K}{Z_0}$ and the magnitude of its reflection coefficient $|\Gamma_K|$ are drawn in Fig. 33a, b, respectively, and are compared to those of the exponential and triangular taper.

Design of Taper Line Impedance Transformers by the Method of Least Squares

Consider a taper line with characteristic impedance $Z(z)$ as a function of distance z connecting a voltage source with internal impedance Z_g to a load impedance Z_L shown in Fig. 34 (Oraizi 1996). Assume that the characteristic impedance of the input section of taper line (Z_0) is equal to the internal impedance of source (Z_g), which is resistive. All impedances are assumed normalized by Z_0, namely,

DOI 10.1007/978-981-4560-75-7_133-1

$$\overline{Z}_0 = 1, \overline{Z}(z) = \frac{Z(z)}{Z_0}, \overline{Z}_L = \frac{Z_L}{Z_0}$$

The reflection coefficient at the taper line input is derived in section "Taper Lines":

$$\Gamma = \frac{1}{2}\int_0^L e^{-j2\beta z}\frac{d}{dz}\left(\ln\overline{Z}(z)\right)dz \tag{123}$$

$$= \frac{1}{2}\exp(-j2\beta L)\ln\overline{Z}(L) + j\beta\int_0^L \exp(-j2\beta z)\ln\overline{Z}(z)dz \tag{124}$$

The last equation is obtained through integration by parts. The boundary conditions at the source and load sides of taper line are

$$\begin{aligned}\overline{Z}(z=0) &= \overline{Z}_0 = 1\\ \overline{Z}(z=L) &= \overline{Z}_L\end{aligned} \tag{125}$$

The method of least squares is used to develop a design algorithm for a step-line and taper line transformer shapes. Accordingly, the functions $\overline{Z}(z)$, $\ln\overline{Z}(z)$, or $\frac{d}{dz}\left(\ln\overline{Z}(z)\right)$ are approximated by a polynomial, pulse or step functions, triangular or piecewise linear functions, and an approximate operator. An error function is then constructed for Γ and minimized to determine the coefficients of the polynomial or the amplitudes of the expansion functions.

Polynomial Expansion of $\frac{d}{dz}\ln\overline{Z}$

Approximate $\frac{d}{dz}\ln\overline{Z}(z)$ by a polynomial of degree N as

$$\frac{d}{dz}\ln\overline{Z}(z) = \sum_{n=0}^{N} a_n z^n \tag{126}$$

$$\overline{Z}(z) = \exp\left[\sum_{n=0}^{N+1}\frac{1}{n+1}a_n z^{n+1} + C\right] \tag{127}$$

where C is a constant. Evaluate a_0 and C by using the boundary conditions in Eq. (124).

$$\begin{aligned}a_0 &= \frac{1}{L}\ln\overline{Z}_L - \sum_{n=1}^{N}\frac{a_n}{n+1}L^n\\ C &= 0\end{aligned} \tag{128}$$

Substitute $v = \left(\frac{\ell}{\sigma S}\right)i = Ri$ from Eq. (127) into Eq. (125) or Eq. (126) to obtain

$$\ln\overline{Z}(z) = \left(\frac{1}{L}\ln\overline{Z}_L - \sum_{n=1}^{N}\frac{a_n}{n+1}L^n\right)z + \sum_{n=1}^{N}\frac{a_n}{n+1}z^{n+1} \tag{129}$$

DOI 10.1007/978-981-4560-75-7_133-1

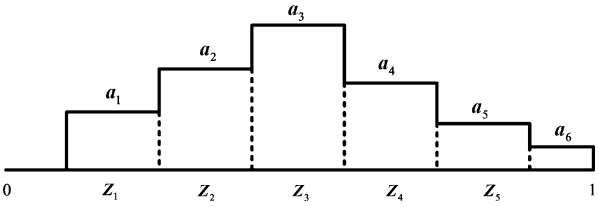


Fig. 35 Representation of $\ln\overline{Z}(z)$ by pulse functions

Equation (128) is then substituted in Eq. (122) to obtain the reflection coefficient:

$$\Gamma = \frac{1}{2L}\ln\overline{Z}_L\int_0^L e^{-j2\beta z}dz + \frac{1}{2}\sum_{n=1}^{N} a_n\left[\int_0^L z^n e^{-j2\beta z}dz - \frac{L^n}{n+1}\int_0^L e^{-j2\beta z}dz\right] \tag{130}$$

Integrate by parts the second integral in Eq. (129) successively or use available integral tables to obtain

$$\Gamma_k = t_k + \sum_{n=1}^{N} l_{nk}a_n \tag{131}$$

where

$$t_k = \frac{1}{2}\ln\overline{Z}_L\frac{\sin\beta_k L}{\beta_k L}e^{-j\beta_k L} \tag{132a}$$

$$l_{nk} = \frac{1}{2}L^{n+1}\left[\frac{n!e^{-j2\beta_k L}}{j2\beta_k L}\sum_{i=0}^{n}\frac{1}{(n-i)!(j2\beta_k L)^i} - \frac{n!}{(j2\beta_k L)^{n+1}} + \frac{1}{n+1}\frac{\sin\beta_k L}{\beta_k L}e^{-j\beta_k L}\right] \tag{132b}$$

Here, the specified bandwidth from f_1 to f_2 is divided into K discrete frequencies, each one is designed by f_k.

Then, construct an error function

$$\text{error} = \sum_{k=1}^{K}|\Gamma_k|^2 \tag{133}$$

which is a quadratic function of polynomial coefficients a_n.

Accordingly, the derivatives of error with respect to a_n^* are evaluated and equated to zero to obtain a set of N linear equations for the evaluation of N unknowns a_n.

$$[a_n] = \left[\text{Re}\sum_{k=1}^{K} l_{nk}l_{mk}^*\right]^{-1}\left[-\text{Re}\sum_{k=1}^{K} l_{mk}^* t_k\right] \tag{134}$$

The error is a non-quadratic function of the taper line length (L).

Its minimization with respect to L may be performed by any optimization procedure or any algorithm available in the MATLAB software. The minimization of error may be pursued simultaneously or separately on $\{a_n\}$ and/or L.

DOI 10.1007/978-981-4560-75-7_133-1

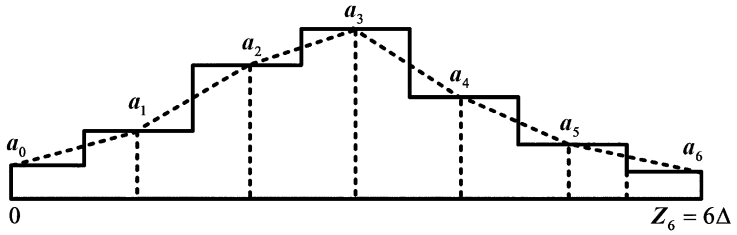


Fig. 36 Representation of $\ln\overline{Z}(z)$ by approximate operator

Approximation of $\ln\overline{Z}(z)$ by Pulse Functions

Assume that the line length L is divided into N + 1 sections, each having a length equal to $\Delta = L/(N+1)$ as shown in Fig. 35. The midpoints of sections are located at $z_n = n\Delta$ for $n = 1, 2, \ldots, N$. The nth section pulse is defined as

$$P_n(z - z_n) = \begin{cases} 1 & |z - z_n| < \dfrac{\Delta}{2} \\ 0 & |z - z_n| > \dfrac{\Delta}{2} \end{cases} \tag{135}$$

The function $\ln\overline{Z}(z)$ is now expanded as pulse functions on the line

$$\ln\overline{Z}(z) = \sum_{n=0}^{N+1} a_n P_n(z - z_n), \qquad 0 \le z \le \mathrm{L} \tag{136}$$

where a_n are the pulse amplitudes. The first P_0 and last P_{N+1} pulses are defined in the intervals $0 \le z \le \frac{\Delta}{2}$ and $L - \frac{\Delta}{2} \le z \le L$, respectively. Using boundary conditions in Eq. (124) gives $a_0 = 0$ and $a_{N+1} = \ln\overline{Z}_L$. The computation of $\frac{d}{dz}\ln[Z(z)]$ requires the derivatives of pulse functions leading to impulse functions at $n\Delta \pm \frac{\Delta}{2}$. Substitute then this function into Eq. (122), perform the integration, and use the sampling property of impulse functions to obtain

$$\Gamma = j\sin(\beta\Delta)\sum_{n=1}^{N} a_n \exp(-j2\pi\beta\Delta) + \frac{1}{2}\ln\overline{Z}_L \exp[-j(2N+1)\beta\Delta] \tag{137}$$

This equation can also be derived using Eq. (133). Equation (136) is then put in the form of Eq. (130), where t_k and ℓ_k are defined.

For the minimization of Γ by Eq. (132b) and Eq. (133), the following relations are needed:

$$\mathrm{Re}\left(\sum_{k=1}^{K} I_{mk}^{*} I_{nk}\right) = \sum_{k=1}^{K} \sin^2(\beta_k\Delta)\cos[2(m-n)\beta_k\Delta] \tag{138}$$

$$\mathrm{Re}\left(\sum_{k=1}^{K} I_{mk}^{*} t_k\right) = \frac{1}{2}\sum_{k=1}^{K} \ln\overline{Z}_L \sin(\beta_k\Delta)\sin[(2m-2N-1)\beta_k\Delta] \tag{139}$$

The minimization of error with respect to the line length L may be performed by any minimization routine needing its derivative or not.

DOI 10.1007/978-981-4560-75-7_133-1

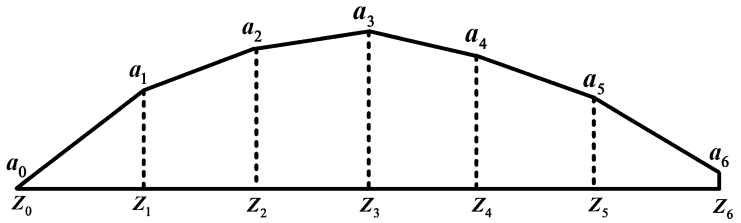


Fig. 37 Representation of $\ln\overline{Z}(z)$ by piecewise linear functions

The characteristic impedance of the multi-section line is then

$$\overline{Z}(z) = \sum_{n=1}^{N} e^{a_n} p_n(z - n\Delta) + \overline{Z}_L P_{N+1}(z - L) \tag{140}$$

Approximate Operator for $\frac{d}{dz}\ln\overline{Z}(z)$

Divide the line length L into $N + 1$ sections, with the width of each section equal to $\Delta = \frac{L}{(N+1)}$. Consider the points in each section at a distance of $z_n = n\Delta$ from the input, as shown in Fig. 36. Represent the characteristic impedance $\overline{Z}(z)$ of the line as a series of impulse functions:

$$\overline{Z}(z) = \sum_{n=0}^{N} a_n \delta(z - n\Delta), \quad 0 \le z \le L \tag{141}$$

The boundary conditions give $a_0 = 1$ for $0 \le z < \frac{\Delta}{2}$ and $a_{N+1} = \overline{Z}_L$ for $L - \frac{\Delta}{2} < z \le L$. Use the approximate operator for $\frac{d}{dz}\ln\overline{Z}(z)$ at the midpoint of each section.

$$\begin{aligned} \frac{d}{dz}\ln\overline{Z}(z)\Big|_{z=\frac{n-1}{2}\Delta} &\approx \frac{1}{\Delta}\left[\ln\overline{Z}(n\Delta) - \ln\overline{Z}(n\Delta - \Delta)\right] \\ &\approx \frac{1}{\Delta}\ln\left(\frac{a_n}{a_{n-1}}\right), \quad (n-1)\Delta < z < n\Delta \end{aligned} \tag{142}$$

Substituting Eq. (141) into Eq. (122) and performing the integration give

$$\Gamma = \frac{\sin\beta\Delta}{\beta\Delta}\left[j\sin\beta\Delta\sum_{n=1}^{N}\ln(a_n)e^{-j2n\beta\Delta} + \frac{1}{2}\ln\left(\overline{Z}_L\right)e^{-j(2N+1)\beta\Delta}\right] \tag{143}$$

Use Eq. (133) to compute ln a_n with the following expressions:

$$\mathrm{Re}\left(\sum_{k=1}^{K} l_{mk}^{*} l_{nk}\right) = \sum_{k=1}^{K} \frac{\sin^4(\beta_k\Delta)}{(\beta_k\Delta)^2}\cos\left[2(m-n)\beta_k\Delta\right] \tag{144}$$

$$\mathrm{Re}\left(\sum_{k=1}^{K} l_{mk}^{*} t_k\right) = \frac{1}{2}\sum_{k=1}^{K} \ln\overline{Z}_L \frac{\sin^3(\beta_k\Delta)}{(\beta_k\Delta)^2}\sin\left[(2m-2N-1)\beta_k\Delta\right] \tag{145}$$

DOI 10.1007/978-981-4560-75-7_133-1

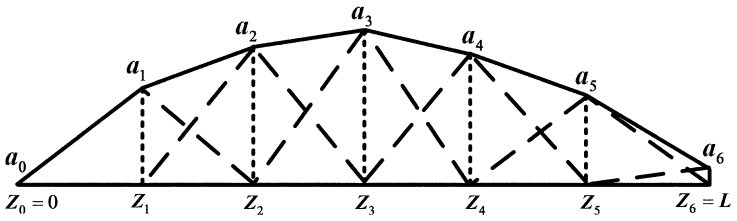


Fig. 38 Representation of $\ln\overline{Z}(z)$ by triangle functions

Table 8 Input data of Examples 9–10 for taper line transformer

Example	L_{in} (cm)	N	K	f_l (GHz)	f_u (GHz)	h (mm)	ε_r	Z_g (Ω)	Z_L (Ω)
9–12	20	10	20	1	4	0.508	2.2	50	100
13–16	20	10	20	1	4	0.787	3.55	120	30

Piecewise Linear Approximation of $\ln\overline{Z}(z)$

Divide the taper line of length L into N sections of equal widths $\Delta = \frac{L}{N}$, as shown in Fig. 37. The section points are at $z_n = n\Delta$ for $n = 0, 1, 2, \ldots, N$. Let $a_n = \ln\overline{Z}(z = z_n)$ and assume the variation of $\ln\overline{Z}(z)$ in each section to be linear. Then,

$$\ln\overline{Z}(z) = \frac{a_n - a_{n-1}}{\Delta} z + a_{n-1} - (n-1)(a_n - a_{n-1}), \quad (n-1)\Delta \le z \le n\Delta, \ \ n = 1, 2 \ldots, N \tag{146}$$

The boundary conditions at $z = 0$ (for $n = 1$ and $\overline{Z}(0) = 1$) and at $z = L$ (for $n = N$ and $\overline{Z}(L) = \overline{Z}_L$) give $a_0 = 1$ and $a_N = \ln\overline{Z}_L$.

Substitutions of Eq. (145) into Eq. (122) or Eq. (123) give Eq. (142), but by replacing $(N - 1)$ for N, since the line here is divided into N sections. The unknown quantities are here a_n. Consequently, the approximate operator for $\frac{d}{dz}\ln\overline{Z}(z)$ and piecewise linear approximation of $\ln\overline{Z}(z)$ are identical. Equations (142) and (143) also apply here, but for $(N - 1)$ replacing N. The characteristic impedance of each section is then

$$\overline{Z}(z) = \exp\left(\frac{a_n - a_{n-1}}{\Delta} z + n a_{n-1} - (n-1)a_n\right), \quad (n-1)\Delta \le z \le n\Delta \tag{147}$$

Alternatively, the function $\ln\overline{Z}(z)$ can be approximated by a triangle function of amplitude a_n as in Fig. 38. This technique gives the same results as above.

Next, design several tapeline transformers according to the specified input data in Table 8. The output parameters are listed in Table 9.

Each example is solved by four methods, namely, approximate operator for $\frac{d}{dz}\ln\overline{Z}$, approximation of $\ln\overline{Z}(z)$ by pulse function, polynomial expansion of $\frac{d}{dz}\ln\overline{Z}$, and approximation of $\ln\overline{Z}$ by triangular function, for which the computer programs are used.

Example 9

Design a taper line according to the data in the first row in Table 8, by the approximate operator for $\frac{d}{dz}\ln\overline{Z}$. Its characteristic impedance, reflection coefficient, and input impedance are drawn in Fig. 39a–c, respectively.

Table 9 Output parameters of Examples 9–10 for taper line transformer

Example		Width (mm)		Length of taper (cm)	Normalized characteristic impedance of sections from source	Error value
		Source	Load			
9–12	Approximate operator for $\Gamma_{in} = \frac{Z_{in}-Z_0}{Z_{in}+Z_0} = \frac{V_1^-}{V_1^+} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1-S_{22}\Gamma_L}$	1.56	0.46	30	[1,1.0245,1.0649,1.1338,1.2262,1.3478,1.4839,1.6311,1.7639,1.878]	0.038
	Approximation of $\Gamma_S = \frac{V_1^+}{V_1^-}$ by pulse function			30	[1,1.0228,1.0638,1.1321,1.2254,1.3471,1.4847,1.6322,1.7666,1.8801]	0.053
	Polynomial expansion of $\Gamma_{out} = \frac{V_2^-}{V_2^+} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1-S_{11}\Gamma_S}$			19.986	[1,1.0303,1.0831,1.1626,1.2755,1.4141,1.5679,1.7202,1.8468,1.9419]	0.021
	Approximation of $V_1 = V_S \frac{Z_{in}}{Z_{in}+Z_S} = V_1^+ + V_1^- = V_1^+(1+\Gamma_{in})$ by triangular function			30	[1,1.0264,1.0763,1.1577,1.2717,1.4142,1.5727,1.7275,1.8583,1.9486,2]	0.038
13–16	Approximate operator for $V_1^+ = \frac{V_S}{2}\frac{1-\Gamma_S}{1-\Gamma_S\Gamma_{in}}$	0.27	3.7	30	[1,0.9528,0.8818,0.7779,0.6651,0.5505,0.4541,0.3759,0.3214,0.2836]	0.151
	Approximation of $Z_{in} = Z_0(1+\Gamma_{in})/(1-\Gamma_{in})$ by pulse function			30	[1,0.9559,0.8837,0.7802,0.666,0.5511,0.4537,0.3754,0.3204,0.2829]	0.211
	Polynomial expansion of $\frac{d}{dz}\ln\overline{Z}$			19.997	[1,0.9422,0.8525,0.7398,0.6147,0.5000,0.4068,0.3379,0.2932,0.2653]	0.084
	Approximation of $\ln\overline{Z}$ by triangular function			30	[1,0.9492,0.8633,0.7461,0.6183,0.5,0.4043,0.3351,0.2896,0.2634,0.25]	0.152

DOI 10.1007/978-981-4560-75-7_133-1

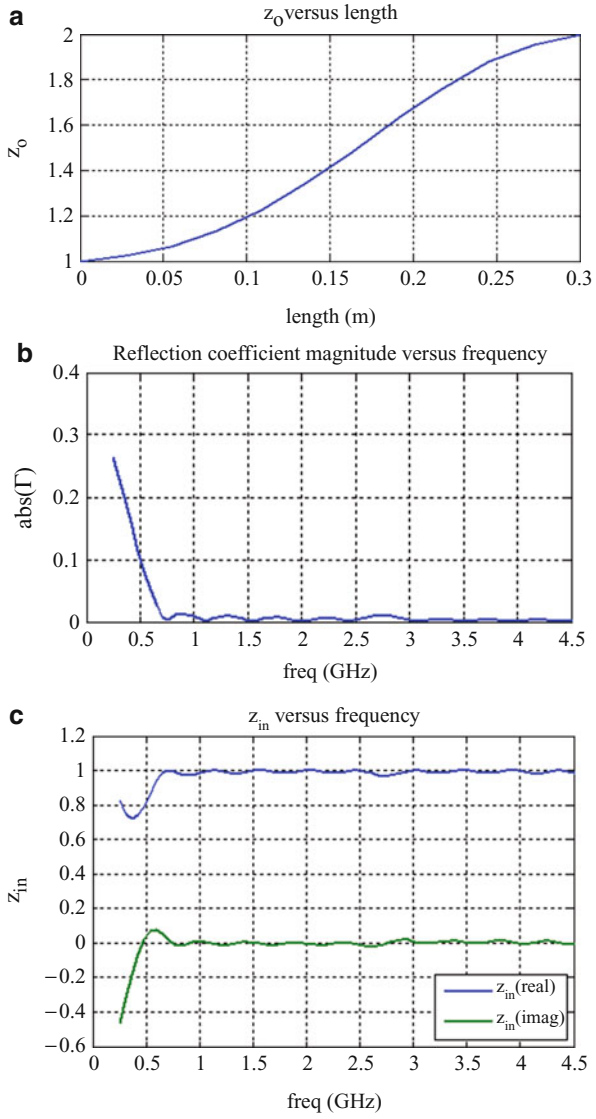


Fig. 39 Example 9. (**a**) Characteristic impedance of tapered line; (**b**) reflection coefficient magnitude versus frequency; (**c**) input impedance versus frequency

Example 10

Design a taper line according to the data in the first row in Table. 8, by the approximation of $\ln\overline{Z}(z)$ by pulse function. Its characteristic impedance, reflection coefficient, and input impedance are drawn in Fig. 40a–c, respectively.

Example 11

Design a taper line according to the data in the first row in Table. 8, by the polynomial expansion of $\frac{d}{dz}\ln\overline{Z}$. Its characteristic impedance, reflection coefficient, and input impedance are drawn in Fig. 41a–c, respectively.

DOI 10.1007/978-981-4560-75-7_133-1

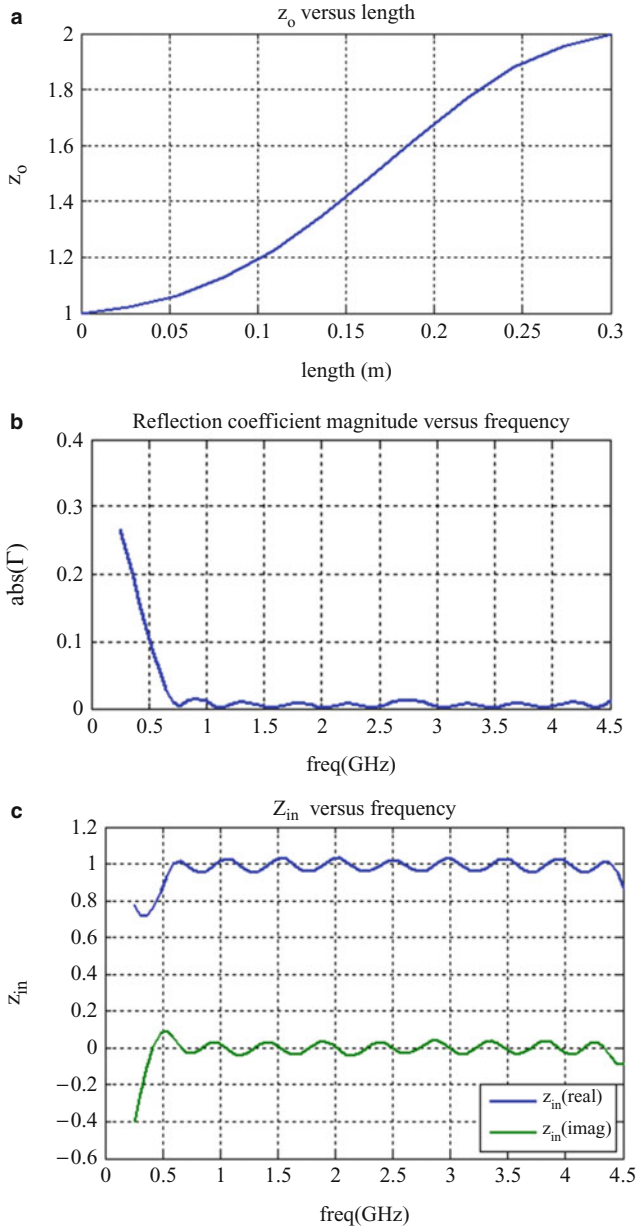


Fig. 40 Example 10. (**a**) Characteristic impedance of tapered line; (**b**) reflection coefficient magnitude versus frequency; (**c**) input impedance versus frequency

Example 12

Design a taper line according to the data in the first row in Table 8, by the approximation of $\ln\overline{Z}$ by triangular function. Its characteristic impedance, reflection coefficient, and input impedance are drawn in Fig. 42a–c, respectively.

Example 13

Design a taper line according to the data in the second row in Table 8, by the approximate operator for $\frac{d}{dz}\ln\overline{Z}$. Its characteristic impedance, reflection coefficient, and input impedance are drawn in Fig. 43a–c, respectively.

DOI 10.1007/978-981-4560-75-7_133-1

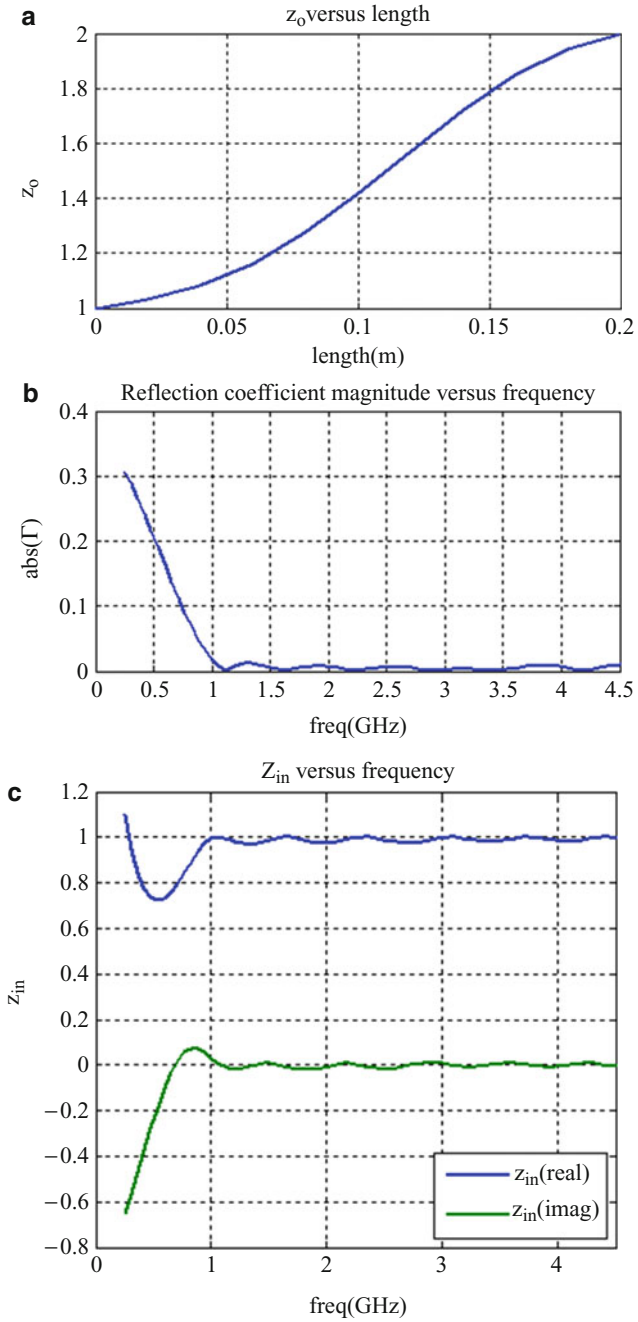


Fig. 41 Example 11. (**a**) Characteristic impedance of tapered line; (**b**) reflection coefficient magnitude versus frequency; (**c**) input impedance versus frequency

Example 14

Design a tapered line according to the data in the second row in Table 8, by the approximation of $\ln\overline{Z}(z)$ by pulse function. Its characteristic impedance, reflection coefficient, and input impedance are drawn in Fig. 44a–c, respectively.

DOI 10.1007/978-981-4560-75-7_133-1

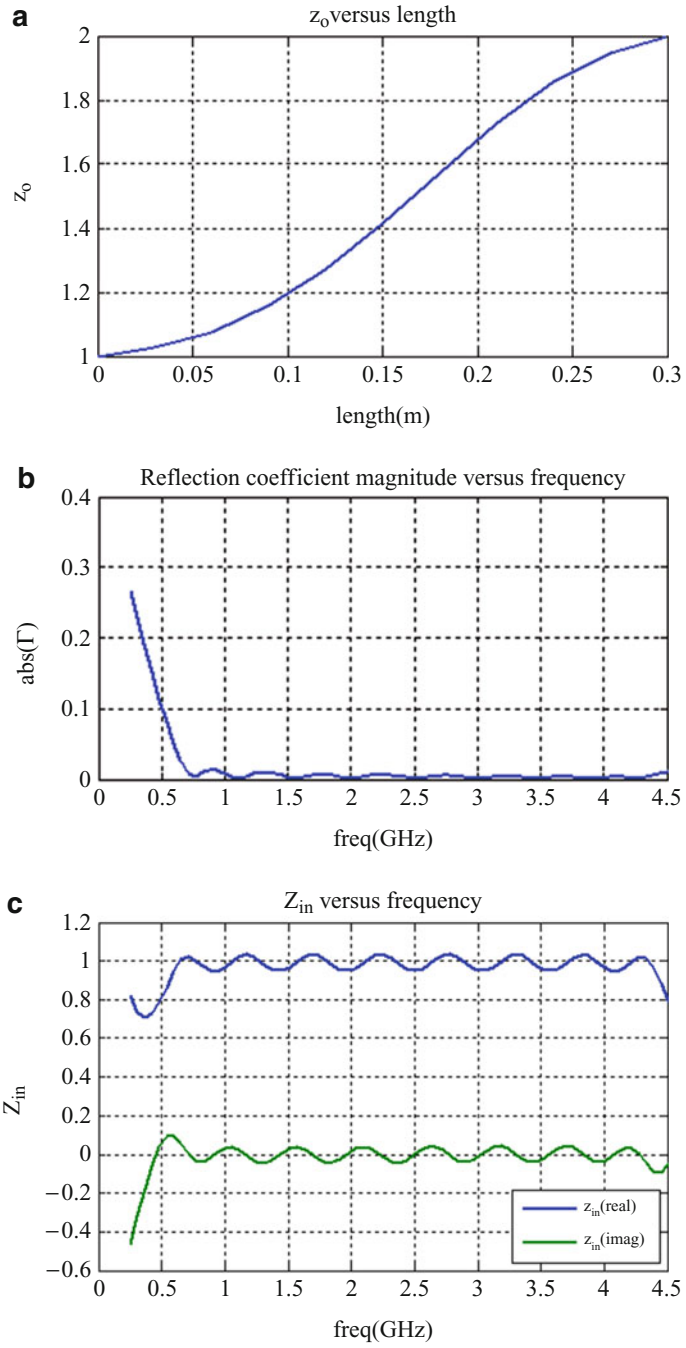


Fig. 42 Example 12. (**a**) Characteristic impedance of tapered line; (**b**) reflection coefficient magnitude versus frequency; (**c**) input impedance versus frequency

Example 15

Design a tapered line according to the data in the second row in Table 8, by the polynomial expansion of $\frac{d}{dz}\ln\overline{Z}$. Its characteristic impedance, reflection coefficient, and input impedance are drawn in Fig. 45a–c, respectively.

DOI 10.1007/978-981-4560-75-7_133-1

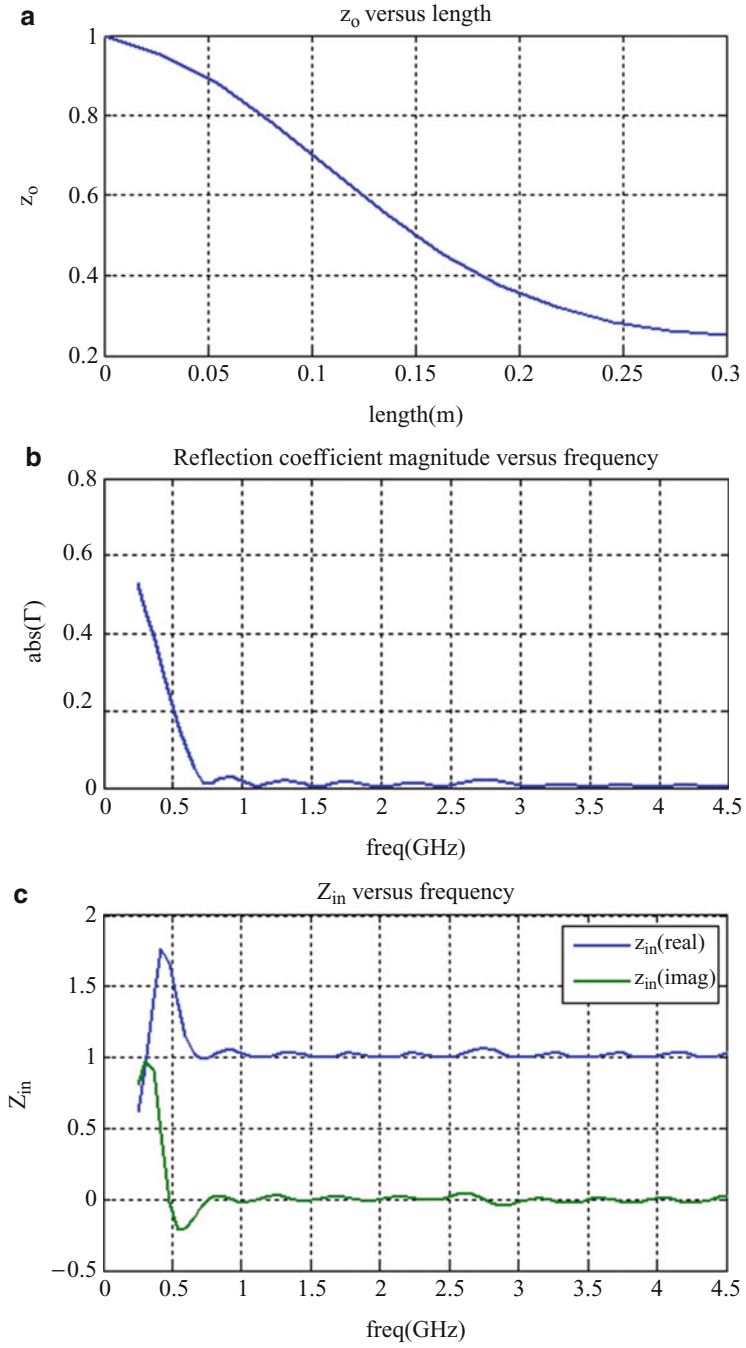


Fig. 43 Example 13. (**a**) Characteristic impedance of tapered line; (**b**) reflection coefficient magnitude versus frequency; (**c**) input impedance versus frequency

Example 16

Design a tapered line according to the data in the second row in Table 8, by the approximation of $\ln\overline{Z}$ by triangular function. Its characteristic impedance, reflection coefficient, and input impedance are drawn in Fig. 46a–c, respectively.

DOI 10.1007/978-981-4560-75-7_133-1

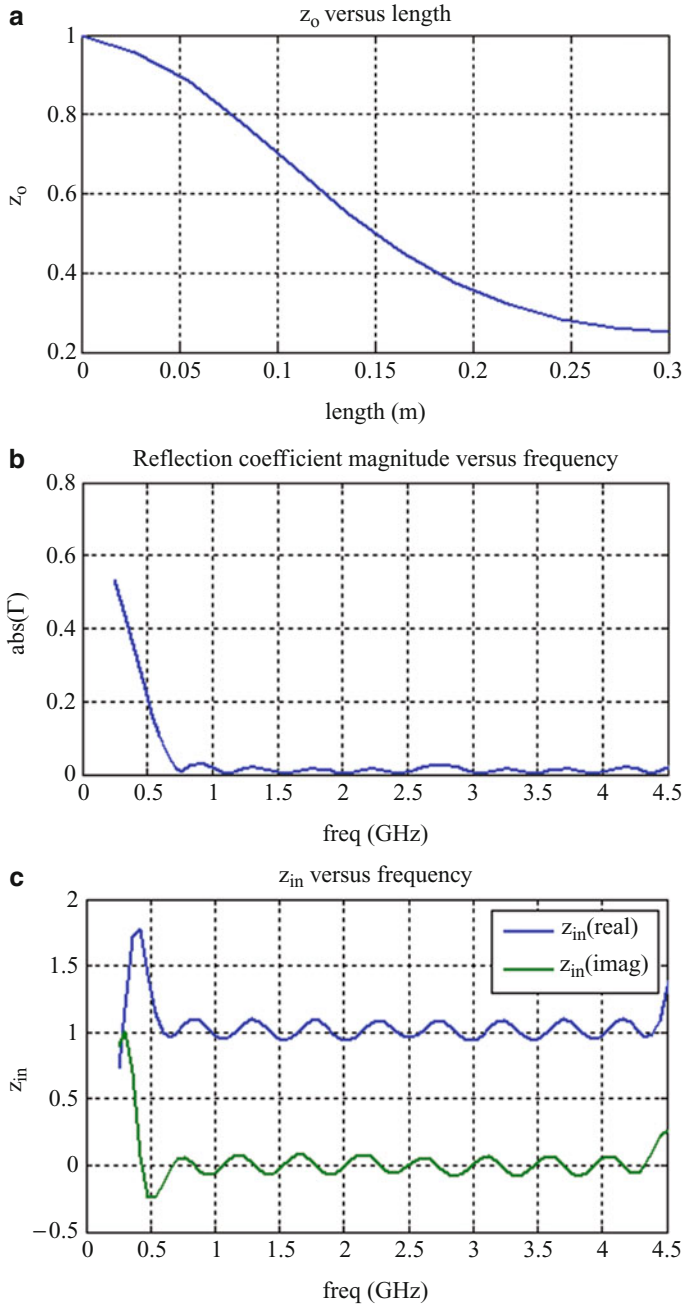


Fig. 44 Example 14. (**a**) Characteristic impedance of tapered line; (**b**) reflection coefficient magnitude versus frequency; (**c**) input impedance versus frequency

Devices and Components for Impedance Matching

Various resistive and reactive components are devised for the realization of impedance matching circuits in different waveguiding systems, such as coaxial waveguides, hollow cylindrical waveguides (rectangular, cylindrical, and elliptical), striplines, microstrips, and various variations and modifications thereof (Collin 2000; Matthaei et al. 1964; Pozar 2011; Reich et al. 1953; Rizzi 1988).

Examples of coaxial step-line transformer, single-stub tuning, double-stub tuning, and movable metal slugs are shown in Fig. 47.

DOI 10.1007/978-981-4560-75-7_133-1

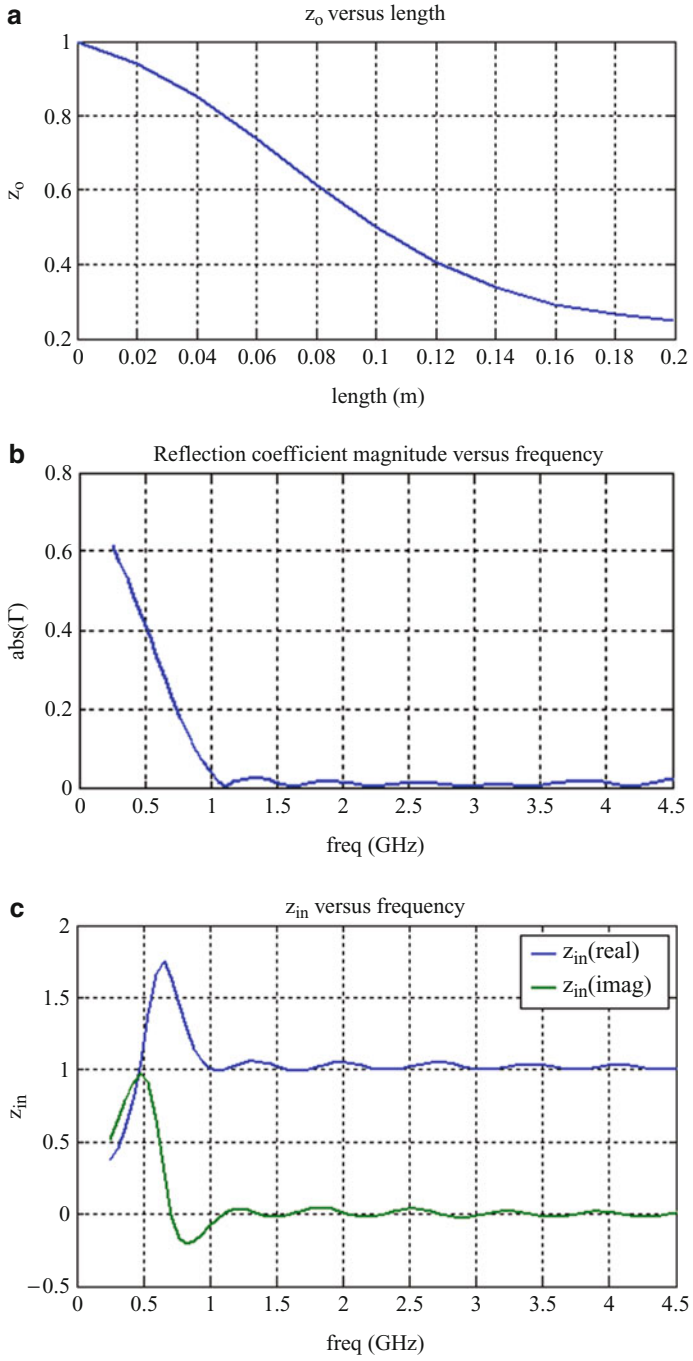


Fig. 45 Example 15. (**a**) Characteristic impedance of tapered line; (**b**) reflection coefficient magnitude versus frequency; (**c**) input impedance versus frequency

Examples of rectangular waveguide components, such as symmetrical and asymmetrical irises for parallel inductive and capacitive reactances, are shown in Fig. 48.

Examples of waveguide tuning parts are shown in Fig. 49.

Examples of microstrip resistors, capacitors, and inductors are shown in Fig. 50.

DOI 10.1007/978-981-4560-75-7_133-1

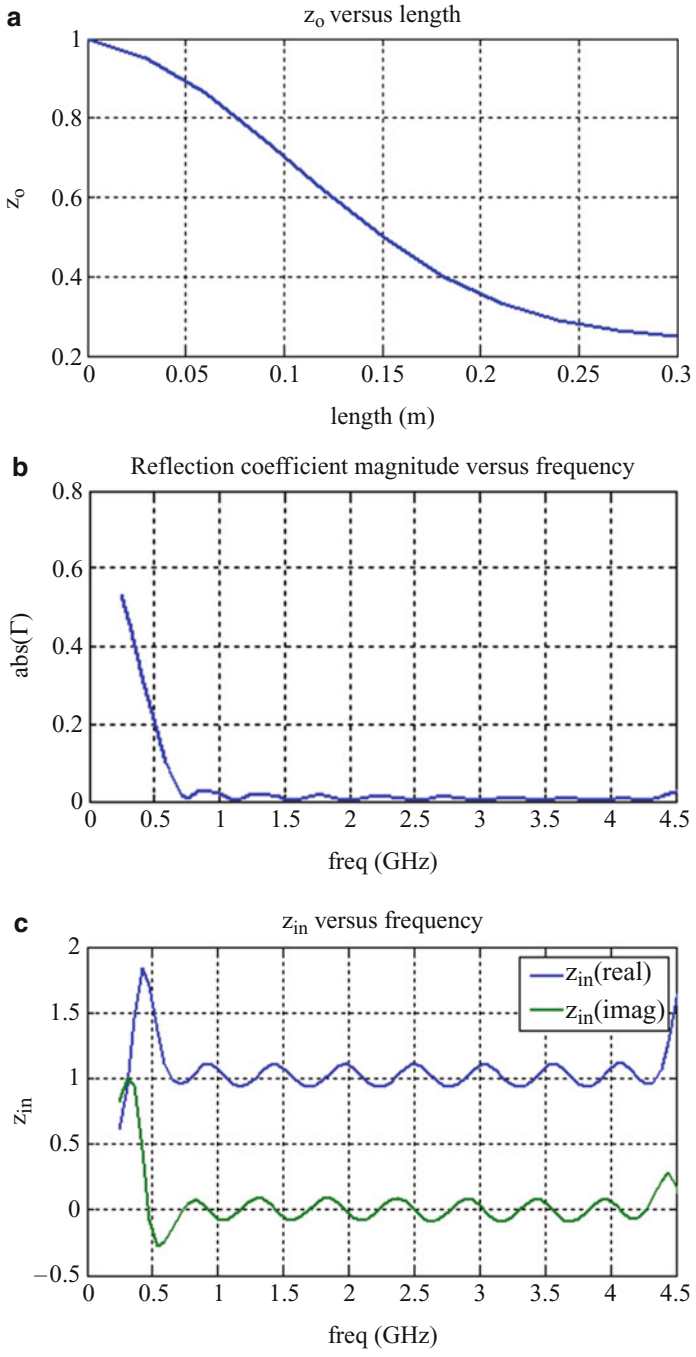


Fig. 46 Example 16. (**a**) Characteristic impedance of tapered line; (**b**) reflection coefficient magnitude versus frequency; (**c**) input impedance versus frequency

BALUN

BALUN (balanced to unbalanced) is a device that converts a balanced signal (flowing on two wires in the absence of a ground path) to an unbalanced signal (carried on a wire together with a ground path) (Reich et al. 1953; Terman 1955; Mongia et al. 1999). A BALUN may also perform impedance matching and transformation, called transformer BALUN. Similarly, UNUN (unbalanced to unbalanced) is a device that transfers signals between unbalanced lines.

DOI 10.1007/978-981-4560-75-7_133-1

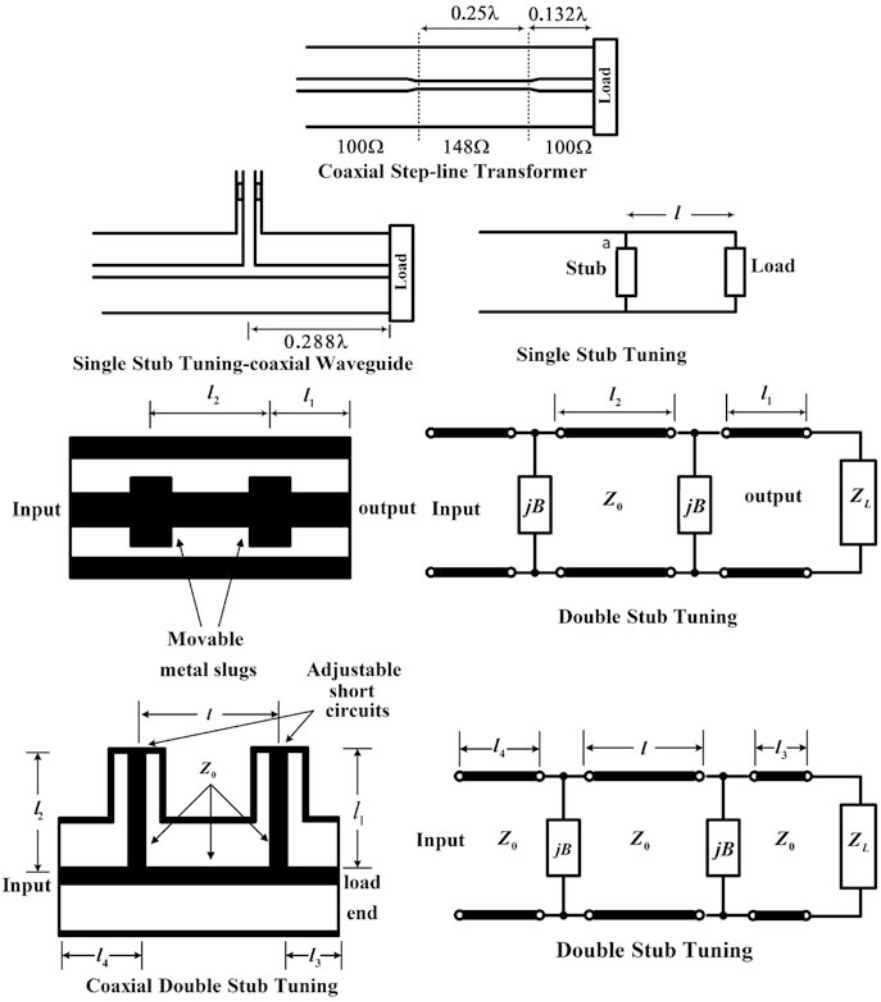


Fig. 47 Resistive and reactive components

BALUN is a device for transition between a balanced or differential transmission line (with opposite currents on the two wires) and an unbalanced or single-ended transmission line where the current path passes through ground. Two conductors having equal and opposite potentials constitute a balanced line, such as shielded two-wire line (where no current flows in the grounded shield). Two conductors of different dimensions constitute an unbalanced line, such as coaxial line, stripline, and microstrip line.

BALUN is a three-port device with an unbalanced matched input (port 1) and differential output ports (ports 2 and 3) with the following scattering parameters:

$$S_{11} = 0$$

$$S_{12} = S_{21} = -S_{13} = -S_{31}$$

$$S_{22} \neq 0,\ S_{33} \neq 0$$

Its two balanced outputs are of equal magnitude but with 180° phase difference in the frequency domain and opposite in sign in the time domain, which may be considered as the two output ports of the BALUN. It is a reciprocal device. Its outputs may not be matched and its output impedances may not

DOI 10.1007/978-981-4560-75-7_133-1

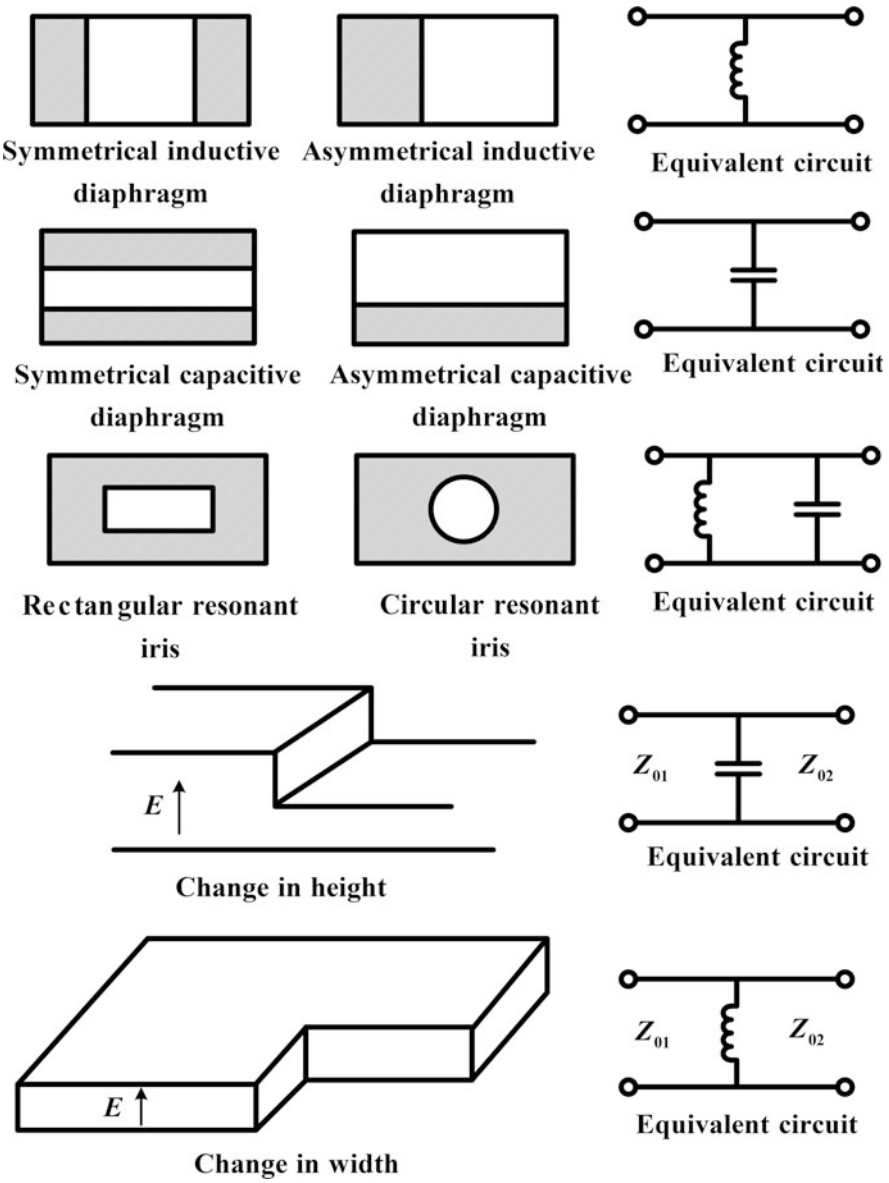


Fig. 48 Reactive waveguide components

be equal to that of the input. The return losses may be different for the differential and common mode signals.

BALUNs have wide applications in RF circuitry, such as balanced mixers, balanced multipliers, balanced modulators, push-pull amplifiers, phase shifters, dipole feeds, and various other antennas. They achieve compatibility between systems. Active BALUNs are also designed for integrated circuits.

In the balanced line, the two wires are symmetrical with respect to the ground, that is, they are not grounded and they have equal impedances with respect to the ground (such as the open two-wire line). In the unbalanced case, one of the lines is directly connected to the ground (such as the coaxial line). Sometimes, it is required to feed a grounded antenna (which is asymmetrical to the ground) through a two-wire line or conversely to feed an ungrounded antenna (which is symmetrical with respect to the ground) by a coaxial line. In such cases, it is necessary to convert between the balanced and unbalanced systems by a BALUN. At lower frequencies, a transformer may be used, as shown in Fig. 51. It transforms the input to output impedances in the ratio of $1:n^2$.

DOI 10.1007/978-981-4560-75-7_133-1

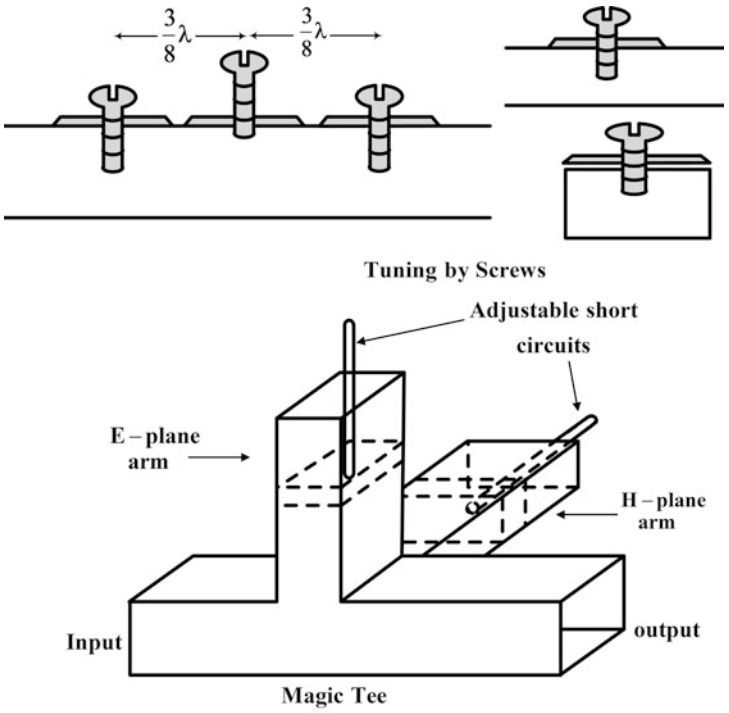


Fig. 49 Waveguide tuning parts

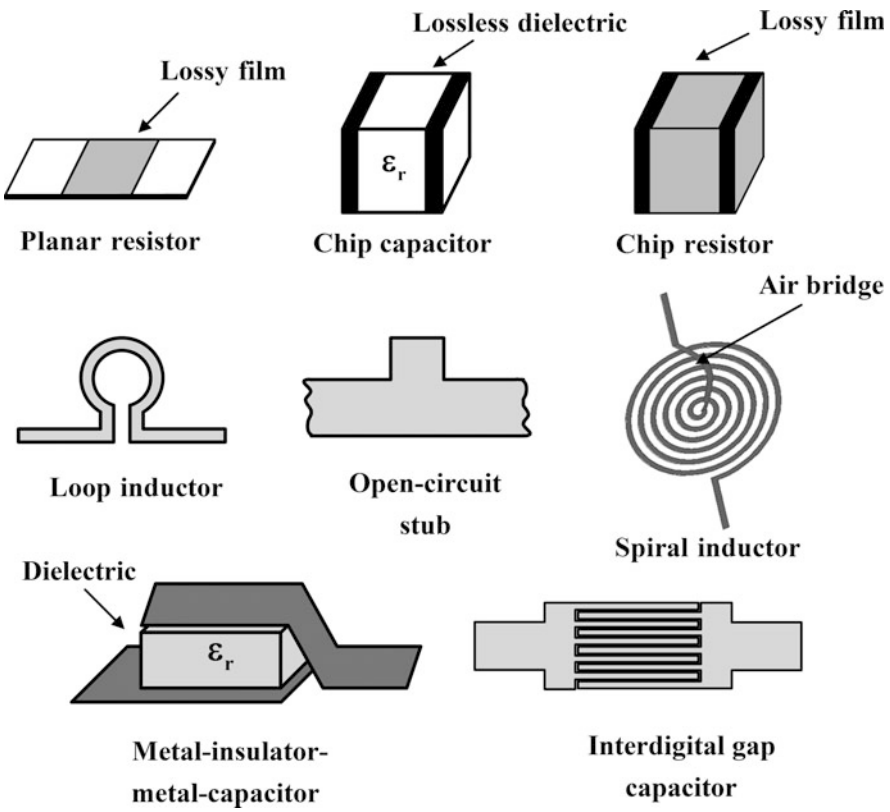


Fig. 50 Microstrip components

DOI 10.1007/978-981-4560-75-7_133-1

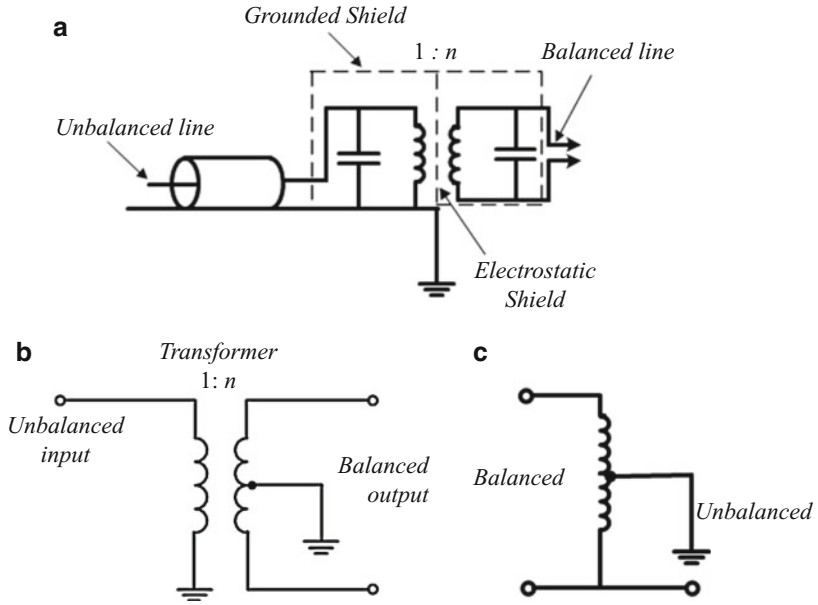


Fig. 51 Transformer BALUN. (**a**) A tuned transformer as BALUN; (**b**) a 1:n transformer BALUN; (**c**) an autotransformer as BALUN (Terman 1955)

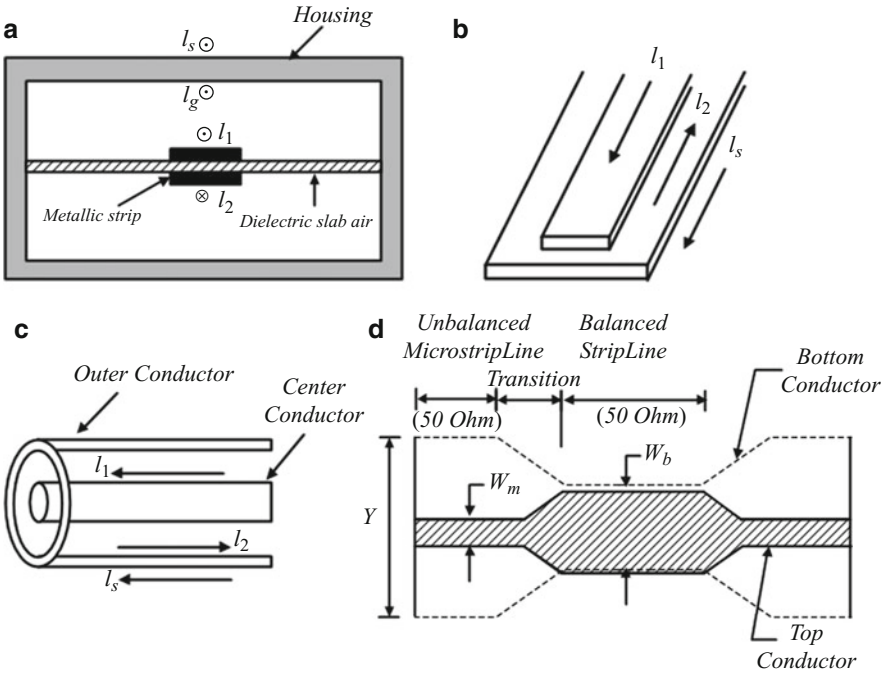


Fig. 52 Examples of balanced and unbalanced lines (Mongia et al. 1999). (**a**) Shielded two-wire line (balanced); (**b**) microstrip line (unbalanced); (**c**) coaxial line (unbalanced); (**d**) back-to-back connection of BALUNs composed of unbalanced microstrip line and balanced strip line

The impedance ratio of BALUN (or turn ratio of transformer) is the ratio of unbalanced impedance to balanced impedance, specified by $1 : n^2$. The differential impedance, measured between the balanced signal lines, is twice that between one signal and the ground. High impedance ratio BALUNs at low frequencies are more easily realized by transformers than at high frequencies by transmission lines.

DOI 10.1007/978-981-4560-75-7_133-1

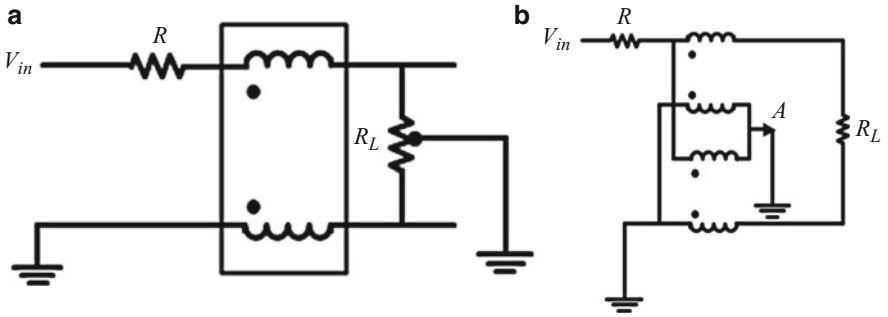


Fig. 53 Guanella BALUNS. (**a**) Guanella 1:1 BALUN; (**b**) Guanella 1:4 BALUN

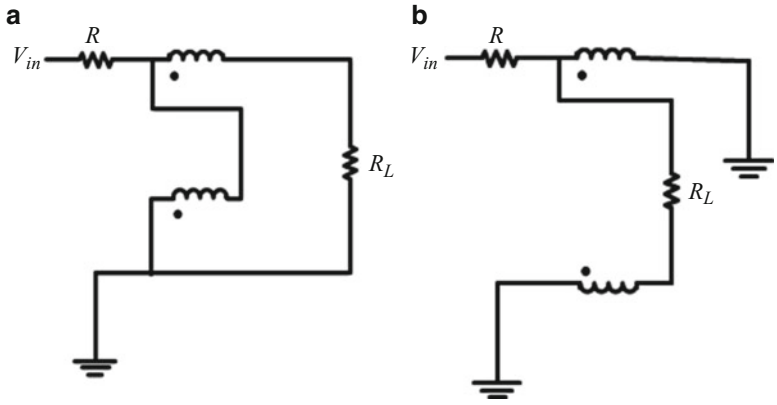


Fig. 54 Ruthroff transformers. (**a**) Ruthroff 1:4 UNUN; (**b**) Ruthroff 1:4 BALUN

A BALUN with no isolation between the two output ports (such as reactive splitter) will have different return losses for common and differential mode signals at the balanced ports. The common mode will be completely reflected, and the differential mode will pass through.

The balanced ports isolation is the insertion loss from one balanced port to the other. The even mode is reflected instead of being absorbed inside BALUN in a resistor.

DC isolation is an indication of how much DC connection exists between the unbalanced and balanced ports. Ground isolation is an indication of the degree of connection of unbalanced ground and balanced signals and ground.

BALUNs are used to interface an unbalanced signal to a balanced transmission line for long distance communication. Differential signals on balanced transmission lines are not affected by noise and interference. BALUNs are used in the feed line of differential antennas for the improvement of common mode rejection ratio (CMRR).

Examples of unbalanced lines are coaxial line and microstrip line. Examples of balanced lines are shielded broadside-coupled stripline and two-wire transmission line (see Fig. 52). Examples of lumped element BALUNs are tuned transformer, center-tap transformer, autotransformer (see Fig. 51), and RF choke, which are usually used at low frequencies. BALUNs made of transformers operate on the basis of magnetic coupling, such as BALUN transformer in Fig. 51a. Various forms of transformer windings could be devised to produce both magnetic and electric couplings, such as Guanella 1:1 BALUN shown in Fig. 53a, 1:4 Guanella BALUN shown in Fig. 53b (Guanella 1944), and Ruthroff BALUN (Ruthroff 1959) shown in Fig. 54. A Guanella 1:1 BALUN configured as a common mode choke is shown in Fig. 55.

DOI 10.1007/978-981-4560-75-7_133-1

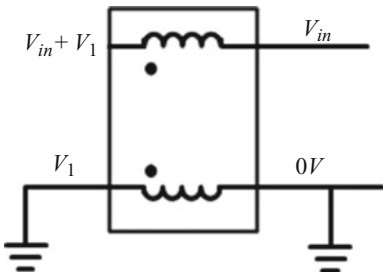


Fig. 55 Guanella 1:1 BALUN configured as a common mode choke

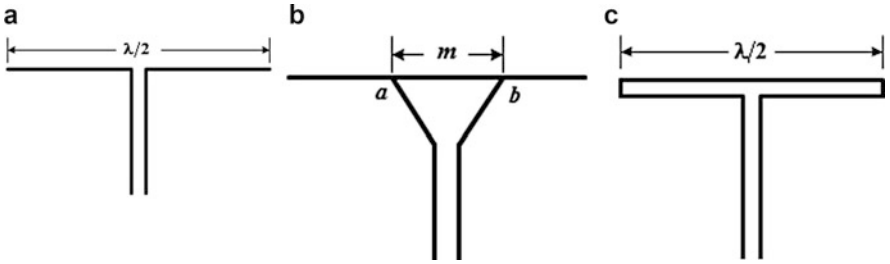


Fig. 56 Three methods of feeding a dipole antenna. (**a**) Two-wire feed line; (**b**) tapered feed (Gamma feed); (**c**) folded dipole

The differential mode (DM) current on a two-wire line (even if one acts as a shield) flows in opposite directions on each line at each end. It is the balanced current component. It neither radiates nor receives any signal. However, the parallel or common mode (CM) current has components flowing in the same direction on the two lines (or feed line of an antenna) which exhibits impedances to the ground. It is the unbalanced current component. It causes all the radiation from the line (or antenna).

The choke BALUN inserts a large common mode impedance in series with the feed line, which decreases the common mode current without disturbing the differential mode current. The combination of proper grounding and choke BALUN will reduce the unwanted common mode currents (see Fig. 55).

Direct feeding of a resonant dipole antenna (having input impedance of about 73 Ω) to a two-wire transmission line (having a characteristic impedance of about 300 Ω) is not appropriate, as shown in Fig. 56a. Two arrangements may be devised for the feed of dipole antennas. One method is to use a taper line (Gamma match) as an impedance transformer at the dipole feed and the other is to use a folded dipole shown in Fig. 56b, c, respectively.

Many antennas such as dipoles are balanced loads, that is, their parts have the same impedance to the ground for equal currents passing through them. They should ideally be fed by a balanced line, such as the two-wire line. However, at high frequencies, an unbalanced line such as a coaxial line is commonly used as a feed line to eliminate any spurious radiation. Generally, the balanced parts of a symmetric transmission line or antenna system carry equal and opposite currents. On the other hand, the current unbalance is the main source of losses of the two-wire transmission line into ground or radiation from it and also pattern distortion of antenna systems. The unbalance of conductor potentials with respect to ground generates the unbalance of current flow.

The capacitance existing between the coax outer conductor and nearby objects or direct connection to ground actually serves to put its conductor surface at ground potential. A two-wire line should be balanced with respect to the ground, so that the two wires carry equal and opposite currents. Direct connection of a coaxial unbalanced line to a balanced two-wire line generates currents on the outside conducting surface

DOI 10.1007/978-981-4560-75-7_133-1

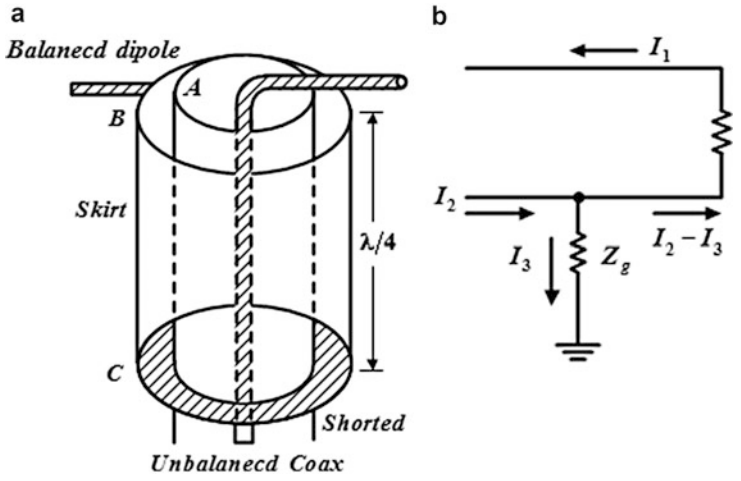


Fig. 57 Direct feeding of a dipole antenna by a coaxial line (Jordan and Balmain 1968). (**a**) Feed-line configuration; (**b**) equivalent circuit

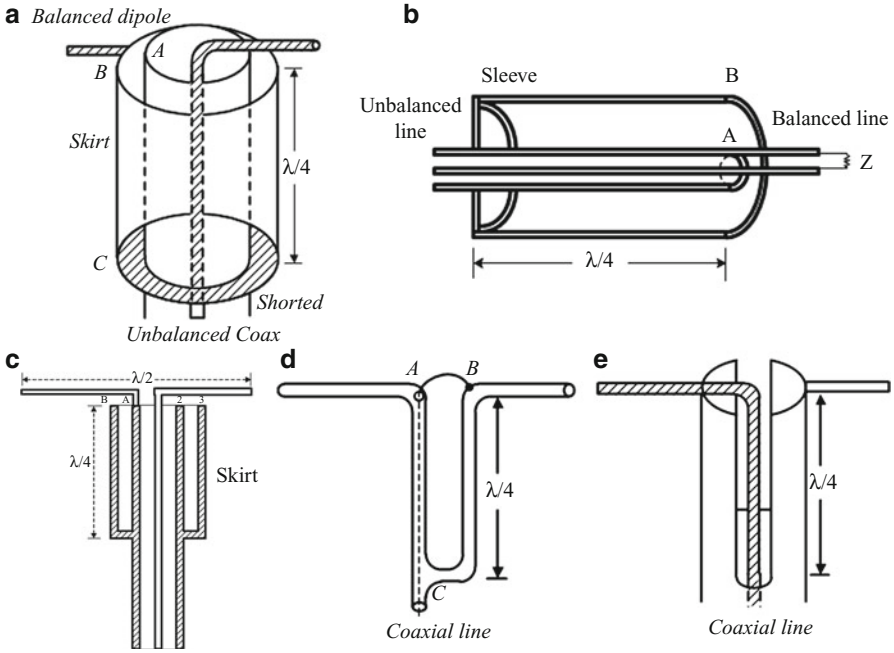


Fig. 58 Bazooka BALUN for the connection of an unbalanced coaxial feed line to a balanced load (Jordan and Balmain 1968; Reich et al. 1953). (**a**) Three-dimensional view for a dipole; (**b**) three-dimensional view for a load Z_L; (**c**) cross-sectional view; (**d**, **e**) a coaxial line feeding a dipole through a quarter-wave line

of the coax producing radiation and losses. Particularly, the development of BALUNs serves to reduce the radiations from the currents on the outer surface of shields.

Figure 57 shows the direct connection of a coaxial feed line (unbalanced) to a dipole antenna (balanced) and its equivalent circuit showing currents flowing in its various parts. Observe that the currents on its two arms are not equal making it unbalanced as the coaxial feed. A dipole is a balanced load, where its two arms have the same impedance to ground provided equal currents flow in its arms. If one arm of a dipole antenna is connected to ground, it would not work properly. It needs a BALUN at its input. Z_g is the effective impedance to ground, which can be made very high by the addition of a sleeve with length $\lambda_g/4$,

DOI 10.1007/978-981-4560-75-7_133-1

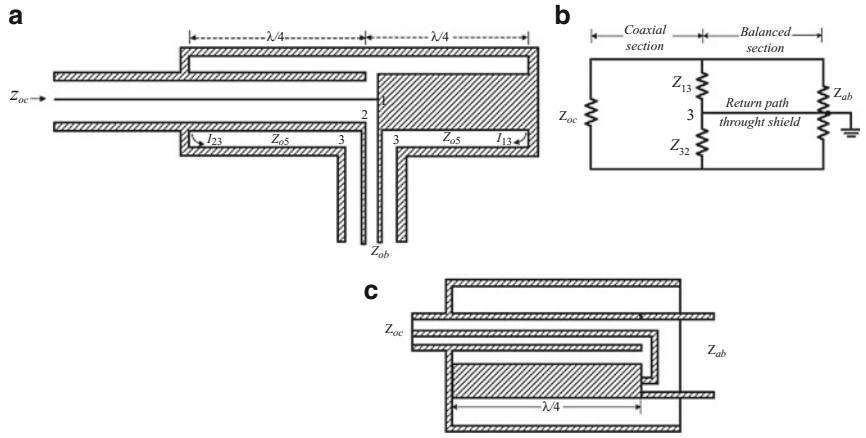


Fig. 59 A wideband BALUN (Reich et al. 1953). (**a**) Cross section of the coaxial configuration, (**b**) equivalent circuit of BALUN, (**c**) folded version of BALUN (Reich et al.)

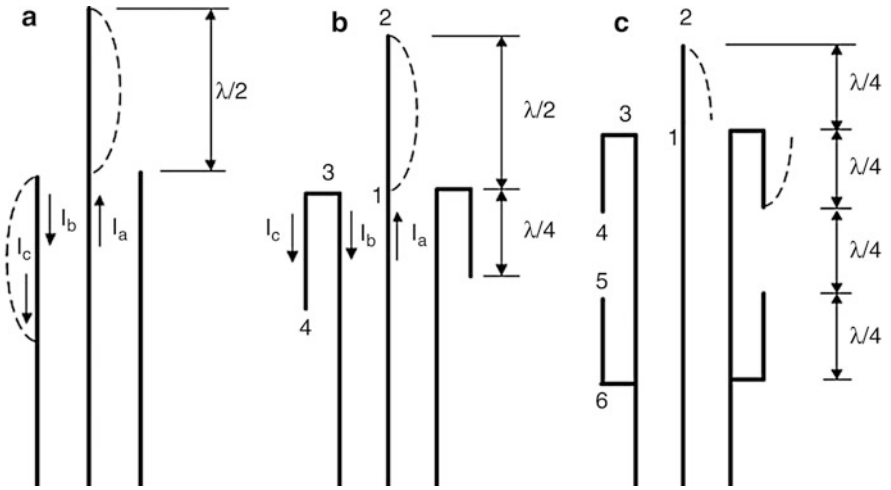


Fig. 60 Coaxial feeding of a monopole of length $\lambda/4$ (Reich et al. 1953). (**a**) Standing waves appearing on the outside shield of coax; (**b**) placement of detuning stub reduces current on the outside surface of coax; (**c**) balanced dipole (1-2 and 3-4) with a stub (5-6)

shown in Fig. 58a. The shorted end transfers to an open circuit at the dipole A. The BALUN provides a high impedance between points A and B relative to the shorted point C through the quarter-wave line. The cross sections of such a BALUN are drawn in Fig. 58b, c; coaxial lines feeding a dipole through a quarter-wave line are shown in Fig. 58e, d.

A wideband BALUN and its equivalent circuit are shown in Fig. 59a, b. Its folded version is shown in Fig. 59c. The performance of these BALUNs is described in the literature.

The bazooka BALUN may be configured in different forms, such as collinear BALUN and folded BALUN. The connection of an unbalanced coaxial feed line to a monopole is shown in Fig. 60. Observe that some current flows on the outer surface of the coaxial line. In order to suppress this spurious current, detuning stubs are used in its structure as shown in Fig. 60b, c.

Another BALUN and its equivalent circuit are drawn in Fig. 61. The characteristic impedance of each leg of coaxial line of length s is equal to Z_0. Therefore, their short-circuit impedances Z_g which shunt each side of the balanced terminal are $Z_g = jZ_0 \tan(\beta s)$.

DOI 10.1007/978-981-4560-75-7_133-1

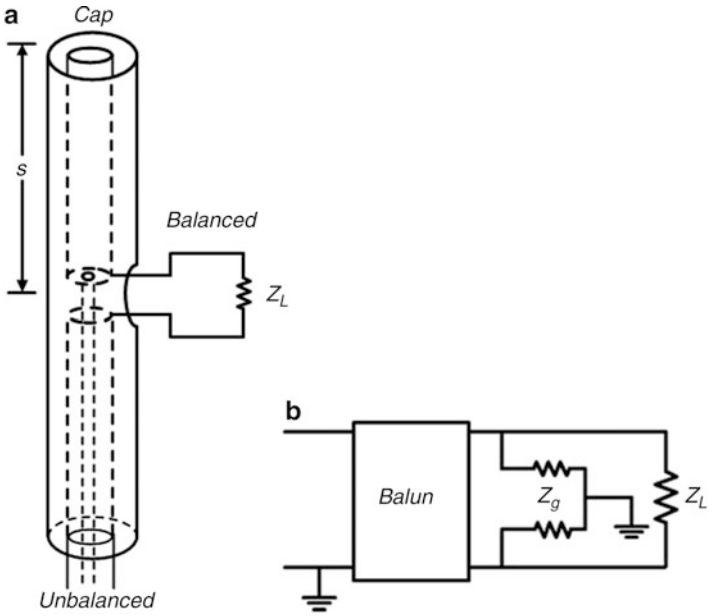


Fig. 61 A sleeve BALUN (Jordan and Balmain 1968)

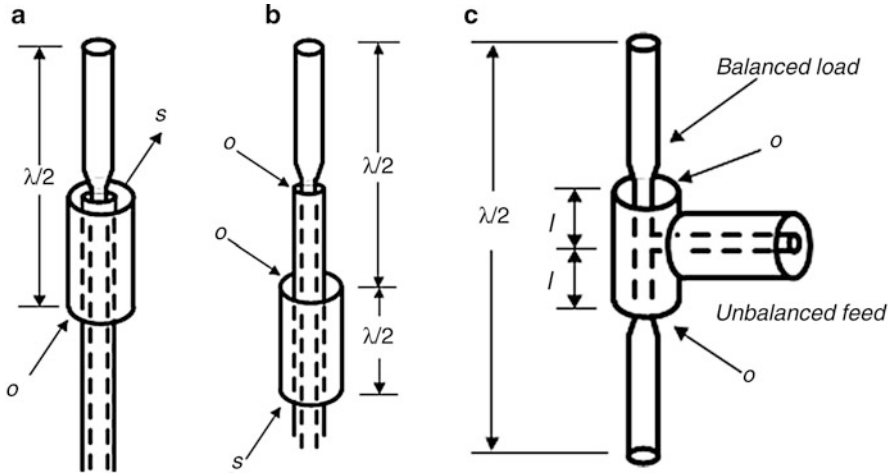


Fig. 62 Several BALUN configurations for the coaxial feed of antenna (*o* open, *s* short) (Terman 1955). (**a**) Flagpole monopole antenna; (**b**) flagpole monopole antenna with a sleeve; (**c**) sleeve dipole antenna

A half-wave dipole antenna may be fed from a coaxial line by devices called flagpole and sleeve antennas, as shown in Fig. 62a–c. Such structures provide both impedance matching and balancing action. The open circuit in the flagpole antenna shown by "o" serves to isolate the antenna from the lower parts. The shorted parts are shown by "s."

A BALUN providing a high impedance in a wide frequency band is a coiled coax as shown in Fig. 63, which may be wound in a coil on a ferrite core to increase the frequency band.

At higher frequencies, where resonant lines may operate, structures called BALUN (balancing unit, balanced to unbalanced, or bazooka) may be used. Consider Fig. 64a, b, where the unbalanced coaxial system at the left end is connected to the balanced system at the right end. Apply a voltage between conductors 1 and 2 at the balanced line. It will be transferred to the left end at the coaxial line 4 without any

DOI 10.1007/978-981-4560-75-7_133-1

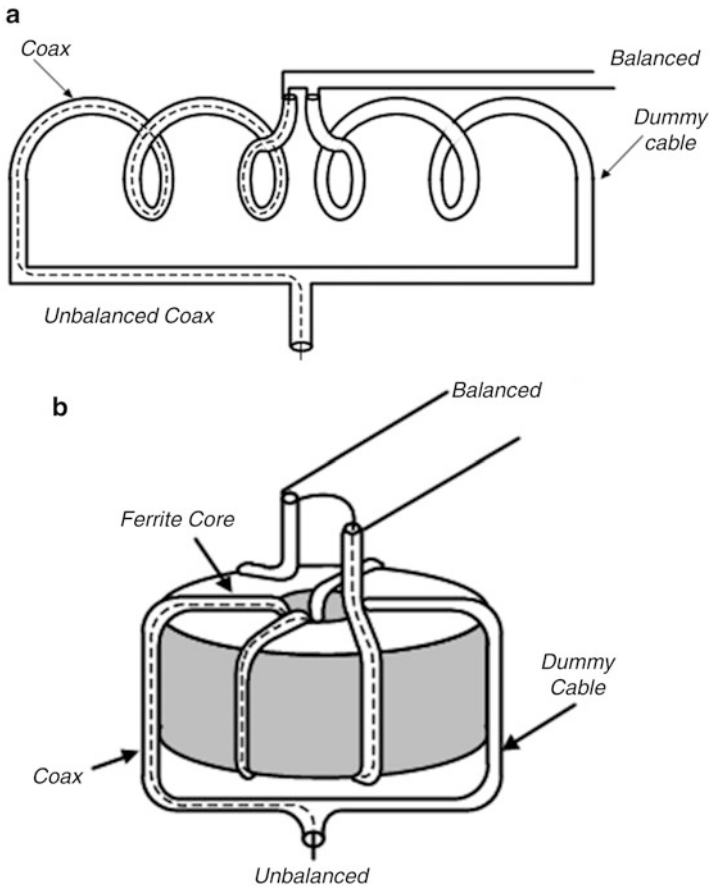


Fig. 63 Coiled coax wideband BALUN (Weeks 1968). (**a**) Coils have a high impedance to ground; (**b**) coax wound on a ferrite toroid

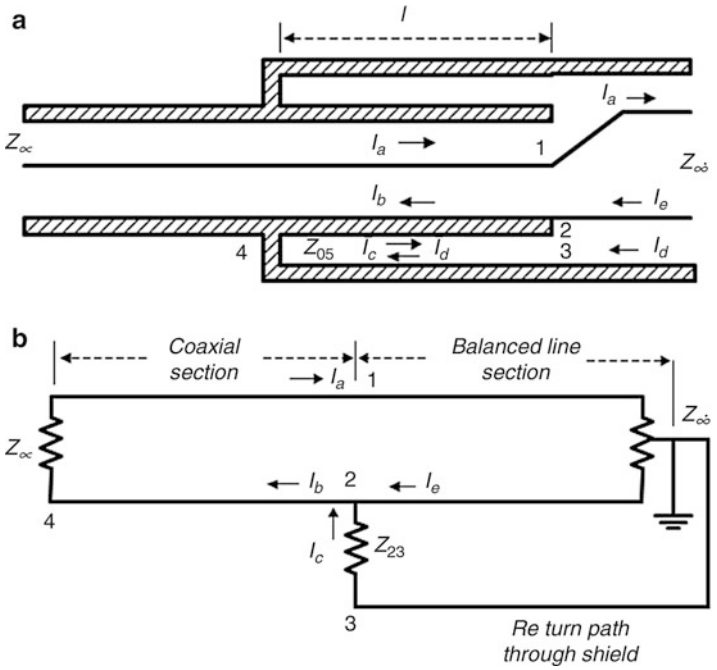


Fig. 64 A narrowband coaxial BALUN (Reich et al. 1953). (**a**) Cross-sectional view; (**b**) equivalent circuit

DOI 10.1007/978-981-4560-75-7_133-1

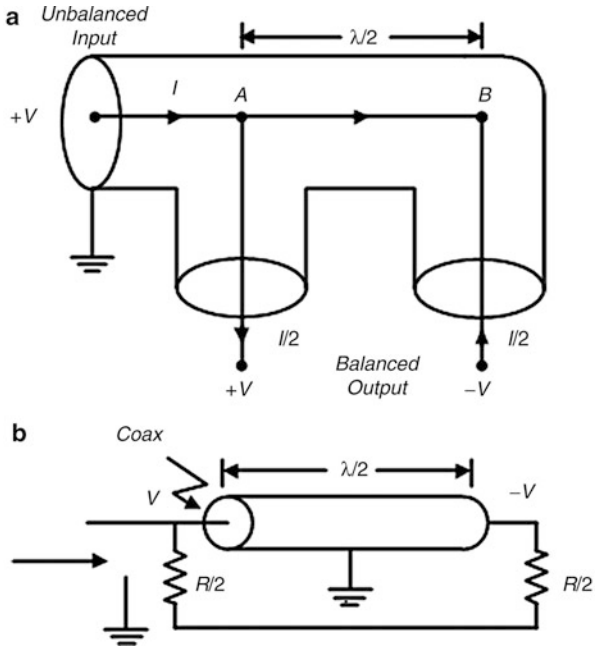


Fig. 65 1:4 BALUNs. (**a**) A 1:4 coaxial BALUN; (**b**) a narrowband 1:4 transmission line BALUN

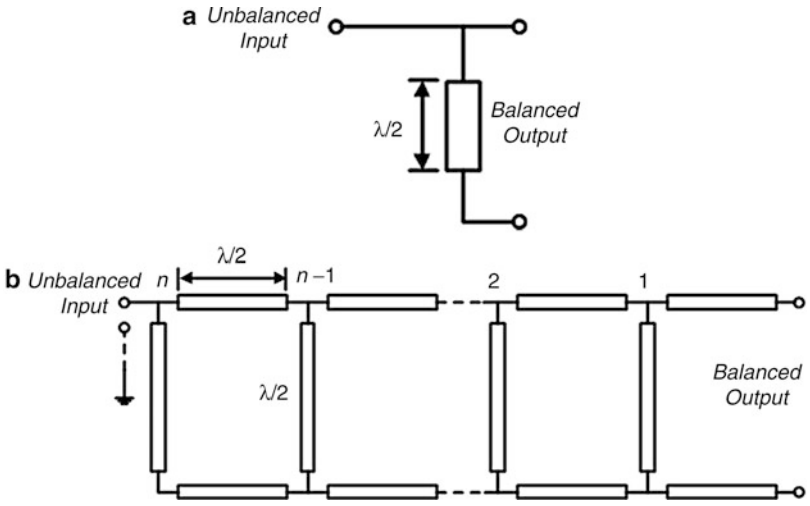


Fig. 66 Planar BALUNs (Mongia et al. 1999). (**a**) A single half-wave line separation splitter; (**b**) multi-section wideband half-wave BALUN (Mongia et al.)

change, because of the direct connection. Now, observe that there will be an open circuit between conductors 2 and 3, because the sleeve is shorted to the outer conductor of coaxial line on the left (unbalanced end) and the sleeve is exactly a quarter-wavelength long. Therefore, no unbalanced effect is generated at the right end. In other words, sleeve 3 is actually an extension of 4 and remains at ground potential. Consequently, conductor 2 may assume any voltage required by the balanced line at the right end. At the coaxial end, conductor 4 is connected to the ground sleeve, but conductors 1 and 2 at the right end are balanced with respect to conductor 3. The current is continuous from the balanced to unbalanced lines.

DOI 10.1007/978-981-4560-75-7_133-1

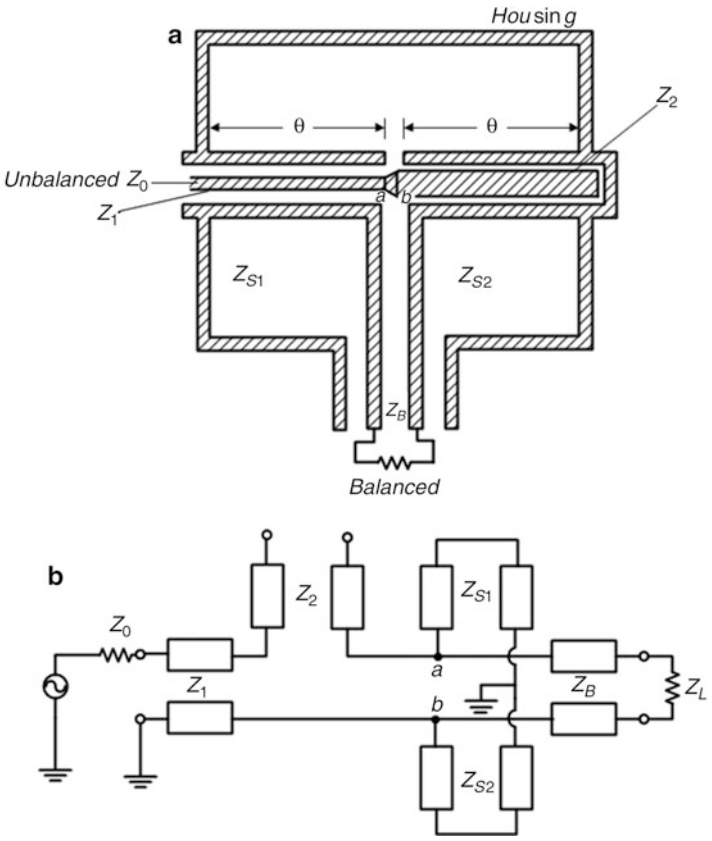


Fig. 67 A compensated Marchand BALUN (Mongia et al. 1999). (**a**) Coaxial version; (**b**) equivalent circuit

Another coaxial BALUN is shown in Fig. 65a. The line section of length $\lambda/2$ produces a phase lag of 180° for both current and voltage and also causes the impedance to repeat itself at points A and B. Therefore, the current splits in half at point A. Consequently, the unbalanced line on the left transforms to the balanced line at the other terminal. Since the voltage on the balanced line is 2 V and its current is I/2, then the impedance is 4 V/I, which is 4 times that at the unbalanced line. Therefore, the BALUN is also an impedance transformer.

A narrowband 1:4 transmission line BALUN is shown in Fig. 65b. A resistor of value R is connected between its input and output terminals. Let it be split in half into two resistors of R/2, where the midpoint at their connection is effectively at ground potential. The balanced port is between terminals at potentials −V and +V across the resistor R. The unbalanced port is between point at potential +V and ground, with a resistor R/2 in parallel with another resistor R/2 transferred by a half-wave line. The effective resistance at unbalanced line is then R/4. Consequently, the BALUN has achieved an impedance transformation of 1:4. The transmission may assume various geometrical shapes to fit the circuitry and possible miniaturization in the microstrip technology.

Two planar BALUNs, namely, a single half-wave line separation splitter and a multi-section half-wave BALUN, are shown in Fig. 66a, b.

A compensated coaxial prototype of Marchand BALUN and its equivalent circuit are shown in Fig. 67a, b.

Coupled microstrip lines may realize a BALUN. A stripline version of the basic Marchand coaxial BALUN is drawn in Fig. 68. It consists of an open-ended half-wave line which has antiphase signals at its two ends. It is coupled to two quarter-wave lines oriented in series. The unbalanced input is applied between one end of half-wave line and the grounded end of one of the quarter-wave lines at port 1. The far

DOI 10.1007/978-981-4560-75-7_133-1

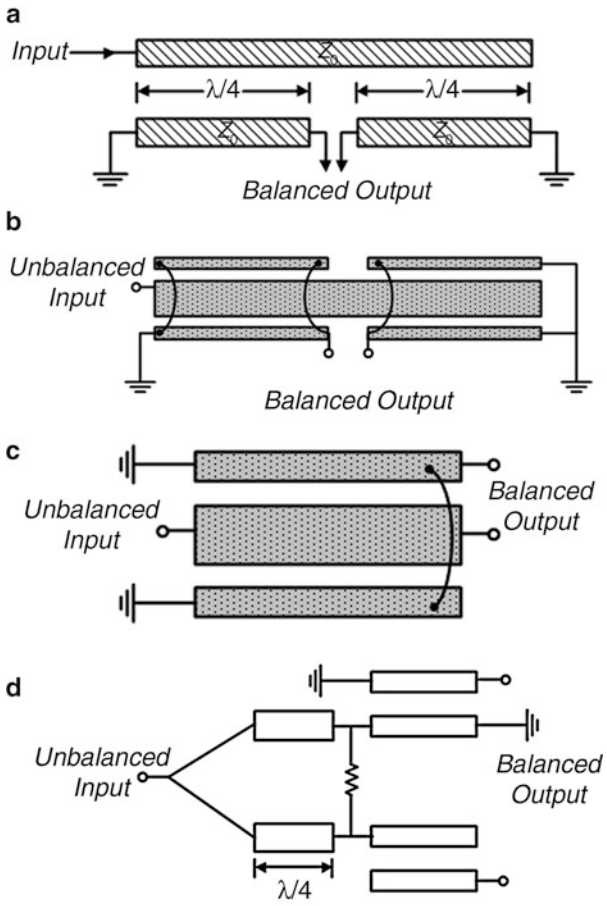


Fig. 68 Various versions of Marchand BALUN (Mongia et al. 1999). (**a**) Basic version; (**b**) coupled-line version 1 (each line section is of length $\lambda/4$); (**c**) coupled-line version 2; (**d**) Marchand BALUN

end of the other quarter-wave line is also grounded. The balanced output is obtained from the adjacent ends of the quarter-wave lines as differential signal at port 2.

A method of analysis for BALUNs is introduced by Marchand, which is also explained in Reich et al. (1953). An analysis of BALUNs is described in Weeks (1968). The interested reader is referred to the extensive literature on planar BALUNs.

Conclusion

The various topics covered in this chapter are applicable to RF and microwave circuits, including the subject matter of antenna matching.

However, there are several other techniques for the realization of matching circuits for antennas, such as the active circuits of non-Foster methods. For the important task of antenna miniaturization, there have always been attempts to incorporate the impedance matching actions into the physical structure of antennas, so that its radiating body and matching parts are all incorporated in a whole unit to operate together.

There are theoretical limitations on impedance matching which should be considered in their physical realization (Fano 1950).

Extensive discussions on antenna matching circuits are provided in Yarman (2008, 2010). Some applications of the method of least squares to electromagnetic problems are given in Oraizi (2006). These and other topics are left over for further study in the literature.

Acknowledgments

The author would like to thank Mr. Mohammad Amin Chaychi-zadeh for his assistance in writing the computer programs and running the numerical examples.

Cross-References

- Foster's Reactance Theorem and Non-Foster Impedance Matching in Antenna Systems
- Microstrip Patch Antennas
- Physical Bounds of Antennas
- Transmission Lines
- Wheeler-Fano Impedance Matching Principles

References

Bex H (1975) New broadband BALUN. Electron Lett 11(2):47–48
Collin RE (2000) Foundations for microwave engineering, 2nd edn. McGraw-Hill, New York (Chap 5)
Edwards TC, Steer MB (2000) Foundations of interconnect and microstrip design, 3rd edn. Wiley, New York
Fano RM (1950) Theoretical limitation on the broad-band matching of arbitrary impedances. J Franklin Inst 249(1):57–83, and pp 139–154
Fatholbab WM, Steer MB (2005) New class of miniaturized planar Marchand BALUN. IEEE Trans Microwave Theory Tech 53(4):1211–1220
Giguere JC (1972) Approximation of a nonuniform transmission line by a cascade of uniform lines. Electron Lett 7(18):511
Gonzalez G (1997) Microwave transistor amplifiers, analysis and design, 2nd edn. Prentice Hall, New Jersy (Chaps 2 and 3)
Guanella G (1944) Novel matching systems for high frequencies. Brown-Boveri Rev 31:327–329
Jordan EC, Balmain KG (1968) Electromagnetic waves and radiating systems. Prentice Hall, New Jersy
Klopfenstein RW (1956) A transmission line taper of improved design. Proc IRE 44(1):31–35
Lin C-H, Wu C-H, Zhou GT, Ma T-G (2013) General compensation method for a Marchand BALUN. IEEE Trans Microwave Theory Tech 61(8):2821–2830
Lu J-C, Lin CC, Chang C-Y (2011) Exact synthesis and implementation of new high-order wideband Marchand Balun. IEEE Trans Microwave Theory Tech TMTT 59(1):80–86
Ludwig R, Bretchko P (2000) RF circuit design, theory and applications. Prentice Hall, New Jersy (Chap 8)
Marchand N (1994) Transmission line conversion transformer. Electronics 17:142–145
Matthaei G, Young L, Jones EMT (1964) Microwave filters, impedance matching and coupling structures. McGraw-Hill, New York
Misra DK (2001) Radio-frequency and microwave communication circuits. Wiley, New York (Chap 5)

Mongia R, Bahl I, Bhartia P (1999) RF and microwave coupled-line circuits. Artech House (Chap 11)
Oltman G (1966) The compensated BALUN. IEEE Trans Microwave Theory Tech TMTT-14 14(3):112–119
Oraizi H (1996) Design of impedance transformers by the method of least squares. IEEE Trans Microwave Theory Tech 44(3):389–399
Oraizi H (2001) Optimum design of stepline transformers of arbitrary length including step discontinuities. Iran J Sci Technol Trans B: Technol 25(1):14
Oraizi H (2006) Application of the method of least squares to electromagnetic engineering problems. Antennas Propagat Mag IEEE 48(1):50–74
Oraizi H, Esfahlan MS (2010) Optimum design of lumped filters incorporating impedance matching by the method of least squares. Progress Electromagn Res 100:83–103
Oraizi H, Seyyed-Esfahlan M (2011) Impedance matching and spurious-response suppression in stepped-impedance low pass filters. Microwave Optical Technol Lett 53(9):2081–2086
Pozar DM (2011) Microwave engineering, 4th edn. Wiley, New York (Chaps 2, 5, and 11)
Reich HJ, Ordung PF, Krauss HL, Skalnik JG (1953) Microwave theory and technique. D. Van Nostrand, New York (Chap 4)
Rizzi PA (1988) Microwave engineering, passive devices. Prentice Hall, New Jersy (Chap 4)
Ruthroff CL (1959) Some broadband transformers. Proc IRE 47:1337–1342
Sevick J (1990) Transmission line transformers, 2nd edn. American Radio Relay League, Newington
Sevick J (2004) A simplified analysis of the broadband transmission line transformer. High Freq Electron 3(2):48–53
Terman FE (1955) Electronic and radio engineering, 4th edn. McGraw-Hill, New York
Trifunovic V, Jakonovic B (1994) Review of printed Marchand and double Y Baluns: characteristics and application. IEEE Trans Microwave Theory Tech 42(8):1454–1462
Valkenburg ME (1991) Network analysis, 3rd edn. Prentice Hall, New Jersy (Chaps 1, 14)
Walker S (1968) Broadband stripline BALUN using quadrature couplers. IEEE Trans Microwave Theory Tech TMTT-16:132–133
Weeks WL (1968) Antenna engineering. McGraw-Hill, New York
Xu Z, MacEachern L (2009) Optimum design of wideband compensated and uncompensated Marchand Baluns with step transformers. IEEE Trans Microwave Theory Tech 57(8):2069–2071
Yarman BS (2008) Design of ultra wideband antenna matching networks, via simplified real frequency techniques. Springer, Berlin
Yarman BS (2010) Design of ultra wideband power transfer networks. Wiley, New York

Handbook of Antenna Technologies
DOI 10.1007/978-981-4560-75-7_134-1

Advanced Antenna Fabrication Processes (MEMS/LTCC/LCP/Printing)

Bijan K. Tehrani[a]*, Jo Bito[a], Jimmy G. Hester[a], Wenjing Su[a], Ryan A. Bahr[a], Benjamin S. Cook[b] and Manos M. Tentzeris[a]
[a]The School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, CA, USA
[b]Georgia Tech. ATHENA Research Lab, Texas Instruments Kilby Labs, Dallas, TX, USA

Abstract

The future of wireless technology is stressing the development of ubiquitous, low-cost, highly specialized wireless devices, where both the design and fabrication of wireless antennas have an essential role. This chapter presents four modern fabrication processes used for the development of advanced wireless antenna structures, both present and emerging: MEMS, LTCC, LCP, and ink-jet/three-dimensional printing. Each fabrication technique is investigated historically and analytically, discussing early achievements, process development, and several state-of-the-art demonstrations. The strengths and challenges of these additive and subtractive methods are highlighted, outlining the cost, processing time, scalability, and dimensional resolution of each method. Throughout the discussion of these electronic fabrication techniques, an emphasis is placed on the realization of highly efficient, robust antenna structures for a variety of applications, including conformal sensor networks, radar systems, low-cost RFID, and on-chip/on-package integration.

Keywords

Ceramic substrate; Flexible substrate; Wearable electronics; Millimeter wave; RFID; Additive manufacturing; Multilayer antenna; On-chip antenna; Tape casting; Lithography; Ink-jet printing; 3-D printing

Introduction

Throughout the past century, wireless technology has developed into a mature staple of modern technology. Beginning in the late 1880s with the discovery by Heinrich Hertz that electromagnetic waves could be used as a means of "telegraphing without wires" (Story 1904), the study of wireless systems has led to a ubiquitous market of electronic devices that are able to broadcast and receive complex data in an efficient, mobile fashion.

Wireless systems can be broken down into two basic components: the radio and the antenna. The wireless radio is responsible for formatting any desired data for wireless communication, typically with the use of oscillators, modulation schemes, and amplification. Second, but just as important, is the wireless antenna. The responsibility of the antenna includes the task of accepting the electromagnetic signal from the radio and radiating it into an environment. Several general factors play a part in how efficiently an antenna radiate in a practical wireless setting, such as impedance, gain, beamwidth, and bandwidth. These factors are typically taken into account during the design of the antenna in order to ensure proper device performance.

*Email: btehrani3@gatech.edu

DOI 10.1007/978-981-4560-75-7_134-1

Though antenna design is integral to the success of wireless systems, antenna fabrication plays an equally significant role. In most cases of antenna development, the two facets go hand in hand, where design criteria face fabrication limits, and fabrication schemes are geared toward practical designs.

Additive and Subtractive Methods

The techniques currently used for antenna fabrication can be separated into two categories: subtractive methods and additive methods. Typically, subtractive methods include the patterning of material through methods that require removing materials from a host, while additive methods include a direct deposition of material onto a host. The general process flows of these two categorical methods are outlined below:

- Subtractive Methods
  - Substrate and material preparation
  - Bulk material deposition
  - Masked patterning
  - Sacrificial etching
- Additive Methods
  - Substrate and material preparation
  - Selective material deposition

In this chapter, a review of several popular methods of electronic fabrication is presented with a specific interest in the realization of highly efficient, robust antenna structures. Processes, challenges, and state-of-the-art demonstrations of these fabrication methods are discussed for a variety of wireless applications, including millimeter-wave, on-chip and on-package integration, wideband flexible RFID systems, and low-cost multilayer radar systems.

MEMS

"Microelectromechanical systems" (MEMS) is a general term for highly miniaturized systems that are created through the integration of many small components, including electrical, thermal, magnetic, optical, and fluidic, with or without moving parts. MEMS devices emerged in the late 1960s following with the development of lithography-based microfabrication techniques including CMOS technologies for highly integrated computer circuit components (Nathanson et al. 1967; Howe and Muller 1982; Senturia 2001). One of the most outstanding features of these micromachining techniques among many other competing fabrication approaches is their ultrahigh precision and very fine resolution capabilities for many different kinds of materials. On the other hand, these microfabrication technologies usually require a clean room environment and are very expensive, especially if the feature size of the structures is very small. Therefore, these technologies have been used mainly on highly integrated circuit core dies including CPUs, memory units, and sensors.

However, in recent years the desire for system miniaturization is getting stronger, especially in the area of mobile electronic devices. In order to realize higher performance and greater functionality with smaller feature sizes, system integration in a larger scale, which can be expressed as system on chip (SOC), stacked ICs and packaging (SIP), and system on packaging (SOP), is becoming more and more important (Tummala 2008). Also, the demand for high frequency antennas following with the exponential increase of data in wireless communication and expansion of millimeter wave use in sensing applications makes MEMS fabrication technologies a reasonable choice for antenna fabrication, a result of the shrinkage of

DOI 10.1007/978-981-4560-75-7_134-1

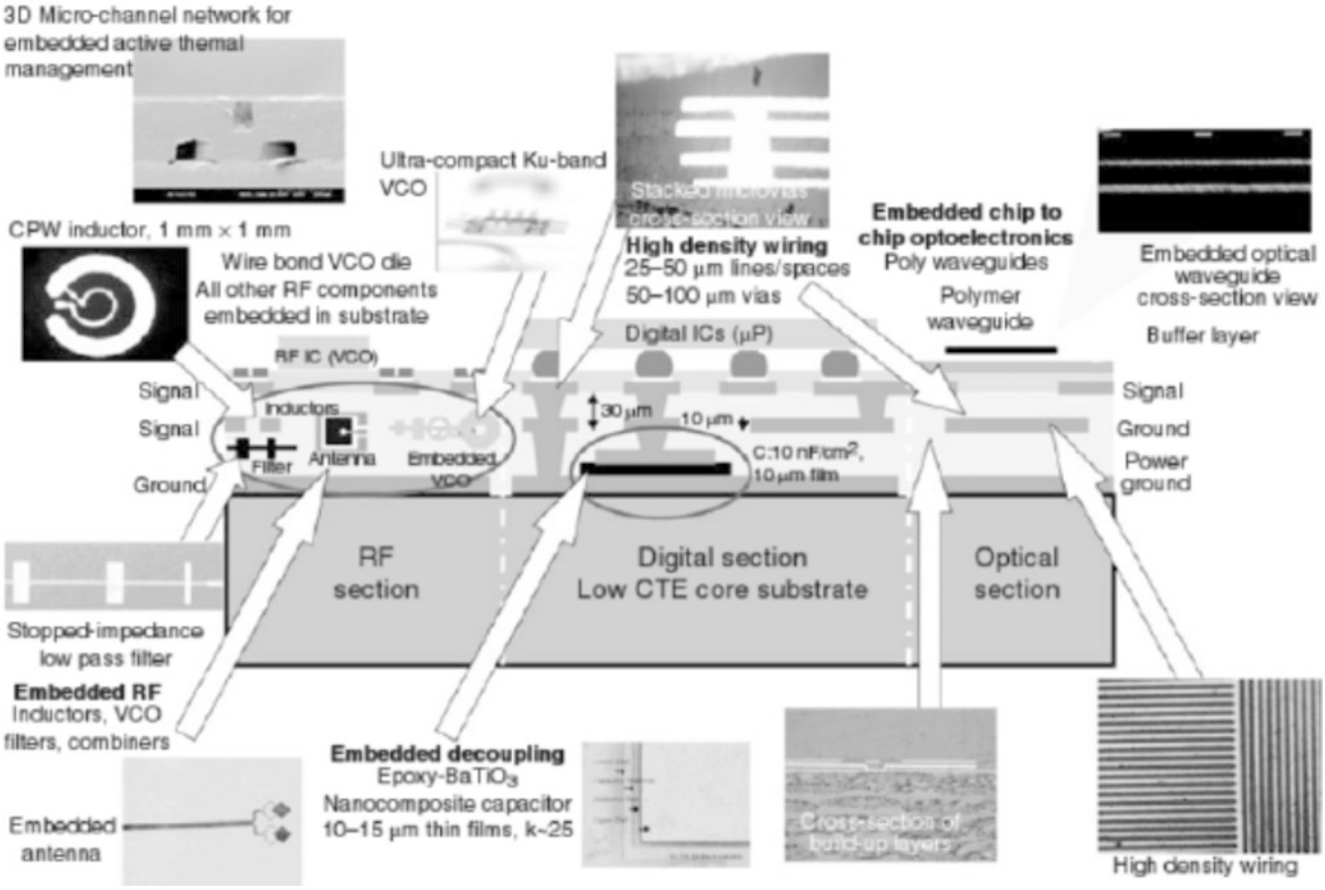


Fig. 1 A cross-sectional view of a typical intelligent network communicator

feature size along with increase in frequency. RF MEMS integration can be conceptualized as intelligent network communicator (INC) as is depicted in Fig. 1.

Microstructures are fabricated with a combination of additive processes, such as chemical vapor deposition (CVD), spin coating, and electroless plating, in conjunction with subtractive processes, such as wet etching and dry etching, depending on the material and thickness of the structure (Madou 2002; Bustillo et al. 1998; Kovacs et al. 1998). With a proper sequential process, it is possible to create 3-D structures applicable for antennas through these well-developed standard micromachining techniques. These methods have the capability to be seamlessly integrated into other CMOS chips and circuitry in the form of on-silicon antennas and on-chip antennas. Generally, micromachining can be categorized into surface micromachining and bulk micromachining with regard to the thickness of the structure.

Surface Micromachining

Surface micromachining is used to create relatively thin-film layers on top of a substrate. The microstructure fabricated is typically released by dry/wet isotropic etching. These thin-film layers could be metal layers for the antenna patterns, dielectric or insulator layers as a substrate or insulation, and polymer layers as sacrificial layers for the metal patterning. Since antennas usually do not require movable parts in them, most of the fabrication can be done only by utilizing surface micromachining. There are several examples of antennas for millimeter-wave frequencies fabricated through CMOS fabrication processes in literature. The top view, side view, and S-parameter frequency response for a 100 GHz quasi-Yagi antenna fabricated with surface micromachining are depicted in Fig. 2 (Sun and Zhang 2007).

DOI 10.1007/978-981-4560-75-7_134-1

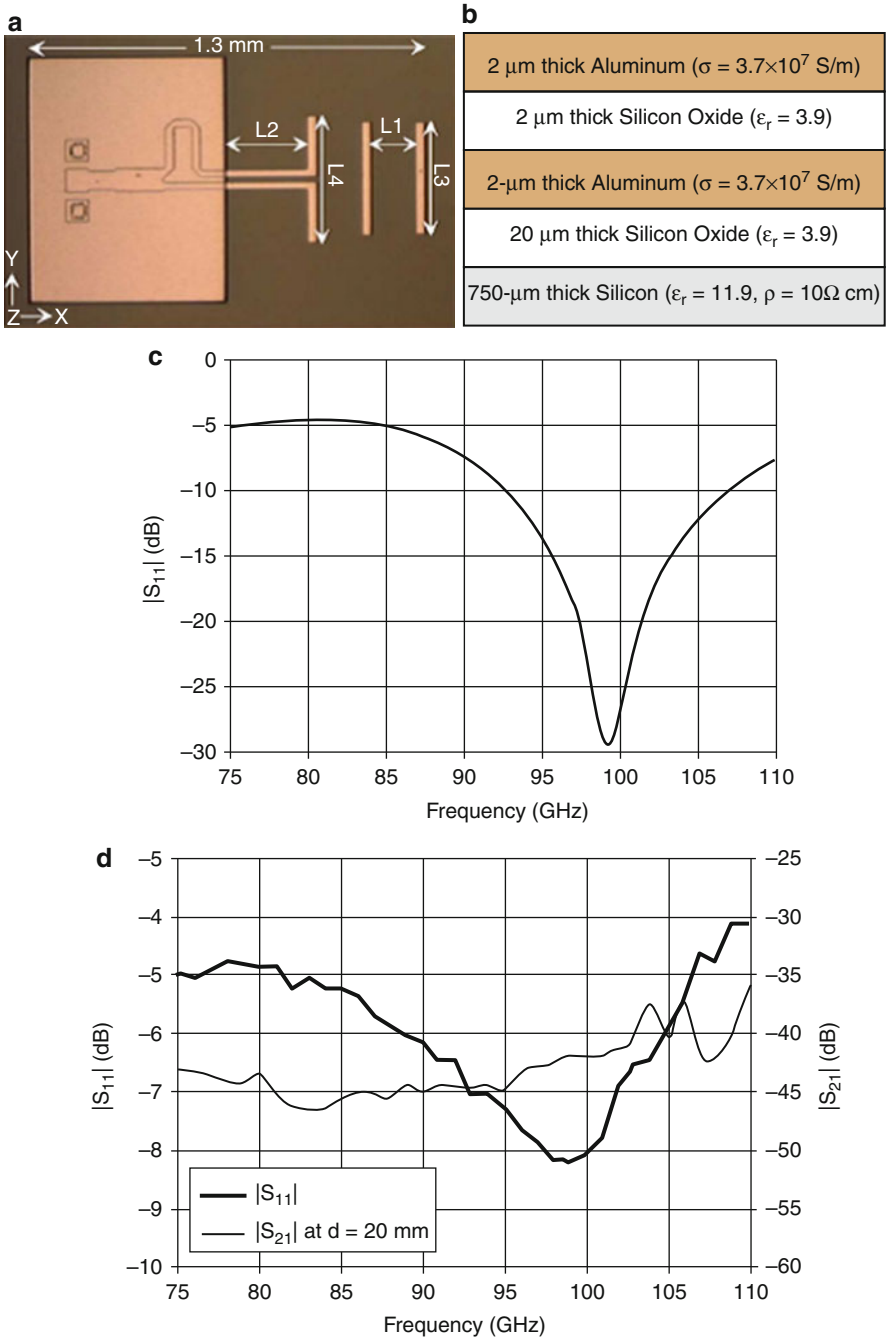


Fig. 2 Quasi-Yagi antenna (**a**) Top view photograph. (**b**) Cross-sectional view illustration. (**c**) Simulated $|S_{11}|$ as a function of frequency. (**d**) Measured $|S_{11}|$ and $|S_{21}|$ as a function of frequency

Bulk Micromachining

Bulk micromachining is utilized to process relatively thick substrates. Usually, in MEMS devices, the substrate is silicon and can be fabricated by utilizing dry/wet and anisotropic/isotropic etching. These processes can be applied for via fabrication in MEMS antenna structures as is depicted in Fig. 2a. In addition to the fabrication of static structures, it is possible to create movable structures by applying bulk micromachining to release a part of surface-micromachined layers. As an application of MEMS structures into antenna topologies, numerous reconfigurable tunable antennas with RF MEMS switches have been

DOI 10.1007/978-981-4560-75-7_134-1

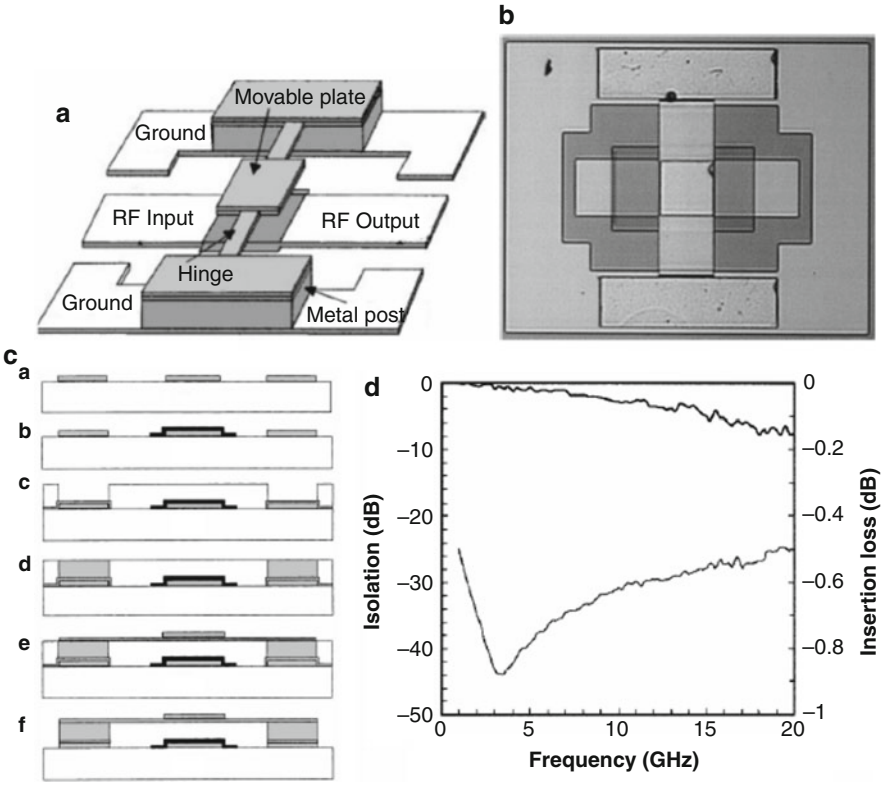


Fig. 3 (**a**) Schematic drawing of RF MEMS capacitive switches. (**b**) Photomicrograph of a fabricated RF MEMS capacitive switch with the hinge structure. (**c**) Fabrication steps of RF MEMS capacitive switches. (**d**) Insertion loss (*upper curve*) and isolation (*lower curve*) of the fabricated RF MEMS capacitive switch

proposed (Weedon et al. 2001; Park et al. 2001). Since both antenna and RF MEMS structures are fabricated through the same CMOS processes, it is possible to realize a highly integrated, seamless antenna configuration using this technology. The design, top view, fabrication processes, isolation, and insertion loss of such an antenna prototype are depicted in Fig. 3.

LTCC

LTCC Technology

The first ceramic technology is believed to have been introduced in the late 1950s at RCA Corporation (Stetson 1965). At the time, the ceramics used required high processing temperatures (1600 °C) for what we now refer to as high-temperature co-fired ceramics (HTCC). These temperatures made it impossible to use low-cost and highly conductive metals (Ag, Cu, etc.) that would melt in such conditions. Furthermore, in the case where LSI components would be connected through the use of flip-chip techniques, problems would arise for ceramics whose thermal expansion coefficient were very different from that of silicon, which features a rather low thermal expansion coefficient ($3.5\ 10^{-6}\ °C^{-1}$). These reasons led, during the 1990s, to the birth of low-temperature co-fired ceramics (LTCC) which only require sintering temperatures lower than 1100 °C, lower than the melting temperature of silver (960 °C) in many cases (Shinohara et al. 1987).

DOI 10.1007/978-981-4560-75-7_134-1

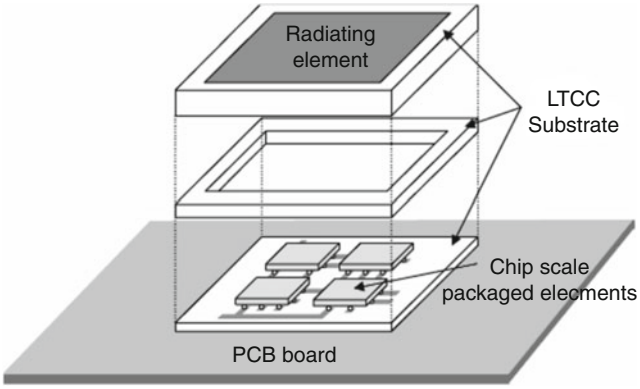


Fig. 4 Package-level integrated antenna integration (Wi et al. 2006)

The fabrication of LTCC components is a five-step process:

1. Slurry, composed of ceramic powder and a binder, is cast using a doctor blade on a thin film to make the green sheet.
2. Holes are then punctured in this green sheet to make vias, cavities, and registration holes.
3. The metal traces are deposited on the sheet to form the conducting pattern, and the via holes are filled with conductive paste. Screen printing of conductive pastes is commonly utilized for this purpose, offering a maximum resolution of 30 pm for line width and spaces. Less conventional methods, such as thin-film deposition technologies, can go down to 10 pm resolution for both lines and spaces (Muller et al. 2006).
4. The different layers are carefully aligned and stacked, using the registration holes or optical methods for improved precision, and laminated with heat and pressure.
5. Finally, the circuit is fired to form the multilayer ceramic substrate component or module.

This fabrication technology offers many advantages because of the high performance of the ceramics, offering a wide range of available permittivities and low dielectric losses ($\epsilon_r = 3.9-487$ and tan $\delta = 0.0007-0.006$ at 1 MHz) (Sebastian and Jantunen 2008). The different processes used for metalization also offer very conductive traces ($3.88-9.66$ m$\Omega/\square$) which can be used to manufacture high-quality RF components (Muller et al. 2006).

Some drawbacks of the technology are the previously mentioned limitations of conductor patterning resolution as well as precise layer alignment difficulties, thickness control, and warping. The major cause for these last issues is the significant shrinkage (9.5−25 %) of the ceramics during sintering. However, given proper material control and reproducibility, these effects can be taken into account and compensated in the design of the components.

Applications

LTCC technology is, by design, a multilayer process. This allows for the low-cost and compact integration of active and passive components as well as the package-level antenna integration, as shown on Fig. 4. These multilayer capabilities and high-performance materials also offer the advantage of making highly efficient multilayer antenna designs possible, as shown in the stacked-patches antenna for GPS applications (reported in Chen et al. 2010) (Fig. 5). As can be seen in Fig. 5b, the fabricated antenna features four very sharp resonances (two near-degenerate modes per patch antenna due to truncated corners) that are very similar to the simulated ones. Only slight shifts in resonant frequencies (from 1241, 1245, 1522, and

DOI 10.1007/978-981-4560-75-7_134-1

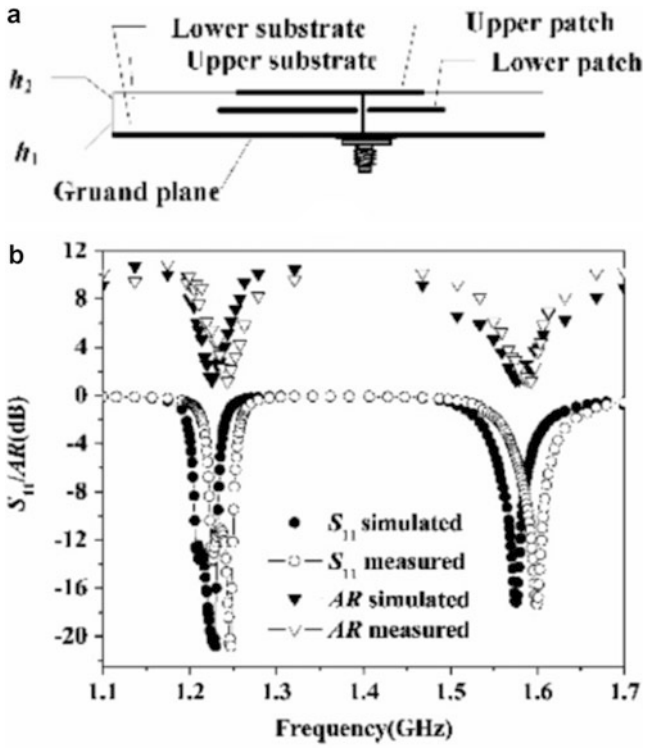


Fig. 5 A stacked-patches antenna manufactured with LTCC technology (**a**) Side view. (**b**) Simulated and measured return loss and axial ratio (Chen et al. 2010)

1584 MHz to 1221, 1233, 1499, and 1577 MHz) can be noticed and are attributed to slight fabrication inaccuracies (Chen et al. 2010).

Antennas can also be readily integrated into complete high-performance and high-complexity systems, even at high operational frequencies. Compact and 3-D integrated filters/duplexers and cross-shaped patch antennas integrated into a complete RF front end for operation in the 60 GHz band are reported in Lee et al. (2006). The antenna component reported excellent performance with a 10 dB bandwidth of 2.4 GHz (4.18 %) at 57.45 GHz and 2.3 GHz (3.84 %) at 59.85 GHz (Fig. 6).

At high frequencies, as the wavelength begins to be on the same scale as the thickness of these rather high-permittivity ceramics, problems can arise from the excitation of surface waves at the ceramic-air interface, resulting in lower gain for the fabricated antennas. This challenge has been tackled in Lamminen and Saily (2008) where an air cavity was placed under a 60 GHz band patch antenna to effectively reduce the overall permittivity of the substrate, thus leading to the beneficial effect. This also had the beneficial effect of increasing the bandwidth of the patch from 5.8 % to 9.5 % and increasing the gain by 2.5 dB. These patches (16 in total) were then used to make a high-performance antenna array (shown in Fig. 7a, b) **that** achieves a maximum gain of 18.2 dBi (Fig. 7c), **which** is a 2.5 dB improvement from its counterpart made of non-air cavity-backed patches.

LCP

A liquid crystal polymer (LCP) gets its name from the fact that its molecules can be mutually aligned and organized (crystal) and that bulk LCP can flow in molten state (liquid). When bulk LCP flows in the liquid state, the rigid segment of the LCP molecules aligns next to each other in the shear flow direction, as

DOI 10.1007/978-981-4560-75-7_134-1

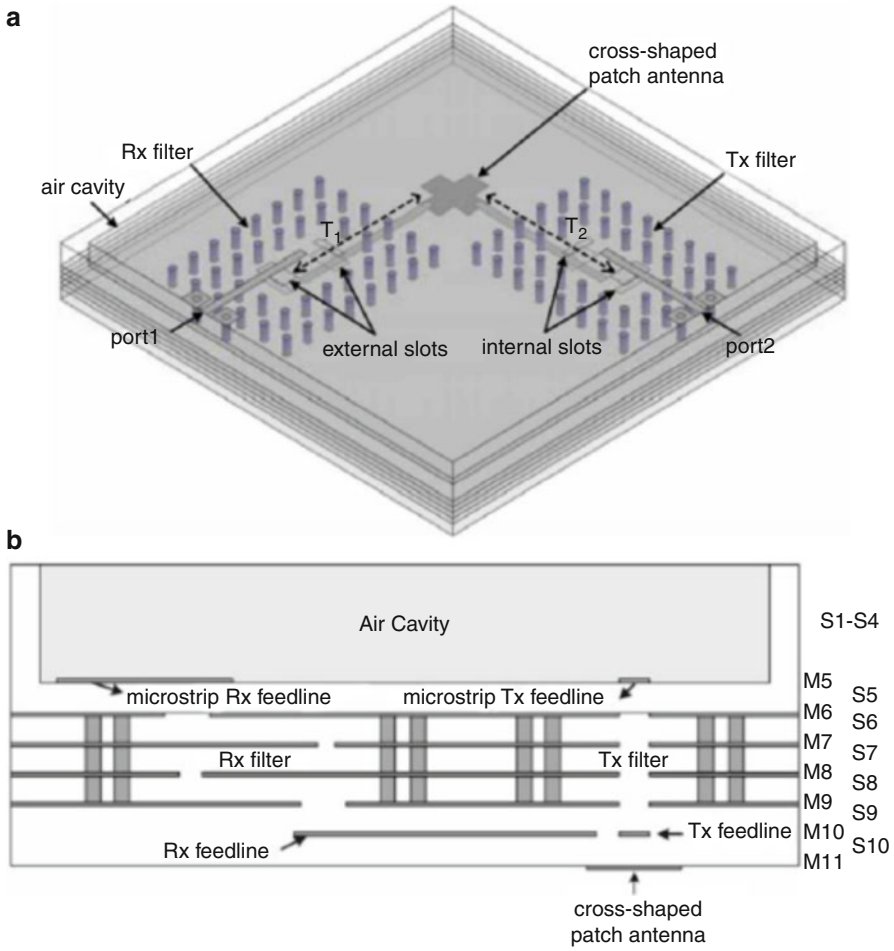


Fig. 6 3-D integrated V-band front-end with cavity filters and antenna, manufactured with LTCC technology (Lee et al. 2006)

shown in Fig. 8, thus creating locally oriented domains that lead to macroscopic-aligned regions. Once formed, the structure and direction of these oriented regions persist (Farrell and St Lawrence 2002). This behavior gives LCP a high mechanical strength at high temperatures, extremely good chemical resistance, inherent flame retardancy, and good durability in rugged ambient conditions (DeJean et al. 2005).

Although thermotropic LCPs possess a variety of properties that make them attractive candidates for electronic substrates, for a long time, standard LCP processing techniques resulted in films with anisotropic in-plane properties. Around 1998, Foster-Miller developed a new process that enables the control of the fibrillary orientation and achieves a biaxial orientation. Currently many electronic designs, especially flexible and conformal ones, including antenna structures, are realized every year using LCP as a host substrate.

LCP as an RF Material

LCPs feature a unique combination of properties that make them ideally suited for antenna fabrication, including: excellent electrical properties at microwave frequencies (low dielectric constant: 2.9, low loss tangent: 0.002 at 20 GHz, as reported by Rogers), low moisture absorption (<0.04 % by weight) and impermeability to moisture, oxygen, and other gases and liquids, and excellent dimensional stability (< 0.1 %).

DOI 10.1007/978-981-4560-75-7_134-1

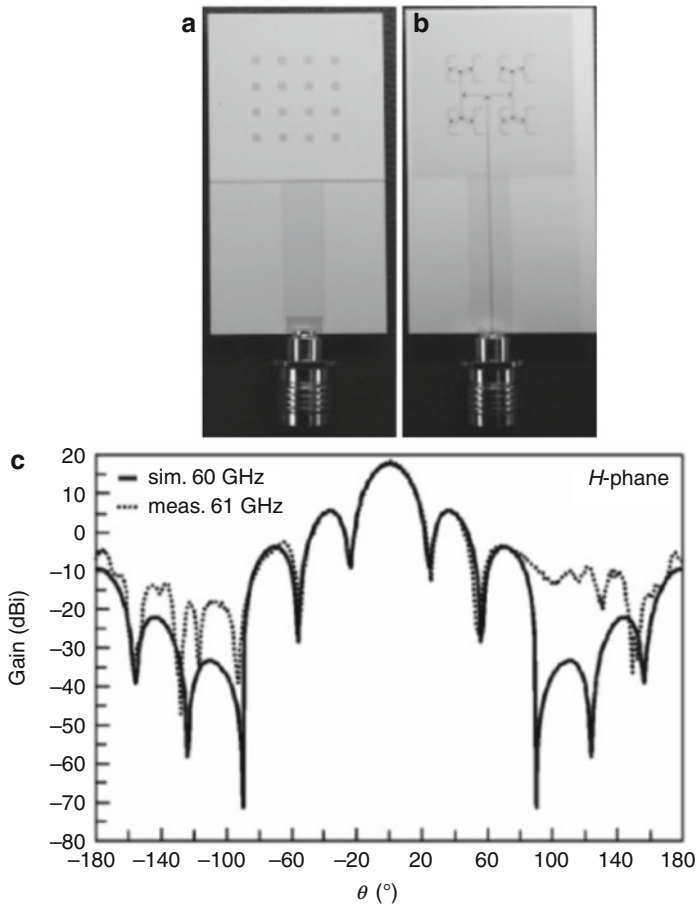


Fig. 7 Air cavity-backed patches antenna array, fabricated using LTCC technology (**a**) Photograph of the antenna array prototype. (**b**) Photograph of the feeding network. (**c**) Simulated and measured gain of the array (Lamminen and Saily 2008)

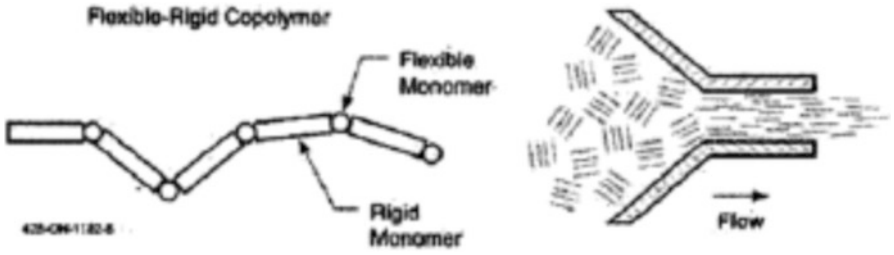


Fig. 8 Liquid crystal polymers have rigid segments which tend to align in shear flow (Farrell and St Lawrence 2002)

Electrical Properties at Microwave Frequencies

Various studies have shown that LCP has excellent electrical properties. Several have reported relative permittivity and dielectric loss parameters are shown in Table 1 with a frequency dependence shown in Figs. 9 and 10, respectively.

Fabrication Techniques

To fabricate antennas on LCP, standard two-sided printed circuit boards (PCBs) are typically used.

DOI 10.1007/978-981-4560-75-7_134-1

Table 1 Summary of reported LCP electrical properties

Source	Frequency (GHz)	ϵ_r	tan $\delta \cdot 10^{-3}$
Rogers Corp.	10	2.9	2
W.L. Gore	10	3	3
K. Jayaraj	6.97−24.66	3.07−3.18	–
G. Zou	3.85−34.55	3.00−3.04	3.4−2.7
D.C. Thompson (2004)	31.53−104.60	3.16 ± 0.05	4−4.5
F. Aryanfar (2010)	20−110	2.4−2.7	–

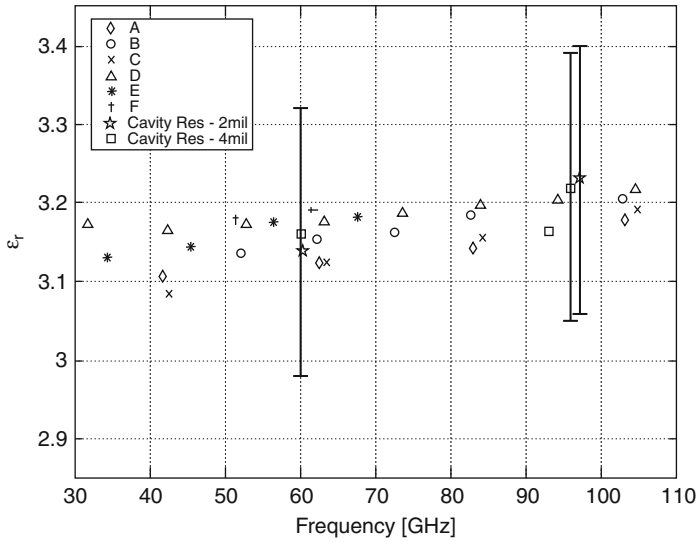


Fig. 9 Extracted dielectric constant using ring resonator designs (AF) and cavity resonators with 2- and 4-mil-thick LCP (Thompson et al. 2004)

- *Imaging and Etching* – LCP is suitable for traditional photolithography processes due to its inherent chemical/moisture resistance that prevents any undercutting at the polymer-copper interface. In this process, a photoresist is first deposited onto the copper clad laminate, followed by exposure (imaging) and stripping to define the desired antenna pattern. Then, the exposed copper is chemically etched out to leave the desired antenna topology.
- *Mechanical Drilling* – For large holes (greater than 8 mils in diameter), mechanical drilling is typically used in standard industry practice. It is shown that lower infeed rates (relatively, in the range of 2–8 mil/in) combined with high retract rates (relatively, in the range of 300–700 IPM) produce a better hole quality (Farrell and St Lawrence 2002).
- *Laser Drilling* – For small holes (less than 8 mils in diameter), lasers are typically used. For LCP materials, the best vias are made with shorter, more frequent pulses along with a higher number of pulses than required for some other standard materials (Farrell and St Lawrence 2002).
- *Desmear* – After drilling, a cleaning step is required and typically achieved by a wet chemical permanganate exposure for most RF substrates. However, this method does not work well with LCP due to its extremely high chemical resistance. Instead, plasma treatment is very effective at cleaning holes.

Applications

LCP has been one of the most preferable conformal RF substrates due to the numerous desirable features previously listed. The dielectric constant is low enough to be used while being conformal and it is easy to

DOI 10.1007/978-981-4560-75-7_134-1

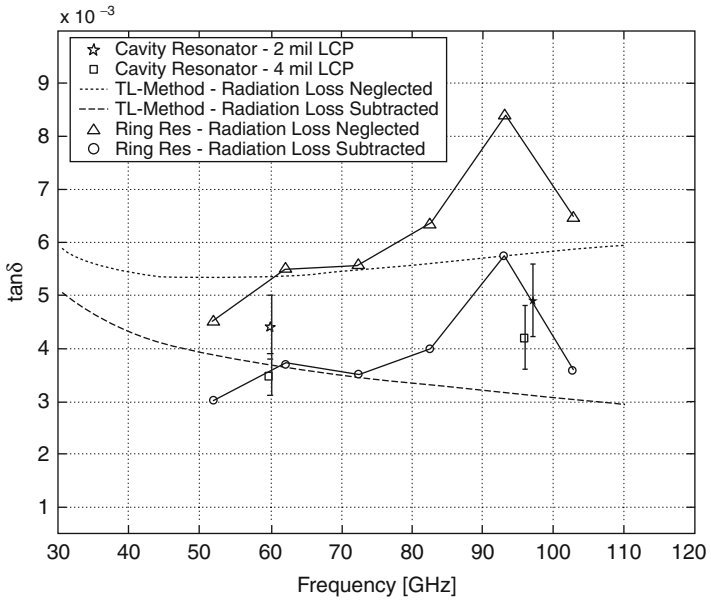


Fig. 10 LCP loss tangent versus frequency for 2- and 4-mil substrate thicknesses measured with a cavity resonator method. Results for the ring resonator method and the TL method on 3-mil LCP substrates are shown with and without subtracting radiation loss (Thompson et al. 2004)

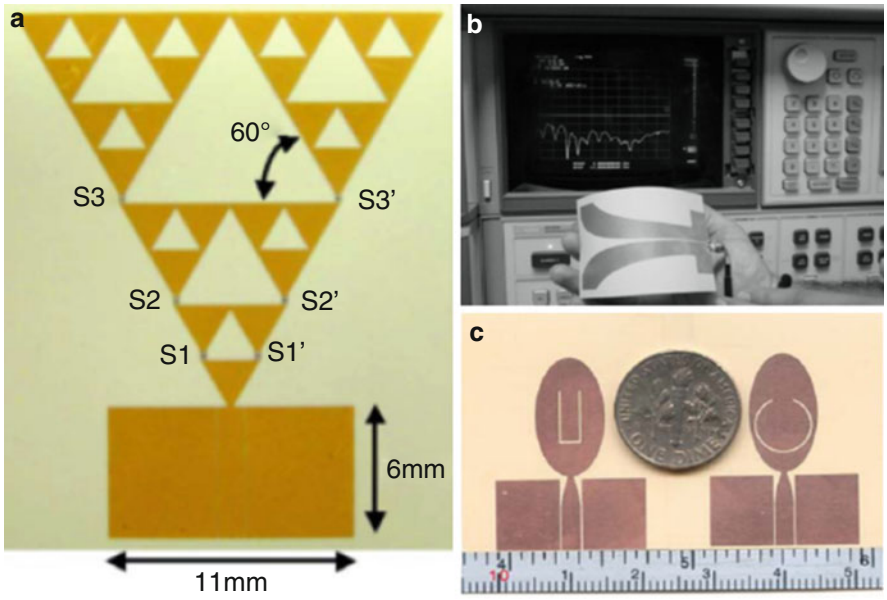


Fig. 11 Photos of several fabricated antenna on LCP: (**a**) a reconfigurable Sierpinski antenna (Kingsley et al. 2007), (**b**) a conformal double exponentially tapered slot antenna (DETSA) on LCP for UWB applications with its VNA measurement during bending (http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1638360), (**c**) two CPW-fed ultra wideband (UWB) monopoles (Nikolaou et al. 2006)

fabricate with an engineered coefficient of thermal expansion. Since the material is a polymer, there are additional packaging and cost advantages. All of these characteristics make it an ideal substrate for flexible antennas, particularly at high frequencies (Fig. 11).

DOI 10.1007/978-981-4560-75-7_134-1

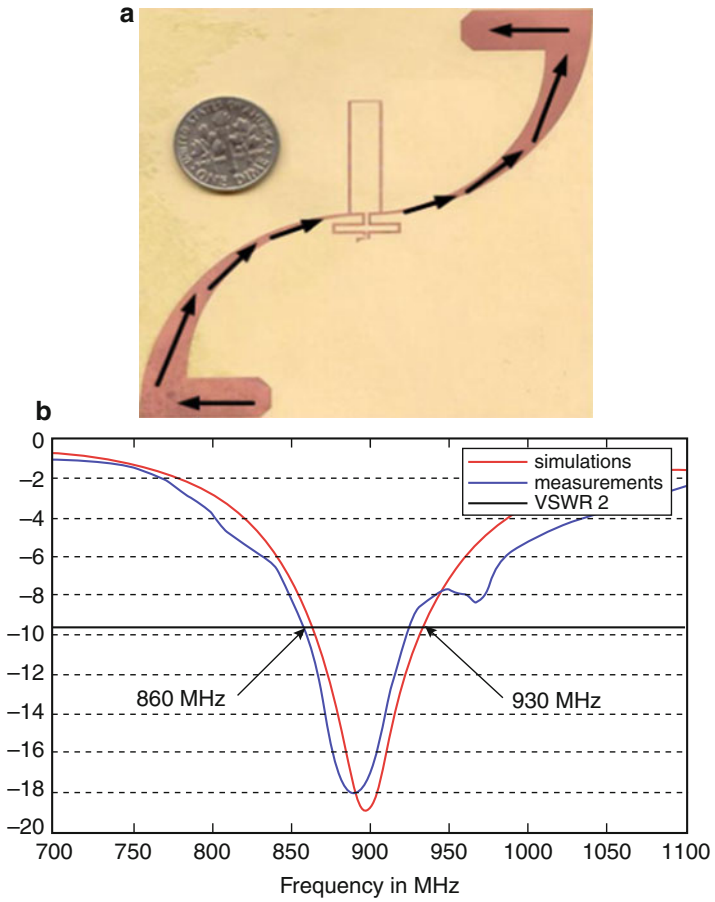


Fig. 12 (**a**) Photo of an RFID tag on LCP. (**b**) Return loss of the RFID (Serkan Basat et al. 2006)

RFID

Requirements in automatic identification in several areas, such as item level tracking, supply chain, and retail applications, have increased the demand for low-cost, flexible radio frequency identification (RFID) tags. LCP as an antenna substrate is considered one of the best candidates for RFID applications, especially in millimeter-wave frequency ranges (mmID's) (Vyas et al. 2007). One of the main required features of most RFID tags is a conformal shape or flexibility in order to be placed on differently shaped objects, such as boxes, cylindrical bins, vehicles, etc. Secondly, since most RFID tags contain no extra packaging and are exposed directly to their surroundings, RFID tags must be able to withstand harsh industrial environments, such as water vapor and high humidity. A third requirement is to have a low dielectric constant value with low dielectric loss for optimal RF power efficiency and transmissivity in embedded objects and optimum power performance, especially in passive RFID systems, where the only power to the tag is the RF power from the reader in its vicinity. A typical passive RFID tag module is shown in Fig. 12 and a Novel 3-D Cubic Antenna is shown in Fig. 13.

Portable Devices

There is a strong need for high-efficiency antennas that can be easily integrated into portable devices where high frequency operation, light weight, and conformability to curved surfaces are required. In recent years there has been an increase in using flexible LCP substrates because of their inherent flexibility and improved loss characteristics. The low loss property of LCPs becomes further important for portable

DOI 10.1007/978-981-4560-75-7_134-1

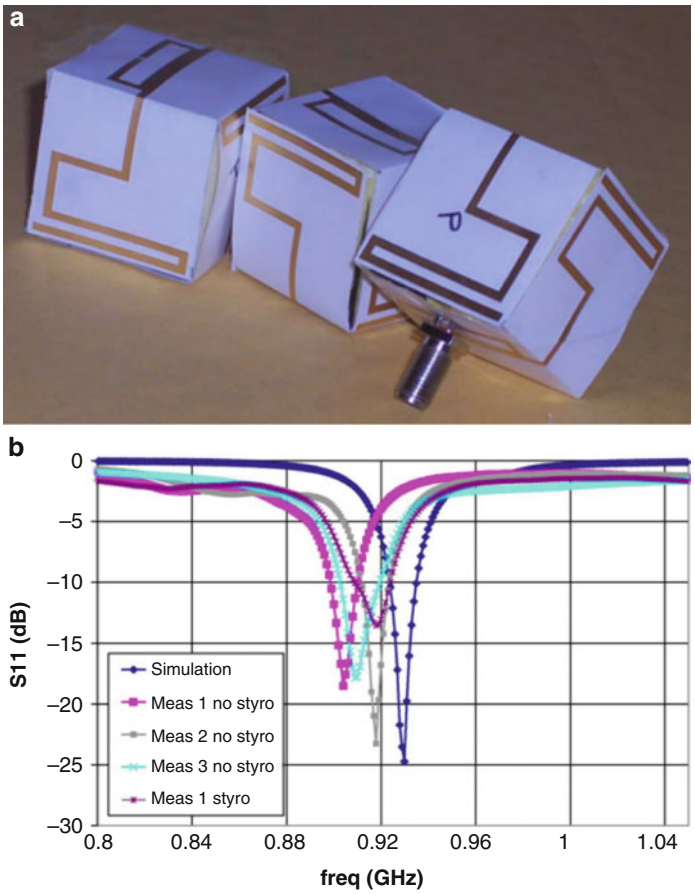


Fig. 13 A novel 3-D cubic antenna for RFID and WSN applications: (**a**) photo, (**b**) S-parameter measurements compared with simulation data (Kruesi et al. 2009)

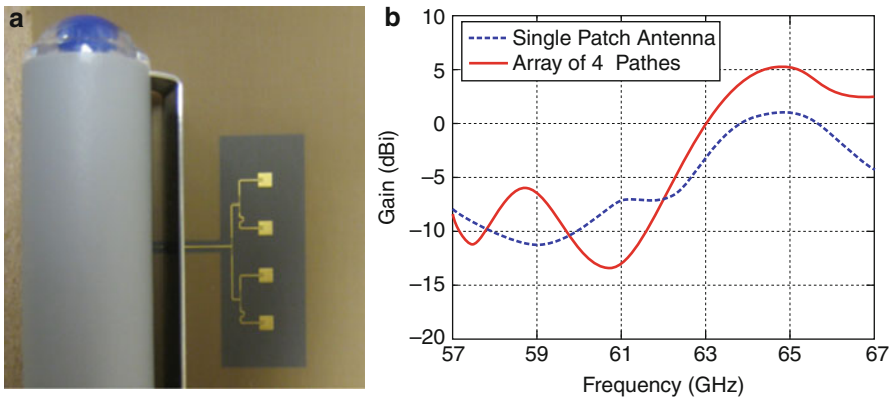


Fig. 14 (**a**) The picture of the fabricated prototype of a 4 × 1 array of patch antennas on LCP for mm-wave applications. (**b**) Measured gain of a single patch and of a 4 × 1 array of patch antennas (Aryanfar and Werner 2010)

devices where available power is scarce and cannot be used to compensate for lack of efficiency in individual RF components (Aryanfar and Werner 2010). A 4 × 1 patch antenna array for high-speed ultrabroadband mm-wave portable modules around 60 GHz can be found in Fig. 14.

DOI 10.1007/978-981-4560-75-7_134-1

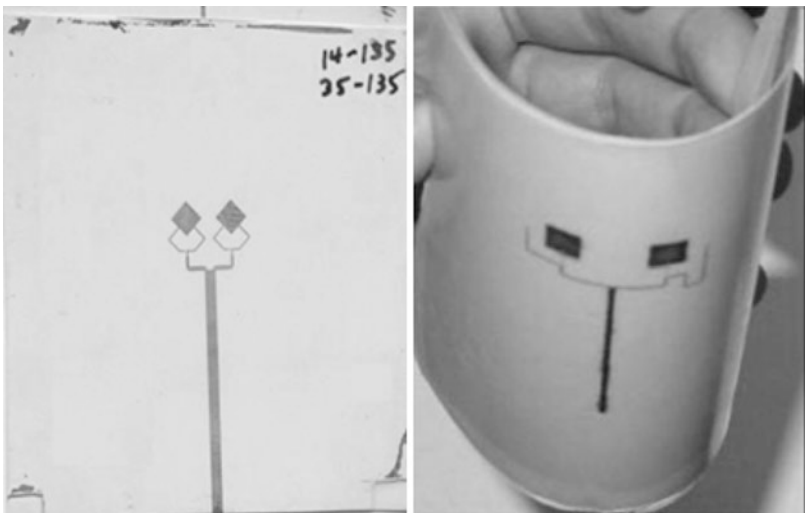


Fig. 15 Photographs of fabricated prototypes of multilayer dual-frequency/dual-polarization flexible antenna arrays in flat and bent configurations (DeJean et al. 2005)

Multilayer Antennas

Due to its inherent capability of multilayer (three-dimensional) vertical integration, good electrical and mechanical properties, and near-hermetic nature, LCP is also a practical choice for the design of low-cost antenna arrays that can be integrated with remote sensing applications operating in the K_u and millimeter-wave frequency bands.

Many thermal and mechanical problems (e.g., stress due to coefficient of thermal expansion mismatches) are inherent in traditional multilayer designs relying on the integration of different materials. However, LCP offers a unique combination of properties with a low dielectric constant and low loss tangent. In addition, LCP has a low CTE (8 or 17) and its CTE can be adjusted through thermal treatments, thus facilitating the integration of integrated circuits in SOP modules. Those properties make LCP an excellent candidate for multilayer structures (Fig. 15).

Printing

The current advancement of wireless technology is pushing toward several improvements within the realm of electronic fabrication, specifically with respect to reducing cost, lowering waste, and increasing functional versatility. Considering the methods of antenna fabrication previously discussed in this chapter, subtractive methods represent a strong majority of antenna fabrication methods typically used in commercially available electronic systems. These methods require the use of patterning masks, electronic plating, planar material growth, chemical etchants, and various other procedures to both deposit and remove materials from host substrates.

Additive material printing offers an alternative approach to the widespread methods of subtractive electronic fabrication. Mass material deposition and subsequent removal is replaced with select, direct patterning, where the only materials used are materials directly constructing the desired electronic patterns and structures.

Through the utilization of conductive and insulating ink/filament materials, direct printing methods such as ink-jet and 3-D printing are capable of fabricating flexible and rigid multilayer antenna structures in a purely additive process (Yang et al. 2007). The entire stack up of a multilayer antenna system, including conductive metallization and thick dielectric substrates, can be achieved with ink-jet printing,

DOI 10.1007/978-981-4560-75-7_134-1

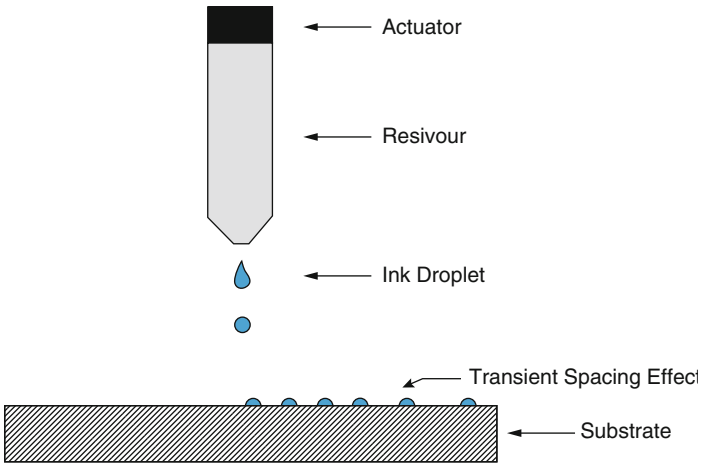


Fig. 16 A typical DOD ink-jet-printing system

removing the need for multi-multilaminate processing, trimming, and bonding while allowing for integration with wireless technologies as a post-processing fabrication method. 3-D printing further extends the versatility of printed systems into an unlimited variety of applications, where flexible plastics and conductors can be used to realize wearable, reconfigurable wireless systems with a feasibly lowered cost and improved functionality.

Ink-Jet Printing

Ink-jet printing is an additive fabrication technology that consists of three general elements: a dispensary apparatus, an ink material, and a host substrate. The dispensary apparatus is typically an ink cartridge consisting of an ink reservoir and a device to eject drops onto a substrate, which is commonly realized through a piezoelectric MEMS device. Ink materials can range in material composition and functionality, including conductive nanoparticle suspensions and polymeric dielectric solutions. The host substrate accepts and is patterned by the ink droplets ejected from the dispensary apparatus, where certain material characteristics such as surface tension and fluidic wetting must be taken into account in order to ensure accurate material patterning.

Ink-Jet Printing Processes

The technology behind ink-jet printing begins with the dispensary methodology. The two main methods of droplet dispensing are known as drop-on-demand (DOD) and continuous. DOD systems utilize a mechanical actuator to feed, form, and eject ink droplets from a dispensing device, where droplets are only dispensed when they are desired to pattern a host substrate. Continuous systems differ in the fact that droplets are always being fed and ejected while a "catcher" structure is used to block or allow dispensed droplets from patterning a substrate. Issues of nozzle clogging are reduced with continuous systems; however, DOD systems are more efficient for small-volume prototyping and allow a larger variety of ink materials, which is especially important for antenna design and fabrication within an experimental setting (Wiederrecht 2009; Magdassi 2010). Figure 16 illustrates a typical DOD ink-jet-printing system (Cook et al. 2014).

The mechanical actuators used within ink-jet printing vary in the physical mechanisms used to dispense ink droplets from a dispensary onto a substrate. For modern ink-jet-printing systems, four main actuators are used:

DOI 10.1007/978-981-4560-75-7_134-1

1. *Thermal* – Ink is heated within a cavity until volume and pressure is increased enough to force a droplet ejection. Very low cost, yet relatively high drop volume (80–100 pL) (Wiederrecht 2009).
2. *Piezoelectric* – Ink is pulled into and ejected in droplets from a cavity with a piezoelectrically actuated membrane. Higher control over droplet formation and wide variety of allowable ink materials.
3. *Electrodynamic* – A large electric field is applied between a nozzle and conductive platten below a host substrate, where ink droplets are forcefully ejected from a nozzle. Capable of producing very small drop volumes (<1 fL), yet requires inks to contain free ions (Choi et al. 2011).
4. *Acoustic* – Ultrasonic waves are focused on an ink meniscus, overcoming surface tension and allowing a droplet to be ejected. Reduced nozzle clogging and increased precision in drop volume variability (Hamazaki and Morita 2009).

Ink Materials

In order to realize robust, multilayer electronic topologies with ink-jet-printing fabrication, a variety of ink materials are required. Specifically for ink-jet-printing antenna structures, extended attention has been drawn to two general ink varieties: conductive and dielectric.

The most efficient conductive inks used within ink-jet printing are metallic in essence but differ in their material composition and processing. Silver nanoparticle solutions of noble metals, such as silver and gold, currently dominate the market of commercially available metallic conducting inks. These inks are composed of a suspension of metallic nanoparticles within an alcoholic solvent solution. Upon printing, the solvent materials are evaporated with a low-temperature thermal treatment and then sintered with either thermal baking (< 250 °C) or laser rastering, offering a method of metallic that does not interfere with low-temperature substrates, such as polymeric or fibrous materials (Cook and Shamim 2012). Non-noble metals, such as copper, are difficult to process as nanoparticles because of their high oxidation rates during sintering. As an alternative, metallic catalyst inks can be used to pattern a latent image of a conductive topology, followed by the electroless chemical plating of metal onto the printed pattern (Cook et al. 2013c). These methods of conductive ink printing are able to achieve conductivity in the range of $1.2 \cdot 10^7$ S/m, 5x lower than that of bulk silver (Cook and Shamim 2012).

Dielectric inks are utilized to pattern insulating structures, where both pattern area and film thickness are variables of interest. Dielectric inks were first developed for ink-jet printing with the purpose of creating thin dielectric films for metal-insulator-metal (MIM) capacitors and transistor gate dielectrics, yielding thicknesses in the ranging from hundreds of nanometers of several micrometers (Cook et al. 2013a; Sanchez-Romaguera et al. 2008; Ko et al. 2007). However, the ability to print thick dielectric structures is integral to the realization of fully printed antenna structures. Recent advancements in ink development have enabled ink-jet printing to effectively pattern thick dielectric substrates for antennas, exceeding 100 μm in thickness (Cook et al. 2013b; Bito et al. 2014; Tehrani et al. 2014). These thick films are achieved with long-chain polymers, such as the SU-8 photoresist, which can yield a high w/w% while maintaining a low viscosity. Per-layer thickness of this printed dielectric ink is within the range of 4–6 μm, which can be used to precisely control the thickness of a desired dielectric substrate, as shown in Fig. 17 (Tehrani et al. 2015).

Ink-Jet-Printed Antenna Structures

Through the utilization of conductive nanoparticle-based and thick dielectric polymer-based inks, ink-jet printing has realized a variety of antenna structures for such applications as millimeter-wave (mm-wave) wireless systems (Cook et al. 2013b; Bito et al. 2014; Tehrani et al. 2015) and microfluidics-enabled tuning/sensing platforms (Su et al. 2014, 2015). Two of the most recent demonstrations of ink-jet-printing antenna fabrication will be presented in this section.

DOI 10.1007/978-981-4560-75-7_134-1

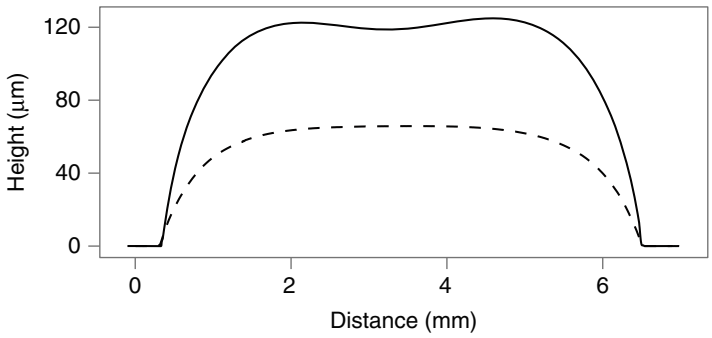


Fig. 17 Film profiles of (– –) 7-layer and (—) 14-layer ink-jet-printed dielectric substrates

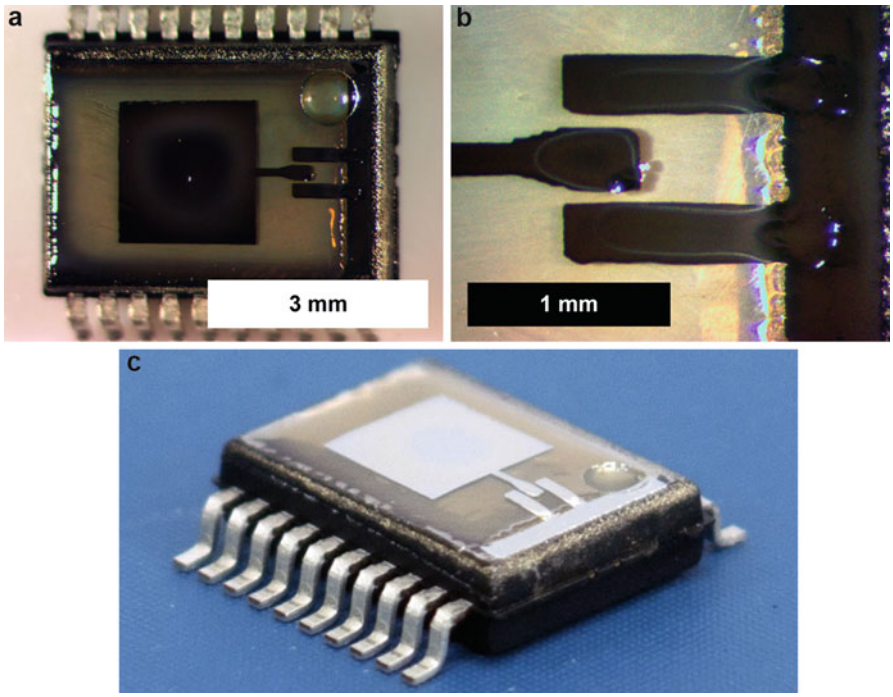


Fig. 18 Ink-jet-printed on-package 30 GHz patch antenna: (**a**) top view, (**b**) coplanar waveguide (CPW) feed and via detail, and (**c**) perspective view

Some of the greatest advantages of ink-jet printing are its fully additive nature and host material flexibility. Conductive and dielectric ink materials can be patterned strictly as needed onto virtually any host substrate, which introduces the application of using ink-jet printing to fabricate multilayer antenna structures in a post-processing fashion into an existing electronic system. An example of this post-processing approach to ink-jet-printing antenna fabrication is demonstrated by B. Tehrani et al. in (Tehrani et al. 2015). In this work, ink-jet printing is used to fabricate a multilayer 30 GHz patch antenna directly onto a wireless integrated circuit (IC) package. Figure 18 shows images of the printed antenna structure, where ground plane, 120 μm thick SU-8 dielectric substrate, and antenna topology are all realized through ink-jet printing. S-parameter measurements are presented in Fig. 19, displaying excellent agreement with simulations and highlighting the predictability and integrity of ink-jet printing as a fabrication process for multilayer antennas.

Another advantage of ink-jet printing is its low-cost nature, which is especially important for applications such as ubiquitous wireless sensor networks and health monitoring platforms. Using a low-cost poly

DOI 10.1007/978-981-4560-75-7_134-1

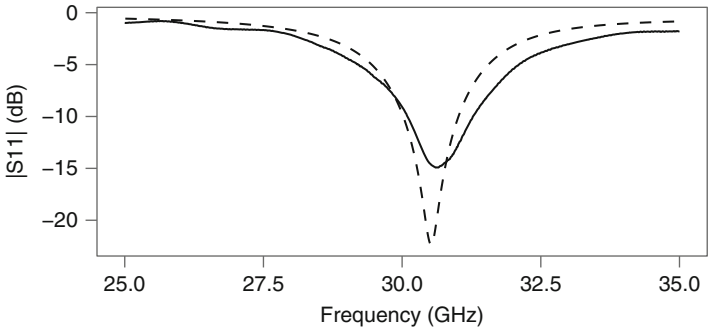


Fig. 19 Reflection coefficient measurements of (– –) simulated and (—) fabricated on-packaged antennas

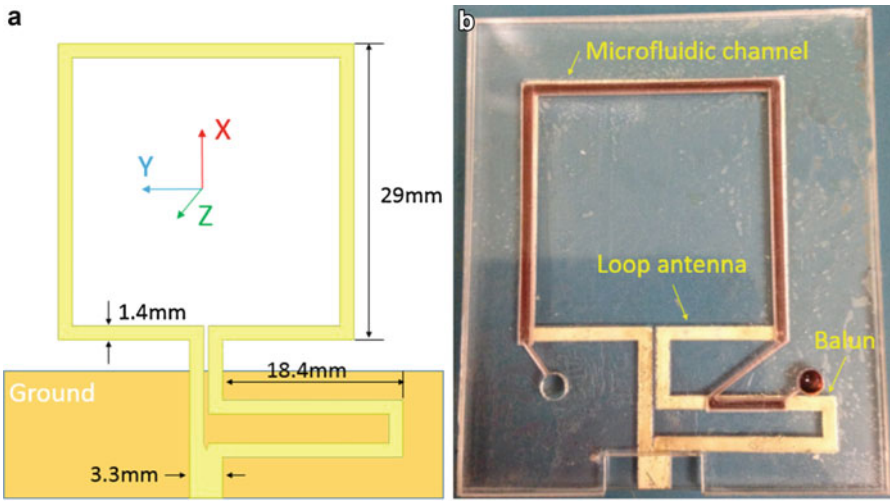


Fig. 20 Ink-jet-printed microfluidic-enabled loop antenna (**a**) model and (**b**) fabricated sample

(methyl methacrylate) (PMMA) host substrate, W. Su et al. demonstrate an ink-jet-printed loop antenna with microfluidics-enabled tuning and matching (Su et al. 2015). Ink-jet-printed antenna topology is stacked with a thicker sheet of PMMA, on which a microfluidic channel is laser etched for fluidic routing and interface, shown in Fig. 20. The bonding between the antenna topology and the microfluidic channel is achieved with the printing and curing of SU-8 polymer ink, acting as a printable epoxy resin. When different fluids are loaded into the microfluidic channel, the dielectric constant of the fluid effectively tunes the antenna as well as the balun, improving matching during fluidic loading. S-parameter measurements are presented in Fig. 21 for the ink-jet-printed antenna with and without microfluidic loading.

3-D Printing

3-D Printing Processes

Almost any additive manufacturing technique that creates three-dimensional objects is considered 3-D printing. While rapid-prototyping techniques involving 3-D printing have existed since the 1980s (Frauenfelder 2013), the recent rapid interest and growth of the technology occurred after the expiration of US patent 5121329 related to fused deposition modeling (FDM) by Stratasys. There exist many 3-D printing processes that vary in materials, resolution, strength, and time. The most commonly used 3-D printing techniques are FDM and stereolithography (SL). Fused deposition modeling is the process of

DOI 10.1007/978-981-4560-75-7_134-1

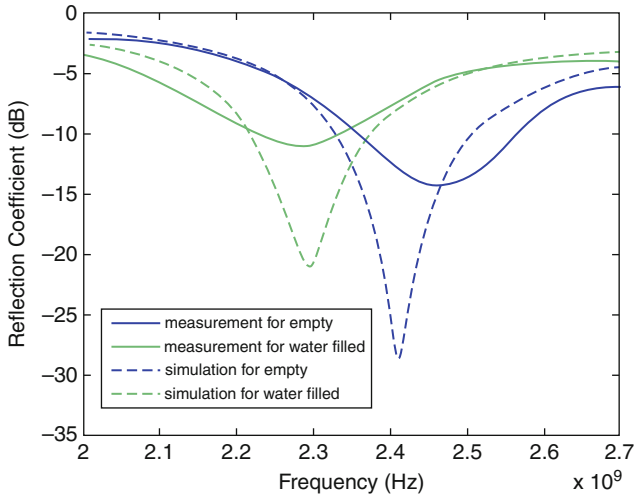


Fig. 21 Reflection coefficient simulations and measurements of an ink-jet-printed microfluidic-enabled loop antenna with and without fluidic loading

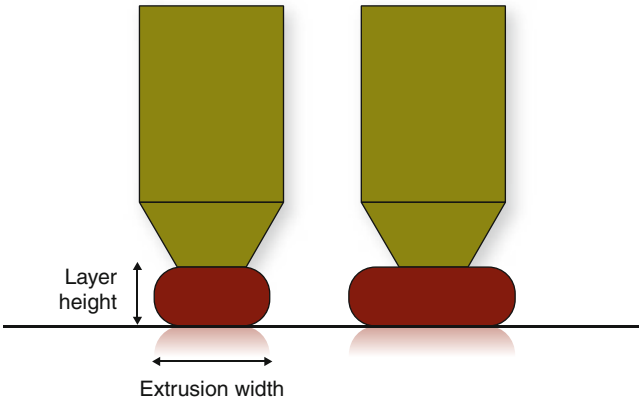


Fig. 22 FDM layer height. The extrusion width is a function of the nozzle diameter (Hodgson 2011)

extruding material, generally a thermoplastic, through a heated nozzle that exceeds the glass transition temperature. FDM techniques, like almost all 3-D printing techniques, print layer by layer. The layer height is a common feature that defines an FDM printer, as shown in Fig. 22, with many consumer-available printers having a layer height of 100 μm (some as low as 20 μm). Stereolithography involves vats of photopolymers that react to laser UV light to harden the polymer at select points. SL printers tend to have higher resolutions, with a common layer height around 30 μm and tend to be more complex with laser or digital light project (DLP)-based systems. Another process is laser sintering/melting, which involves small particles of materials, even metals, which are fused together by a laser.

As this is a relatively new technology that has garnered much public interest recently, many new methods are being developed. There are printers combining ink-jet or paste dispensing techniques that allow conductive materials to be printed simultaneously with the main substrate material, for example, with the electronics printing system from Voxel8 (Voxel8 (13 April 2015). The ability to print highly conductive materials with substrates is a desired feature that is currently rapidly evolving.

DOI 10.1007/978-981-4560-75-7_134-1

3-D Printing Materials

Most 3-D printers involve a polymer-based material. Acrylonitrile butadiene styrene (ABS) and polylactic acid (PLA) are common for FDM-based printers. Multiple materials can be printed simultaneously with an FDM printer with relative ease. A wide variety of materials are available, usually a mixture of a polymer such as ABS or PLA, such as: ceramic, wood mixture, copper mixture, thermoplastic elastomers, conductive mixtures (approx. 1 Ω/cm), PET polymer, iron-based mixture, steel-based mixture, carbon fiber-based mixture, and a variety of other materials. Photopolymers for SL systems are generally a single material print, with much less variety currently. There are SL systems that combine alternative manufacturing technologies to allow multi-material prints (Wicker et al. 2004). PolyJet printing from Stratasys is similar to ink-jet printing, down to 16 μm with a wide variety of polymer materials. A variety of materials have been characterized on multiple types of systems (Deffenbaugh et al. 2013b; Deffenbaugh 2014).

3-D-Printed Antenna Structures

Considerations must be made when using a manufacturing method intended for 3-D prototypes to create dielectric-dependent designs, a necessity for many antennas. Dielectric constants may vary between brand and color of the material, as well as the settings on the 3-D printer. One of the most important settings that will change the dielectric constant is the infill density. Infill density versus dielectric constant has been characterized for several materials (Deffenbaugh 2014). Another issue that may arise, depending on the metallization method, is the surface roughness. This directly relates to the 3-D printing method. SL systems generally have layers below 30 μm, which results in a surface variation less than that. A concern with FDM is the ability for gaps to occur on the material surface, which may cause shorts during metallization of thin layers. This can be overcome and is generally an occurrence of poor manufacturing. With these limitations in mind, several designs are demonstrated utilizing both SL and FDM printing techniques (Deffenbaugh 2014).

Two examples of 3-D printed RF structures are shown in Figs. 23 and 24. The microstrip transmission line is designed on an SL system. The complex structure, in spite of the material having a higher loss tangent than an FR4 laminate, demonstrates lower loss than a planar FR4 structure. The metallization for the 2.4 GHz antennas designed on an FDM substrate is realized with an nScrypt 3Dn micro dispensing system (Deffenbaugh et al. 2013a), to show the feasibility of 3-D printed antennas comparable to typical manufacturing methods.

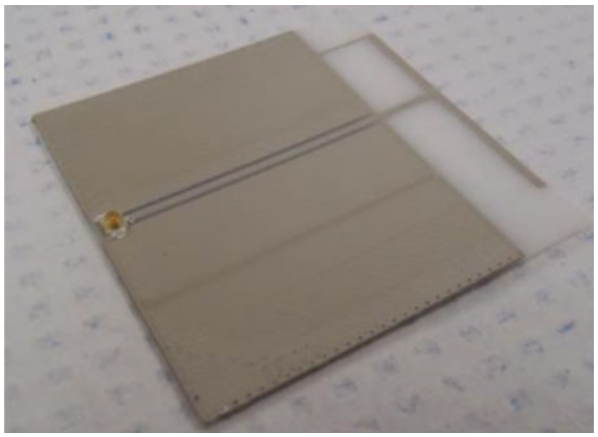


Fig. 23 3-D-printed 2.4 GHz antenna printed using FDM printer

DOI 10.1007/978-981-4560-75-7_134-1

Fig. 24 3-D-printed low loss microstrip line

Future Directions and Open Problems

Throughout the four general methods of antenna fabrication presented in this chapter, there exist fundamental trade-offs between cost, minimum feature size, processing time, and scalability. Lithographic methods are a widespread standard throughout all forms of electronic fabrication and offer the lowest minimum feature sizes; however, high equipment costs, long processing times, and a lack of large-area scalability are factors that must be considered. Additive techniques such as ink-jet and 3-D printing offer relatively quick prototyping, lower cost, and seamless integration with large-scale fabrication; however, an increase in minimum feature size must be considered, where printing techniques offer minimum features nearly $10-100\times$ larger than lithographic techniques.

With the current push toward the development of low-cost, ubiquitous wireless systems throughout emerging technologies, the need for improvement within the realm of antenna fabrication is greatly stressed. This improvement should be realized with methods that reduce material waste, increase productive throughput, and exhibit an ease of reconfigurability, while highlighting current trends in miniaturization and functional diversity. Current advancements in additive manufacturing techniques show great promise for the realization of these goals for efficient antenna fabrication. The growing interest present within both industrial and academic research throughout the areas of ink-jet and 3-D printing offer a glimpse into the future of low-cost, vertically integrated wireless antenna fabrication.

Acknowledgment

The authors would like to recognize the National Science Foundation (NSF) and the Defense Threat Reduction Agency (DTRA) for their support with this work.

Cross-References

- ► Antennas in Body-Centric Sensor Network Devices
- ► Conformal Array Antennas
- ► Low-Profile Antennas
- ► Metamaterials and Antennas
- ► Milestones of Antenna Patents in History
- ► Millimeter-Wave Antennas and Arrays
- ► On-chip Antennas
- ► Physical Bounds of Antennas

DOI 10.1007/978-981-4560-75-7_134-1

References

Aryanfar F, Werner CW (2010) Exploring liquid crystal polymer (lcp) substrates for mm-wave antennas in portable devices. In: Antennas and propagation society international symposium (APSURSI), 2010 IEEE, IEEE, Toronto, ON, pp 1–4

Bito J, Tehrani B, Cook B, Tentzeris M (2014) Fully inkjet-printed multilayer microstrip patch antenna for Ku-band applications. In: Antennas and propagation society international symposium (APS/URSI), 2014 IEEE, Memphis, TN, pp 854−855

Bustillo JM, Howe RT, Muller RS (1998) Surface micromachining for microelectromechanical systems. Proc IEEE 86(8):1552–1574

Chen S, Liu G, Chen X, Lin T (2010) Compact dual-band GPS microstrip antenna using multilayer LTCC substrate. IEEE Antennas Wirel Propag Lett 9:421–423

Choi K-H, Rahman K, Muhammad NM, Khan A, Kwon K-R, Doh Y-H, Kim H-C (2011) Recent advances in nanofabrication techniques and applications. InTech, Rijeka

Cook B, Shamim A (2012) Inkjet printing of novel wideband and high gain antennas on low-cost paper substrate. IEEE Trans Antennas Propag 60(9):4148–4156

Cook B, Cooper J, Tentzeris M (2013a) Multilayer RF capacitors on flexible substrates utilizing inkjet printed dielectric polymers. IEEE Microwave Wireless Compon Lett 23(7):353–355

Cook B, Tehrani B, Cooper J, Tentzeris M (2013b) Multilayer inkjet printing of millimeter-wave proximity-fed patch arrays on flexible substrates. IEEE Antennas Wirel Propag Lett 12:1351–1354

Cook BS, Fang Y, Kim S, Le T, Goodwin B, Sandhage KH, Tentzeris MM (2013c) Inkjet catalyst printing and electroless copper deposition for low-cost patterned microwave passive devices on paper. Electron Mater Lett 9:669–676

Cook B, Tehrani B, Cooper J, Kim S, Tentzeris M (2014) Integrated printing for 2d/3d flexible organic electronic devices. In: Handbook of flexible organic electronics: materials, manufacturing and applications, Woodhead Publishing, Sawston, Cambridge, pp 199

Deffenbaugh PI (2014) 3D printed electromagnetic transmission and electronic structures fabricated on a single platform using advanced process integration techniques. PhD thesis, The University of Texas at El Paso

Deffenbaugh PI, Goldfarb J, Chen X, Church K (2013a) Fully 3d printed 2.4 ghz bluetooth/wi-fi antenna. In: IMAPS, 46th international symposium on microelectronics, Orlando

Deffenbaugh PI, Rumpf RC, Church KH (2013b) Broadband microwave frequency characterization of 3-d printed materials. IEEE Trans Compon Packag Manuf Technol 3(12):2147–2155

DeJean G, Bairavasubramanian R, Thompson D, Ponchak G, Tentzeris M, Papapolymerou J (2005) Liquid crystal polymer (lcp): a new organic material for the development of multilayer dual-frequency/dual-polarization flexible antenna arrays. IEEE Antennas Wirel Propag Lett 4:22–26

Farrell B, St Lawrence M (2002) The processing of liquid crystalline polymer printed circuits. In: Electronic components and technology conference, 2002. Proceedings. 52nd, IEEE, San Diego, CA, pp 667−671

Frauenfelder M (2013) Make: ultimate guide to 3D printing 2014. Maker Media, Sebastopol

Hamazaki T, Morita N (2009) Ejection charamteristics and drop modulation of acoustic inkjet printing using fresnel lens. J Fluid Sci Tech 2:25–36

Hodgson G (2011) Slic3r manual. Aleph Objects

Howe RT, Muller RS (1982) Polycrystalline silicon micromechanical beams. In: Electrochemical society extended abstracts, vol 82, Montreal

DOI 10.1007/978-981-4560-75-7_134-1

Kingsley N, Anagnostou DE, Tentzeris M, Papapolymerou J (2007) Rf mems sequentially reconfigurable sierpinski antenna on a flexible organic substrate with novel dc-biasing technique. J Microelectromech Syst 16(5):1185–1192

Ko SH, Pan H, Grigoropoulos CP, Luscombe CK, Frchet JMJ, Poulikakos D (2007) All-inkjet-printed flexible electronics fabrication on a polymer substrate by low-temperature high-resolution selective laser sintering of metal nanoparticles. Nanotechnology 18(34):345202

Kovacs GT, Maluf NI, Petersen KE (1998) Bulk micromachining of silicon. Proc IEEE 86(8):1536–1551

Kruesi CM, Vyas RJ, Tentzeris MM (2009) Design and development of a novel 3-d cubic antenna for wireless sensor networks (wsns) and rfid applications. IEEE Trans Antennas Propag 57(10):3293–3299

Lamminen A, Saily J (2008) 60-GHz patch antennas and arrays on LTCC with embedded-cavity substrates. IEEE Trans Antennas Propag 56(9):2865–2874

Lee J, Kidera N, Gerald D, Pinel S, Laskar J, Tentzeris MM (2006) A v-band front-end with 3-D integrated cavity filters/duplexers and antenna in LTCC technologies. IEEE Trans Microwave Theory Tech 54(7):2925–2936

Madou MJ (2002) Fundamentals of microfabrication: the science of miniaturization. CRC Press, Boca Raton

Magdassi S (2010) The chemistry of inkjet inks. World Scientific Publishing, Singapore

Muller J, Perrone R, Thust H, Drue K (2006) Technology benchmarking of high resolution structures on LTCC for microwave circuits. In: Electronics system integration technology conference, 2006. 1st, Dresden, pp 111–117

Nathanson HC, Newell WE, Wickstrom RA, Davis JR Jr (1967) The resonant gate transistor. IEEE Trans Electron Devices 14(3):117–133

Nikolaou S, Kim B, Kim YS, Papapolymerou J, Tentzeris MM (2006) Cpw-fed ultra wideband (uwb) monopoles with band rejection characteristic on ultra thin organic substrate. In: Microwave conference, 2006. APMC 2006. Asia-Pacific, IEEE, pp 2010−2013

Park JY, Kim GH, Chung KW, Bu JU (2001) Monolithically integrated micromachined rf mems capacitive switches. Sensors Actuators A Phys 89(1):88–94

Sanchez-Romaguera V, Madec M-B, Yeates SG (2008) Inkjet printing of 3d metal-insulator-metal crossovers. React Funct Polym 68(6):1052–1058

Sebastian M, Jantunen H (2008) Low loss dielectric materials for LTCC applications: a review. Int Mater Rev 53(2):57–90

Senturia SD (2001) Microsystem design, vol 3. Kluwer, Boston

Serkan Basat S, Bhattacharya S, Li Y, Rida A, Tentzeris MM, Laskar J (2006) Design of a novel high-efficiency uhf rfid antenna on flexible lcp substrate with high read-range capability. In: Antennas and propagation society international symposium (APS/URSI), 2006 IEEE, pp 1031−1034

Shinohara H, Ushifusa N, Nagayama K, Ogihara S (1987) Multilayer ceramic circuit board. US Patent 4,672,152

Stetson H (1965) Method of making multilayer circuits. US Patent 3,189,978

Story AT (1904) A story of wireless telegraphy. D. Appleton, New York

Su W, Cook B, Tentzeris M, Mariotti C, Roselli L (2014) A novel inkjet-printed microfluidic tunable coplanar patch antenna. In: Antennas and propagation society international symposium (APSURSI), 2014 IEEE, pp 858−859

Su W, Cook B, Tentzeris M (2015) Low-cost microfluidics-enabled tunable loop antenna using inkjet-printing technologies. In: 9th European conference on antennas and propagation (EuCAP), 2015 IEEE, Lisbon, Portugal

Sun M, Zhang YP (2007) 100-ghz quasi-yagi antenna in silicon technology. IEEE Electron Device Lett 28(5):455–457

Tehrani B, Bito J, Cook B, Tentzeris M (2014). Fully inkjet-printed multilayer microstrip and T-resonator structures for the RF characterization of printable materials and interconnects. In: Microwave symposium (IMS), 2014 I.E. MTT-S international, pp 1−4

Tehrani B, Cook B, Tentzeris M (2015) Postprocess fabrication of multilayer mm-wave on-package antennas with inkjet printing. In: Antennas and propagation society international symposium (APS/URSI), 2015 IEEE

Thompson D, Tantot O, Jallageas H, Ponchak GE, Tentzeris MM, Papapolymerou J (2004) Characterization of liquid crystal polymer (lcp) material and transmission lines on lcp substrates from 30 to 110 ghz. IEEE Trans Microwave Theory Tech 52(4):1343–1352

Tummala R (2008) System on package. McGraw-Hill Professional

Voxel8 (13 April 2015). http://www.voxel8.co/

Vyas R, Rida A, Bhattacharya S, Tentzeris MM (2007) Liquid crystal polymer (lcp): the ultimate solution for low-cost rf flexible electronics and antennas. In: Antennas and propagation society international symposium, 2007 IEEE, IEEE, Honolulu, HI, pp 1729−1732

Weedon WH, Payne WJ, Rebeiz GM (2001) Mems-switched reconfigurable antennas. In: Antennas and propagation society international symposium, 2001. IEEE, vol 3, IEEE, Boston, MA, pp 654−657

Wi S, Sun Y, Song I, Choa S (2006) Package-level integrated antennas based on LTCC technology. IEEE Trans Antennas Propag 54(8):2190–2197

Wicker RB, Medina F, Elkins C (2004) Multiple material micro-fabrication: extending stereolithography to tissue engineering and other novel applications. In: Proceedings of 15th annual solid freeform fabrication symposium, Austin, pp 754−64

Wiederrecht G (2009) Handbook of nanofabrication. Elsevier, Amsterdam

Yang L, Rida A, Vyas R, Tentzeris M (2007) RFID tag and RF structures on a paper substrate using inkjet-printing technology. IEEE Trans Microwave Theory Tech 55(12):2894–2901